# Economics 620, Lecture 17: SEM Miscellaneous 

Nicholas M. Kiefer

Cornell University

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

Note that the 2SLS method is a limited information method.
In this case, what are the sources of inefficiency?

1. 2 SLS does not exploit error correlation across equations.
2. 2SLS does not impose restrictions on $\Pi$ which is the matrix of reduced form coefficients.

Remember that

$$
\hat{\delta}=\left[\begin{array}{cc}
\hat{Y}_{2}^{\prime} \hat{Y}_{2} & \hat{Y}_{2}^{\prime} X_{1} \\
X_{1}^{\prime} \hat{Y}_{2} & X_{1}^{\prime} X_{1}
\end{array}\right]^{-1}\left[\begin{array}{c}
\hat{Y}_{2}^{\prime} y_{1} \\
X_{1}^{\prime} y_{1}
\end{array}\right]
$$

Note: $\quad \hat{Y}_{2}=X \hat{\Pi}_{2}=\bar{M} Y_{2}$.
If we fit each equation by $2 S L S$, we get all the structural parameters.
These imply reduced form parameters $\Pi$, 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 $\delta$ and $\Pi_{2}$ (other equations in unrestricted reduced form).

## Three Stage Least Squares (3SLS):

Stack the $G$ equations:

$$
\left[\begin{array}{c}
y_{1} \\
y_{2} \\
\bullet \\
\bullet \\
y_{G}
\end{array}\right]=\left[\begin{array}{cccc}
z_{1} & 0 & \ldots & 0 \\
0 & z_{2} & \ldots & 0 \\
\bullet & \bullet & \ldots & \bullet \\
\bullet & \bullet & \ldots & \bullet \\
0 & 0 & \ldots & z_{G}
\end{array}\right]\left[\begin{array}{c}
\delta_{1} \\
\delta_{2} \\
\bullet \\
\bullet \\
\delta_{G}
\end{array}\right]+\left[\begin{array}{c}
\varepsilon_{1} \\
\varepsilon_{2} \\
\bullet \\
\bullet \\
\varepsilon_{G}
\end{array}\right]
$$

In compact notation, $y=Z \delta+\varepsilon$.
$V(\varepsilon)=\sum \otimes I$ where $V(\varepsilon)$ is $G N \times G N, \sum$ is $G \times G$ and $I$ is $N \times N$.

Look at $\left(I \otimes X^{\prime}\right) y=\left(I \otimes X^{\prime}\right) Z \delta+\left(I \otimes X^{\prime}\right) \varepsilon$.
Written out:

$$
\begin{aligned}
{\left[\begin{array}{c}
X^{\prime} y_{1} \\
X^{\prime} y_{2} \\
\bullet \\
\bullet \\
X^{\prime} y_{G}
\end{array}\right]=} & {\left[\begin{array}{cccc}
X^{\prime} Z_{1} & 0 & \cdots & 0 \\
0 & X^{\prime} Z_{2} & \cdots & 0 \\
\bullet & \bullet & \cdots & \bullet \\
\bullet & \bullet & \cdots & \bullet \\
0 & 0 & \ldots & X^{\prime} Z_{G}
\end{array}\right]\left[\begin{array}{c}
\delta_{1} \\
\delta_{2} \\
\bullet \\
\bullet \\
\delta_{G}
\end{array}\right] } \\
& +\left[\begin{array}{c}
X^{\prime} \varepsilon_{1} \\
X^{\prime} \varepsilon_{2} \\
\bullet \\
\bullet \\
X^{\prime} \varepsilon_{G}
\end{array}\right]
\end{aligned}
$$

There are $K G$ equations - $G$ blocks that look like the equations giving 2 SLS.
Let's do GLS.

$$
\begin{aligned}
\operatorname{Var}\left(\left(I \otimes X^{\prime}\right) \varepsilon\right)= & \left(I \otimes X^{\prime}\right) V(\varepsilon)\left(I \otimes X^{\prime}\right)^{\prime} \\
& =\left(I \otimes X^{\prime}\right)\left(\sum \otimes I\right)\left(I \otimes X^{\prime}\right)^{\prime} \\
& =\left(\sum \otimes X^{\prime} X\right) \\
\hat{\delta}= & {\left[Z^{\prime}(I \otimes X)\left(\sum \otimes X^{\prime} X\right)^{-1}\left(I \otimes X^{\prime}\right) Z\right]^{-1} Z^{\prime}(I \otimes X) } \\
= & \left(\sum^{\prime} \otimes X^{\prime} X\right)^{-1}\left(I \otimes X^{\prime}\right) y
\end{aligned}
$$

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, $S_{i j}$, is given by

$$
S_{i j}=\left(y_{i}-z_{i} \hat{\delta}_{i}\right)^{\prime}\left(y_{j}-z_{j} \hat{\delta}_{j}\right) / N .
$$

Note that this is not

$$
\left(y_{i}-\hat{z}_{i} \hat{\delta}_{i}\right)^{\prime}\left(y_{j}-\hat{z}_{j} \hat{\delta}_{j}\right) .
$$

Proposition: $\hat{\delta}=\left(Z^{\prime}\left(S^{-1} \otimes \bar{M}\right) Z\right)^{-1} Z^{\prime}\left(S^{-1} \otimes \bar{M}\right) y$ is the 3SLS estimator.
What are the stages of 3SLS estimation?
3SLS:

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

This suggests comparison with 2 SLS 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 | $y_{1}, Y_{2}$ | $L y_{1}, L Y_{2}(*)$ |
| exogenous | $X(*)$ | $L X(*)$ |

The starred terms are predetermined variables 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

$$
\begin{aligned}
& y_{1}=\gamma y_{2}+\varepsilon_{1} \\
& y_{2}=\pi x+\varepsilon_{2} \\
& E \varepsilon_{1} \varepsilon_{2}=\sigma_{12}
\end{aligned}
$$

OLS gives

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

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 $Z_{1}=\left[\begin{array}{ll}Y_{2} & X_{1}\end{array}\right]$.
We have $\delta_{I V}=\left(Z_{1}^{\prime} \bar{M} Z_{1}\right)^{-1} Z_{1}^{\prime} \bar{M} y_{1}$ and

$$
\delta_{O L S}=\left(Z_{1}^{\prime} Z_{1}\right)^{-1} Z_{1}^{\prime} y_{1}
$$

$$
\begin{aligned}
\delta_{I V}-\delta_{O L S} & =\left(Z_{1}^{\prime} \bar{M} Z_{1}\right)^{-1}\left[Z_{1}^{\prime} \bar{M}-Z_{1}^{\prime} \bar{M} Z_{1}\left(Z_{1}^{\prime} Z_{1}\right)^{-1} Z_{1}^{\prime}\right] y_{1} \\
& =\left(Z_{1}^{\prime} \bar{M} Z_{1}\right)^{-1} Z_{1}^{\prime} \bar{M} M_{Z} y_{1} .
\end{aligned}
$$

where $M_{Z}=I-Z_{1}\left(Z_{1}^{\prime} Z_{1}\right)^{-1} Z_{1}^{\prime}$.
The first factor is irrelevant. Does the second have mean zero?

The second factor is $Z_{1}^{\prime} \bar{M} M_{z} y_{1}$.
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 not specified.

Example: Schooling and Earnings

$$
\ln y_{i}=X_{i} \beta+\delta S_{i}+u_{i}
$$

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 $\delta$ (and $\beta$ ) 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 choice 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 $S_{i}$ and $T_{i}$ 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?

