15 Irrelevant and Weak Instruments

15.1 Irrelevant Instruments

Occurs when $E(w_i x_i')$ is not of full column rank, violating IV1. In this case, parameter β is not identified. Consider:

$$y_i = x_i\beta + \varepsilon_i$$

$$x_i - w_i\gamma + e_i$$

with one-dimernsional endogenous variable x_i, and instrument w_i. Satisfying $E(w_i\varepsilon_i) = 0$, but fails the relevance assumption, so that $\gamma = 0$ and hence $E(w_i x_i) = 0$.
The system of equations (moment conditions):

$$E(w_i\varepsilon_i) = 0$$

$$E(w_i x_i) = 0$$

is equivalent to

$$\begin{cases} E(w_i(y_i - x_i\beta)) = 0 \\ E(w_i x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} E(w_i y_i) \\ E(w_i x_i) \end{cases}$$

which tells us nothing about β and so it is not identified.
We can still compute IV estimator as unlikely $W'X$ exactly zero in a finite sample:

Proposition 15.1.1. Under non-identifiability:

- $\hat{\beta}_{IV}$ does not converge in probability to a limit. Instead it converges in distribution to a RV. In particular, $\hat{\beta}_{IV}$ is not consistent.
- The limiting distribution of $\hat{\beta}_{IV}$ is not centered at β but instead has its median at $\beta + \rho$, like β_{OLS}.
- $\hat{\beta}_{IV}$ will have very wild fluctuations in finite samples, since the ratio ξ_0/ξ_2 is Cauchy distributed and has no moments, due its fat tails.

Proof. For simplicity we assume homoskedasticity and suppose:

$$E(e_i|w_i) = E(\varepsilon_i|w_i) = 0$$

$$Var(e_i|w_i) = Var(\varepsilon_i|w_i) = 1, Cov(e_i, \varepsilon_i|w_i) = \rho \neq 0$$

$$Ew_i = 0, Ew_i^2 = 1$$

By CLT:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} w_i\varepsilon_i \\ w_i e_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Note: $\xi_0 = \xi_1 - \rho\xi_2$ is a normal random variable, independent from ξ_2

$$E(\xi_0|\xi_2) = E(\xi_1 - \rho\xi_2|\xi_2) = E(\xi_1|\xi_2) - \rho\xi_2 = 0$$

Then (using $\gamma = 0 \Rightarrow x_i = e_i$)

$$\hat{\beta}_{OLS} - \beta = \frac{\frac{1}{n}\sum x_i \varepsilon_i}{\frac{1}{n}\sum x_i^2} = \frac{\frac{1}{n}\sum e_i \varepsilon_i}{\frac{1}{n}\sum e_i^2} \xrightarrow{p} \rho \neq 0$$

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}}\sum w_i \varepsilon_i}{\frac{1}{\sqrt{n}}\sum w_i x_i} = \frac{\frac{1}{\sqrt{n}}\sum w_i \varepsilon_i}{\frac{1}{\sqrt{n}}\sum w_i e_i} \xrightarrow{d} \frac{\xi_1}{\xi_2} = \rho + \frac{\xi_0}{\xi_2}$$

□

Proposition 15.1.2. Under non-identifiability, the β t-statistics will diverge, such that we may conclude statistically significant estimates when they are in fact useless. (we prove this explicitly below in the case when $\rho \to 1$, i.e. lots of endogeneity)

$$|t| \xrightarrow{p} \infty$$

Proof.

$$\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - x_i\hat{\beta}_{IV})^2 = \frac{1}{n}\sum(\varepsilon_i - x_i(\hat{\beta}_{IV} - \beta))^2$$

$$= \frac{1}{n}\sum(\varepsilon_i - e_i(\hat{\beta}_{IV} - \beta))^2$$

$$\frac{1}{n}\sum \varepsilon_i^2 - 2(\hat{\beta}_{IV} - \beta)\frac{1}{n}\sum \varepsilon_i e_i + (\hat{\beta}_{IV} - \beta)^2\frac{1}{n}\sum e_i^2$$

$$\xrightarrow{d} 1 - 2\rho\left(\rho + \frac{\xi_0}{\xi_2}\right) + \left(\rho + \frac{\xi_0}{\xi_2}\right)^2$$

$$= 1 - \rho^2 + \left(\frac{\xi_0}{\xi_2}\right)^2$$

Therefore, the t-statistic for $H_0 : \beta = \beta_0, H1 : \beta \neq \beta_0$ has the asymptotic distribution:

$$t = \frac{\hat{\beta}_{IV} - \beta_0}{\sqrt{V\hat{a}r(\hat{\beta}_{IV})}}; \qquad Var(\hat{\beta}_{IV}) = \sigma^2(D'C^{-1}D)^{-1}$$

Replacing with sample analogues:

$$V\hat{a}r(\hat{\beta}_{IV}) = \hat{\sigma}^2(\hat{D}'\hat{C}^{-1}\hat{D})^{-1}$$

In the 1D case:

$$\hat{D} = \frac{1}{\sqrt{n}}\sum w_i x_i$$

$$\hat{C} = \frac{1}{n}\sum w_i^2$$

$$\therefore t = \frac{\hat{\beta}_{IV} - \beta}{\sqrt{\hat{\sigma}^2\frac{1}{n}\sum w_i^2 / \frac{1}{\sqrt{n}}|\sum w_i x_i|}} = \frac{\hat{\beta}_{IV} - \beta}{\sqrt{\hat{\sigma}^2\frac{1}{n}\sum w_i^2 / \frac{1}{\sqrt{n}}|\sum w_i e_i|}}$$

$$\xrightarrow{d} \frac{\rho + \frac{\xi_0}{\xi_2}}{\sqrt{1 - \rho^2 + \left(\frac{\xi_0}{\xi_2}\right)^2 / |\xi_2|}}$$

This distribution is non-normal. Note when $\rho \to 1$

$$Var(\xi_0) = Var(\xi_1 - \rho\xi_2) = 1 - 2\rho^2 + \rho^2 \xrightarrow{0}$$

so that

$$\xi_0 \xrightarrow{p} 0 \text{ and } 1 - \rho^2 + (\frac{xi_0}{\xi_2})^2 \xrightarrow{p} 0$$

This implies that

$$|t| \xrightarrow{p} \infty$$

□

15.2 Weak Instruments

When $E(w_i x_i')$ is of full column rank but $E(w_i w_i')^{-1}E(w_i x_i')$ (the coefficient matrix in the first stage regression) is small, the instruments, although relevant, are **weak**.

Proposition 15.2.1. Under weak instruments, $\hat{\beta}_{IV}$ will still not be consistent for β and will again have a non normal distribution.

Proof. Consider the same 1D model as previous, except instead of $E(w_i x_i) = 0$ We assume it is 'small', modelled by the 'local-to-zero' assumption:

$$E(w_i x_i) = \gamma = \frac{1}{\sqrt{n}}\mu$$

where μ is fixed, which will yield useful asymptotic approximations for $\hat{\beta}_{IV}$. Large μ corresponds to relatively strong instruments, whereas small μ corresponds to almost irrelevant instruments.

For $\hat{\beta}_{OLS}$ as before:

$$\hat{\beta}_{OLS} - \beta = \frac{\frac{1}{n}\sum x_i\varepsilon_i}{\frac{1}{n}\sum x_i^2} = \frac{\frac{1}{n}\sum(\frac{\mu}{\sqrt{n}}w_i + e_i)\varepsilon_i}{\frac{1}{n}\sum(e_i^2 + \frac{2\mu}{\sqrt{n}}w_i e_i + \frac{\mu^2}{n}w_i^2)} \xrightarrow{p} \rho \neq 0$$

For $\hat{\beta}_{IV}$:

$$\hat{\beta}_{IV} - \beta = \frac{\sum w_i\varepsilon_i}{\sum w_i x_i} = \frac{\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i}{\frac{1}{\sqrt{n}}\sum w_i(\frac{\mu}{\sqrt{n}}w_i + e_i)}$$

$$= \frac{\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i}{\mu\frac{1}{n}\sum w_i^2 + \frac{1}{\sqrt{n}}\sum w_i e_i} \xrightarrow{d} \frac{\xi_1}{\mu + \xi_2} = \frac{\xi_0}{\mu + \xi_2} + \rho\frac{\xi_2}{\mu + \xi_2}$$

Thus as in the case of irrelevant instruments, $\hat{\beta}_{IV}$ is not consistent for β and has a non-normal asymptotic distribution. When $\mu \to \infty$, so the instruments become strong, the consistency is restored because $\frac{\xi_1}{\mu+\xi_2} \xrightarrow{p} 0$ as $\mu \to \infty$. □

Proposition 15.2.2. The t-statistic for β will inflate as instruments become weaker and the distribution will become closer to $\chi^2(1)/\|\mu|$.

Proof. First consider $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - x_i\hat{\beta}_{IV})^2 = \frac{1}{n}\sum(\varepsilon_i - x_i(\hat{\beta}_{IV} - \beta))^2$$

$$\frac{1}{n}\sum\varepsilon_i^2 - 2(\hat{\beta}_{IV} - \beta)\frac{1}{n}\sum\varepsilon_i x_i + (\hat{\beta}_{IV} - \beta)^2\frac{1}{n}\sum x_i^2$$

But

$$\frac{1}{n}\sum\varepsilon_i x_i = \frac{1}{n}\sum\varepsilon_i\left(\frac{\mu}{\sqrt{n}}w_i + e_i\right) \xrightarrow{p} \rho \text{ and}$$

$$\frac{1}{n}\sum x_i^2 \xrightarrow{p} 1$$

Hence,

$$\begin{aligned}\hat{\sigma}^2 &\xrightarrow{d} 1 - 2\rho\frac{\xi_1}{\mu+\xi_2} + \left(\frac{\xi_1}{\mu+\xi_2}\right)^2 \\ &= 1 - \rho^2 + \left(\rho - \frac{\xi_1}{\mu+\xi_2}\right)^2 \\ &= 1 - \rho^2 + \left(\frac{\rho\mu - \xi_0}{\mu+\xi_2}\right)^2\end{aligned}$$

Thus for the t-statistic:

$$t = \frac{\hat{\beta}_{IV} - \beta}{\sqrt{\hat{\sigma}^2\frac{1}{n}\sum w_i^2}/\frac{1}{\sqrt{n}}|\sum w_i x_i|} \xrightarrow{d} \frac{\xi_1/(\mu+\xi_2)}{\sqrt{1-\rho^2+(\frac{\rho\mu-\xi_0}{\mu+\xi_2})^2}/|\mu+\xi_2|}$$

$$= \frac{\xi_1}{sgn(\mu+\xi_2)\sqrt{(1-\rho^2)+(\frac{\rho\mu-\xi_0}{\mu+\xi_2})^2}}$$

This has a non-normal distribution.
Again we consider the extreme case where $\rho = 1$ (lots of endogeneity in x_i):
$\rho = 1 \Rightarrow \xi_1 = \xi_2 = \xi \sim N(0,1)$, $\xi_0 = \xi - \rho\xi = 0$

$$t \xrightarrow{d} \frac{\xi(\mu+\xi)}{|\mu|}$$

With $|\mu|$ very large (strong instruments), t is almost normal, but when $|\mu|$ is small, t is almost $\chi^2(1)/|\mu|$.
As $\mu \to 0$, $t \xrightarrow{p} \infty$. Thus there is a multiplicity of possibilities when μ varies. □

15.2.1 Classifying Weak Instruments

Stock and Yogo (2005) classify strength of instruments based on the size distortion of the nominal 5% significance asymptotic t-test, which rejects the null hypothesis iff $|t| > 1.96$.

No distortion implies:

$$Pr(|t| > 1.96) \to 0.05$$

But with $\mu < \infty$, such convergence will not take place:

$$Pr(|t| > 1.96) = Pr\left(|\frac{\xi(\mu+\xi)}{|\mu|}| > 1.96\right) \nrightarrow 0.05$$

Thus the actual (as opposed to nominal(i.e. intended)) size of the test will not be 5%.

Stock and Yogo suggested that a 'tolerable' actual size should be perhaps not larger than 15%.

Let τ^2 be such that, whenever $\mu^2 \geq \tau^2$ the actual size of the t-test is below 15%. τ^2 is found through simulating the $\frac{\xi(\mu+\xi)}{|\mu|}$ distribution.

Then proposed to use the F-stat from first stage regression (normally used to test hypothesis $\mu = 0$) to test hypothesis that $\mu^2 \leq \tau^2$.

By simulation they found the appropriate critical value is **approx 10**, though this is dependent on choosing 15% as the benchmark distortion and number of instruments used.
But for not too many instruemnts critical value remains around 10 so this is used in empirical literature often to determine if instruments weak.