M1 INTERMEDIATE ECONOMETRICS

Instrumental-variable estimation

Koen Jochmans   François Poinas

2025 — 2026

Problem statement

Interested in

$$Y = X'\beta + e, \qquad \mathbb{E}(Xe) \neq 0.$$

So, β cannot be recovered via a least-squares procedure.

We have instrumental variables Z that satisfy two conditions:

C1 $\mathbb{E}(Ze) = 0$ (exclusion),

C2 $\operatorname{rank} \mathbb{E}(ZX') = k$ (relevance).

By C1 (Exclusion), $\mathbb{E}(Ze) = 0$, which together with the model means that

$$\mathbb{E}(Z(Y - X'\beta)) = 0.$$

Furthermore,

$$\mathbb{E}(ZX')\,\beta = \mathbb{E}(ZY).$$

If $\dim Z < \dim X$ we have less equations than unknown and cannot learn β; we say it is underidentified.

If $\dim Z = \dim X$, because of C2 (Relevance) $\mathbb{E}(ZX')$ is invertible and so

$$\beta = \mathbb{E}(ZX')^{-1}\mathbb{E}(ZY)$$

is (just) identified.

If $\dim Z > \dim X$ we have more equations than unknown parameters.

The matrix $\mathbb{E}(ZX')$ is rectangular but, for any $\dim Z \times \dim X$ matrix A with full column rank the matrix

$$A' \, \mathbb{E}(ZX')$$

is $\dim X \times \dim X$ and invertible.

So,

$$A' \, \mathbb{E}(Z(Y - X'\beta)) = 0$$

are $\dim X$ equations that can be solved to find

$$\beta = \mathbb{E}(A'ZX')^{-1}\mathbb{E}(A'ZY).$$

This is the overidentified case.

Instrumental-variable estimator

Given a random sample on (Y, X, Z) the implied estimator takes the form

$$\hat{\beta}_A = \left(\sum_{i=1}^{n} A' Z_i X_i' \right)^{-1} \left(\sum_{i=1}^{n} A' Z_i Y_i \right).$$

Think about A as constructing 'new' instrumental variables as linear combinations of Z_i.

Any set of linearly-independent linear combinations 'works'! But some are 'better' than others.

Here, we are finding $\hat{\beta}_A$ by setting to zero

$$\sum_{i=1}^{n} (A' Z_i)\, \hat{e}_i = 0, \qquad \hat{e}_i = Y_i - X_i' \hat{\beta}_A.$$

Note that we do not have that $\sum_{i=1}^{n} Z_i \hat{e}_i = 0$ in the overidentified case!

Write everything in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$$

with instrument matrix $\boldsymbol{Z}$.

An alternative interpretation comes from minimizing

$$(\boldsymbol{Y} - \boldsymbol{X}b)'\boldsymbol{Z}W\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b) = \|\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b)\|_W^2$$

with respect to b.

Here, W is a chosen weight matrix that defines the distance being minimized. It could for example be set to equal the identity matrix, yielding Euclidean distance.

Thus, we obtain our estimator by forming residuals $\hat{\boldsymbol{e}}$ such that the 'length' $\|\boldsymbol{Z}'\hat{\boldsymbol{e}}\|_W$ is as small as possible.

The connection between the matrices A and W follows from inspection of the first-order condition for the quadratic form:

$$(\boldsymbol{X}'\boldsymbol{Z})W\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}b) = 0.$$

Thus,

$$A = W\boldsymbol{Z}'\boldsymbol{X}.$$

Can then write

$$\hat{\beta}_A = (\boldsymbol{X}'\boldsymbol{Z}W\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}W\boldsymbol{Z}'\boldsymbol{Y}).$$

The question of constructing the ‘optimal’ instrument is thus a question of finding the optimal weight matrix.

Optimal weight matrix

It can be shown that the optimal choice is

$$W = \text{var}(Ze)^{-1} = \mathbb{E}(ZZ'e^2)^{-1} = \mathit{\Omega}^{-1},$$

in that it makes the asymptotic variance of the estimator as small as possible.

In practice this requires a two-step procedure, where we first compute

$$\tilde{e}_i = Y_i - X_i'\tilde{\beta},$$

where $\tilde{\beta}$ uses some chosen (non-optimal) W, to construct the estimator

$$\hat{\mathit{\Omega}} = n^{-1}\sum_{i=1}^{n} Z_i Z_i' \tilde{e}_i^2$$

and then calculate $(\boldsymbol{X}'\boldsymbol{Z}\hat{\mathit{\Omega}}^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}\hat{\mathit{\Omega}}^{-1}\boldsymbol{Z}'\boldsymbol{Y})$.

Two-stage least squares

If

$$\mathbb{E}(e^2|Z = z) = \sigma^2$$

then

$$\Omega = \mathbb{E}(ZZ'e^2) = \sigma^2\, \mathbb{E}(ZZ')$$

so the optimal (infeasible) instrumental-variable estimator becomes

$$(\boldsymbol{X}'\boldsymbol{Z}\,\mathbb{E}(ZZ')^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}\,\mathbb{E}(ZZ')^{-1}\boldsymbol{Z}'\boldsymbol{Y}).$$

An estimator of $\mathbb{E}(ZZ')$ is $\boldsymbol{Z}'\boldsymbol{Z}/n$ and so the feasible version of the optimal estimator is

$$\hat{\beta}_{2SLS} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}).$$

Note that $\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}' = \boldsymbol{P_Z}$ so

$$\hat{\beta}_{2SLS} = (\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{P_Z}\boldsymbol{Y}),$$

explaining the estimator's name.

Alternatively,

$$A = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$$

so the optimal instrument is

$$A'Z_i = (\boldsymbol{X}'\boldsymbol{Z})(\boldsymbol{Z}'\boldsymbol{Z})^{-1}Z_i.$$

These are fitted values of a regression of $\boldsymbol{X}$ on $\boldsymbol{Z}$.

So the optimal linear combination here is the regression line.

Assumptions:

C3. Random sampling

C4. $\mathbb{E}(|Y|^4) < +\infty$, $\mathbb{E}(\|X\|^4) < +\infty$, and $\mathbb{E}(\|Z\|^4) < +\infty$.

C5. $\Omega = \mathbb{E}(ZZ'e^2)$ is positive definite.

Consistency

Write

$$\hat{\beta} - \beta = \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n} W \frac{\boldsymbol{Z}'\boldsymbol{X}}{n} \right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n} W \frac{\boldsymbol{Z}'\boldsymbol{e}}{n} \right).$$

Here,

$$\frac{\boldsymbol{X}'\boldsymbol{Z}}{n} \underset{p}{\to} \mathbb{E}(XZ'), \qquad \frac{\boldsymbol{Z}'\boldsymbol{e}}{n} \underset{p}{\to} \mathbb{E}(Ze) = 0,$$

by law of large numbers.

The inverse is well-defined because of our rank conditions, so the continuous-mapping theorem gives

$$\hat{\beta} \underset{p}{\to} \beta,$$

so that we have consistency.

Can replace W by some $\hat{W}$ for which $\hat{W} \underset{p}{\to} W$.

Again,

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n} W \frac{\boldsymbol{Z}'\boldsymbol{X}}{n} \right)^{-1} \frac{\boldsymbol{X}'\boldsymbol{Z}}{n} W \right) \frac{\boldsymbol{Z}'\boldsymbol{e}}{\sqrt{n}}.$$

Here, the first right-hand side term converges in probability to

$$Q = (\mathbb{E}(XZ') \, W \, \mathbb{E}(ZX'))^{-1} \, \mathbb{E}(XZ') \, W,$$

from before.

For the second term,

$$\frac{\boldsymbol{Z}'\boldsymbol{e}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i e_i \underset{d}{\rightarrow} N(0, \Omega).$$

Hence,

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\rightarrow} N(0, Q\Omega Q')$$

Written-out in full, $Q\Omega Q'$ is

$$(\mathbb{E}(XZ')\,W\,\mathbb{E}(ZX'))^{-1}\,\mathbb{E}(XZ')\,W\,\Omega W\mathbb{E}(ZX')(\mathbb{E}(XZ')\,W\,\mathbb{E}(ZX'))^{-1}.$$

Inference based on this is fine, but not efficient.

If we use $\hat{W} = \hat{\Omega}^{-1}$ for

$$\hat{\Omega} \underset{p}{\rightarrow} \Omega$$

then $Q\Omega Q'$ simplifies to

$$(\mathbb{E}(XZ')\,\Omega^{-1}\,\mathbb{E}(ZX'))^{-1}.$$

Inference

Because

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\rightarrow} N(0, V_\beta)$$

for $V_\beta = Q\Omega Q'$, we can test

$$\mathbb{H}_0 : \theta = r(\beta) = \theta_0$$

in the same way as before.

For $\hat{V}_\theta = V_\theta = R'(\hat{\beta})\hat{V}_\beta R(\hat{\beta}) \underset{p}{\rightarrow} V_\theta$, under the null

$$W = n\,(\hat{\theta} - \theta_0)'\hat{V}_\theta^{-1}(\hat{\theta} - \theta_0) \underset{d}{\rightarrow} \chi^2_m,$$

where m is the number of restrictions being tested (i.e., the dimension of r).

J-test

Recall that optimally-weighted (and normalized) objective function was

$$\frac{(\boldsymbol{Y}-\boldsymbol{X}b)'\boldsymbol{Z}\hat{\Omega}^{-1}\boldsymbol{Z}'(\boldsymbol{Y}-\boldsymbol{X}b)}{n}.$$

At $b=\beta$ we have $\boldsymbol{Y}-\boldsymbol{X}\beta=\boldsymbol{e}$ and so

$$\frac{\boldsymbol{e}'\boldsymbol{Z}\hat{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{e}}{n}=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_ie\right)'\hat{\Omega}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_ie\right)\underset{d}{\rightarrow}\chi^2_{\dim(Z)}$$

when the exclusion restrictions in C1, that is, $\mathbb{E}(Ze)=0$ hold for all instruments.

Under homoskedasticity, with 2SLS we have $\hat{\Omega}=\hat{\sigma}^2\boldsymbol{Z}'\boldsymbol{Z}/n$ and so we obtain

$$\frac{\boldsymbol{e}'\boldsymbol{Z}\hat{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{e}}{n}=\frac{\boldsymbol{e}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{e}}{\hat{\sigma}^2}=\frac{\boldsymbol{e}'\boldsymbol{P_Z}\boldsymbol{e}}{\hat{\sigma}^2}=n\,R^2$$

for $\hat{\sigma}^2=\boldsymbol{e}'\boldsymbol{e}/n$.

Sounds too good to be true, which it also is.

We do not know β (or, equivalently $\boldsymbol{e}$) and so need to replace this by an estimator $\hat{\beta}$.

Using the optimally-weighted estimator for $\hat{\beta}$ we obtain

$$\frac{\hat{\boldsymbol{e}}'\boldsymbol{Z}\hat{\Omega}^{-1}\boldsymbol{Z}'\hat{\boldsymbol{e}}}{n} \underset{d}{\rightarrow} \chi^2_{\dim(Z)-\dim(X)}.$$

With 2SLS this corresponds to a regression of $\hat{e}_i$ on Z_i.

When $\dim(Z) = \dim(X)$ (the just-identified case) this yields an exact fit of the moments: Cannot test anything!

Intuition is that we need $\dim(X)$ moments to be able to estimate $\hat{\beta}$.

So need overidentification.

Can select the moments we 'trust' by using an incremental J-test. This channels power.