LECTURE 3

SIMULTANEOUS EQUATIONS II: MULTIPLE-EQUATION GMM, 3SLS.

In this lecture, we consider joint GMM estimation of more than one simultaneous equation. As we will see, joint estimation can lead to efficiency gains.

Multiple-equation GMM estimator

Suppose that all m equations (in (1) in Lecture 2) are identified. Adopting the notation of Lecture 2, equation (7), we can write the system of m equations as follows:

$$\begin{aligned} y_{1i} &= X_{1,i}'\delta_1 + u_{1i}, \\ &\dots \\ y_{mi} &= X_{m,i}'\delta_m + u_{mi}, \end{aligned}$$

where for all $j = 1, \dots, m$, $\delta_j \in R^{k_j}$, $k_j = m_j + l_j$, and the random l-vector Z_i is such that

$$\begin{aligned} \text{rank}\left(EZ_i X_{j,i}'\right) &= k_j, \\ EZ_i u_{ji} &= 0. \end{aligned}$$

Equivalently, the system can be re-written in the matrix notation as

$$\begin{gathered} y_1 = X_1\delta_1 + u_1, \\ \dots \\ y_m = X_m\delta_m + u_m, \end{gathered}$$

where, for $j = 1, \dots, m$, X_j collects the n observations on the right-hand side variables in the j-th equation:

$$X_j = \begin{pmatrix} X_{j,1}' \\ \vdots \\ X_{j,n}' \end{pmatrix},$$

y_j collects the n observations on the left-hand side variable in the j-th equation:

$$y_j = \begin{pmatrix} y_{j1} \\ \vdots \\ y_{jn} \end{pmatrix},$$

and u_j is defined similarly. Further, define

$$\begin{aligned} X &= \begin{pmatrix} X_1 & & 0 \\ & \dots & \\ 0 & & X_m \end{pmatrix}, \\ Y &= \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}, \\ U &= \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \\ \delta &= \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_m \end{pmatrix}. \end{aligned}$$

Note that X is $(nm) \times k$, where $k = k_1 + \ldots + k_m$, Y and U are $(nm) \times 1$, and δ is $k \times 1$. The system can now be compactly written as

$$Y = X\delta + U.$$

In this system, we have ml population moment conditions:

$$E \begin{pmatrix} Z_i u_{1i} \\ \vdots \\ Z_i u_{mi} \end{pmatrix} = E(U_i \otimes Z_i) = 0,$$

where $U_i = (u_{1i}, \ldots, u_{mi})'$, and $A \otimes B$ denotes the Kronecker product of A and B.[1] To define sample moment conditions that can be used for estimation, consider the ml-vector of sample covariations between the exogenous variables and errors:

$$\begin{pmatrix} Z'u_1 \\ \vdots \\ Z'u_m \end{pmatrix} = \begin{pmatrix} Z'(y_1 - X_1\delta_1) \\ \vdots \\ Z'(y_m - X_m\delta_m) \end{pmatrix},$$

where Z denotes the $n \times l$ matrix of observations on the exogenous variables. Using the Kronecker product notation, this can be conveniently written as:

$$(I_m \otimes Z)'U = (I_m \otimes Z)'(Y - X\delta),$$

Let A_n be an $(ml) \times (ml)$ weight matrix. The system or multiple-equation GMM estimator is obtained by solving

$$\min_{d \in \mathbb{R}^k} (Y - Xd)' (I_m \otimes Z) A_n' A_n (I_m \otimes Z)' (Y - Xd).$$

Thus, the system GMM estimator is given by[2]

$$\hat{\delta} = (X'(I_m \otimes Z)A_n'A_n(I_m \otimes Z)'X)^{-1} X'(I_m \otimes Z)A_n'A_n(I_m \otimes Z)'Y.$$

Define

$$W_n = A_n' A_n,$$

and introduce a partition

$$W_n = \begin{pmatrix} W_{11,n} & \ldots & W_{1m,n} \\ \ldots & \ldots & \ldots \\ W_{m1,n} & \ldots & W_{mm,n} \end{pmatrix},$$

where each $W_{ij,n}$ is an $l \times l$ symmetric matrix. The system GMM estimators for the m equations can be

[1] Suppose that A is $k \times l$ and B is $m \times n$. Then $A \otimes B$ is a $(km) \times (ln)$ matrix given by

$$A \otimes B = \begin{pmatrix} a_{11} & \ldots & a_{1l} \\ & \ldots & \\ a_{k1} & \ldots & a_{kl} \end{pmatrix} \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1l}B \\ & \ldots & \\ a_{k1}B & \ldots & a_{kl}B \end{pmatrix}.$$

The properties of the Kronecker product include: $(A \otimes B)' = A' \otimes B'$, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ when A and B are invertible, and $(A \otimes B)(C \otimes D) = AC \otimes BD$ for properly defined matrices C and D.

[2] Recall that in the single equation case, the GMM estimator solves $\min_b (Y - Xb)' Z A_n' A_n Z' (Y - Xb)$, and the solution is given by $\hat{\beta} = (X'ZA_n'A_nZ'X)^{-1}X'ZA_n'A_nZ'Y$.

written as

$$
\begin{pmatrix} \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_m \end{pmatrix} =
$$

$$
\begin{aligned}
= & \left(\begin{pmatrix} X_1'Z & & 0 \\ & \dots & \\ 0 & & X_m'Z \end{pmatrix} \begin{pmatrix} W_{11,n} & \dots & W_{1m,n} \\ \dots & \dots & \dots \\ W_{m1,n} & \dots & W_{mm,n} \end{pmatrix} \begin{pmatrix} Z'X_1 & & 0 \\ & \dots & \\ 0 & & Z'X_m \end{pmatrix} \right)^{-1} \\
& \times \begin{pmatrix} X_1'Z & & 0 \\ & \dots & \\ 0 & & X_m'Z \end{pmatrix} \begin{pmatrix} W_{11,n} & \dots & W_{1m,n} \\ \dots & \dots & \dots \\ W_{m1,n} & \dots & W_{mm,n} \end{pmatrix} \begin{pmatrix} Z'y_1 \\ \vdots \\ Z'y_m \end{pmatrix} \qquad (1) \\
= & \begin{pmatrix} X_1'ZW_{11,n}Z'X_1 & \dots & X_1'ZW_{1m,n}Z'X_m \\ \dots & \dots & \dots \\ X_m'ZW_{m1,n}Z'X_1 & \dots & X_m'ZW_{mm,n}Z'X_m \end{pmatrix}^{-1} \begin{pmatrix} X_1'ZW_{11,n}Z'y_1 + \ldots + X_1'ZW_{1m,n}Z'y_m \\ \dots \\ X_m'ZW_{m1,n}Z'y_1 + \ldots + X_m'ZW_{mm,n}Z'y_m \end{pmatrix} \qquad (2)
\end{aligned}
$$

We can compare the above expression with that for equation-by-equation GMM:

$$
\begin{pmatrix} \tilde{\delta}_1 \\ \vdots \\ \tilde{\delta}_m \end{pmatrix} = \begin{pmatrix} X_1'ZA_{1n}'A_{1n}Z'X_1 & & 0 \\ & \dots & \\ 0 & & X_m'ZA_{mn}'A_{mn}Z'X_m \end{pmatrix}^{-1} \begin{pmatrix} X_1'ZA_{1n}'A_{1n}Z'y_1 \\ \dots \\ X_m'ZA_{mn}'A_{mn}Z'y_m \end{pmatrix}.
$$

From the comparison, it is apparent that the equation-by-equation GMM estimator is a particular case of the system GMM estimator with weighting matrices $W_{ij,n} = 0$ for $i \neq j$.

Large-sample properties of the multiple-equation GMM estimator

From (1), we can write

$$
\begin{pmatrix} \hat{\delta}_1 - \delta_1 \\ \vdots \\ \hat{\delta}_m - \delta_m \end{pmatrix} =
$$

$$
\begin{aligned}
= & \left(\begin{pmatrix} \sum_{i=1}^n X_{1,i}Z_i' & & 0 \\ & \dots & \\ 0 & & \sum_{i=1}^n X_{m,i}Z_i' \end{pmatrix} \begin{pmatrix} W_{11,n} & \dots & W_{1m,n} \\ \dots & \dots & \dots \\ W_{m1,n} & \dots & W_{mm,n} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Z_iX_{1,i}' & & 0 \\ & \dots & \\ 0 & & \sum_{i=1}^n Z_iX_{m,i}' \end{pmatrix} \right)^{-1} \\
& \times \begin{pmatrix} \sum_{i=1}^n X_{1,i}Z_i' & & 0 \\ & \dots & \\ 0 & & \sum_{i=1}^n X_{m,i}Z_i' \end{pmatrix} \begin{pmatrix} W_{11,n} & \dots & W_{1m,n} \\ \dots & \dots & \dots \\ W_{m1,n} & \dots & W_{mm,n} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Z_iu_{1i} \\ \vdots \\ \sum_{i=1}^n Z_iu_{mi} \end{pmatrix}.
\end{aligned}
$$

In addition to the previous assumptions, we assume:

- $\{(Y_i', Z_i') : i \geq 1\}$ are iid.
- $W_n \to_p W$ positive definite (and symmetric).
- The elements of $U_i = (u_{1i}, \ldots, u_{mi})'$ and Z_i have finite second moments (which together with the reduced form equations implies that $EZ_iX_{j,i}'$ is finite for all $j = 1, \ldots, m$).

Under these assumptions we have consistency of the system GMM estimator: $\hat{\delta}_j \to_p \delta_j$ for all $j = 1, \ldots, m$.

Next, for asymptotic normality we also assume that:

- The elements of U_i and Z_i have finite fourth moments (which implies that the elements of Y_i have finite fourth moments as well).

Under these assumptions,

$$n^{-1/2}\begin{pmatrix}\sum_{i=1}^n Z_i u_{1i}\\ \vdots \\ \sum_{i=1}^n Z_i u_{mi}\end{pmatrix} \to_d N(0,\Omega),$$

where

$$\begin{aligned}\Omega &= E\begin{pmatrix}Z_i u_{1i}\\ \vdots \\ Z_i u_{mi}\end{pmatrix}\begin{pmatrix}Z_i u_{1i}\\ \vdots \\ Z_i u_{mi}\end{pmatrix}' \\ &= E(U_i\otimes Z_i)(U_i\otimes Z_i)' \\ &= E(U_iU_i'\otimes Z_iZ_i').\end{aligned}$$

Then, we have that

$$n^{1/2}\begin{pmatrix}\hat{\delta}_1-\delta_1\\ \vdots \\ \hat{\delta}_m-\delta_m\end{pmatrix} \to_d N(0,V(W)),$$

where

$$\begin{aligned}V(W) &= (C'WC)^{-1}C'W\Omega WC(C'WC)^{-1},\\ C &= \begin{pmatrix}Q_1 & & 0\\ & \dots & \\ 0 & & Q_m\end{pmatrix},\\ Q_j &= EZ_iX_{j,i}',\ \text{for } j=1,\dots,m.\end{aligned}$$

Let's assume that

- Ω is positive definite.

As usual, the efficient GMM estimator corresponds to W_n that satisfies

$$W_n \to_p \Omega^{-1}.$$

For example,

$$\begin{aligned}W_n &= \widehat{\Omega}_n^{-1}\\ &= \left(n^{-1}\sum_{i=1}^n\left(\widehat{U}_i\widehat{U}_i'\otimes Z_iZ_i'\right)\right)^{-1},\end{aligned}$$

where $\widehat{U}_i$ is constructed using some preliminary consistent estimators of δ_j's, for example, equation-by-equation 2SLS estimators:

$$\widehat{U}_i = \begin{pmatrix}y_{1i}-X_{1,i}'\widetilde{\delta}_1^{2SLS}\\ \vdots \\ y_{mi}-X_{m,i}'\widetilde{\delta}_m^{2SLS}\end{pmatrix}.$$

The asymptotic variance of the efficient GMM estimator is given by

$$V\left(\Omega^{-1}\right) = \left(C'\Omega^{-1}C\right)^{-1}. \tag{3}$$

Homoskedastic errors

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

$$E\left(U_iU_i'|Z_i\right) = \Sigma,$$

where Σ is some positive definite $m \times m$ matrix, we have

$$\begin{aligned} \Omega &= E\left(U_iU_i' \otimes Z_iZ_i'\right) \\ &= E\left(E\left(U_iU_i'|Z_i\right) \otimes Z_iZ_i'\right) \\ &= \Sigma \otimes EZ_iZ_i'. \end{aligned}$$

In this case,

$$\widehat{\Omega}_n = \widehat{\Sigma}_n \otimes n^{-1}\sum_{i=1}^n Z_iZ_i',$$

where

$$\widehat{\Sigma}_n = n^{-1}\sum_{i=1}^n \widehat{U}_i\widehat{U}_i'.$$

A GMM estimator with

$$W_n = \widehat{\Sigma}_n^{-1} \otimes \left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right)^{-1}$$

is called the three-stage LS estimator (3SLS).

The 3SLS estimator can be written in the matrix notation as follows:

$$\begin{aligned} \hat{\delta}^{3SLS} &= \left(X'\left(I_m \otimes Z\right)\left(\widehat{\Sigma}_n^{-1} \otimes \left(Z'Z\right)^{-1}\right)\left(I_m \otimes Z\right)'X\right)^{-1} X'\left(I_m \otimes Z\right)\left(\widehat{\Sigma}_n^{-1} \otimes \left(Z'Z\right)^{-1}\right)\left(I_m \otimes Z\right)'Y \\ &= \left(X'\left(\widehat{\Sigma}_n^{-1} \otimes \left(Z\left(Z'Z\right)^{-1}Z'\right)\right)X\right)^{-1} X'\left(\widehat{\Sigma}_n^{-1} \otimes \left(Z\left(Z'Z\right)^{-1}Z'\right)\right)Y \\ &= \left(X'\left(\widehat{\Sigma}_n^{-1} \otimes P_Z\right)X\right)^{-1} X'\left(\widehat{\Sigma}_n^{-1} \otimes P_Z\right)Y, \end{aligned}$$

where $P_Z = Z(Z'Z)^{-1}Z'$. Further, let $\widehat{\sigma}^{ij}$ be the (i,j)-th element of $\widehat{\Sigma}_n^{-1}$, and define

$$\begin{aligned} \widehat{X}_j &= P_ZX_j, \\ \widehat{X} &= \begin{pmatrix} \widehat{X}_1 & & 0 \\ & \ldots & \\ 0 & & \widehat{X}_m \end{pmatrix}. \end{aligned}$$

We have,

$$\begin{aligned} \begin{pmatrix} \hat{\delta}_1^{3SLS} \\ \vdots \\ \hat{\delta}_m^{3SLS} \end{pmatrix} &= \begin{pmatrix} \widehat{\sigma}^{11}X_1'P_ZX_1 & \ldots & \widehat{\sigma}^{1m}X_1'P_ZX_m \\ \ldots & \ldots & \ldots \\ \widehat{\sigma}^{m1}X_m'P_ZX_1 & \ldots & \widehat{\sigma}^{mm}X_m'P_ZX_m \end{pmatrix}^{-1} \\ &\quad \times \begin{pmatrix} \widehat{\sigma}^{11}X_1'P_Zy_1 + \ldots + \widehat{\sigma}^{1m}X_1'P_Zy_m \\ \vdots \\ \widehat{\sigma}^{m1}X_m'P_Zy_1 + \ldots + \widehat{\sigma}^{mm}X_m'P_Zy_m \end{pmatrix} \\ &= \begin{pmatrix} \widehat{\sigma}^{11}\widehat{X}_1'X_1 & \ldots & \widehat{\sigma}_m^{1m}\widehat{X}_1'X_m \\ \ldots & \ldots & \ldots \\ \widehat{\sigma}^{m1}\widehat{X}_m'X_1 & \ldots & \widehat{\sigma}^{mm}\widehat{X}_m'X_m \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\sigma}^{11}\widehat{X}_1'y_1 + \ldots + \widehat{\sigma}^{1m}\widehat{X}_1'y_m \\ \vdots \\ \widehat{\sigma}^{m1}\widehat{X}_m'y_1 + \ldots + \widehat{\sigma}^{mm}\widehat{X}_m'y_m \end{pmatrix} \\ &= \left(\widehat{X}'\left(\widehat{\Sigma}_n^{-1} \otimes I_n\right)X\right)^{-1} \widehat{X}'\left(\widehat{\Sigma}_n^{-1} \otimes I_n\right)Y. \end{aligned}$$

Using (3), the asymptotic variance of the 3SLS estimator is given by the inverse of

$$\begin{aligned}
&\begin{pmatrix} Q_1 & & 0 \\ & \dots & \\ 0 & & Q_m \end{pmatrix}' \left(\Sigma^{-1} \otimes (EZ_iZ_i')^{-1}\right) \begin{pmatrix} Q_1 & & 0 \\ & \dots & \\ 0 & & Q_m \end{pmatrix} \\
= &\begin{pmatrix} Q_1 & & 0 \\ & \dots & \\ 0 & & Q_m \end{pmatrix}' \begin{pmatrix} \sigma^{11} (EZ_iZ_i')^{-1} & \dots & \sigma^{1m} (EZ_iZ_i')^{-1} \\ \dots & \dots & \dots \\ \sigma^{m1} (EZ_iZ_i')^{-1} & \dots & \sigma^{mm} (EZ_iZ_i')^{-1} \end{pmatrix} \begin{pmatrix} Q_1 & & 0 \\ & \dots & \\ 0 & & Q_m \end{pmatrix},
\end{aligned}$$

where

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \dots & \sigma^{1m} \\ \dots & \dots & \dots \\ \sigma^{m1} & \dots & \sigma^{mm} \end{pmatrix}.$$

Thus, the asymptotic variance of the 3SLS estimator is

$$V\left(\Omega^{-1}\right) = \begin{pmatrix} \sigma^{11} EX_{1,i}Z_i' (EZ_iZ_i')^{-1} EZ_iX_{1,i}' & \dots & \sigma^{1m} EX_{1,i}Z_i' (EZ_iZ_i')^{-1} EZ_iX_{m,i}' \\ \dots & \dots & \dots \\ \sigma^{m1} EX_{m,i}Z_i' (EZ_iZ_i')^{-1} EZ_iX_{1,i}' & \dots & \sigma^{mm} EX_{m,i}Z_i' (EZ_iZ_i')^{-1} EZ_iX_{m,i}' \end{pmatrix}^{-1}.$$

Single- vs. multiple-equation GMM

As we have seen, the equation-by-equation GMM corresponds to the case where W_n has a diagonal structure. Since in general Ω^{-1} does not take this form, even the "efficient" version of the equation-by-equation GMM estimator is inefficient when compared to the multiple-equation GMM. The reason for this is that the single equation estimator for equation j ignores the information about that equation contained in other equations. However, there are two exceptions to that rule.

First, let's assume conditional homoskedasticity. Suppose further that the errors are uncorrelated across the equations, i.e. Σ is diagonal:

$$\begin{aligned}
\Sigma &= \begin{pmatrix} Eu_{1i}^2 & \dots & Eu_{1i}u_{mi} \\ \dots & \dots & \dots \\ Eu_{mi}u_{1i} & \dots & Eu_{mi}^2 \end{pmatrix} \\
&= \begin{pmatrix} Eu_{1i}^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & Eu_{mi}^2 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_m^2 \end{pmatrix}.
\end{aligned}$$

In this case,

$$\begin{aligned}
&V\left(\Omega^{-1}\right) \\
= &\left(\begin{pmatrix} Q_1' & & 0 \\ & \dots & \\ 0 & & Q_m' \end{pmatrix} \begin{pmatrix} \sigma_1^2 EZ_iZ_i' & & 0 \\ & \dots & \\ 0 & & \sigma_m^2 EZ_iZ_i' \end{pmatrix}^{-1} \begin{pmatrix} Q_1 & & 0 \\ & \dots & \\ 0 & & Q_m \end{pmatrix}\right)^{-1} \\
= &\begin{pmatrix} \sigma_1^2 \left(Q_1' (EZ_iZ_i')^{-1} Q_1\right)^{-1} & & 0 \\ & \dots & \\ 0 & & \sigma_m^2 \left(Q_m' (EZ_iZ_i')^{-1} Q_m\right)^{-1} \end{pmatrix}.
\end{aligned}$$

First, one can see that the estimators are asymptotically independent across the equation. Second, the asymptotic variance of the multiple-equation efficient GMM estimator is the same as that of the equation-by-equation 2SLS estimator. Hence, in the homoskedastic case and when the errors are uncorrelated across the equations (conditional on Z's), the equation-by-equation 2SLS estimator is efficient.

The equation-by-equation estimator is also efficient when all equations are exactly identified. In fact, single- and multiple-equations estimators are the same when the system is exactly identified:

$$\begin{aligned}
\begin{pmatrix} \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_m \end{pmatrix} &= \left(X' \left(I_m \otimes Z\right) W_n \left(I_m \otimes Z\right)' X\right)^{-1} X' \left(I_m \otimes Z\right) W_n \left(I_m \otimes Z\right)' Y \\
&= \left(\left(I_m \otimes Z\right)' X\right)^{-1} W_n^{-1} \left(X' \left(I_m \otimes Z\right)\right)^{-1} X' \left(I_m \otimes Z\right) W_n \left(I_m \otimes Z\right)' Y \\
&= \left(\left(I_m \otimes Z\right)' X\right)^{-1} \left(I_m \otimes Z\right)' Y \\
&= \begin{pmatrix} \left(Z' X_1\right)^{-1} Z' y_1 \\ \vdots \\ \left(Z' X_m\right)^{-1} Z' y_m \end{pmatrix}.
\end{aligned}$$

Thus, when the system is exactly identified, all the estimators discussed so far reduce to the equation-by-equation IV estimator.