Simultaneous Equation Model

Class 7

Wonmun Shin

(wonmun.shin@sejong.ac.kr)

Department of Economics, Sejong University

* This lecture note is written based on Professor Chang Sik Kim's lecture note.

Introduction: Simultaneous Equation Model

Introduction: Simultaneous Equation Model

- The regression models we have considered so far are all single equation regression models. A single dependent variable (Y) is expressed as a function of one or more variables (independent variables : X's).
- Underlying assumption for a single equation model is that there exists a causal relationship, and we treat Y as dependent variable and X as the determining or causal variables.
- However, there are many situations in economic models in which such a uni-directional relationship between Y and X's can not be maintained.
- To take into account the bilateral relationship between Y and X, we need more than one equations in which there are feedback relationships among variables. We call this set of equations as ***Simultaneous Equation Model***.

Example: Simple Keynesian Model

- Consider the model

$$C_t = \beta_1 + \beta_2 Y_t + e_t \quad (1)$$
$$Y_t = C_t + I_t \quad (2)$$

  - for $t = 1, \cdots, n$
  - where C_t: Consumption, Y_t: Income, and I_t: Investment
  - e_t is the error term that satisfies $E(e_t) = 0$, homoskedasticity, and no autocorrelation.

- Assume that I_t is determined exogenously by outside of system.

- From the equation (1), we know that income affects the consumption expenditure, and from the equation (2), we also find that the consumption is a component for the determination of income.

Example: Simple Keynesian Model [cont'd]

- Therefore, consumption and income are interdependent (or jointly dependent), and we have to find out how Y_t and C_t can be determined in this SEM.
- Usually, we call two variables Y_t and C_t as ***endogenous*** variables, and I_t as ***exogenous*** variable.
- We also have estimation problem in the equation (1) because Y_t and e_t are not uncorrelated, and generally simple OLS estimator is not consistent in the system. Often we call this as *endogeneity problem*.

Example: Demand-Supply Model

- Supply and demand equation is a classical example of SEM. Consider the following model:

$$\begin{aligned} \textit{Demand}: \quad & Q_i = \alpha_1 P_i + \alpha_2 Y_i + e_i^d & (3) \\ \textit{Supply}: \quad & P_i = \beta_1 Q_i + e_i^s & (4) \end{aligned}$$

  - where Q_i is quantity, P_i is price, and Y_i is income.
  - Assume Y_i is exogenous $\rightarrow$ Y_i is uncorrelated to e_i^d and e_i^s.
  - Assume that $E\left(e_i^d\right) = E\left(e_i^s\right) = 0$, $Var\left(e_i^d\right) = \sigma_d^2$, $Var\left(e_i^s\right) = \sigma_s^2$, and no autocorrelation.
  - Furthermore, we assume $Cov\left(e_i^d, e_i^s\right) = 0$.

Example: Demand-Supply Model [cont'd]

- In this model, the equilibrium P^* and Q^* should be determined simultaneously, and those two variables (P_i, Q_i) are endogenous variables.
- We see the feedback between P_i and Q_i because they are jointly determined.
  - The random errors $\left(e_i^d, e_i^s\right)$ affect both P_i and Q_i.
  - Consider a small change in e_i^s. It affects the value of P_i through the supply equation (4). At the same time, the price change will induce the quantity change through the demand equation (3). $\rightarrow$ There exists a correlation between e_i^s and Q_i.
  - Likewise, there is a correlation between e_i^d and P_i.
  - Therefore, we face the *endogeneity problem*.

Failure of OLS estimation in SEM

- In SEM, there are correlations between the random error and the endogenous variables.
- By the endogeneity problem, OLS estimators in SEM are not consistent, that is, the estimator does not converge to the true parameter value even if we increase the sample size.
- Then, how do we resolve the endogeneity problem in SEM?
  1. **Indirect Least Squares (ILS)**
  2. **Two-Stage Least Squares (2SLS)**

Reduced Form Model

Reduced Form Equations

- Recall, our problem comes from that endogenous variables exist on the right-hand side of the regression equation. $\rightarrow$ Let's do not leave any endogenous variable on the RHS.
- The simple Keynesian model and the demand-supply model are called **Structural Form Model** (which are based on some underlying theoretical economic model).
- We can express these structural models as functions of exogenous variables. This reformulation of the model is called as **Reduced Form Model**.

Reduced Form Equations: Examples [cont'd]

- Reduced form of the simple Keynesian model:

$$C_t = \frac{\beta_1}{1-\beta_2} + \frac{\beta_2}{1-\beta_2} I_t + \frac{1}{1-\beta_2} e_t$$
$$Y_t = \frac{\beta_1}{1-\beta_2} + \frac{1}{1-\beta_2} I_t + \frac{1}{1-\beta_2} e_t$$

$$\Longrightarrow \quad C_t = \gamma_1 + \gamma_2 I_t + \varepsilon_t$$
$$Y_t = \omega_1 + \omega_2 I_t + v_t$$

- where $\gamma_1 = \frac{\beta_1}{1-\beta_2}$, $\gamma_2 = \frac{\beta_2}{1-\beta_2}$, $\omega_1 = \frac{\beta_1}{1-\beta_2}$, $\omega_2 = \frac{1}{1-\beta_2}$, $\varepsilon_t = \frac{1}{1-\beta_2} e_t$, and $v_t = \frac{1}{1-\beta_2} e_t$

Reduced Form Equations: Examples [cont'd]

- Note that the reduced form errors ε_t and v_t satisfies the classical assumptions.
- Moreover, the regressor I_t does not correlated with the reduced form errors.
- As a result, we can obtain consistent estimators for the *reduced form parameters* $(\gamma_1, \gamma_2, \omega_1, \omega_2)$ through the OLS estimation.
- Question   Can we obtain the estimates of the original *structural form parameters* (β_1, β_2) from the reduced form coefficients?
- Answer   Yes! In this case, we call the method **Indirect Least Squares (ILS)**.

Reduced Form Equations: Examples [cont'd]

- Reduced form of the demand-supply model

$$Q_i = \frac{\alpha_2}{1-\alpha_1\beta_1} Y_i + \frac{\alpha_1 e_i^s + e_i^d}{1-\alpha_1\beta_1}$$
$$P_i = \frac{\alpha_2\beta_1}{1-\alpha_1\beta_1} Y_i + \frac{e_i^s + \beta_1 e_i^d}{1-\alpha_1\beta_1}$$

$$\Longrightarrow \quad Q_i = \pi_1 Y_i + u_{1,i}$$
$$P_i = \pi_2 Y_i + u_{2,i}$$

- where $\pi_1 = \frac{\alpha_2}{1-\alpha_1\beta_1}$, $\pi_2 = \frac{\alpha_2\beta_1}{1-\alpha_1\beta_1}$, $u_{1,i} = \frac{\alpha_1 e_i^s + e_i^d}{1-\alpha_1\beta_1}$, and $u_{2,i} = \frac{e_i^s + \beta_1 e_i^d}{1-\alpha_1\beta_1}$

Reduced Form Equations: Examples [cont'd]

- Note that the reduced form errors $u_{1,i}$ and $u_{2,i}$ satisfies the classical assumptions.
- Also we know that the regressor Y_i is exogenous and determined outside this system.
- Therefore, in the reduced form equations, two parameters (π_1, π_2) can be consistently estimated by the OLS.

- Question Can we obtain the estimates of the original structural form coefficients $(\alpha_1, \alpha_2, \beta_1)$ from the reduced form coefficients?
- Answer Unfortunately not because there are three unknowns but we have only two equations $\rightarrow$ Indeterminacy
  - We can obtain $\hat{\beta}_1$ but we do not identify $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

Identification Problem

The Identification Problem

- By the **identification problem** we mean whether numerical estimates of the parameters of a structural equation can be obtained from the estimated reduced form coefficients.
  - If this can be done, we say that the particular equation is *identified*.
  - If this cannot be done, then we say that the equation under consideration is *unidentified* (or *underidentified*).

Intuition: Demand-Supply Model

- Recall the following model:

$$\begin{aligned} Demand: \quad & Q_i = \alpha_1 P_i + \alpha_2 Y_i + e_i^d \\ Supply: \quad & P_i = \beta_1 Q_i + e_i^s \end{aligned}$$

- Specifically, suppose that we want to get a good fit of supply curve from the observations of (Q_i, P_i).
- We cannot know what is the supply curve from the equilibrium points since there are a lot of possibilities of a combination of demand and supply curve.
  - In other words, we cannot identify the coefficient β_1 from (Q_i, P_i) since both Q_i and P_i are endogenous variables.
- Clearly, some additional information about the nature of demand and supply curves is needed.
- Fortunately, we can know the shifts of the demand schedule with respect to change in income (Y_i), then the scatter-points (Q_i, P_i) can trace out the supply curve!
  - We use Y_i as an instrument to identify the supply curve.

Intuition: Demand-Supply Model [cont'd]

- Note that the right-hand side of supply curve has an endogenous variable Q_i.
  - In order to identify the supply curve (or β_1), Q_i needs helps from *something outside of the equation*. $\rightarrow$ exogenous variable Y_i in the demand curve can provide endogenous variable Q_i with the help!
  - That is, we need 1 instrument to estimate β_1 ($=$ 1 endogenous variable on the RHS) and we have 1 available instrument ($=$ 1 exogenous variable outside of the equation).
  - In case, we can identify β_1.
- Note that we can't identify the demand curve because there is no exogenous variable in the SEM except Y_i (which is already included in the demand equation).

Order Condition

- A necessary (but not sufficient) condition of identification is the **order condition**.
- For the identification of a equation in the SEM, the number of excluded exogenous variables should be greater than the number of included endogenous variables (on the right-hand side) in the equation.

$$\# \text{ of excluded exogenous variables} \geq \# \text{ of included endogenous variables (on RHS)}$$

- If there are K exogenous variables in the SEM, and the equation of interest has k exogenous variables, the number of excluded exogenous variables is $K - k$.

$$K - k \geq \# \text{ of included endogenous variables (on RHS)}$$

- If there are m endogenous variables in the equation, the number of included endogenous variables on the RHS is $m-1$ because the LHS has 1 endogenous variable:

$$K - k \geq m - 1$$

- If $K - k < m - 1$, the equation is **under-identified** (or unidentified).
- If $K - k = m - 1$, the equation is **just-identified** (or exactly identified).
- If $K - k > m - 1$, the equation is **over-identified**.
- We can obtain the estimates of structural form equation parameters when the equation is just-identified or over-identified.
- ***(cf)*** The necessary and sufficient condition of identification is the *rank condition*, but its discussion is beyond our course. Please refer to *Gujarati Chapter 19* if you are interest in it.

Example: Demand-Supply Model

$$\begin{aligned} Demand: \quad & Q_i = \alpha_1 P_i + \alpha_2 Y_i + e_i^d \\ Supply: \quad & P_i = \beta_1 Q_i + e_i^s \end{aligned}$$

- The supply curve equation is just-identified.

$$(K - k =)\, 1 = 1\, (= m - 1)$$

- The demand curve equation is under-identified.

$$(K - k =)\, 0 < 1\, (= m - 1)$$

SEM Estimation Methods

Recursive Model: OLS

- We learned that the OLS method is inappropriate for the SEM because of the interdependence between the endogenous explanatory variables and the error term.
- However, there is one situation where OLS can be applied appropriately even in the context of the SEM.
- This is the case of the **recursive model** (or triangular model).

$$Y_{1t} = \beta_{10} + \gamma_{11} X_{1t} + \gamma_{12} X_{2t} + e_{1t} \tag{5}$$

$$Y_{2t} = \beta_{20} + \beta_{21} Y_{1t} + \gamma_{21} X_{1t} + \gamma_{22} X_{2t} + e_{2t} \tag{6}$$

$$Y_{3t} = \beta_{30} + \beta_{31} Y_{1t} + \beta_{32} Y_{2t} + \gamma_{31} X_{1t} + \gamma_{32} X_{2t} + e_{3t} \tag{7}$$

Recursive Model: OLS [cont'd]

- Equation (5) contains only exogenous variables on the right-hand side and they are uncorrelated with the error term $\rightarrow$ OLS can be applied.
- Equation (6) contains the endogenous variable Y_{1t} as an explanatory variable. But, Y_{1t} is a predetermined variable insofar as Y_{2t} is concerned! $\rightarrow$ OLS can be applied.
- Likewise, OLS can be applied in the equation (7) because Y_{1t} and Y_{2t} are predetermined variables insofar as Y_{3t} is concerned.
- In other words, there are no endogeneity problems in the case of the recursive model.
  - Unfortunately, however, most simultaneous equation model do not exhibit such a unilateral cause-and-effect relationship.
  - Therefore, OLS, in general, is inappropriate to estimate a single equation in the context of the SEM.

Just-Identified Equation: ILS

- For a just-identified equation, the method of obtaining the estimates of the structural coefficients from the OLS estimates of the reduced form coefficients is known as the method of **Indirect Least Squares (ILS)**.

- **[Step 1]** Check whether the equation is just-identified or not.

- **[Step 2]** Obtain the reduced form equations.

- **[Step 3]** Apply OLS to the reduced form equations individually.

- **[Step 4]** Derive the estimates of the structural coefficients.
  - If an equation is just-identified, there is one-to-one correspondence between the structural and reduced-form coefficients.
  - Therefore, we can calculate the estimates of the original structural coefficients from the estimated reduced-form coefficients.

ILS Example: Simple Keynesian Model

$$C_t = \beta_1 + \beta_2 Y_t + e_t$$
$$Y_t = C_t + I_t$$

- We would like to obtain estimates of β_2.
- **[Step 1]** Order condition: $(K - k =)\, 1 = 1\, (= m - 1) \rightarrow$ *Just-identified*
- **[Step 2]** Reduced form (Slide $\#9$)
- **[Step 3]** OLS estimator $\hat{\gamma}_2$ and $\hat{\omega}_2$
- **[Step 4]** ILS estimator $\hat{\beta}_2 = \hat{\omega}_2 / \hat{\gamma}_2$

ILS Example: Demand-Supply Model

$$\begin{aligned} Demand: \quad & Q_i = \alpha_1 P_i + \alpha_2 Y_i + e_i^d \\ Supply: \quad & P_i = \beta_1 Q_i + e_i^s \end{aligned}$$

- We would like to obtain estimates of β_1.
- **[Step 1]** Order condition: $(K - k =)\, 1 = 1\, (= m - 1) \rightarrow$ *Just-identified*
- **[Step 2]** Reduced form (Slide $\#11$)
- **[Step 3]** OLS estimator $\hat{\pi}_1$ and $\hat{\pi}_2$
- **[Step 4]** ILS estimator $\hat{\beta}_1 = \hat{\pi}_2 / \hat{\pi}_1$

Just- and Over-Identified Equation: 2SLS

- Recall that the main problem in estimating the SEM is the endogeneity issue, and we know that the endogeneity problem can be solved by IV estimation.
- Consider the following model:

$$Y_{1i} = \alpha_1 + \alpha_2 Y_{2i} + \alpha_3 X_{1i} + e_{1i} \tag{8}$$

$$Y_{2i} = \beta_1 + \beta_2 Y_{1i} + \beta_3 X_{2i} + e_{2i} \tag{9}$$

  - Y_{1i}, Y_{2i}: endogenous variables, X_{1i}, X_{2i}: exogenous variables
  - Identification: Each equation in the model given by (8) and (9) is just-identified.
  - Endogeneity problem: $Cov\left(Y_{2i}, e_{1i}\right) \neq 0$ for (8), $Cov\left(Y_{1i}, e_{2i}\right) \neq 0$ for (9)
  - If we can find an **instrument** for Y_{2i} in (8) and that for Y_{1i} in (9), we apply the OLS.

Just- and Over-Identified Equation: 2SLS [cont'd]

- Recall the conditions of a good instrument:

$$Cov\left(Instrument,\ Error\ Term\right) = 0 \quad \text{(Exogeneity)}$$

$$Cov\left(Instrument,\ Origianl\ Regressor\right) \neq 0 \quad \text{(Relevance)}$$

- Then, how do we obtain such a good instrument for Y_{2i} in (8) and that for Y_{1i} in (9)?
- The answer is provided by the **Two-Stage Least Squares (2SLS)**.

- As the name of 2SLS indicates, the method involves two successive applications of OLS.
- **[Step 1]** Check whether the equation is identified (just- or over-identified) or not.
- **[Step 2]** Transform the structural form to the reduced from.

$$Y_{1i} = \pi_{11} + \pi_{12}X_{1i} + \pi_{13}X_{2i} + \varepsilon_{1i} \tag{10}$$

$$Y_{2i} = \pi_{21} + \pi_{22}X_{1i} + \pi_{23}X_{2i} + \varepsilon_{2i} \tag{11}$$

  - Note that endogenous variables are expressed in terms of exogenous variables and error terms only.
  - Newly defined error terms satisfies the classical assumptions.

Just- and Over-Identified Equation: 2SLS [cont'd]

- **[Step 3]** ***First Stage***
  - Run the OLS regression in (11) and obtain the fitted value

$$\hat{Y}_{2i} = \hat{\pi}_{21} + \hat{\pi}_{22}X_{1i} + \hat{\pi}_{23}X_{2i}$$

- **[Step 4]** ***Second Stage***
  - Use this fitted value to estimate (8) as

$$Y_{1i} = \alpha_1 + \alpha_2\hat{Y}_{2i} + \alpha_3X_{1i} + e_{1i}$$

  - That is, $\hat{Y}_{2i}$ from the first stage is used as the instrument for Y_{2i}!
  - There is no endogeneity problem, so we can estimate α_1, α_2, α_3 by OLS.

- The same procedure is applied to the estimation of (9).

Just- and Over-Identified Equation: 2SLS [cont'd]

- Multicollinearity issue?
  - When plugging $\hat{Y}_{2i}$ into (8):

$$\begin{aligned} Y_{1i} &= \alpha_1 + \alpha_2 \hat{Y}_{2i} + \alpha_3 X_{1i} + e_{1i} \\ &= \alpha_1 + \alpha_2 \left(\hat{\pi}_{21} + \hat{\pi}_{22} X_{1i} + \hat{\pi}_{23} X_{2i} \right) + \alpha_3 X_{1i} + e_{1i} \\ &= \left(\alpha_1 + \alpha_2 \hat{\pi}_{21} \right) + \left(\alpha_2 \hat{\pi}_{22} + \alpha_3 \right) X_{1i} + \alpha_2 \hat{\pi}_{23} X_{2i} + e_{1i} \end{aligned}$$

  - We can identify all parameters. $\rightarrow$ No problem

2SLS Example: Demand-Supply Model

$$\begin{aligned} Demand: \quad Q_i &= \alpha_1 P_i + \alpha_2 Y_i + e_i^d \\ Supply: \quad P_i &= \beta_1 Q_i + e_i^s \end{aligned}$$

- We would like to obtain estimates of β_1.
- **[Step 1]** Order condition: $(K - k =)\, 1 = 1\, (= m - 1) \rightarrow$ *Just-identified*
- **[Step 2]** Reduced form (Slide #11)
- **[Step 3]** First stage: $\hat{Q}_i = \hat{\pi}_1 Y_i$
- **[Step 4]** Second stage: $P_i = \beta_1 \hat{Q}_i + e_i^s$

2SLS Example: Demand-Supply Model [cont'd]

$$\begin{aligned} Demand: \quad & Q_i = \alpha_1 P_i + \alpha_2 Y_i + e_i^d \\ Supply: \quad & P_i = \beta_1 Q_i + e_i^s \end{aligned}$$

- How about the demand curve?
- **[Step 1]** Order condition: $(K - k =)\, 0 < 1\, (= m - 1) \rightarrow$ *Under-identified*
  - So we can't use the 2SLS, but let's investigate what happens when using it.
- **[Step 2]** Reduced form (Slide #11)
- **[Step 3]** First stage: $\hat{P}_i = \hat{\pi}_2 Y_i$
- **[Step 4]** Second stage: $Q_i = \alpha_1 \hat{P}_i + \alpha_2 Y_i + e_i^d$
  - We can't identify $\hat{\alpha}_1 \because Q_i = \alpha_1 \hat{\pi}_2 Y_i + \alpha_2 Y_i + e_i^d = (\alpha_1 \hat{\pi}_2 + \alpha_2)\, Y_i + e_i^d$ *(multicollinearity issue)*