Advanced Econometrics I

Jürgen Meinecke

Lecture 8 of 12

Research School of Economics, Australian National University

Roadmap

What are we to do?

In some undergraduate textbooks, the so-called *rule of thumb* is often recommended

That rule originated from Staiger and Stock (1997)

Their main goal was to come up with a quick and easy and robust way for practitioners to rule out the weak IV case

Staiger and Stock had the idea to propose a rule that is solely based on the first stage F statistic

The first stage F statistic is the F statistic in the regression of the endogenous regressor(s) on the instruments

The rule of thumb can be viewed as a decision rule:

- if first stage $F > 10$, then decide that instruments are strong
- if first stage $F < 10$, then decide that instruments are weak

If you decide that your instruments are strong, then $\hat{\beta}^{\mathrm{IV}}$ and $\hat{\beta}^{\mathrm{2SLS}}$ can be used safely and statistical inference based on asymptotic normality is fine

The rule of thumb turns out to be a bit crude

Stock and Yogo (2005) offer a somewhat more sophisticated way of testing for weak IV

While more complex than the rule of thumb, their recommendation is still quite easy to implement

Their test is still entirely based on the first stage F statistic

Stock and Yogo's idea is the following:

Unless you've got very strong instruments, $S(\rho, \tau)$ will be non-normal (recall from last week: S is the limiting distribution of the t-statistic)

Statistical inference based on $S(\rho, \tau)$ will be misleading

In particular, the *actual size* of the test will differ from the *nominal size*

Recall: the nominal size of a well-behaved statistical test is the probability to reject the null when it's true

By trusting the normal approximation and setting the critical value to 1.96 you are fixing the nominal size at 5%

Here we're not dealing with a well-behaved statistical test, so we admit that we may have to accept a higher actual size

Stock and Yogo's idea is to pin down a worst case scenario for the actual size and provide the corresponding critical values for the first stage F statistic

It's best to look at an example

- You understand that an actual size of 5% is unrealistic
- Instead you are willing to live with an actual size of 20%
- This is your worst case scenario; the largest actual size that you are willing to tolerate
- For the case of one endogenous regressor and one instrument, Stock and Yogo suggest to use the critical value 6.66
- This means that your F-statistic in the first stage regression of X_i on Z_i must exceed 6.66 to achieve an actual size that is at most 20%

Excerpt from Stock and Yogo (2005),
"Testing for Weak Instruments in Linear IV Regression"

Table 5.2. *Critical values for the weak instrument test based on TSLS size* (Significance level is 5%)

	$n = 1, r =$				$n = 2, r =$			
K_2	0.10	0.15	0.20	0.25	0.10	0.15	0.20	0.25
1	16.38	8.96	6.66	5.53				
2	19.93	11.59	8.75	7.25	7.03	4.58	3.95	3.63
3	22.30	12.83	9.54	7.80	13.43	8.18	6.40	5.45
4	24.58	13.96	10.26	8.31	16.87	9.93	7.54	6.28
5	26.87	15.09	10.98	8.84	19.45	11.22	8.38	6.89
6	29.18	16.23	11.72	9.38	21.68	12.33	9.10	7.42
7	31.50	17.38	12.48	9.93	23.72	13.34	9.77	7.91
8	33.84	18.54	13.24	10.50	25.64	14.31	10.41	8.39
9	36.19	19.71	14.01	11.07	27.51	15.24	11.03	8.85
10	38.54	20.88	14.78	11.65	29.32	16.16	11.65	9.31

Note: K_2: number of IV; n: number of endogenous variables

Stock and Yogo suggest to use their weak instrument test as a decision rule:

- if the first stage F-statistic lower than the critical value, conclude that the instruments are weak
- otherwise conclude that they are strong

What to do when you conclude that instruments are weak?
Lot's of research still underway, but options include: Anderson-Rubin test, Kleibergen's K-statistic, Moreira's conditional likelihood test

Depending on your application the Staiger and Stock (1997) *rule of thumb* that sets the critical value simply to 10 is either too restrictive or too relaxed

If you want your actual size to be at most 10%, then the rule of thumb is too relaxed

Roadmap

Let's simulate the following toy model:

$$\begin{aligned} Y_i &= X_i\beta + e_i \\ X_i &= Z_i\pi + v_i, \end{aligned}$$

where all variables are scalars, and we assume

- instrument generated according to $Z_i \sim N(0,1)$
- errors generated according to
$$\begin{pmatrix} e_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

To generate a pseudo random sample (X_i, Y_i) we need to pin down values for

- sample size N
- degree of endogeneity ρ
- structural coefficient β
- reduced form coefficient π

Let's go with these ones:
(could easily play around with different values)

- sample size $N = 1,000$
- degree of endogeneity $\rho \in \{0.10, 0.90\}$
  (low and high degree)
- structural coefficient $\beta = 0$

What about π?

We want to be clever when pinning down the value for π

The value of π is related to the strength of the instrument

There is a dependency between π and the first stage F-statistic

π will be determined by setting the first stage F

How do we make this connection between π and F?

For $\hat{\pi}$ it can be shown that

$$R^2 = (\widehat{\mathrm{Corr}}(X, Z))^2 = s^2_{XZ}/s^2_{X} = \hat{\pi}^2/(1 + \hat{\pi}^2),$$

where R^2 is the ‘R squared’ (measure of fit) in the reduced form regression of X_i on Z_i

It follows that $\hat{\pi} = \sqrt{R^2/(1 - R^2)}$

The R^2, in turn, is related to the F-statistic:
In the simple linear regression model $F = N \cdot R^2/(1 - R^2)$

Now we can make the connection between $\hat{\pi}$ and F: $F = N \cdot \hat{\pi}^2$

For practical purposes we take this to mean that $\pi \approx \sqrt{\frac{F}{N}}$

From previous slide: $\pi \approx \sqrt{\frac{F}{N}}$

So what we will be doing here is setting F to pin down π

Stock and Yogo have given us a good collection of values for F (see their Table 5.2 shown earlier)

Let's focus on two values:

- relatively weak instrument: $F = 5.53$
  (actual size allowed to blow out to 25%)
- relatively strong instrument: $F = 16.38$
  (actual size restricted to be at most 10%)

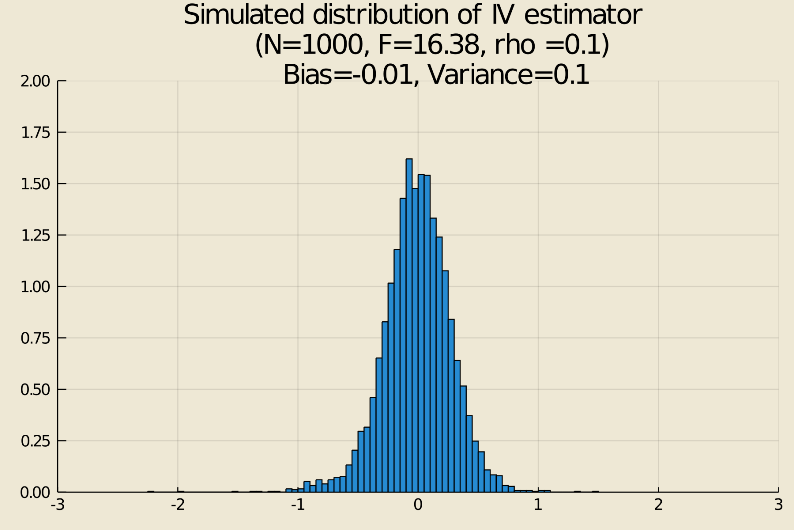
Now that we have discussed how to set π, we've closed the model (no more undetermined parameters)

On the next 8 slides I will show you

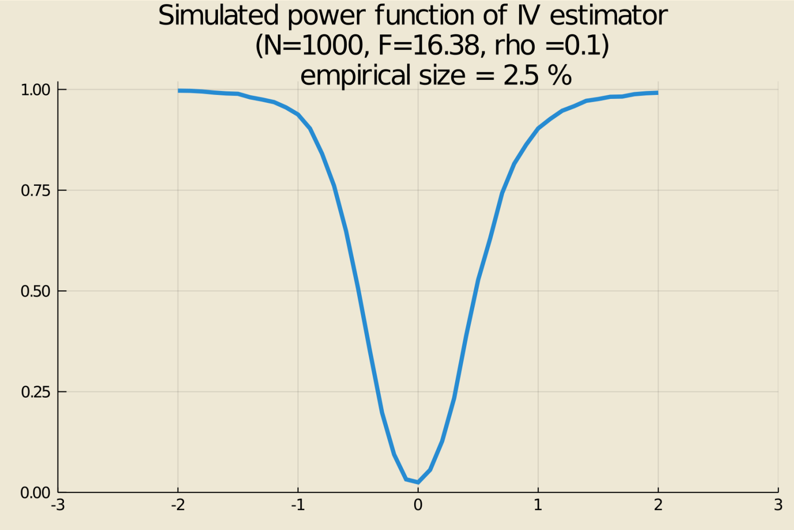
- simulated distributions
- power functions

for the following four parameter combinations:

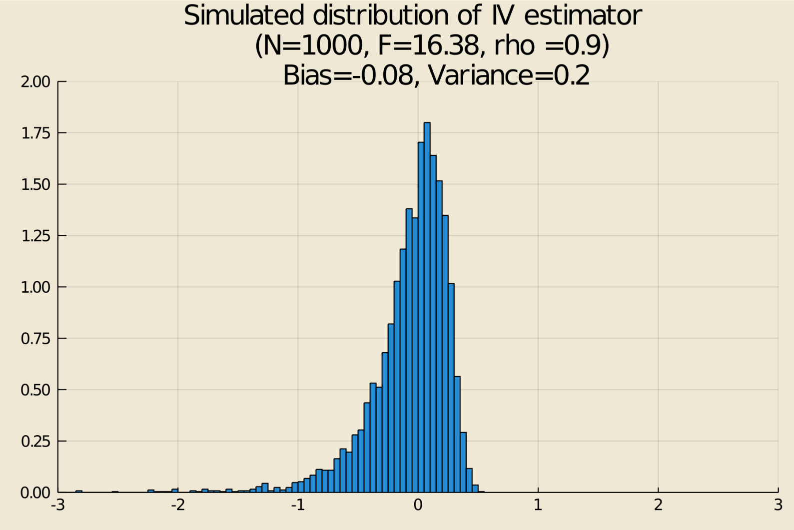
- degree of endogeneity $\rho \in \{0.10, 0.90\}$ (low and high degree)
- strength of instrument $F \in \{5.53, 16.38\}$ (weak and strong)



Strong instrument, low degree of endogeneity
well behaved distribution

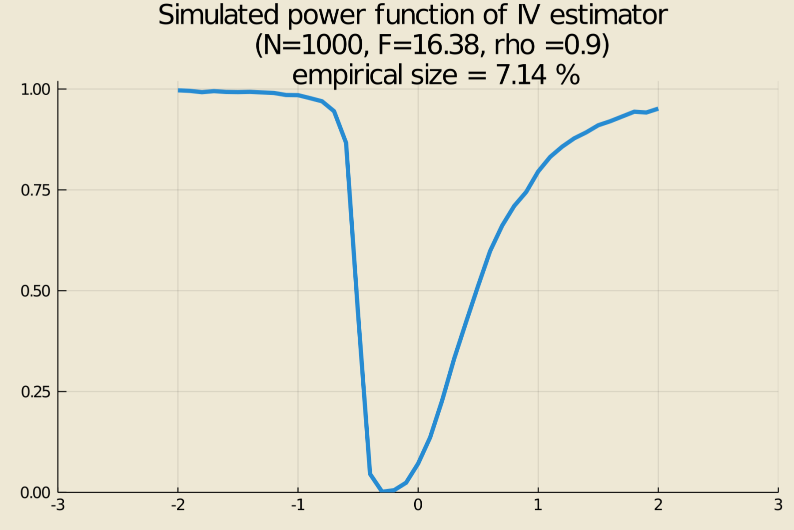


Strong instrument, low degree of endogeneity
nice looking power function

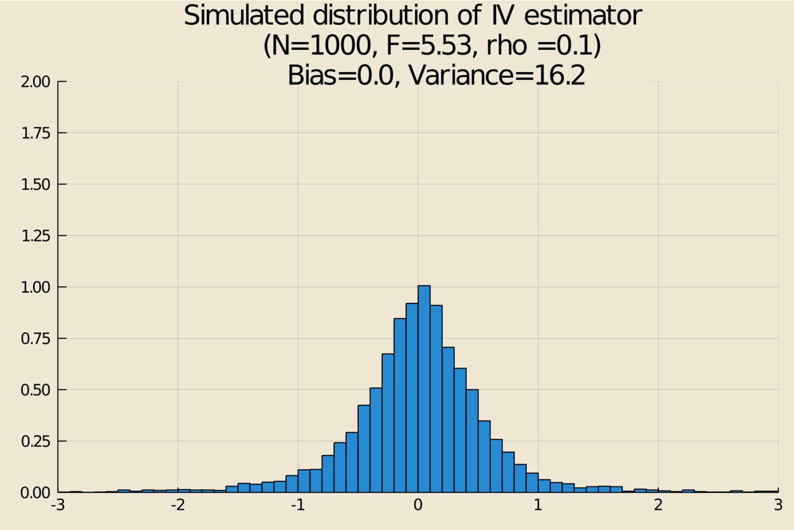


Strong instrument, high degree of endogeneity

Can you spot the distortion and the resulting leftward bias?

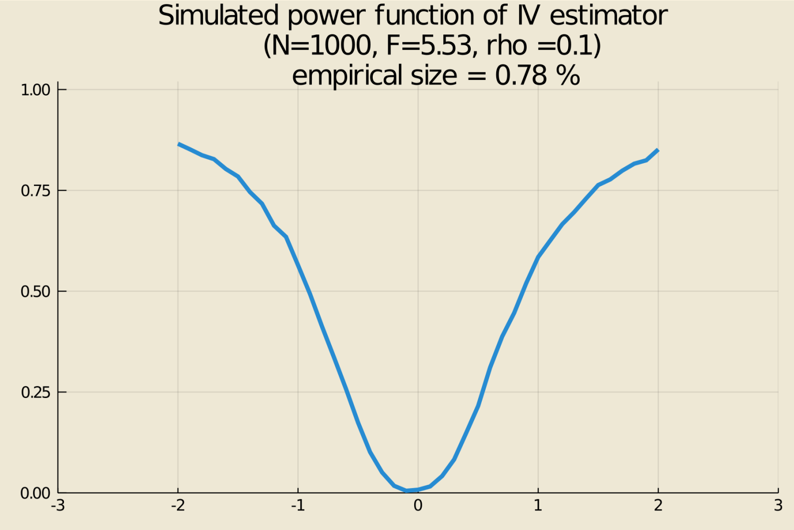


Strong instrument, high degree of endogeneity



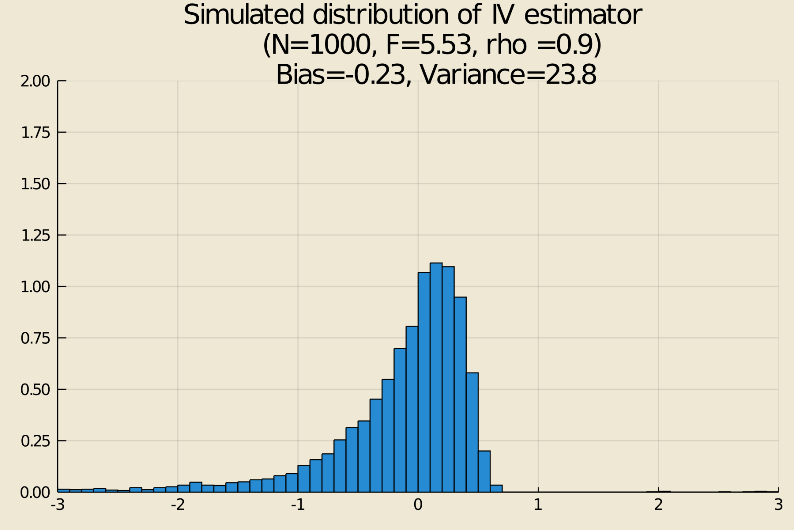
Weak instrument, low degree of endogeneity

Looks symmetric, but awful variance

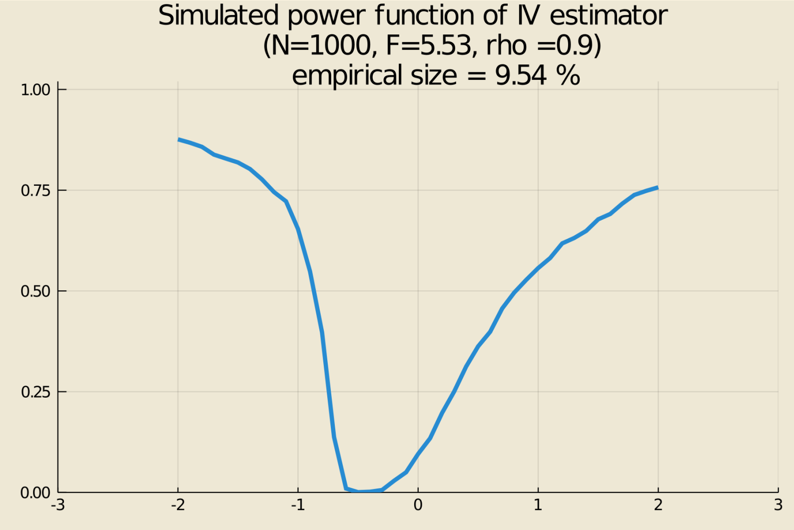


Weak instrument, low degree of endogeneity

The probability to reject far away from zero is terribly low



Weak instrument, high degree of endogeneity
Not symmetric, and awful variance



Weak instrument, high degree of endogeneity

Things only get worse here