LECTURE 13
SIMULTANEOUS EQUATIONS

Demand-supply system

In this lecture, we discuss endogeneity problem that arises due to simultaneity, i.e. the left-hand side variable and some of the right-hand side variables are determined simultaneously. A major example is the demand-supply system of equations:

$$\begin{aligned} Q_i^d &= \gamma_1 P_i + U_{1i}, \\ Q_i^s &= \gamma_2 P_i + U_{2i}, \end{aligned}$$

where Q_i^d and Q_i^s are quantities demanded and supplied respectively, and P_i is the price (we can assume that $\gamma_1 < 0$ and $\gamma_2 > 0$). The system also includes the following *identity* or equilibrium condition:

$$Q_i^d = Q_i^s = Q_i.$$

Naturally, econometrician does not observe Q_i^d and Q_i^s, but only Q_i determined in the equilibrium together with P_i. As a result, a simple regression of Q_i against P_i is meaningless, since Q_i comes from both equations. Further, we can show that P_i is correlated with both U_{1i} and U_{2i}. First, we solve the system in terms of U_{1i} and U_{2i}. Subtract the demand equation from the supply and use the equilibrium condition to obtain:

$$P_i = \frac{U_{1i} - U_{2i}}{\gamma_2 - \gamma_1},$$

and

$$Q_i = \frac{\gamma_2 U_{1i} - \gamma_1 U_{2i}}{\gamma_2 - \gamma_1}.$$

Therefore, assuming that $E(U_{1i} U_{2i}) = 0$,

$$\begin{aligned} E(P_i U_{1i}) &= \frac{EU_{1i}^2}{\gamma_2 - \gamma_1} \\ &\neq 0. \end{aligned}$$

Similarly, we can show that other three covariances $E(P_i U_{2i})$, $E(Q_i U_{1i})$, $E(Q_i U_{2i})$, and, therefore, both Q_i and P_i are endogenous, which violates one of the critical assumptions of the regression analysis. As a result, it is impossible to estimate consistently γ_1 and γ_2.

Next, assume that the demand equation includes another variable, say, income (I_i). Further, assume that I_i is excluded from the supply equation and *predetermined,* i.e. is not affected by Q_i and P_i. In fact, we assume that I_i is exogenous:

$$E(I_i U_{1i}) = E(I_i U_{2i}) = 0.$$

Now, the system is given by

$$\begin{aligned} Q_i^d &= \gamma_1 P_i + \beta_1 I_i + U_{1i}, & (1) \\ Q_i^s &= \gamma_2 P_i + U_{2i}, & (2) \\ Q_i^d &= Q_i^s = Q_i. & (3) \end{aligned}$$

Again, we can solve the system in terms of the predetermined variable I_i and the *shocks* U_{1i} and U_{2i} :

$$\begin{aligned} P_i &= \frac{\beta_1}{\gamma_2 - \gamma_1} I_i + \frac{U_{1i} - U_{2i}}{\gamma_2 - \gamma_1}, \\ Q_i &= \frac{\gamma_2 \beta_1}{\gamma_2 - \gamma_1} I_i + \frac{\gamma_2 U_{1i} - \gamma_1 U_{2i}}{\gamma_2 - \gamma_1}, \end{aligned}$$

or

$$\begin{aligned} P_i &= \pi_1 I_i + V_{1i}, \quad (4)\\ Q_i &= \pi_2 I_i + V_{2i}, \quad (5) \end{aligned}$$

where

$$\begin{aligned} \pi_1 &= \frac{\beta_1}{\gamma_2 - \gamma_1},\\ \pi_2 &= \frac{\gamma_2\beta_1}{\gamma_2 - \gamma_1},\\ V_{1i} &= \frac{U_{1i} - U_{2i}}{\gamma_2 - \gamma_1},\\ V_{2i} &= \frac{\gamma_2 U_{1i} - \gamma_1 U_{2i}}{\gamma_2 - \gamma_1}. \end{aligned}$$

The system of equations (1)-(3) is called *structural* its parameters are referred as the *structural coefficients.* Equations (4) and (5) are called the *reduced form equations*, and π_1, π_2 - *reduced form coefficients.* Note that the reduced form errors V_{1i} and V_{2i} are correlated even if the demand and supply shocks U_{1i} and U_{2i} are independent.

Since I_i is exogenous, one can consistently estimate reduced form equations by the usual LS estimation. However, the economists are usually interested in the structural equations. The structural equation is called identified if its coefficients can be recovered from the reduced form parameters. In the above example, we have that

$$\gamma_2 = \pi_2/\pi_1. \quad (6)$$

Note that we assume that $\pi_1 \neq 0$ or $\beta_1 \neq 0$. Thus, the supply equation is identified while the demand is not. Identification of the supply equation is possible because variation in I_i introduces exogenous shifts of the demand equation, which allows us to "see" the points on the supply line.

One can consistently estimate γ_2 by *Indirect LS* (ILS). Let $\widehat{\pi}_1$ and $\widehat{\pi}_2$ be the LS estimators of the reduced coefficients π_1 and π_2 respectively. The ILS estimator of γ_2 is given by

$$\widehat{\gamma}_2^{ILS} = \frac{\widehat{\pi}_2}{\widehat{\pi}_1}.$$

The estimator $\widehat{\gamma}_2^{ILS}$ is consistent if $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are consistent as follows from (6). Further, it is easy to show that the ILS estimator is, in fact, identical to the IV estimator. We have that

$$\begin{aligned} \widehat{\pi}_1 &= \frac{\sum_{i=1}^n I_i P_i}{\sum_{i=1}^n I_i^2},\\ \widehat{\pi}_2 &= \frac{\sum_{i=1}^n I_i Q_i}{\sum_{i=1}^n I_i^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \widehat{\gamma}_2^{ILS} &= \frac{\sum_{i=1}^n I_i Q_i}{\sum_{i=1}^n I_i P_i}\\ &= \widehat{\gamma}_2^{IV}. \end{aligned}$$

The asymptotic distribution of the IV estimator has been discussed in Lecture 10.

Next, consider the following system:

$$\begin{aligned} Q_i^d &= \gamma_1 P_i + \beta_1 I_i + U_{1i},\\ Q_i^s &= \gamma_2 P_i + \beta_2 I_i + U_{2i}, \quad (7)\\ Q_i^d &= Q_i^s = Q_i. \end{aligned}$$

The reduced form equations are the same as in (4)-(5), however, now we have

$$\begin{aligned}\pi_1 &= \frac{\beta_1 - \beta_2}{\gamma_2 - \gamma_1},\\ \pi_2 &= \frac{\gamma_2\beta_1 - \gamma_1\beta_2}{\gamma_2 - \gamma_1},\end{aligned}$$

which cannot be solved for either of the structural coefficients. The system is not identified because changes in I_i shift both equations. Thus, in order for the supply to be identified, I_i must be *excluded* from the supply equation ($\beta_2 = 0$).

Next, suppose that the demand equation contains two exogenous variables excluded from the supply equation:

$$\begin{aligned}Q_i^d &= \gamma_1 P_i + \beta_{11} I_i + \beta_{12} W_i + U_{1i},\\ Q_i^s &= \gamma_2 P_i + U_{2i},\\ Q_i^d &= Q_i^s = Q_i,\end{aligned}$$

where W_i is exogenous. The reduced form is

$$\begin{aligned}P_i &= \pi_{11} I_i + \pi_{12} W_i + V_{1i},\\ Q_i &= \pi_{21} I_i + \pi_{22} W_i + V_{2i},\end{aligned}$$

with

$$\begin{aligned}\pi_{11} &= \frac{\beta_{11}}{\gamma_2 - \gamma_1},\\ \pi_{12} &= \frac{\beta_{12}}{\gamma_2 - \gamma_1},\\ \pi_{21} &= \frac{\gamma_2\beta_{11}}{\gamma_2 - \gamma_1},\\ \pi_{22} &= \frac{\gamma_2\beta_{12}}{\gamma_2 - \gamma_1}.\end{aligned}$$

Now, there are two solutions for γ_2:

$$\gamma_2 = \frac{\pi_{21}}{\pi_{11}} \text{ and } \gamma_2 = \frac{\pi_{22}}{\pi_{12}}.$$

As a result, the ILS will produce two different estimates of γ_2. In this case, we say that the model is overidentified, and a better approach is the 2SLS. Define

$$Z_i = \begin{pmatrix} I_i \\ W_i \end{pmatrix}.$$

The 2SLS estimator of γ_2 is given by

$$\widehat{\gamma}_2^{2SLS} = \frac{\sum_{i=1}^n P_i Z_i' \left(\sum_{i=1}^n Z_i Z_i'\right)^{-1} \sum_{i=1}^n Z_i Q_i}{\sum_{i=1}^n P_i Z_i' \left(\sum_{i=1}^n Z_i Z_i'\right)^{-1} \sum_{i=1}^n Z_i P_i}.$$

If U_{2i}'s are heteroskedastic (conditional on Z_i), one can use the two step procedure to obtain the efficient GMM estimator, as discussed in Lecture 12:

$$\widehat{\gamma}_2^{GMM} = \frac{\sum_{i=1}^n P_i Z_i' \left(\sum_{i=1}^n \widehat{U}_{2i}^2 Z_i Z_i'\right)^{-1} \sum_{i=1}^n Z_i Q_i}{\sum_{i=1}^n P_i Z_i' \left(\sum_{i=1}^n \widehat{U}_{2i}^2 Z_i Z_i'\right)^{-1} \sum_{i=1}^n Z_i P_i},$$

where $\widehat{U}_{2i} = Q_i - \widehat{\gamma}_2^{2SLS} P_i$.

Identification and estimation

We will use the following notation to describe the system of m simultaneous equations. Let $y_{1i}, \ldots, y_{mi}$ be the m random endogenous variables, and $Z_i = (z_{1i}, \ldots, z_{li})'$ be the random l-vector of exogenous variables. The variable y_{ji} appears on the left-hand side of equation j, $j = 1, \ldots, m$. Let Y_{ji} to denote the m_j-vector of the right-hand side endogenous variables included in the j-th equation. Similarly, the random l_j-vector Z_j denotes the right-hand side exogenous variables included in equation j. Let u_{ji} be the random shock to equation j. Thus, we can write the j-th equation as

$$y_{ji} = Y_{ji}'\gamma_j + Z_{ji}'\beta_j + u_{ji},$$

where $E(Z_i u_{ji}) = 0$, $\gamma_j \in R^{m_j}$, and $\beta_j \in R^{l_j}$. This equation describes an IV regression model. Define further,

$$\begin{aligned} X_{ji} &= \begin{pmatrix} Y_{ji} \\ Z_{ji} \end{pmatrix}, \\ \delta_j &= \begin{pmatrix} \gamma_j \\ \beta_j \end{pmatrix}. \end{aligned}$$

Then, the above equation can be written as

$$y_{ji} = X_{ji}'\delta_j + u_{ji},$$

where we know that m_j out of $m_j + l_j$ regressors are endogenous. We have total l instrumental variables in Z_i available to us. The moment condition for equation j is given by

$$\begin{aligned} 0 &= EZ_i\left(y_{ji} - X_{ji}'\delta_j\right) \\ &= E\left(Z_i y_{ji} - Z_i X_{ji}'\delta_j\right). \end{aligned} \tag{8}$$

The GMM estimation requires that the $l \times (m_j + l_j)$ matrix $EZ_i X_{ji}'$ has the full column rank $m_j + l_j$ (the rank condition). Thus, the (necessary) order condition for identification of equation j is $l \geq m_j + l_j$, or

$$l - l_j \geq m_j.$$

The order condition says, that for equation j to be identified, the number of *exogenous* regressors *excluded* from that equation must be at least as large as the number of *included endogenous* regressors. If the order and rank conditions are satisfied, one can estimate δ_j by GMM as

$$\widetilde{\delta}_j^{GMM} = \left(\sum_{i=1}^{n} X_{ji}Z_i'\left(A_{jn}'A_{jn}\right)\sum_{i=1}^{n} Z_i X_{ji}'\right)^{-1}\sum_{i=1}^{n} X_{ji}Z_i'\left(A_{jn}'A_{jn}\right)\sum_{i=1}^{n} Z_i y_{ji}.$$

The efficient single equation GMM estimator is such that

$$A_{jn}'A_{jn} \to_p Eu_{ji}^2 Z_i Z_i'$$

In the case of homoskedastic errors, i.e. when

$$E\left(u_{ji}^2 | Z_i\right) = \sigma_{jj} \text{ for all } i, \tag{9}$$

we need that

$$A_{jn}'A_{jn} \to_p \left(EZ_i Z_i'\right)^{-1}$$

We can set

$$A_{jn}'A_{jn} = \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1},$$

and the efficient GMM reduces to the 2SLS estimator:

$$\widetilde{\delta}_j^{2SLS} = \left(\sum_{i=1}^n X_{ji}Z_i'\left(\sum_{i=1}^n Z_iZ_i'\right)^{-1}\sum_{i=1}^n Z_iX_{ji}'\right)^{-1}\sum_{i=1}^n X_{ji}Z_i'\left(\sum_{i=1}^n Z_iZ_i'\right)^{-1}\sum_{i=1}^n Z_iy_{ji}.$$

The 2SLS procedure has been discussed in Lecture 12. The 2SLS estimators for the identified equations are consistent and *jointly* asymptotically normal. The asymptotic variance of the j-th 2SLS estimator is given by

$$\sigma_{jj}\left(EX_{ji}Z_i'\left(EZ_iZ_i'\right)^{-1}EZ_iX_{ji}'\right)^{-1}.$$

Let's assume further that, in addition to (9), the errors satisfy

$$E\left(u_{r,i}u_{s,i}|Z_i\right) = \sigma_{rs} \text{ for all } i. \tag{10}$$

The 2SLS estimators are *asymptotically correlated* across equations if $\sigma_{rs} \neq 0$. Assuming that equations r and s are both identified, the asymptotic covariance of $\widetilde{\delta}_r^{2SLS}$ and $\widetilde{\delta}_s^{2SLS}$ is given by the following result

$$n^{1/2}\begin{pmatrix}\left(\widetilde{\delta}_r^{2SLS}-\delta_r\right)\\ \left(\widetilde{\delta}_s^{2SLS}-\delta_s\right)\end{pmatrix} \to_d N\left(0,\begin{pmatrix}\sigma_{rr}\left(Q_r'\Sigma_Z^{-1}Q_r\right)^{-1} & \sigma_{rs}\left(Q_r'\Sigma_Z^{-1}Q_r\right)^{-1}Q_r'\Sigma_Z^{-1}Q_s\left(Q_s'\Sigma_Z^{-1}Q_s\right)^{-1}\\ \ldots & \sigma_{rs}\left(Q_s'\Sigma_Z^{-1}Q_s\right)^{-1}\end{pmatrix}\right),$$

where

$$\begin{aligned}Q_r &= EZ_iX_{r,i}',\\ Q_s &= EZ_iX_{s,i}',\\ \Sigma_Z &= EZ_iZ_i'.\end{aligned}$$

The asymptotic covariances between equations are required if one wants to test restrictions on parameters across equations, and $\sigma_{rs} \neq 0$.

When the errors are uncorrelated across equations (given Z_i's), the individual 2SLS estimators are asymptotically uncorrelated and, therefore, *asymptotically independent*, since the asymptotic distribution is normal.

System estimation

If the errors correlated across equations, the system GMM estimation can be viewed as the GLS procedure for simultaneous equations. In this case, the system estimation is efficient, while 2SLS estimation of individual equations is not.

Suppose that all m equations are identified. The moments conditions for the system are given by

$$\begin{aligned}0 &= E\begin{pmatrix}Z_i\left(y_{1i}-X_{1i}'\delta_1\right)\\ \vdots\\ Z_i\left(y_{mi}-X_{mi}'\delta_m\right)\end{pmatrix}\\ &= E\left(\begin{pmatrix}Z_iy_{1i}\\ \vdots\\ Z_iy_{mi}\end{pmatrix}-\begin{pmatrix}Z_iX_{1i}'\delta_1\\ \vdots\\ Z_iX_{mi}'\delta_m\end{pmatrix}\right)\\ &= E\left(\begin{pmatrix}Z_iy_{1i}\\ \vdots\\ Z_iy_{mi}\end{pmatrix}-\begin{pmatrix}Z_iX_{1i}' & & 0\\ & \ldots & \\ 0 & & Z_iX_{mi}'\end{pmatrix}\begin{pmatrix}\delta_1\\ \vdots\\ \delta_m\end{pmatrix}\right).\end{aligned} \tag{11}$$

Let A_{jn} be a $l \times l$ matrix of full rank that gives the weights assigned to the moment conditions associated with equation j. Comparing (8) and (11), we deduce that the system GMM estimator is given by

$$\begin{pmatrix} \widehat{\delta}_1^{GMM} \\ \vdots \\ \widehat{\delta}_m^{GMM} \end{pmatrix}$$

$$= \left(\begin{pmatrix} \sum_{i=1}^n X_{1i}Z_i' & & 0 \\ & \cdots & \\ 0 & & \sum_{i=1}^n X_{mi}Z_i' \end{pmatrix} \begin{pmatrix} A_{1n}'A_{1n} & \dots & A_{1n}'A_{mn} \\ \dots & \dots & \dots \\ A_{mn}'A_{1n} & \dots & A_{mn}'A_{mn} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Z_iX_{1i}' & & 0 \\ & \cdots & \\ 0 & & \sum_{i=1}^n Z_iX_{mi}' \end{pmatrix} \right)^{-1}$$

$$\times \begin{pmatrix} \sum_{i=1}^n X_{1i}Z_i' & & 0 \\ & \cdots & \\ 0 & & \sum_{i=1}^n X_{mi}Z_i' \end{pmatrix} \begin{pmatrix} A_{1n}'A_{1n} & \dots & A_{1n}'A_{mn} \\ \dots & \dots & \dots \\ A_{mn}'A_{1n} & \dots & A_{mn}'A_{mn} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Z_iy_{1i} \\ \vdots \\ \sum_{i=1}^n Z_iy_{mi} \end{pmatrix}.$$

Provided that homoskedasticity conditions (9) and (10) hold, the optimal weight matrices have to satisfy

$$\begin{pmatrix} A_{1n}'A_{1n} & \dots & A_{1n}'A_{mn} \\ \dots & \dots & \dots \\ A_{mn}'A_{1n} & \dots & A_{mn}'A_{mn} \end{pmatrix} \to_p \begin{pmatrix} \sigma_{11}^2 EZ_iZ_i' & \dots & \sigma_{1m}^2 EZ_iZ_i' \\ \dots & \dots & \dots \\ \sigma_{1m}^2 EZ_iZ_i' & \dots & \sigma_{mm}^2 EZ_iZ_i' \end{pmatrix}^{-1}$$

Let's assume that the errors are uncorrelated across the equations conditional on Z_i, i.e.

$$\sigma_{rs} = 0 \text{ for all } r, s = 1, \dots, m, \text{ and } r \neq s.$$

Then, the above condition for optimal weight matrices becomes

$$\begin{pmatrix} A_{1n}'A_{1n} & \dots & A_{1n}'A_{mn} \\ \dots & \dots & \dots \\ A_{mn}'A_{1n} & \dots & A_{mn}'A_{mn} \end{pmatrix} \to_p \begin{pmatrix} \sigma_1^2 EZ_iZ_i' & & 0 \\ & \cdots & \\ 0 & & \sigma_m^2 EZ_iZ_i' \end{pmatrix}^{-1}.$$

In this case, one can set

$$\begin{aligned} A_{jn}'A_{jn} &= \left(n^{-1} \sum_{i=1}^n Z_iZ_i' \right)^{-1}, \text{ and} \\ A_{r,n}'A_{s,n} &= 0 \text{ for } r \neq s, \end{aligned}$$

to obtain that the efficient system GMM estimator and 2SLS estimators for individual equations are identical.