Lecture 17: SEM Miscellaneous

Last time, we saw that the 2SLS estimator is given by $\hat{\delta} = (Z'\bar{M}Z)^{-1}Z'\bar{M}y_1$.

Note that the 2SLS method is a <u>limited information</u> method.

In this case, what are the sources of inefficiency?

1.   2SLS does not exploit error correlation across equations.

2.   2SLS does not impose restrictions on Π which is the matrix of reduced form coefficients.

Remember that

$$\hat{\delta} = \left[\begin{array}{cc} \hat{Y}_2'\hat{Y}_2 & \hat{Y}_2'X_1 \\ X_1'\hat{Y}_2 & X_1'X_1 \end{array} \right]^{-1} \left[\begin{array}{c} \hat{Y}_2'y_1 \\ X_1'y_1 \end{array} \right].$$

Note: $\hat{Y}_2 = X\hat{\Pi}_2 = \bar{M}Y_2$.

If we fit each equation by 2SLS, we get all the structural parameters.   These imply reduced form parameters Π, but the unrestricted $\hat{\Pi}$ is used in the first stage.

Single equation IV methods are limited information methods.   So is LIML (limited information maximum likelihood) which is maximum likelihood estimation of δ and Π_2 (other equations in unrestricted reduced form).
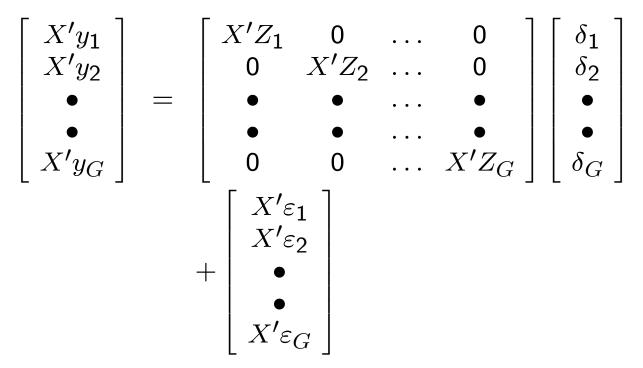
Three Stage Least Squares (3SLS):

Stack the G equations:

$$
\begin{bmatrix} y_1 \\ y_2 \\ \bullet \\ \bullet \\ y_G \end{bmatrix} = \begin{bmatrix} Z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & \ldots & 0 \\ \bullet & \bullet & \ldots & \bullet \\ \bullet & \bullet & \ldots & \bullet \\ 0 & 0 & \ldots & Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \bullet \\ \bullet \\ \delta_G \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \bullet \\ \bullet \\ \varepsilon_G \end{bmatrix}
$$

In compact notation, $y = Z\delta + \varepsilon$.

$V(\varepsilon) = \sum \otimes I$ where $V(\varepsilon)$ is $GN \times GN$, $\sum$ is $G \times G$ and I is $N \times N$.

Look at $(I \otimes X')y = (I \otimes X')Z\delta + (I \otimes X')\varepsilon$.

Written out:

$$
\begin{bmatrix} X'y_1 \\ X'y_2 \\ \bullet \\ \bullet \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 & \ldots & 0 \\ 0 & X'Z_2 & \ldots & 0 \\ \bullet & \bullet & \ldots & \bullet \\ \bullet & \bullet & \ldots & \bullet \\ 0 & 0 & \ldots & X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \bullet \\ \bullet \\ \delta_G \end{bmatrix}
$$
$$
+ \begin{bmatrix} X'\varepsilon_1 \\ X'\varepsilon_2 \\ \bullet \\ \bullet \\ X'\varepsilon_G \end{bmatrix}
$$

There are KG equations - G blocks that look like the equations giving 2 SLS.

Let's do GLS.

$$
\begin{aligned}
Var((I \otimes X')\varepsilon) &= (I \otimes X')V(\varepsilon)(I \otimes X')' \\
&= (I \otimes X')(\textstyle\sum \otimes I)(I \otimes X')' \\
&= (\textstyle\sum \otimes X'X)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\delta} &= [Z'(I \otimes X)(\textstyle\sum \otimes X'X)^{-1}(I \otimes X')Z]^{-1}Z'(I \otimes X) \\
&\quad (\textstyle\sum \otimes X'X)^{-1}(I \otimes X')y \\
&= (Z'(\textstyle\sum^{-1} \otimes \bar{M})Z)^{-1}Z'(\textstyle\sum^{-1} \otimes \bar{M})y
\end{aligned}
$$

Note that $\sum$ is unknown and must be estimated.

Let S be the matrix of mean squares and cross products of 2 SLS residuals. The ijth entry of this matrix, S_{ij}, is given by

$$
S_{ij} = (y_i - Z_i\hat{\delta}_i)'(y_j - Z_j\hat{\delta}_j)/N.
$$

Note that this is not

$$
(y_i - \hat{Z}_i\hat{\delta}_i)'(y_j - \hat{Z}_j\hat{\delta}_j).
$$

Proposition: $\hat{\delta} = (Z'(S^{-1} \otimes \bar{M})Z)^{-1}Z'(S^{-1} \otimes \bar{M})y$
is the <u>3SLS</u> estimator.

What are the stages of 3SLS estimation?

3SLS:

-     is more efficient than 2SLS
-     can spread effects of misspecification

This suggests comparison with 2SLS estimates as a specification check.

FIML (full information maximum likelihood) method is the joing ML estimation of all parameters.

What is the difference between FIML and 3SLS?

Lagged Endogenous Variables:

The classical approach:

	current	lagged
endogenous	y_1, Y_2	$Ly_1, LY_2(*)$
exogenous	$X(*)$	$LX(*)$

The starred terms are <u>predetermined variables</u> which can be used as instruments. Recall that lagged endogenous variables require IV estimation when there is autocorrelation. It is usually wise to treat lagged endogenous variables as endogenous.

Testing "Endogeneity"- Wu-Hausman Tests

$$y_1 = \gamma y_2 + \varepsilon_1$$

$$y_2 = \pi x + \varepsilon_2$$

$$E\varepsilon_1\varepsilon_2 = \sigma_{12}$$

OLS gives

$$
\begin{aligned}
\widehat{\gamma} &= \frac{\widehat{\sum} y_{1i} y_{2i}}{\sum y_{2i} y_{2i}} = \frac{\sum(\gamma(x\pi + \varepsilon_2) + \varepsilon_1) y_2}{\sum y_2^2} \\
&\rightarrow \gamma + \frac{n\sigma_{12}}{\left(\sum(x\pi)^2 + n\sigma_{22}\right)}
\end{aligned}
$$

The IV estimator is consistent. This suggests testing "endogeneity" by comparing the OLS and IV estimators. This works.

Returning to $y_1 = Z_1 \delta + \varepsilon_1$, with $Z_1 = [Y_2 \ X_1]$.

We have $\delta_{IV} = (Z_1' \bar{M} Z_1)^{-1} Z_1' \bar{M} y_1$ and

$$\delta_{OLS} = (Z_1' Z_1)^{-1} Z_1' y_1,$$

$$\begin{aligned} \delta_{IV} - \delta_{OLS} &= (Z_1' \bar{M} Z_1)^{-1} [Z_1' \bar{M} - Z_1' \bar{M} Z_1 (Z_1' Z_1)^{-1} Z_1'] y \\ &= (Z_1' \bar{M} Z_1)^{-1} Z_1' \bar{M} M_Z y_1. \end{aligned}$$

where $M_Z = I - Z_1 (Z_1' Z_1)^{-1} Z_1'$.

The first factor is irrelevant. Does the second have mean zero?

The second factor is $Z_1' \bar{M} M_Z y_1$.

This r.v. has some identically zero elements. (Why?) What are all these projections?

With some thought, the r.v. we are considering is proportional to the regression coefficient on $\hat{Y}_2$, in the regression of y_1 on Z_1 and $\hat{Y}_2$.

Thus endogeneity of Y_2 can be tested by running the OLS regression of y_1 on X_1, Y_2, and $\hat{Y}_2$, and testing the significance of $\hat{Y}_2$.

Intuition? Under the null, $\hat{Y}_2$, is irrelevant, since the systematic variation in Y_2 is picked up by Y_2 itself. But if Y_2 is endogenous, $\hat{Y}_2$ can pick up the systematic variation and Y_2 can pick up the effect of error correlation (recall the example).

The following example shows that the SEM tecnhiques, especially limited information techniques, are useful even when a full system is <u>not</u> specified.

Example: <u>Schooling and Earnings</u>

$$\ln y_i = X_i \beta + \delta S_i + u_i$$

Now suppose earnings are affected by ability A_i as well as by experience X_i and schooling S_i.

Then $u_i = \gamma A_i + \varepsilon_i$. If S and A are correlated the LS estimates of δ (and β) will be biased. (*Why*?)

Can we fit this equation by IV or 2 SLS? Using what instruments?

Suppose instead that we try to measure "ability" by a test score.

We have in mind that earnings and the test score are jointly endogenous, both determined by "ability".

$$\ln y_i = X_i\beta + \delta S_i + \gamma A_i + \varepsilon_i$$

$$T_i = \alpha_0 + \alpha_1 A_i + \eta_i$$

Think of this as a reduced form; but it is a little stronger since A is unobserved.

The equation we can fit is $\ln y_i = X_i\beta + \delta S_i + T_i\tilde{\gamma} + \tilde{\varepsilon}_i$ but T_i is endogenous.

- This suggests using IV.

- What instruments?

The result using family background variables as instruments is usually to reduce the schooling coefficient slightly.

Now: What about schooling itself? It is probably measured with error - are years of schooling homogeneous?

It is a <u>choice</u> variable, and thus should be determined within the model.

What if we consider S to be endogenous?

Then we fit $\ln y_i = X_i\beta + \delta S_i + \gamma T_i + \varepsilon_i$ by IV with S_i <u>and</u> T_i endogenous.

What are the instruments?

(This often raises $\hat{\delta}$ relative to coefficient with S_i exogenous and T_i endogenous.)

Point: IV is useful even without specifying a complete model.

Other issues:

1.   Identification and the size of the system?

Identification is easier for large systems.

2.   Estimation and the size of the system?

Estimation is harder for large systems.

3.   Are there really identified models?