ECON 61001: *Lecture 8*

Alastair R. Hall

The University of Manchester

Instrumental Variables estimation

- Instrument relevance
- Construction of estimator

- Large sample properties and inference
- Weak instruments
- 2SLS

- Empirical application

Relevance condition

Last time discussed how IV estimation is based on the information about the parameter vector in the population moment condition:

$$E[z_i u_i(\beta_0)] = 0 \qquad (\sharp)$$

As we will see, for estimation to be "successful" also need $(\sharp)$ to represent unique information about β_0 that is,

$$E[z_i u_i(\beta)] \neq 0 \qquad \text{for all } \beta \neq \beta_0.$$

This condition is equivalent to

$$rank\left\{E[z_i x_i']\right\} = k$$

This is known as the identification condition for β_0.

Three key conditions:

- $E[z_i u_i(\beta_0)] = 0$
  - known as the orthogonality condition

- $rank\{E[z_i x_i']\} = k$
  - known as relevance condition ⇒ instruments are "sufficiently related" to regressors.

- $rank\{E[z_i z_i']\} = q$, where z_i is $q \times 1$.
  - uniqueness condition ⇒ each moment condition provides some unique information.

Relationship between q and k

- $q < k$: not enough information $\longrightarrow \beta_0$ is under-identified.

- $q = k$: same $\#$ of pieces of information as unknowns $\longrightarrow \beta_0$ is just-identified.

- $q > k$: more pieces of information as unknowns $\longrightarrow \beta_0$ is over-identified.

Just-identified case: $q = k$

In this case can apply MoM principle and IV estimator $\hat{\beta}_{IV}$ is defined as solution to sample moment conditions that is,

$$N^{-1}\sum_{i=1}^{N} z_i u_i(\hat{\beta}_{IV}) \;=\; N^{-1}Z'u(\hat{\beta}_{IV}) \;=\; 0$$

where Z is $N \times q$ matrix with i^{th} row z_i'.

And so:

$$\hat{\beta}_{IV} \;=\; (Z'X)^{-1}Z'y$$

Over-identified case: $q > k$

In this case, MoM does not work as have more equations than unknowns.

Instead, define $\hat{\beta}_{IV}$ as value of β that is closest to solving sample moment conditions.

How do we measure how far sample moment function is from zero?

Answer via

$$Q_{IV}(\beta) = u(\beta)'Z(Z'Z)^{-1}Z'u(\beta),$$

where (we have assumed) $rank(Z) = q$ and so $(Z'Z)^{-1}$ is p.d. $\Rightarrow$ $Q_{IV}(\beta)$ satisfies

- $Q_{IV}(\beta) \geq 0$ for all β
- $Q_{IV}(\beta) = 0$ iff $Z'u(\beta) = 0$.

Define IV estimator to be

$$\hat{\beta}_{IV} = argmin_{\beta \in \mathcal{B}} Q_{IV}(\beta)$$

$$\rightarrow \quad \hat{\beta}_{IV} = \left\{X'Z(Z'Z)^{-1}Z'X\right\}^{-1} X'Z(Z'Z)^{-1}Z'y$$

(See Tutorial 8 Question 2)

As examples illustrate IV can be applied in cross-section or time series data.

Here we concentrate on IV in cross-section data. For times series case see Lecture Notes.

As with OLS need to impose certain assumptions. So start with those.

- *CS1-IV:* $y_i = x_i'\beta_0 + u_i$
- *CS2-IV:* $\{\ (u_i, x_i', z_i'),\ \ i = 1, 2, \ldots N\}$ forms an independent and identically distributed sequence.
- *CS3-IV:* (i) $E[z_i z_i'] = Q_{zz}$, finite, p.d.; (ii) $E[z_i x_i'] = Q_{zx}$, $rank\{Q_{zx}\} = k$.
- *CS4-IV:* $E[u_i|z_i] = 0$
- *CS5-IV:* $Var[u_i|z_i] = h(z_i) > 0$.

Notice:

- *CS4-IV:* $\Rightarrow E[z_i u_i] = 0$ (via LIE), the orthogonality condition.
- *CS3-IV(ii):* is relevance condition; *CS3-IV(i):* is uniqueness condition;.
- it is now properties of u_i conditional on z_i that matter.

Consider just-identified case ($q = k$) - over-identified case ($q > k$) in Lecture Notes.

We have

$$\hat{\beta}_{IV} - \beta_0 = \left(N^{-1} \sum_{i=1}^{N} z_i x_i' \right)^{-1} N^{-1} \sum_{i=1}^{N} z_i u_i,$$

and using the WLLN, we have

$$N^{-1} \sum_{i=1}^{N} z_i x_i' \xrightarrow{p} E[z_i x_i'] = Q_{zx},$$

$$N^{-1} \sum_{i=1}^{N} z_i u_i \xrightarrow{p} E[z_i u_i] = 0.$$

So using Slutsky's Theorem: $\hat{\beta}_{IV} \xrightarrow{p} \beta_0 + Q_{zx}^{-1} \times 0 = \beta_0$.

$$N^{1/2}(\hat{\beta}_{IV} - \beta_0) = \left(N^{-1} \sum_{i=1}^{N} z_i x_i' \right)^{-1} N^{-1/2} \sum_{i=1}^{N} z_i u_i$$

As in Lecture 6: CLT $\rightarrow$

$$N^{-1/2} \sum_{i=1}^{N} z_i u_i \xrightarrow{d} N(0, \Omega), \text{ where } \Omega = lim_{N \to \infty} \Omega_N$$

and (with slight abuse of notation)

$$\Omega_N = Var\left[N^{-1/2} \sum_{i=1}^{N} z_i u_i \right]$$

What is Ω here?

Assumption *CS2-IV* $\Rightarrow \{z_i u_i;\ i = 1, 2, \ldots N\}$ are i.i.d. and so $Cov[z_i u_i, z_j u_j] = 0$ $(i \neq j)$.

$\Rightarrow \Omega_N = Var[z_i u_i]$.

Using $E[z_i u_i] = 0$

$Var[z_i u_i] = E[u_i^2 z_i z_i'] = E\left[E[u_i^2 | z_i] z_i z_i' \right] = E[h(z_i) z_i z_i'] = \Omega_h,$ say.

Therefore, under our assumptions, we have:

$$N^{-1/2} \sum_{i=1}^{N} z_i u_i \xrightarrow{d} N(0, \Omega_h).$$

Under Assumptions *CS1-IV-CS4-IV* and *CS5-IV*:

$$N^{1/2}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} N(0, V_{IV}).$$

where $V_{IV} = Q_{zx}^{-1}\Omega_h(Q_{zx}^{-1})'$.

To use this result as basis for inference, need a consistent estimator of V_{IV} and so Ω_h.

Can adapt ideas from discussion of OLS, and show that:

$$\hat{\Omega}_h = N^{-1} \sum_{i=1}^{N} e_i^2 z_i z_i' \xrightarrow{p} \Omega_h,$$

where have (re-)defined $e_i = y_i - x_i'\hat{\beta}_{IV}$.

Set $\hat{Q}_{zx} = N^{-1}Z'X$ then

$$\hat{V}_{IV} = \hat{Q}_{zx}^{-1}\hat{\Omega}_h(\hat{Q}_{zx}^{-1})' \xrightarrow{p} V_{IV}.$$

Can then perform inference using same techniques as in Lecture 6 provided we use modified variance estimator.

For example, an approximate $100(1-\alpha)\%$ confidence interval for $\beta_{0,\ell}$ is given by,

$$\left(\hat{\beta}_{IV,\ell} \pm z_{1-\alpha/2}\sqrt{\hat{V}_{IV,\ell,\ell}/N}\right).$$

Recall that large sample is used as an approximation to the finite sampling distribution of test statistics.

It has been realized that if relevance condition holds but almost fails then large sample distribution theory derived above can be a very poor approximation even in very large samples.

In this scenario, instruments are said to be weak.

Example: Angrist & Krueger (1991) study of returns to education (example 2 above)

- used Q_i, quarter of birth, as instrument.
- but ed_i is only very weakly related to Q_i.

Often IV viewed through the lens of simultaneous equations model. We explore this in case where only one endogenous regressor.

$$
\begin{aligned}
y_{1,i} &= z_{1,i}'\gamma_0 + \alpha_0 y_{2,i} + u_{1,i} \\
y_{2,i} &= z_{1,i}'\delta_{1,0} + z_{2,i}'\delta_{2,0} + u_{2,i}
\end{aligned}
$$

where

- $$\begin{bmatrix} u_{1,i} \\ u_{2,i} \end{bmatrix} \sim iid\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{bmatrix} \right), \quad \sigma_{1,2} \neq 0$$

- $E[z_{\ell,i} u_{j,i}] = 0$ for $\ell, j = 1, 2$.

Equation of interest:

$$y_{1,i} = z_{1,i}'\gamma_0 + \alpha_0 y_{2,i} + u_{1,i} = x_i'\beta_0 + u_{1,i}$$

Estimate β_0 via IV based on:

$$E[z_i u_{1,i}(\beta_0)] = 0$$

where

- $z_i = [z_{1,i}', z_{2,i}']'$
- $u_{1,i}(\beta) = y_{1,i} - x_i'\beta$.

Note:

- $z_{1,i}$ are instruments for themselves and $z_{2,i}$ are instruments for $y_{2,i}$

- $z_{2,i}$ does not appear on rhs of equation of interest.

IV can be implemented via a two-step procedure:

- regress $y_{2,i}$ on z_i (via OLS) and obtain $\hat{y}_{2,i}$.
- regress $y_{1,i}$ on $(z_{1,i}', \hat{y}_{2,i})$ (via OLS) $\rightarrow \tilde{\beta}_N$.

$\tilde{\beta}_N$ is known as Two Stage Least squares (2SLS) estimator of β_0.

It can be shown that $\tilde{\beta}_N = \hat{\beta}_{IV}$.

In this context, relevance relates to relationship between $y_{2,i}$ and $z_{2,i}$ controlling for $z_{1,i}$. Can be assessed by looking at the first stage regression

$$y_{2,i} \;=\; z_{1,i}'\delta_{1,0} \;+\; z_{2,i}'\delta_{2,0} \;+\; u_{2,i}$$

Instruments are relevant if $\delta_{2,0} \neq 0$.

This can be tested using F test described in Lectures 3/4:
$H_0 : \delta_{2,0} = 0$ vs $H_A : \delta_{2,0} \neq 0$.

- $H_0 \Rightarrow$ instruments do not satisfy relevance condition.
- $H_A \Rightarrow$ relevance condition satisfied.

Acemoglu *et al* (2001):

$$ln[y_i] = \mu_0 + r_i\alpha_0 + w_i\gamma_0 + u_i$$

where

- y_i is income per capita in developing country i
- r_i is quality of institutions in developing country i
- w_i is life expectancy

r_i is likely correlated with u_i due to reverse causality.

Base estimation on the population moment condition:

$$E[z_i \mathbf{u}_i] = 0$$

where $z_i = [1, w_i, z_{1,i}, z_{2,i}, z_{3,i}, z_{4,i}]'$ and

- $z_{1,i}$ is log settler mortality of country i,
- $z_{2,i}$ is the absolute latitude of country i,
- $z_{3,i}$ is the mean temperature of country i,
- $z_{4,i}$ is the proportion of land area within 100km of the seacoast.

Example: Economic development and institutions

	$\hat{\alpha}$	95% CI for α_0	$\hat{\gamma}$	95% CI for γ_0
OLS	0.287	(0.186,0.387)	0.0496	(0.036,0.063)
2*SLS*	0.744	(0.335,1.153)	0.016	(-0.018,0.051)

- point estimate of α_0 is higher with 2SLS
- s.e.'s are larger for 2SLS than OLS

Are instruments relevant? Assess this using first stage regression:

$$r = \delta_0 + \delta_1 * w + \delta_2 * z_1 + \delta_3 * z_2 + \delta_4 * z_3 + \delta_5 * z_4 + \text{“error”}$$

Test:

- $H_0 : \delta_2 = 0, \delta_3 = 0, \delta_4 = 0, \delta_5 = 0 \Rightarrow$ instruments not relevant
- $H_A : \delta_i \neq 0$ for at least one $i = 2, 3, 4, 5 \Rightarrow$ instruments relevant

$F = 2.27$ with p-value of 0.0740 $\Rightarrow$ only marginal evidence in support of instrument relevance (may be in weak instrument territory)

Further reading

- Notes: Chap 5.
- Greene:

  - 8.1 (models with endogenous regressors)
  - 8.2 (assumptions)
  - 8.3.1 (OLS)
  - 8.3.2 (IV)