LECTURE 16: ESTIMATION OF SIMULTANEOUS EQUATIONS MODELS

Consider \( y_1 = Y_2 \gamma + X_1 \beta \delta \epsilon + \epsilon_1 \) which is an equation from a system.

We can rewrite this as \( y_1 = Z \delta + \epsilon_1 \) where \( Z = [Y_2 \ X_1] \) and \( \delta = [\gamma \ \beta] \).

Note that \( Y_2 \) is jointly determined with \( y_1 \), so \( \text{plim} (1/N)Z' \epsilon_1 \neq 0 \) (usually).

IV ESTIMATION:

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

\[
p \lim \frac{W' \epsilon_1}{N} = 0
\]

and

\[
p \lim \frac{W'Z}{N} = Q
\]

where \( Q \) is nonsingular.
The IV estimator is \((WZ)^{-1}Wy_1\). As in the lecture on dynamic models, multiplying the model by the transpose of the matrix of instruments yields \(W'y_1 = W'Z + W'\varepsilon_1\) which gives \(\hat{\delta}_{IV}\)

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

Note that \(\hat{\delta}_{IV} - \delta = (WZ)^{-1}W\varepsilon_1\). Assume that

\[
\frac{W'\varepsilon_1}{\sqrt{N}} \to N\left(0, \frac{\sigma^2WW}{N}\right).
\]

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

Then

\[
\sqrt{N}(\hat{\delta}_{IV} - \delta) \to N(0, \sigma^2\Sigma_{\delta})
\]
where
\[ \Sigma_\delta = N(W'Z)^{-1}W'W(W'Z)^{-1} = (1/N) Q^{-1}W'WQ^{-1}. \]

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 \( NxK \), \( X_1 \) is \( NxK_1 \) and \( Y_2 \) is \( Nx(G_1 - 1) \). 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 zero 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 GLS is more efficient since $V(X'\varepsilon_1) = \sigma^2(X'X) (\neq \sigma^2I)$.

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),

$$\hat{\delta} = (X'Z)^{-1}X'X(Z'X)^{-1}Z'X(X'X)^{-1}X'y_1$$
$$= (X'Z)^{-1}X'y_1 = \hat{\delta}_{IV} \text{ with } W = X.$$
TWO-STAGE LEAST SQUARES:

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 \( \bar{M} = X(X'X)^{-1}X' = I - M. \)

Then \( \hat{\delta} = (Z\bar{M} Z)^{-1}Z\bar{M} y_1. \)
We will write out the expression for $\hat{\delta}$

$$\hat{\delta} = \begin{bmatrix} \hat{Y}_2 \hat{Y}_2 & \hat{Y}_2 X_1 \\ X_1' \hat{Y}_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2 y_1 \\ X_1' y_1 \end{bmatrix}. $$

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

$$Z'MZ = \begin{bmatrix} Y_2'MY_2 & Y_2'MX_1 \\ X_1'MY_2 & X_1'MX_1 \end{bmatrix}$$

Note that $X_1'MX_1 \supseteq X_1'X_1$. $(R[X_1] \subset R[X] \Rightarrow M X_1 = X_1; \ M X = X$.

Also: $Y_2'MY_2 = Y_2'MMY_2 = \hat{Y}_2 \hat{Y}_2$

So, $$\hat{\delta} = \begin{bmatrix} \hat{Y}_2 \hat{Y}_2 & \hat{Y}_2 X_1 \\ X_1' \hat{Y}_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2 y_1 \\ X_1' y_1 \end{bmatrix}. $$
\( \hat{\delta} \) 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.
Asymptotic distribution of $\hat{\delta}$:
We know this from IV results.

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

$$\text{Var}(u) = N\sigma^2(Z'MZ)^{-1}Z'M \left[ M'Z(Z'MZ)^{-1} \right] = N\sigma^2(Z'MZ)^{-1}.$$

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?

\( \hat{Y}_2 \) 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 \( \hat{Y}_2 \) 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? (For example, $K_1 = 0$, that is, no included regressors)

What if underidentified? (For example, $K_2 = 2$, that is, no excluded regressors)