Computing

Always use proper IV regression software AND DO NOT do 2SLS in two stages.

STATA

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

$$X = [\underset{\sim}{Y}\ \underset{\sim}{Z}]$$

Y is endogenous. k_2 $Y_1, \ldots Y_{k_2}$

Z is exog. a predetermined k_1 $Z_1, \ldots, Z_{k_1}$

$\underset{\sim}{W}$ are instruments. $\ell \geq k_2$

ivregress 2SLS y $(Y_1\ Y_2 = w_1\ w_2 \ldots w_\ell)$ $z_1\ z_2 \ldots z_{k_1}$ $r \ldots Y_{k_2}$

you can use, VCE(robust) if desired

estat firststage, all

This computes the test for weak instruments.

estat# overid

This computes the Hausman test.
(Regression version)

To compute the "classic" version

regress y $Y_1 Y_2 \cdots Y_{k2}$ $z_1 \cdots z_{k1}$

estimates store LS

ivregress 2sls y $(Y_1 \cdots Y_{k2} = w_1 \cdots w_n)$ $z_1 \cdots z_k$

estimates store IV

Hausman IV LS, constant sigmamore.

constant includes constant = 1 in computation

Sigmamore uses LS estimate of σ^2 in both Est Cov Matricies

Small Sample Properties

In LARGE Samples

$\hat{\beta}_{IV}$ is consistent and asymptotically distributed as long as the instruments are exogenous and strong enough.

In small samples, the $\hat{\beta}_{IV}$ has no mean when $k = \ell$

When $\ell > k$, it is BIASED.
(often substantially)

This happens because in finite samples it is
W will be correlated with u.
(But not asymptotically)

Other situations when $\hat{\beta}_{IV}$ Performs poorly in small samples

- l is ~~large~~ LARGE relative to k
- 1 or more of the instruments is weak

Overidentification can make bias better or worse depending on the degree and strength of the surplus instruments.