Economics 620, Lecture 16: Estimation of Simultaneous Equations Models

Nicholas M. Kiefer

Cornell University
Consider \( y_1 = Y_2\gamma + X_1\beta + \varepsilon_1 \) which is an equation from a system.

We can rewrite this at \( y_1 = Z\delta + \varepsilon_1 \) where \( Z = [Y_2 \ X_1] \) and \( \delta = [\gamma' \ \beta']' \).

Note that \( Y_2 \) is jointly determined with \( y_1 \), so

\[
\text{plim}(1/N)Z'\varepsilon_1 \neq 0 \ (\text{usually}).
\]

**IV Estimation:**

The point of IV estimation is to find a matrix of instruments \( W \) so that

\[
\text{plim} \frac{W'\varepsilon_1}{N} = 0
\]

and

\[
\text{plim} \frac{W'Z}{N} = Q
\]

where \( Q \) is nonsingular.
The IV estimator is \((W'Z)^{-1}W'y_1\). As in the lecture on dynamic models, multiplying the model by the transpose of the matrix of instruments yields 
\[ W'y_1 = W'Z\delta + W'\varepsilon_1 \]
which gives \(\hat{\delta}_{IV}\).

**Asymptotic distribution of \(\hat{\delta}_{IV}\):**

Note that 
\[ \hat{\delta}_{IV} - \delta = (W'Z)^{-1}W'\varepsilon_1. \]
Assume that 
\[ W'\varepsilon_1 \overset{\text{d}}{\rightarrow} N \left(0, \sigma^2 \frac{W'W}{N}\right). \]

*(Is this a sensible assumption? Recall the CLT.)*

Then 
\[ \sqrt{N} (\hat{\delta}_{IV} - \delta) \overset{\text{d}}{\rightarrow} N(0, \sigma^2 \sum_{\delta}) \]
where

$$\sum_\delta = N(W'Z)^{-1}W'W(W'Z)^{-1} = (1/N)Q^{-1}W'WQ^{-1}.$$

The question is what to use for $W$. Suppose we use $X$.

Multiplying by the transpose of the matrix of instruments gives

$$X'y_1 = X'Z\delta + X'\varepsilon_1.$$

For this system of equations to have a solution, $X'Z$ has to be square and nonsingular. When is this possible?

Note the following dimensions: $X$ is $N \times K$, $X_1$ is $N \times K_1$ and $Y_2$ is $N \times (G_1 - 1)$. This, of course, requires $K = K_1 + G_1 - 1$. *(Recall the order condition: $K \geq K_1 + G_1 - 1$).*
Thus, the above procedure works when the equation is just identified.

The resulting IV estimates are indirect least squares which we saw last time.

Suppose $K < K_1 + G_1 - 1$. Then what happens? Consider the supply and demand example. This is the underidentified case.

Suppose $K > K_1 + G_1 - 1$. Then $X' y_1 = X' Z \delta + X' \varepsilon_1$ is $K$ equations in $K_1 + G_1 - 1$ unknowns (setting $X' \varepsilon_1$ to 0 which is its expectation). We could choose $K_1 + G_1 - 1$ equations to solve for $\delta$- there are many ways to do this, typically leading to different estimates. This is the overidentified case.
Another way to look at this case is as a regression model - with $K$ “observations” on the dependent variable.

We could apply the LS method, but the GLS is more efficient since $V(X' \varepsilon_1) = \sigma^2(X'X)(\neq \sigma^2 I)$.

The observation matrix is $X'y_1$ and $X'Z$. GLS gives the estimator

$$\hat{\delta} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'y_1.$$ 

In the just-identified case (where $X'Z$ is invertible),

$$\hat{\delta} = (X'Z)^{-1}X'X(Z'X)^{-1}Z'X(X'X)^{-1}X'y_1$$

$$= (X'Z)^{-1}X'y_1 = \hat{\delta}_{IV} \text{ with } W = X.$$
Return to the overidentified case:

\[ \hat{\delta} = [Z'X(X'X)^{-1}X'Z]^{-1} Z'X(X'X)^{-1}X'y_1. \]

**Proposition**: The estimator

\[ \hat{\delta} = [Z'X(X'X)^{-1}X'Z]^{-1} Z'X(X'X)^{-1}X'y_1 \]

is the two-stage least squares (2SLS or TSLS) estimator.

Why is \( \hat{\delta} \) called the TSLS estimator?

Let \( \tilde{M} = X(X'X)^{-1}X' = I - M. \)
Then \( \hat{\delta} = (Z'\tilde{MZ})^{-1} Z'\tilde{M}y_1. \)
We will write out the expression for $\hat{\delta}$.

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

Now: \( \tilde{M} Y_2 = X (X' X)^{-1} X' Y_2 = \hat{Y}_2 = X \tilde{\Pi}_2 \) which is the LS predictor of \( Y_2 \).

\[
Z' \tilde{M} Z = \left[ \begin{array}{cc}
Y_2' \tilde{M} Y_2 & Y_2' \tilde{M} X_1 \\
X_1' \tilde{M} Y_2 & X_1' \tilde{M} X_1 
\end{array} \right]
\]

Note that \( X_1' \tilde{M} X_1 = X_1' X_1. (R[X_1] \subset R[X] \Rightarrow \tilde{M} X_1 = X_1; \tilde{M} X = X) \).

Also: \( Y_2' \tilde{M} Y_2 = Y_2' \tilde{M} \tilde{M} Y_2 = \hat{Y}_2' \hat{Y}_2 \).

So,

\[
\hat{\delta} = \left[ \begin{array}{cc}
\hat{Y}_2' \hat{Y}_2 & \hat{Y}_2' X_1 \\
X_1' \hat{Y}_2 & X_1' X_1 
\end{array} \right]^{-1} \left[ \begin{array}{c}
\hat{Y}_2' y_1 \\
X_1' y_1 
\end{array} \right].
\]
\( \hat{\delta} \) is the coefficient vector from a regression of \( y_1 \) on \( \hat{Y}_2 \) and \( X_1 \).

**Interpretation as 2SLS? Interpretation as IV?**

**Proposition:** 2SLS is IV estimation with \( W = [\hat{Y}_2 X_1] \).

**Proof:** Note that

\[
W'Z = \begin{bmatrix}
\hat{Y}_2' Y_2 & \hat{Y}_2' X_1 \\
X_1' \hat{Y}_2 & X_1' X_1
\end{bmatrix} = \begin{bmatrix}
\hat{Y}_2' \hat{Y}_2 & \hat{Y}_2' X_1 \\
X_1' \hat{Y}_2 & X_1' X_1
\end{bmatrix}.
\]

This is the matrix appearing inverted in \( \hat{\delta} \). ■
Asymptotic distribution of $\hat{\delta}$: We know this from IV results.

Note that $\hat{\delta} = \delta + (Z'\widetilde{M}Z)^{-1}Z'\widetilde{M}\varepsilon_1$. The asymptotic variance of $N^{1/2}(\hat{\delta} - \delta)$ is the asymptotic variance of $N^{1/2}(Z'\widetilde{M}Z)^{-1}Z'\widetilde{M}\varepsilon_1 = u$.

$$\text{Var}(u) = N\sigma^2(Z'\widetilde{M}Z)^{-1}Z'\widetilde{M}Z(Z'\widetilde{M}Z)^{-1}$$

$$= N\sigma^2(Z'\widetilde{M}Z)^{-1}.$$

Remember to remove the $N$ in calculating estimated variance for $\hat{\delta}$. (Why?)
Estimation of $\sigma^2$:

$$\hat{\sigma}^2 = (y_1 - Z\hat{\delta})'(y_1 - Z\hat{\delta})/N.$$ 

Note that $Z = [Y_2 X_1]$ appears in the expressions for $\hat{\sigma}^2$, not $[\hat{Y}_2 X_1]$. 

If you regress $y_1$ on $\hat{Y}_2$ and $X_1$, you will get the right coefficients but the wrong standard errors.
GEOMETRY OF 2SLS:

Take:
$N = 3$ (observations)
$K = 2$ (exogenous variables),
$K_1 = 1$ (included exogenous variables) and
$G_1 = 2$ (included endogenous variables - one is normalized).

How many parameters?
Plane in $\mathbb{R}^3$ spanned by $x_1$ and $x_2$

$x_1, y_1, \text{ etc. are 3 x 1 vectors}$
\( \hat{Y}_2 \) is in the plane spanned by \( X_1 \) and \( X_2 \). \( y_1 \) is projected to the plane spanned by \( \hat{Y}_2 \) and \( X_1 \).

Note that \( X_1 \) and \( X_2 \) and \( X_1 \) and \( \hat{Y}_2 \) span the same plane. \( (Why?) \)

Model is just identified (projection of both stages is to the same plane).

What happens if the model is overidentified? \( (\text{For example, } K_1 = 0, \text{ that is, no included regressors}) \).

What if underidentified? \( (\text{For example, } K_2 = 2, \text{ that is, no excluded regressors}) \).