Instrumental Variables

If we have

$$y = \beta x + u$$
$$Cov(x,u) \neq 0$$

then the OLS estimate

$$\hat{\beta} = \frac{\sum xy}{\sum x^2}$$

will be inconsistent.

Now if we had a variable z where $Cov(z,u) = 0$ then

$$\hat{\beta} = \frac{\sum zy}{\sum zx} = \frac{\sum z(\beta x + u)}{\sum zx}$$

Now $plim\,\hat{\beta} = \beta$

$$\hat{\beta} = \beta + \frac{\sum zu/N}{\sum zx/N}$$
$$= \beta + \frac{Cov(z,u)}{Cov(z,x)}$$

as $N \to \infty$

$$= \beta + 0$$

Note require $Cov(z,x) \neq 0$

Consider

$$y_1 = \alpha_{11}y_2 + \alpha_{12}z_1 + \alpha_{13}z_2 + u_1$$
$$y_2 = \alpha_{21}y_1 + \alpha_{22}z_3 + u_2$$

For the first equation $Cov(z_1,u_1) = 0, Cov(z_2,u_1) = 0$ but $Cov(y_2,u_1) \neq 0$
Have three coefficients to estimate
Have to find a variable independent of $u_1$and z_3 fits the bill as $Cov(z_3,u_1) = 0$
So z_3 can act as an instrumental variable for y_2

For the second $Cov(z_3,u_2) = 0, Cov(y_1,u_2) \neq 0$ but $Cov(y_2,u_1) \neq 0$
Have a choice of z_1 and z_2 which means we have an overidentified model by the order condition

Order rule $\Rightarrow$have we enough exogenous variables elsewhere in the system to use as instruments for the endogenous variables in the equation with unknown own coefficients
Underidentified $\Rightarrow$ not enough IVs
Overidentified $\Rightarrow$ too many IVs: So use a weighted average of the available IVs. Compute these to get most efficient estimators (minimum asymptotic variance)

The efficient IV variables can be shown to be constructed by regressing the endogenous variables on all the exogenous variables in the system i.e. estimating the reduced form equations. So for this example

$$y_1 = \alpha_{11} y_2 + \alpha_{12} z_1 + \alpha_{13} z_2 + u_1$$
$$y_2 = \alpha_{21} y_1 + \alpha_{22} z_3 + u_2$$

which implies estimating the following reduced form by OLS

$$\hat{y}_1 = \gamma_{11} z_1 + \gamma_{12} z_2 + \gamma_{13} z_3 + u_1$$
$$\hat{y}_2 = \gamma_{21} z_1 + \gamma_{22} z_2 + \gamma_{23} z_3 + u_2$$

then in the first equation use $\hat{y}_2, z_1, z_2$ as instrumental variables. NB this is exactly the same as using z_1, z_2, z_3 as instruments because exactly identified so no choice with instruments

For second equation choice is between z_1, z_2 but using $\hat{y}_1$gives the optimum weighting, the optimal weight for z_1, z_2 are γ_{11}, γ_{12}

Using this approach $\Rightarrow$ problems with R^2, which will be lower than OLS and can be negative, though this would imply misspecification

Note we can have a very general view of what are instruments

using levels as endogenous

using lagged values as exogenous -even when correlated

Problem of arbitrary nature of classification

Two Stage Least Squares

Differs from IV in that $\hat{y}$s are used as regressors rather than as instruments.

Consider

$$y_1 = \alpha_{11} y_2 + \alpha_{12} z_1 + u_1$$

with other exogenous variables $z_{2,} z_3, z_4$

Estimate

$$y_2 = \gamma_{11} z_1 + \gamma_{12} z_2 + \gamma_{13} z_3 + \gamma_{14} z_4 + v_1$$

implies

$$y_2 = \hat{y}_2 + v_1$$

where the error is uncorrelated with the exogenous regressors

The efficient IV method is to replace $y_2$with $\hat{y}_2$

Now if we estimate

$$y_1 = \alpha_{11} \hat{y}_2 + \alpha_{12} z_1 + u_1$$

Then we have our TSLS estimates

NB we could replace all of the endogenous variables and not just the right hand side ones by their predicted value. That is estimate

$$\hat{y}_1 = \alpha_{11}\hat{y}_2 + ...$$

which only affects the standard errors
This method is general, can use for much more complex models

Standard errors need correcting and this is normally done by the programme
What we want is

$$\sqrt{\frac{\sigma_u^2}{\sum \hat{y}_t^2}}$$

from

$$y_1 = \beta y_2 + u$$

but as do

$$y_1 = \beta \hat{y}_2 + w$$

where

$$\hat{y}_2 = y_2 - v_2$$

from reduced form

$$y_2 = \hat{y}_2 + v_2$$

so the error is not u but $(-\beta v_2 + u_2)$ in the second stage so we have to correct.

Limited Information Maximum Likelihood

Another method you might come across is LIML -again this is limited to a single equation

rather than the whole sample

$$y_1 = \alpha_{11}y_2 + \alpha_{12}z_1 + \alpha_{12}z_2 + u_1$$

rewrite as

$$y^* = y_1 - \alpha_{11}y_2 = \alpha_{12}z_1 + \alpha_{13}z_2 + u_1$$

Take

$$y^* = \gamma_{12}z_1 + \gamma_{13}z_2 \quad \Rightarrow \quad RSS1$$
$$y^* = \delta_{12}z_1 + \delta_{12}z_2 + \delta_{13}z_3 \quad \Rightarrow \quad RSS2$$

Using all the exogenous variables in the system for the second equation. Adding z_3 should have little effect if the original equation was OK.

Now choose α_{11} such that $RSS1/RSS2$ is minimised (NB can show 2SLS minimises the difference)

After α_{11} is determined estimate

$$y^* = \alpha_{12}z_1 + \alpha_{13}z_2$$

to get estimates of α_{12} and α_{13}
Note: invariant to normalisation
if equation is exactly identified TSLS/LIML give the same estimates
asymptotically variances and covariances are same by both methods but not the standard errors
NB LIML uses variances and covariances among the endogenous variable TSLS does not