Moshe Buchinsky
Department of Economics
UCLA

Economics 203C
Spring 2003

Lecture Note 6

System Estimation and Three-Stage Least Squares

I. MULTIVARIATE REGRESSION WITH m ENDOGENOUS REGRESSORS

$$y_{ji} = x'_{ji}\beta_j + \epsilon_{ji} \qquad (j = 1, \ldots, m;\; i = 1, \ldots, n).$$

Let

$$Y_j = \begin{pmatrix} y_{j1} \\ \vdots \\ y_{jn} \end{pmatrix}, \quad X_j = \begin{pmatrix} x'_{j1} \\ \vdots \\ x'_{jn} \end{pmatrix}, \quad \epsilon_j = \begin{pmatrix} \epsilon_{j1} \\ \vdots \\ \epsilon_{jn} \end{pmatrix},$$

then

$$Y_j = X_j\beta_j + \epsilon_j \qquad (j = 1, \ldots, m).$$

Since X_j includes some y's from the other equations, we expect that $E[x_{ji}\epsilon_{ji}] \neq 0$.

INSTRUMENTAL VARIABLES:

$$Z = (z_1, \ldots, z_n)'$$

an $n \times l$ matrix, where each z_i is an $l \times 1$ vector.

For the instrumental variables:

1. Since z_i are exogenous, they are uncorrelated with ϵ_{ji}. That is

$$E[z_i\epsilon_{ji}] = 0.$$

2. $E[z_i x'_{ji}] = \Sigma_{zx_j}$, and $\text{rank}(\Sigma_{zx_j}) = k_j \leq l$.

II. STACKED MODEL

Can stack the Y_j's and their corresponding X_j's as in the SUR model:

$$y = X\beta + \epsilon,$$

where

$$y = \text{vec}(Y) = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}, \qquad X = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & X_m \end{pmatrix},$$

$$\beta = \text{vec}(B) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \qquad \epsilon = \text{vec}(E) = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}.$$

Note, y and ϵ are $nm \times 1$ vectors, X is an $mn \times k$ matrix, β is a $k \times 1$ vector and $k = \sum_{j=1}^{m} k_j$.

This is merely a different representation of the SEM, but the simultaneity problem has not been solved. We still have $E[X\epsilon] \neq 0$, in general.

HOMOSKEDASTIC ERRORS:

Assume that the SUR usual assumptions hold:

1. $E[\epsilon_{ji} \mid z_i] = 0$.

2. The disturbance term covariance matrix:

$$E[\epsilon_{ji}\epsilon_{ki'}] = \begin{cases} \sigma_{jk} & \text{if } i = i', \\ 0 & \text{otherwise.} \end{cases}$$

   Hence,

$$E[\epsilon\epsilon' \mid Z] = \Sigma \otimes I = V_0,$$

   where $\Sigma \equiv [\sigma_{jk}]$ an $m \times m$ matrix.

The rest, is the same as the SUR model.

III. TWO-STAGE LEAST SQUARES ESTIMATION

This estimation ignores the covariance structure of SUR disturbance term.

Stage 1: Get fitted LS values of X given Z:

$$\begin{aligned} \hat{X} &= \left(I_m \otimes Z(Z'Z)^{-1}Z'\right) X \\ &= \begin{pmatrix} Z(Z'Z)^{-1}Z'X_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Z(Z'Z)^{-1}Z'X_m \end{pmatrix}. \end{aligned}$$

Stage 2: Use $\hat{X}$ from stage 1 as an IV for X:

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{X}'X)^{-1}\hat{X}'y \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\ &= \left(X'(I \otimes Z(Z'Z)^{-1}Z')X\right)^{-1} X'(I \otimes Z(Z'Z)^{-1}Z')y. \end{aligned}$$

$$\Longrightarrow \quad \hat{\beta}_{j2SLS} = \left(X_j'Z(Z'Z)^{-1}Z'X_j\right)^{-1} X_j'(Z(Z'Z)^{-1}Z')Y_j,$$

as before.

IV. GLS VERSION OF TWO-STAGE LEAST SQUARES

Use $\hat{X}$ as an IV for X, but use V_0^{-1} as a weight matrix. That is,

$$\tilde{\beta}_{G2SLS} = (\hat{X}'V_0^{-1}X)^{-1}\hat{X}'V_0^{-1}y,$$

where $V_0^{-1} = \Sigma^{-1} \otimes I_n$.

$$\begin{aligned} \Longrightarrow \quad \tilde{\beta}_{G2SLS} &= (X'(I_m \otimes H_z)'(\Sigma^{-1} \otimes I_n)X)^{-1} X'(I_m \otimes H_z)'(\Sigma^{-1} \otimes I_n)y \\ &= (X'(\Sigma^{-1} \otimes H_z)X)^{-1} X'(\Sigma^{-1} \otimes H_z)y. \end{aligned}$$

That is, the identity matrix I_m in the formula of the 2SLS estimator is simply replaced with the matrix Σ^{-1}.

V. THREE-STAGE LEAST SQUARES

Stage 1: Get $\hat{X} = (I \otimes H_z)X$.

Stage 2: Get the 2SLS estimator $\tilde{\beta}_{2SLS}$ and compute the residuals:

$$\hat{\epsilon}_j = Y_j - X_j\tilde{\beta}_j \qquad (j = 1, \dots, m),$$

where each $\hat{e}_j$ is an $n \times 1$ vector.

Let

$$E = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_m)' = (e_1, \dots, e_n),$$

where e_i contains the m residuals for a given observation i, e_i is an $m \times 1$ vector.

Now set

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} e_i e_i' \xrightarrow{p} \Sigma.$$

Stage 3:

$$\hat{\beta}_{3SLS} = \hat{\beta}_{FG2SLS} = (\hat{X}\hat{V}^{-1}X)^{-1}\hat{X}\hat{V}^{-1}y,$$

where

$$\hat{V} = \hat{\Sigma} \otimes I_n.$$

VI. CONCLUDING REMARKS:

1. We can show that

$$\begin{aligned}\hat{\beta}_{3SLS} &= \left(\hat{X}'\hat{V}^{-1}X\right)^{-1}\hat{X}'\hat{V}^{-1}y \\ &= \left(\hat{X}'\hat{V}^{-1}\hat{X}\right)^{-1}\hat{X}'\hat{V}^{-1}y \\ &= \left(\hat{X}'\hat{V}^{-1}\hat{X}\right)^{-1}\hat{X}'\hat{V}^{-1}\hat{y},\end{aligned}$$

   where $\hat{y} = (I \otimes H_z)y$. Hence, the interpretation given in the previous class note applies here as well.

2. Under the assumption that $E[\epsilon\epsilon' \mid Z] = \Sigma \otimes I$, it is easy to verify that

$$\sqrt{n}(\hat{\beta}_{3SLS} - \beta) \xrightarrow{D} N(0, C_0^{-1}),$$

where

$$C_0 = \operatorname*{plim}_{n\to\infty} \frac{1}{n} X'(\Sigma^{-1} \otimes H_z) X.$$

Hence,

$$\hat{\beta}_{3SLS} \overset{A}{\sim} N(\beta, (\hat{X}'\hat{V}^{-1}\hat{X})^{-1}).$$

Also,

$$\hat{\beta}_{2SLS} \overset{A}{\sim} N(\beta, (\hat{X}'\hat{X})^{-1}(\hat{X}'(\hat{\Sigma} \otimes I)\hat{X})^{-1}(\hat{X}'\hat{X})^{-1})$$

and

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \left(\hat{X}'V^{-1}\hat{X}\right) \geq \operatorname*{plim}_{n\to\infty} \frac{1}{n} (\hat{X}'\hat{X})^{-1}\hat{X}'(\Sigma \otimes I)\hat{X}(\hat{X}'\hat{X})^{-1}.$$

That is, the 3SLS estimator is, ingeneral, more efficient than the 2SLS estimator.

3. If $\epsilon \sim N(0, \Sigma \otimes I)$, then the 3SLS estimator is asymptotically equivalent to an ML estimator. That is, the 3SLS is asymptotically efficient.

4. If all equations, but the j^{th} equation, are just identified, then for the j^{th} equation, 3SLS estimator is identical to the 2SLS estimator.

5. If all equations are just identified, then all estimators: 3SLS, 2SLS and IV are identical.

VII. OTHER ESTIMATORS

VII.1. INDIRECT LEAST SQUARES

This method is applicable for the j^{th} equation, only if the equation is just identified.

$$Y_j = X_j\beta_j + \epsilon_j, \quad E[\epsilon_j \mid Z] = 0.$$

Reduced form for the X_j:

$$X_j = Z\Pi_j + V_j, \quad E[V_j \mid Z] = 0.$$

So,

$$\begin{aligned} Y_j &= (Z\Pi_j + V_j)\beta_j + \epsilon_j \\ &= Z\Pi_j\beta_j + V_j\beta_j + \epsilon_j \\ &= Z\pi_j + u_j, \end{aligned}$$

where $u_j = \epsilon_j + V_j\beta_j$ and $\pi_j = \Pi_j\beta_j$.

If the j^{th} equation is just identified, then $k_j = l$. Therefore,

$$\beta_j = \Pi_j^{-1}\pi_j.$$

Hence, we can estimate β_j in two steps:

Step 1: Estimate Π_j and π_j by LS.

Step 2: Estimate β_j by

$$\hat{\beta}_j = \hat{\Pi}_j^{-1}\hat{\pi}_j.$$

For just identified equation: $\hat{\beta}_{LS} = \hat{\beta}_{2SLS} = \hat{\beta}_{IV}$.

VII.2. LIMITED INFORMATION MAXIMUM LIKELIHOOD (LIML)

This is an equation-by-equation ML estimation.

Simultaneous equation:

$$Y_j = X_j\beta_j + \epsilon_j.$$

Reduced form:

$$X_j = Z\Pi_j + V_j,$$

with

$$\begin{pmatrix} \epsilon_{ji} \\ v_{ji} \end{pmatrix} \sim \text{i.i.d. } N(0, \Sigma)$$

and Σ may be singular.

The estimator $\hat{\beta}_{jLIML}$ is an ML estimator for β_j and $\hat{\pi}_j$ an ML estimator for π_j, ignoring any other restriction on π_j from the other equations.

One can show that

$$\sqrt{n}(\hat{\beta}_{LIML} - \hat{\beta}_{2SLS}) \xrightarrow{p} 0,$$

so that both estimator have the same asymptotic distribution.

VII.3. FULL INFORMATION MAXIMUM LIKELIHOOD (FIML)

This is a multi-equation ML estimation.

SEM:

$$Y\Gamma = XB + E, \quad \epsilon = \text{vec}(E) \sim N(0, \Sigma \otimes I).$$

This procedure obtains the usual ML estimator. That is, we get

$$\hat{\beta}_{FIML} = \text{vec}\begin{pmatrix} I - \hat{\Gamma}_{FIML} \\ \hat{B}_{FIML} \end{pmatrix}$$

by ML, ignoring any zero restrictions.

One can show that

$$\sqrt{n}(\hat{\beta}_{FIML} - \hat{\beta}_{3SLS}) \xrightarrow{p} 0,$$

so that $\hat{\beta}_{FIML}$ and $\hat{\beta}_{3SLS}$ have the same asymptotic distribution. Hence, the 3SLS estimator is asymptotically efficient as was claimed before.

VII.4. BEST THREE-STAGE LEAST SQUARES ESTIMATOR (SYSTEM-WIDE GMM)

Like in the GMM and best 2SLS, we can allow for heteroskedasticity and serial correlation. Stacked form:

$$y = X\beta + \epsilon.$$

Premultiplying by $(I \otimes Z)$ gives,

$$(I \otimes Z)y = \begin{pmatrix} Z'Y_1 \\ \vdots \\ Z'Y_m \end{pmatrix} = \begin{pmatrix} Z'X_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Z'X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} Z'\epsilon_1 \\ \vdots \\ Z'\epsilon_m \end{pmatrix},$$

or

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon},$$

where

$$E[\tilde{\epsilon} \mid Z] = 0$$

$$\mathrm{Var}(\tilde{\epsilon} \mid Z) = \mathrm{Var}\left((I \otimes Z')\epsilon \mid Z\right) \equiv C_0 = [C_{jk}]_{j,k=1\dots,m}$$

and

$$C_{jk} = E[Z'\epsilon_j\epsilon_k' Z \mid Z].$$

Note, $\tilde{y}$ and $\tilde{\epsilon}$ are $lm \times 1$ vectors, $\tilde{X}$ is an $lm \times k$ matrix and β is an $k \times 1$ vector (where $k = \sum_{j=1}^{m} k_j$).

The covariance matrix C_0 can be estimated consistently by the Newey-West estimator (accounting for both serial correlation and heteroskedasticity), or by Eicker-White estimator (accounting only for heteroskedasticity). Then,

$$\hat{\beta}_{B3SLS} = \left(\tilde{X}'\hat{C}^{-1}\tilde{X}\right)^{-1}\tilde{X}'\hat{C}^{-1}\tilde{y}.$$

Consequently,

$$\hat{\beta}_{B3SLS} \overset{A}{\sim} N\left(,\beta, \left(\tilde{X}'\hat{C}^{-1}\tilde{X}\right)^{-1}\right).$$

Note, if $\text{Var}((\epsilon \mid Z) = \Sigma \otimes I$, then

$$\begin{aligned}\text{Var}(\tilde{\epsilon}) &= (I \otimes Z')\text{Var}(\epsilon \mid Z)(I \otimes Z) \\ &= (I \otimes Z')(\Sigma \otimes I)(I \otimes Z) \\ &= \Sigma \otimes Z'Z\end{aligned}$$

and

$$\sqrt{n}(\hat{\beta}_{B3SLS} - \hat{\beta}_{3SLS}) \xrightarrow{p} 0.$$

Otherwise, $\hat{\beta}_{B3SLS}$ is more efficient than $\hat{\beta}_{3SLS}$. Furthermore, the usual standard errors for 3SLS estimator are inconsistent.