ECMA31100 Introduction to Empirical Analysis II

Winter 2022, Week 4: Discussion Session

Conroy Lau

The University of Chicago

February 2, 2022

Some logistics

Problem sets

- Please compile your file before submitting the problem set
- Do not include your code as a pdf
- Submit your code as a separate script that I can get all the results by running it

Topics for this and next TA session

- Last time we discussed blocking and a bit on heterogeneous treatment effect
- I will defer the discussion of Abadie (2003) to next TA session
- Today we will focus on weak instruments
- I will also walk through some of my code snippets to illustrate the implementation

Today's theme: weak instruments

Outline

- Test inversion
- Anderson-Rubin test statistic
- k-class estimators and LIML
- Effective F-statistic

Recommended readings

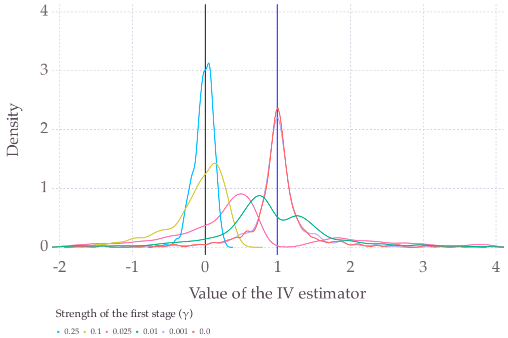
- Stock and Yogo (2005)
- Andrews et al. (2019)

Please let me know if you have any suggestions or if you find any typos in the slides

Motivation (again)

- As γ decreases, the IV estimator is centered at the OLS estimator
- You have seen a similar figure in Prof. Hardwick's week 4 slides

Figure 1: Distribution of OLS and IV estimators (similar to that in Nelson and Startz (1990))



Contents

1. Test inversion
2. Anderson-Rubin test statistic
3. *k*-class estimators
4. Effective *F*-statistic

Test inversion

Main idea

- Suppose we are interested in testing the null of $H_0 : \theta = \theta_0$, where $\Theta \equiv \text{supp}(\theta)$
- Consider a test statistic $T(\theta)$ and a critical value c_α for a given $\alpha \in (0, 1)$
- A $(1 - \alpha)$-confidence region is $\mathcal{C} \equiv \{\theta : H_0 \text{ is not rejected}\}$ with $\lim_{\theta \in \Theta} \mathbb{P}[\theta \in \mathcal{C}] \geq 1 - \alpha$

Implementation

- One way to conduct test inversion is via grid testing:
  1. Partition Θ into N points $\widehat{\Theta} \equiv \{\theta_1, \ldots, \theta_N\}$
  2. For each $\theta_i \in \widehat{\Theta}$, it belongs to $\mathcal{C}$ if $T(\theta_i) \leq c_\alpha$
- Easy to implement but it can be impractical and computationally expensive

Test inversion

Example

- DGP: $Y_i = X_i\beta_0 + U_i$ where $X_i, U_i \sim \mathcal{N}(0,1)$ and $\beta_0 = 1$
- Goal: Construct a 95%-confidence interval for β_0
- Test statistic: $T_n(\beta) = \frac{\widehat{\beta}_n - \beta}{\text{se}(\widehat{\beta}_n)}$

  $\implies$ We immediately know that our confidence interval is $\widehat{\beta}_n \pm \text{se}(\widehat{\beta}_n) z_{1-\frac{\alpha}{2}}$

Implementations

1. Use built-in functions (e.g., `lm` and `confint` in `R`)
2. Use test inversion with grid testing to find $\mathcal{C} = \{\beta : |T_n(\beta)| \leq z_{1-\frac{\alpha}{2}}\}$
   - We know $T_n(\beta)$ is monotonic decreasing here, so let's search over $\beta \in [-50, 50]$
   - $z_{0.975} = 1.96$ (e.g., `qnorm(.975)` in `R` or `quantile(Normal(), .975)` in `Julia`)

Test inversion

DGP

```
alpha <- .05
n <- 1e3           # Sample size
b <- 1             # beta
# Simulate data
set.seed(31100)
x <- rnorm(n)
u <- rnorm(n)
df <- data.frame(y = x * b + u, x = x, u = u)
```

Implementation 1

```
lmout <- lm(y ~ x - 1, data = df)
confint(lmout)
#>          2.5 %   97.5 %
#>    x   0.9516435 1.076949
```

Test inversion

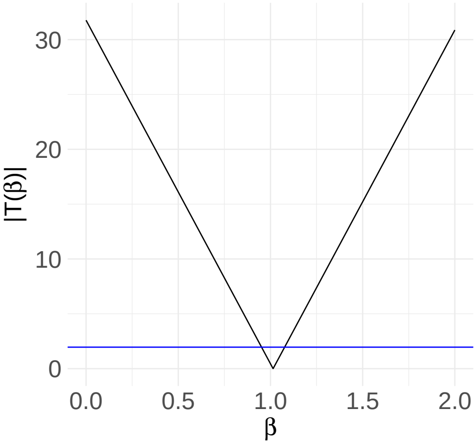
Implementation 2

```r
bs <- seq(-50, 50, .001)      # A sequence of grid points
cv <- qnorm(1 - alpha/2)   # Critical value
ts <- function(b) (lmout$coef[1] - b)/(summary(lmout)$coef[2]) # Test statistic
tout <- unlist(lapply(bs, FUN = ts))  # Evaluate `ts` at each `bs`
cs <- bs[abs(tout) <= cv]             # Confidence interval
cat(sprintf("The confidence interval is [%s, %s]",
            round(min(cs), digits = 3),
            round(max(cs), digits = 3)))
#> The confidence interval is [0.952, 1.076]
```

- The result is different from the previous slide since I chose the grid size as 0.001
- You can obtain a more accurate result by using finer grids

Test inversion

Figure 2: Plot of $|T_n(\beta)|$ against β over $[0, 2]$



Any questions?

Contents

1. Test inversion

2. Anderson-Rubin test statistic

3. *k*-class estimators

4. Effective *F*-statistic

Anderson-Rubin test statistic

Linear IV model without covariates

Let Y be the outcome, X be the endogenous variables, and Z be the excluded IVs

$$\begin{aligned} \underset{(N\times 1)}{Y} &= \underset{(N\times d_X)}{X}\underset{(d_X\times 1)}{\beta} + \underset{(N\times 1)}{U} \\ \underset{(N\times d_X)}{X} &= \underset{(N\times d_Z)}{Z}\underset{(d_Z\times d_X)}{\Pi} + \underset{(N\times d_X)}{V} \end{aligned} \tag{1}$$

Test statistic (Anderson and Rubin, 1949)

$$\mathrm{AR}(\beta) \equiv \frac{N-d_Z}{d_Z}\frac{(Y-X\beta)'P_Z(Y-X\beta)}{(Y-X\beta)'M_Z(Y-X\beta)},$$

where $P_Z \equiv Z(Z'Z)^{-1}Z'$ and $M_Z \equiv I_{d_Z} - P_Z$

- Asymptotic distribution: $\mathrm{AR}(\beta) \xrightarrow{d} \frac{1}{d_Z}\chi^2_{d_Z}$

Anderson-Rubin test statistic

AR test statistic is nonlinear

- The AR test statistic cannot be written in a form like the *t*-statistic
- One way to construct confidence region is via test inversion

Example

- DGP:

$$Y_i = X_i\beta_0 + U_i$$

$$X_i = Z_i\pi_0 + V_i$$

  - $\beta_0 = 1$ and $\pi_0 = 0.2$
  - $X_i, Z_i \sim \mathcal{N}(0, 1)$
  - $\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}\right)$
- Critical value: $\chi^2_{1,0.95} = 3.841$

Anderson-Rubin test statistic

DGP

```r
b <- 1
pi <- .2
rho <- -.9

# Simulate data
set.seed(31100)
z <- rnorm(n)
uv <- MASS::mvrnorm(n, mu = c(0, 0), Sigma = matrix(c(1, rho, rho, 1), 2, 2))
x <- pi * z + uv[1, ]
y <- b * x + uv[2, ]
df <- data.frame(y = y, x = x, z = z)
```

Anderson-Rubin test statistic

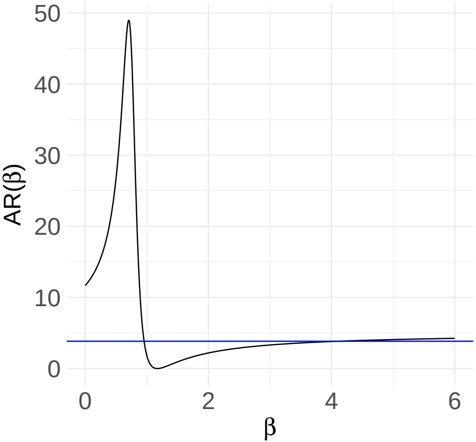
AR test statistic

```r
# AR test statistic
ar <- function(b) {
  # Generate the residual maker and projection makers
  pz <- df$z %*% solve(t(df$z) %*% df$z) %*% t(df$z)
  qz <- diag(n) - pz

  # Compute the quantities
  res <- df$y - df$x * b
  out1 <- t(res) %*% pz %*% res
  out2 <- t(res) %*% qz %*% res
  dz <- ncol(df$z)
  return((n - dz)/dz * out1/out2)
}
```

Anderson-Rubin test statistic

Figure 3: Plot of AR(β) against β over $[0, 6]$



Anderson-Rubin test statistic

AR confidence intervals (Dufour and Taamouti, 2005; Andrews et al., 2019)

- There can be three forms of AR confidence regions
  1. A bounded interval $[a, b]$
  2. The real line $(-\infty, \infty)$
  3. The real line excluding an interval $(-\infty, a] \cup [b, \infty)$
- Can you think of some intuitions?

Behavior as $|\beta| \longrightarrow \infty$ (Kleibergen, 2007)

- As $|\beta| \longrightarrow \infty$, AR$(\beta) \longrightarrow$ Wald statistic for testing that $\Pi = 0$
- AR(β) at level α has infinite length if and only if a robust F-test cannot reject $\Pi = 0$

Any questions?

Contents

1. Test inversion
2. Anderson-Rubin test statistic
3. k-class estimators
4. Effective F-statistic

k-class estimators

Model

- Consider the same model as in (1)
- The reduced-form is $Y = (Z\Pi + V)\beta + U = Z\Pi\beta + (V\beta + U) \equiv Z\delta + U_2$

***k*-class estimators**

- Theil (1953):

$$\widehat{\beta}_k \equiv [X'(I_n - kM_Z)X]^{-1}[X'(I_n - kM_Z)Y]$$

  where $M_Z \equiv I_n - Z(Z'Z)^{-1}Z'$
- When $k = 0$, we have $\widehat{\beta}_0 = (X'X)^{-1}(X'Y) \longrightarrow$ OLS
- When $k = 1$, we have $\widehat{\beta}_1 = [X'(I_n - M_Z)X]^{-1}[X'(I_n - M_Z)Y] \longrightarrow$ TSLS

k-class estimators

Limited information maximum likelihood (LIML) estimators

- Maximizes the likelihood of subject to $\Pi\beta - \delta = 0$ with a normal assumption
- In model (1) under a normal assumption on (U, V), the loglikelihood function is

$$L(\beta, \Pi, \Omega) = -\frac{N}{2}\log\Omega - \frac{1}{2}\sum_{i=1}^{N} g_i(\beta, \Pi)'\Omega^{-1}g_i(\beta, \Pi),$$

  where Ω is the covariance matrix of $(U_i, V_i)'$ and $g_i(\beta, \Pi) \equiv (Y_i - Z_i\Pi\beta, X_i - Z_i\Pi)'$
- LIML can be written as a k-class estimator with $k_{\text{LIML}} \equiv \min_{\beta} \frac{(Y-X\beta)'(Y-X\beta)}{(Y-X\beta)'M_Z(Y-X\beta)}$

Some remarks on LIML estimators

- Some simulation evidence suggest that LIML performs well in weak instruments
- However, the variance is usually higher than TSLS

Any questions?

Contents

1. Test inversion

2. Anderson-Rubin test statistic

3. *k*-class estimators

4. Effective *F*-statistic

Review of Nagar approximation

Model

- Consider model (1) again, but with $d_X = d_Z = 1$ (and I now write π instead of Π)
- Assume homoskedasticity so that $\mathbb{E}[U_i^2] = \sigma_U^2$, $\mathbb{E}[\sigma_V^2]$ and $\mathbb{E}[U_i V_i] = \sigma_{UV}$

Nagar (1959) approximation (see week 3 lecture slides for details)

- Derive the asymptotic distribution of $\widehat{\beta}_{\text{IV}} - \beta$ and probability limit of $\widehat{\beta}_{\text{OLS}} - \beta$:
  - $\widehat{\beta}_{\text{IV}} - \beta \xrightarrow{d} \frac{A}{\pi \mathbb{E}[Z_i^2] + B}$ where $\binom{A}{B} \sim \mathcal{N}\left(0_2, \mathbb{E}[Z_i^2]\begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix}\right)$
  - $\widehat{\beta}_{\text{OLS}} - \beta \xrightarrow{p} \frac{\mathbb{E}[U_i V_i]}{\mathbb{E}[V_i^2]} = \frac{\sigma_{UV}}{\sigma_V^2}$
- The relative asymptotic bias is $\frac{\sigma_V^2}{\sigma_{UV}} \mathbb{E}\left[\frac{A}{\pi \mathbb{E}[Z_i^2] + B}\right]$
- Compute $\mathbb{E}\left[\frac{A}{\pi \mathbb{E}[Z_i^2] + B} \middle| B = b\right]$ (conditional on $B = b$), and take Taylor approximation
- The relative asymptotic bias is $\frac{1}{\mu^2}$, where μ^2 is the concentration parameter

Montiel Olea and Pflueger (2013)

This paper

- Allows for errors that are heteroskedastic and with autocorrelation
- Consider IVs to be weak when the TSLS/LIML is bias is relative to a benchmark τ
- Their `Stata` package for implementation: `weakivtest` (Pflueger and Wang, 2015)

My plan

- Due to time constraints, I will only discuss the homoskedastic error and TSLS case
- Show how they use Nagar approximation — relate to what you learnt in lecture
- I will follow the notations of the paper — they are slightly different from lecture
- I am happy to discuss more about this paper in my office hours

Montiel Olea and Pflueger (2013)

Pflueger and Wang (2015)

- `weakivtest` is a post-estimation command in `Stata`
- Run `weakivtest` after using `ivreg2` and `ivregress`
- Returns the effective F-statistic, as well as critical values at various thresholds

Example (based on the help file of `weakivtest`)

```
webuse set http://fmwww.bc.edu/repec/bocode/d
webuse Data_USAQ.dta, clear

ivregress 2sls dc (rrf = z1 z2 z3 z4), vce(hac nw 6)
weakivtest
```

Montiel Olea and Pflueger (2013)

Model

We observe $\{y_i, Y_i, Z_i\}_{i=1}^{S}$ and consider the linear IV model with $y_i, Y_i \in \mathbb{R}$

$$y = Z\Pi\beta + v_1$$

$$Y = Z\Pi + v_2$$

Assumptions

1. Local to zero: $\Pi = \Pi_S = \frac{C}{\sqrt{S}}$ where $C \in \mathbb{R}^K$ is a fixed vector
2. High-level assumptions:
   (a) $\frac{1}{\sqrt{S}} \begin{pmatrix} Z'v_1 \\ Z'v_2 \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2K}(0, W)$, where $W \equiv \begin{pmatrix} W_1 & W_{12} \\ W_{12}' & W_2 \end{pmatrix}$ is positive definite
   (b) $\frac{1}{S} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \begin{pmatrix} v_1' & v_2' \end{pmatrix} \xrightarrow{p} \Omega$ where $\Omega \equiv \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}$ is positive definite
3. Normalization: $\frac{1}{S} Z'Z = I_K$
4. Homoskedasticity: $W = \Omega \otimes I_K$

Montiel Olea and Pflueger (2013)

Two intermediate results

1. Under assumptions 1 and 2 from the previous slide, we have

$$\frac{1}{\sqrt{S}}\begin{pmatrix} Z'y \\ Z'Y \end{pmatrix} = \begin{pmatrix} \beta C + \frac{1}{\sqrt{S}} Z'v_1 \\ C + \frac{1}{\sqrt{S}} Z'v_2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim \mathcal{N}_{2K}\left(\begin{pmatrix} \beta C \\ C \end{pmatrix}, W \right)$$

2. The asymptotic distribution of the TSLS estimator is

$$\widehat{\beta}_{\text{TSLS}} - \beta_{\text{TSLS}} \xrightarrow{d} \beta^{\star}_{\text{TSLS}} \equiv \frac{\gamma_2'(\gamma_1 - \beta\gamma_2)}{\gamma_2'\gamma_2}$$

Montiel Olea and Pflueger (2013)

Proposition 1.1 of Montiel Olea and Pflueger (2013)

The Nagar approximation of the TSLS bias is

$$N_{\text{TSLS}}(\beta, C, W, \Omega) \equiv \frac{1}{\mu^2} n_{\text{TSLS}}(\beta, C_0, W, \Omega) \equiv \frac{1}{\mu^2} \frac{\text{tr}(S_{12})}{\text{tr}(S_2)} \left[1 - \frac{2C_0' S_{12} C_0}{\text{tr}(S_{12})} \right]$$

Nagar approximation under homoskedasticity

The above becomes

$$N_{\text{TSLS}}(\beta, C, W, \Omega) = \frac{1}{\mu^2} \frac{\sigma_{12}}{\sigma_2^2} \left(1 - \frac{2\sigma_{12} C_0' I_K C_0}{\sigma_{12} K} \right) = \frac{1}{\mu^2} \frac{\sigma_{12}}{\sigma_2^2} \left(1 - \frac{2}{K} \right)$$

Montiel Olea and Pflueger (2013)

Test statistic

- The test is based on the effective F-statistic $\widehat{F}_{\text{eff}}$ where

$$\widehat{F}_{\text{eff}} \xrightarrow{d} F^{\star}_{\text{eff}} \equiv \frac{\gamma_2'\gamma_2}{\text{tr}(W_2)}$$

- The null is rejected when

$$\widehat{F}_{\text{eff}} > c\left(\alpha, \widehat{W}_2, \frac{1}{\tau} B_{\text{TSLS}}(\widehat{W}, \widehat{\Omega})\right) \equiv \sup_{C \in \mathbb{R}^K} \left\{ F^{-1}_{C, \widehat{W}_2}(\alpha) \, \mathbb{1}\left[\frac{C'C}{\text{tr}(\widehat{W}_2)} < \frac{B_{\text{TSLS}}(\widehat{W}, \widehat{\Omega})}{\tau} \right] \right\}$$

Critical values

- Monte Carlo methods
- Curve-fitting methodology by Patnaik (1949)

Any questions?

References 1

ABADIE, A. (2003): "Semiparametric instrumental variable estimation for causal effects," *Journal of Econometrics*, 113, 231–263.

ANDERSON, T. W. AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46 – 63.

ANDREWS, I., J. H. STOCK, AND L. SUN (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

DUFOUR, J.-M. AND M. TAAMOUTI (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

References 2

KLEIBERGEN, F. (2007): "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139, 181–216, endogeneity, instruments and identification.

MONTIEL OLEA, J. L. AND C. PFLUEGER (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General K-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.

NELSON, C. R. AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

PATNAIK, P. B. (1949): "The non-central chi2- and F-distributions and their applications," *Biometrika*, 36, 202–232.

References 3

PFLUEGER, C. E. AND S. WANG (2015): "A Robust Test for Weak Instruments in Stata," *The Stata Journal*, 15, 216–225.

STOCK, J. AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*.

THEIL, H. (1953): "Repeated least-squares applied to complete equation systems," *The Hague: Central Planning Bureau*.