ECON0064
MSc Econometrics

Part 8, Two Stage Least Squares (2SLS)

UCL

Autumn 2024

Notation

[Reference for IV and 2SLS: **Wooldridge: Ch.5**]

- Model: $y_i = x_i\beta + u_i$, (this is the structural equation) where $x_i = (x_{i1}, \ldots, x_{iK})$ is vector of K regressors
- We observe instruments a vector of L instruments $z_i = (z_{i1}, \ldots, z_{iL})$. All exogenous regr. x_i are included in z_i.
- To define the 2SLS estimator we have to assume $L \geq K$.
- Vector-matrix notation: $y = X\beta + u$, where y and u are $n \times 1$ vectors, X is $n \times K$ matrix (as before), and we also define the $n \times L$ matrix

$$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} z_{11} & \cdots & z_{1L} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{nL} \end{pmatrix}$$

$L \geq K$

$$\underline{\underset{L\times 1}{0} = E[z_i' u_i]} = E[z_i'(y_i - x_i \overset{K\times 1}{\beta})]$$

Exact / Just Identified IV

$L = K$ $$\underline{0 = \frac{1}{n} \sum z_i'(y_i - x_i \hat{\beta}_{IV})}$$

$$\hat{\beta}_{IV} = \left(\frac{1}{n} \sum z_i' x_i\right)^{-1} \frac{1}{n} \sum z_i' y_i$$

$L > K$ $$0_{L\times 1} \neq \frac{1}{n} \sum z_i'(y_i - x_i \hat{\beta})$$

$x_i - \tilde{x}_i = \varepsilon_i$

Two Stage Least Squares Estimator (2SLS)

First Stage:

- For each regressors apply OLS to estimate

pop. projection

$$x_{ik} = z_i\Gamma_k + \varepsilon_{ik}$$

$$\underset{n\times 1}{\hat{x}_k} = \underset{n\times L}{Z}\,\underset{L\times 1}{\hat{\Gamma}_k}$$

We obtain $\hat{\Gamma}_k = (Z'Z)^{-1}Z'x_k$, $k = 1, \ldots, K$.

- For the $L \times K$ matrix $\hat{\Gamma} = (\hat{\Gamma}_1, \ldots, \hat{\Gamma}_K)$ we have

$$\underset{L\times K}{\hat{\Gamma}} = (Z'Z)^{-1}Z'X = (Z'Z)^{-1}Z'\overset{X}{[x_1 \ldots x_K]}$$

$\hat{\Gamma}_1 \quad \hat{\Gamma}_K$

- Calculate the predicted values of this regression:

$$[\hat{x}_1 \ldots \hat{x}_K] = \underset{n\times K}{\hat{X}} = Z\hat{\Gamma} = \underbrace{Z(Z'Z)^{-1}Z'}_{= P_Z}X = P_Z X$$

$Z(Z'Z)^{-1}Z'x_1, \ Z(Z'Z)^{-1}Z'x_K$

- We discussed the orthogonal projector M_Z which when applied to a n-vectors gives the part of that vector that is not explained by Z. Conversely, the projector $P_Z = Z(Z'Z)^{-1}Z'$ gives the part $\hat{X}$ of X that is explained by Z.

$$\hat{X} = P_Z X$$

$$P_Z' = P_Z$$

$$P_Z P_Z = P_Z$$

Two Stage Least Squares Estimator (2SLS) (cont.)

Second Stage:

- We got $\hat{X} = Z\hat{\Gamma} = P_Z X$ from the first stage.
  Let $\hat{x}_i = z_i\hat{\Gamma}$ be the rows of $\hat{X}$.
- z_i and thus $z_i\Gamma$ are exogenous
  (wrt to both u_i and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{1K})$).
  $\hat{x}_i = z_i\hat{\Gamma}$ is not exogenous at finite sample
  (because endogenous x_i enters into $\hat{\Gamma}$),
  but as $n \to \infty$ we have $\hat{\Gamma} \to_p \Gamma$ and thus $\hat{x}_i \to_p z_i\Gamma$,
  i.e. $\hat{x}_i$ is "asymptotically exogenous".
- This suggest that we can obtain a consistent estimator for β by applying OLS to the equation

$$y_i = \hat{x}_i\beta + v_i,$$

where $v_i = u_i + (x_i - \hat{x}_i)\beta \to_p u_i + (x_i - z_i\Gamma)\beta = u_i + \varepsilon_i\beta$.

$\tilde{X}_i = Z_i\Gamma$

$\hat{X}_i = Z_i\hat{\Gamma} \xrightarrow[p]{} \underline{Z_i\Gamma}$

MODEL : $y_i = x_i\beta + u_i$

$$y_i = \tilde{X}_i\beta + \tilde{V}_i = \tilde{X}_i\beta + \underbrace{(\overset{\varepsilon_i}{(X_i - \tilde{X}_i)}\beta + u_i)}_{\tilde{V}_i}$$

$Z_i \perp \varepsilon_i$ Projection

$Z_i \perp u_i$ Exogeneity

$Z_i \perp \tilde{V}_i = \varepsilon_i\beta + u_i$

$$y_i = \underbrace{\tilde{X}_i}_{Z_i\Gamma}\beta + \tilde{V}_i$$

Two Stage Least Squares Estimator (2SLS) (cont.)

- From this we obtain the 2SLS estimator:

$$\begin{aligned}\hat{\beta}_{\mathrm{2SLS}} &= \underline{(\hat{X}'\hat{X})^{-1}}\hat{X}'y \\ &= \underline{(X'P_ZX)^{-1}X'P_Zy} \\ &= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'y.\end{aligned}$$

  Here we used that $P_ZP_Z = P_Z$ (idempotent) and that $P_Z' = P_Z$ (symmetric).

- Thus, we can define $\hat{\beta}_{\mathrm{2SLS}}$ equivalently as $(\hat{X}'\hat{X})^{-1}\hat{X}'y$, i.e. using the two-stage procedure, or as $(X'P_ZX)^{-1}X'P_Zy$.

$$\hat{X} = P_Z X$$

$$\hat{X}'\hat{X} = X'\overbrace{P_Z'P_Z}^{P_Z}X = X'P_ZX$$

$$\hat{X}'\hat{y} = X'P_Zy$$

Two Stage Least Squares Estimator (2SLS) (cont.)

- [Side comment: the structure of the 2SLS estimator is very similar to the structure of the GLS estimator, with Ω^{-1} replaced by P_Z. However, both the math and the interpretation is different here. In particular, $\mathrm{rank}(\Omega) = n$ and $\mathrm{rank}(P_Z) = L \ll n$].
- Remember: All exogenous regressors are also included in z_i. For those regressors we have $\hat{x}_{ik} = x_{ik}$, because they are perfectly predicted by themselves. For example if there is no endogenous regressors and no additional instrument, then $Z = X$ and therefore

  $\hat{X}_k = X_k$ for exog. X_k

  $$\hat{X} = P_Z X = P_X X = X(X'X)^{-1}X'X = X.$$

  In that case we have $\hat{\beta}_{\mathrm{2SLS}} = \hat{\beta}_{\mathrm{OLS}}$. $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$
- Another, somewhat more general, special case is $L = K$ (exactly identified case). We then obtain the following simplification: $\hat{\beta}_{\mathrm{2SLS}} = \hat{\beta}_{\mathrm{IV}} = (Z'X)^{-1} Z'y$.

$K = L$

$$\Big(\underbrace{X'Z}_{K\times K}\underbrace{(Z'Z)^{-1}}_{K\times K}\underbrace{Z'X}_{K\times K}\Big)^{-1}$$

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

$$\hat{\beta}_{2SLS} = (Z'X)^{-1} \cancel{Z'Z} \cancel{(X'Z)^{-1}} \cancel{X'Z} \cancel{(Z'Z)^{-1}} Z'y$$

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

2SLS for only one endogenous regressor

- Consider the special case where only one regressors is endogenous, say the K'th regressor x_{iK}.
- Then we have $z_i = (x_{i1}, \ldots, x_{i,K-1}, w_i)$, where w_i is a vector of additional instruments (at least one).
- We then also have $\hat{x}_{ik} = x_{ik}$ for $k = 1, \ldots, K-1$, i.e. we only need to run the first stage regression for x_{iK}.
- In that case:
  - First stage:
    Estimate $x_{iK} = z_i\gamma + \varepsilon_i$ by OLS, calculate $\hat{x}_{iK} = z_i\hat{\gamma}$.
  - Second stage:
    $+\hat{x}_{i,} \quad .\hat{x}_{i,K-1}$
    Estimate $y_i = x_{i1}\beta_1 + \ldots x_{i,K-1}\beta_{K-1} + \hat{x}_{i,K}\beta_K + v_i$ by OLS.
    The resulting estimator is $\hat{\beta}_{\mathrm{2SLS}}$.
- Thus, in that case calculating $\hat{\beta}_{\mathrm{2SLS}}$ requires running two OLS regressions.
- But in the following we continue to discuss the general case.

Large Sample Properties of 2SLS

- Analogous to the OLS estimator we now analyze the asymptotic properties of the 2SLS estimator as $n \to \infty$.
- We mostly follow the presentation in **[Wooldridge, Ch.5.2]**
- The most convenient form to write the 2SLS estimator is

$$\hat{\beta}_{\text{2SLS}} = (X'P_ZX)^{-1}X'P_Zy$$

where $P_Z = Z(Z'Z)^{-1}Z'$.

- Analogous to the OLS case we can use the model $Y = X\beta + u$ to obtain

$$\underline{\hat{\beta}_{\text{2SLS}} = \beta} + (X'P_ZX)^{-1}\underline{X'P_Zu}$$

- If $x_i = z_i$, then we have $\hat{\beta}_{\text{2SLS}} = \hat{\beta}_{\text{OLS}}$, and all of the following assumptions and results become the same as for OLS.

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

$$= (X'P_Z X)^{-1} X'P_Z (X\beta + u)$$

$$= \beta + (X'P_Z X)^{-1} X'P_Z u$$

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

Large Sample Properties of 2SLS (cont.)

Assumptions:

(A1) Exogeneity of Instruments:
$\mathbb{E}(z_i'u_i) = 0$

$$\mathbb{E}[z_i'(y_i - x_i\beta)] = 0$$

$$\text{rank}\,[\mathbb{E}[z_i'x_i]] = K$$

(A2) Non-Collinearity of Instruments:
$\mathbb{E}(z_i'z_i)$ exists, and $\text{rank}\,\mathbb{E}(z_i'z_i) = L$, i.e. $\mathbb{E}(z_i'z_i)$ is invertible

(A3) Relevance of Instruments:
$\mathbb{E}(z_i'x_i)$ exists, and $\text{rank}\,\mathbb{E}(z_i'x_i) = K$

$$\text{rank}\,[\Sigma_{zx}] = k$$

Theorem (Consistency of 2SLS)

Assume data are iid draws, that the the linear model $y_i = x_i\beta + u_i$ holds, and that assumptions A1, A2 and A3 are satisfied. Then $\hat{\beta}_{2\text{SLS}} \to_p \beta$ as $n \to \infty$.

$(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1} \leftarrow$ invertible by A3

A1

proof!

$$\hat{\beta}_{2SLS} - \beta = \underbrace{\left(\tfrac{1}{n}X'P_Z X\right)^{-1}}_{\to_p\,(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}} \underbrace{\tfrac{1}{n}X'P_Z u}_{\to_p\,0}$$

$\to_p \mathbb{E}[z_i'u_i] = 0$

$$\frac{1}{n}X'P_Z u = \frac{1}{n}X'Z\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{n}Z'u$$

$$\text{e.g. } \frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i'u_i \xrightarrow[p]{} \mathbb{E}[z_i'u_i]$$

$$\frac{1}{n}X'P_Z u = \underbrace{\frac{1}{n}X'Z}_{\substack{\frac{1}{n}\sum x_i'z_i \\ \downarrow p \\ \Sigma_{xz} = \mathbb{E}[z_i'z_i]}} \underbrace{\left(\frac{1}{n}Z'Z\right)^{-1}}_{\substack{\downarrow p \\ \Sigma_{zz} = \mathbb{E}[z_i'z_i]^{-1}}} \underbrace{\frac{1}{n}Z'u}_{\substack{\downarrow p \\ \mathbb{E}[z_iu_i] \\ = 0 \text{ (A1)}}}$$

$$\xrightarrow[p]{} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zu} = 0$$

$$\frac{1}{n}X'P_Z X = \underbrace{\frac{1}{n}X'Z}_{\substack{\downarrow p \\ \Sigma_{xz}}} \underbrace{\left(\frac{1}{n}Z'Z\right)^{-1}}_{\substack{\downarrow p \\ \Sigma_{zz}}} \underbrace{\frac{1}{n}Z'X}_{\substack{\downarrow p \\ \Sigma_{zx}}} \xrightarrow[p]{} \underbrace{\overbrace{\Sigma_{xz}\Sigma_{zz}^{-1}}^{\substack{W \\ k \times k}}\Sigma_{zx}}_{k \times k}$$

$L \times k$

Large Sample Properties of 2SLS (cont.)

Comments on relevance assumption (A3):

- Remember reduced form equation: $x_i = z_i\Gamma + \varepsilon_i$, with $\mathbb{E}(\varepsilon_i' z_i) = 0$.
- From our OLS analysis we know that the this reduced form parameter Γ satsfies $\boxed{\overset{L\times K}{\Gamma} = \mathbb{E}(z_i' z_i)^{-1}\mathbb{E}(z_i' x_i)}$, i.e. under (A2) we have $\text{rank}(\Gamma) = \text{rank}\mathbb{E}(z_i' x_i)$.
- Thus, the relevance condition (A3) can equivalently we written as $\text{rank}(\Gamma) = K$.
- (A3) implies that $L \geq K$.

single endogenous

Large Sample Properties of 2SLS (cont.)

One Additional assumption:

(A4) $\mathbb{E}(u_i^2 z_i' z_i)$ exists.

Theorem (Asymptotic Normality of 2SLS)

Assume data are iid draws, that the the linear model $y_i = x_i\beta + u_i$ holds, and that assumptions A1, A2, A3 and A4 are satisfied. Then as $n \to \infty$

$$\sqrt{n}\left(\hat{\beta}_{\mathrm{2SLS}} - \beta\right) \Rightarrow \mathcal{N}\left(0,\ W^{-1}\ V\, W^{-1}\right),$$

where

$$W = \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}, \quad V = \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{u^2zz}\Sigma_{zz}^{-1}\Sigma_{zx},$$

$$\Sigma_{xz} = \Sigma_{zx}' = \mathbb{E}(x_i' z_i), \quad \Sigma_{zz} = \mathbb{E}(z_i' z_i), \quad \Sigma_{u^2zz} = \mathbb{E}(u_i^2 z_i' z_i)$$

$Var(z_i' u_i)$

$$\sqrt{n}(\hat\beta_{2SLS}-\beta) = \underset{\xrightarrow{p} W^{-1} > 0}{\left(\frac{1}{n}X'P_ZX\right)^{-1}} \frac{1}{\sqrt{n}}X'P_Z u$$

$$\frac{1}{\sqrt{n}}X'P_Z u = \underset{\xrightarrow{p}\Sigma_{xz}}{\frac{1}{n}X'Z} \underset{\xrightarrow{p}\Sigma_{zz}^{-1}}{\left(\frac{1}{n}Z'Z\right)^{-1}} \underset{\xrightarrow{d} N(0, Var(z_i'u_i))}{\frac{1}{\sqrt{n}}Z'u}$$

$$\frac{1}{\sqrt{n}}Z'u = \frac{1}{\sqrt{n}}\sum z_i'u_i \Rightarrow N(0, Var(z_i'u_i))$$

$$\sqrt{n}(\hat\beta_{2SLS}-\beta) = \underset{\xrightarrow{p} W^{-1}}{\left(\frac{1}{n}X'P_ZX\right)^{-1}} \underset{\xrightarrow{p}\Sigma_{xz}}{\frac{1}{n}X'Z} \underset{\xrightarrow{p}\Sigma_{zz}^{-1}}{\left(\frac{1}{n}Z'Z\right)^{-1}} \underset{\xrightarrow{d} N(0, Var(z_i'u_i))}{\frac{1}{\sqrt{n}}Z'u}$$

$$\Rightarrow W^{-1}\Sigma_{xz}\Sigma_{zz}^{-1} N(0, Var(z_i'u_i))$$

$$= N\left(0,\ W^{-1}\overbrace{\Sigma_{xz}\Sigma_{zz}^{-1}Var(z_i'u_i)\Sigma_{zz}^{-1}\Sigma_{xz}}^{V}W^{-1}\right)$$

Large Sample Properties of 2SLS (cont.)

$$\sqrt{n}\left(\hat{\beta}_{\mathrm{2SLS}} - \beta\right) \Rightarrow \mathcal{N}\left(0, \Sigma_{\beta_{\mathrm{2SLS}}}\right), \quad \Sigma_{\beta_{\mathrm{2SLS}}} = W^{-1}\; V\, W^{-1}$$

- Need to estimate W and V. Consistent estimators are

$$\hat{W} = \hat{\Sigma}_{xz}\hat{\Sigma}_{zz}^{-1}\hat{\Sigma}_{zx}, \quad \hat{V} = \hat{\Sigma}_{xz}\hat{\Sigma}_{zz}^{-1}\hat{\Sigma}_{u^2zz}\hat{\Sigma}_{zz}^{-1}\hat{\Sigma}_{zx},$$

where

$$\hat{\Sigma}_{xz} = \hat{\Sigma}'_{zx} = \frac{1}{n}\sum_{i=1}^{n} x'_i z_i, \quad \hat{\Sigma}_{zz} = \frac{1}{n}\sum_{i=1}^{n} z'_i z_i, \quad \underline{\hat{\Sigma}_{u^2zz} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z'_i z_i},$$

where $\underline{\hat{u}_i = y_i - x_i\hat{\beta}_{\mathrm{2SLS}}}$ $\xrightarrow[p]{} u_i$

- Inference is based on the following approximation:

$$\boxed{\hat{\beta}_{\mathrm{2SLS}} \overset{a}{\sim} \mathcal{N}\left(\beta, \frac{1}{n}\hat{\Sigma}_{\beta_{\mathrm{2SLS}}}\right)}, \quad \hat{\Sigma}_{\beta_{\mathrm{2SLS}}} = \hat{W}^{-1}\; \hat{V}\, \hat{W}^{-1}$$

Large Sample Properties of 2SLS (cont.)

$W^{-1} V W^{-1} = \sigma^2 W^{-1}$

Homoscedasticity:

- If we also assume homoscedasticity, i.e. $\mathbb{E}(u_i^2 | z_i) = \sigma^2$, then

$= \sigma^2 \cancel{\Sigma_{zz}}$

$$V = \sigma^2 W$$

$V = \Sigma_{xz} \Sigma_{zz}^{-1} \mathbb{E}[z_i z_i' u_i^2] \cancel{\Sigma_{zz}} \Sigma_{zx} = \sigma^2 \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx}$

$= \sigma^2 W$

In this case we find

$$\Sigma_{\beta_{2SLS}} = W^{-1}\ V\, W^{-1} = \sigma^2 W^{-1}$$

- Can (but never should in practice) estimate $\Sigma_{\beta_{2SLS}}$ by

$$\hat{\Sigma}_{\beta_{2SLS}} = \hat{\sigma}^2 \hat{W}^{-1}, \quad \hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^{n} \hat{u}_i^2$$

Only valid under homoscedasticity!

Large Sample Properties of 2SLS (cont.)

In either way,

$$\hat{\beta}_{\mathrm{2SLS}} \overset{a}{\sim} \mathcal{N}\left(\beta, \frac{1}{n}\hat{\Sigma}_{\beta_{\mathrm{2SLS}}}\right)$$

Testing restrictions on β:

- All our large sample testing results for the t-test and Wald test remain unchanged.

- When we discussed large sample testing we only assumed that we have an estimator for β, which is now the 2SLS estimator, which is asymptotically normal and unbiased, and that we have a consistent estimator for the variance matrix of this estimator.

Finite Sample Theory for 2SLS?

- For the OLS estimator we discussed the finite sample properties and showed that under appropriate conditions the estimator is unbiased and normally distributed at finte *n*, and we also justified our variance estimator at finite sample.

- It is not possible to give such a nice finite sample justification for the 2SLS estimator. In fact, for $K = L$ and under standard distributional assumptions, the expected value of $\hat{\beta}_{\text{2SLS}}$ does not even exist.

$L = K$

$$\mathbb{E}[|\hat{\beta}_{2SLS}|] = \infty$$

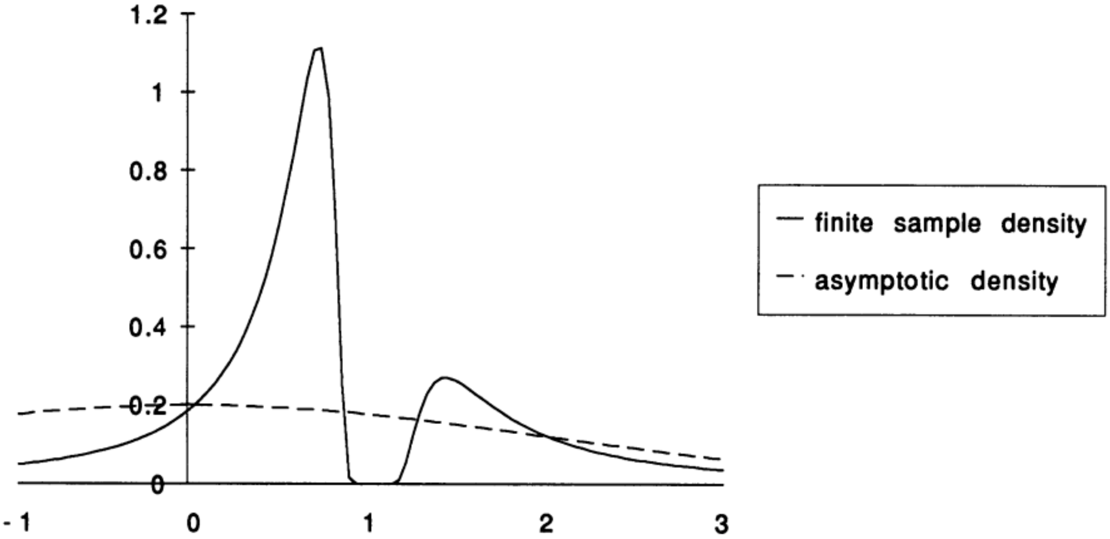
$L-K$ moments exist

Relevance of Instruments, Weak Instruments

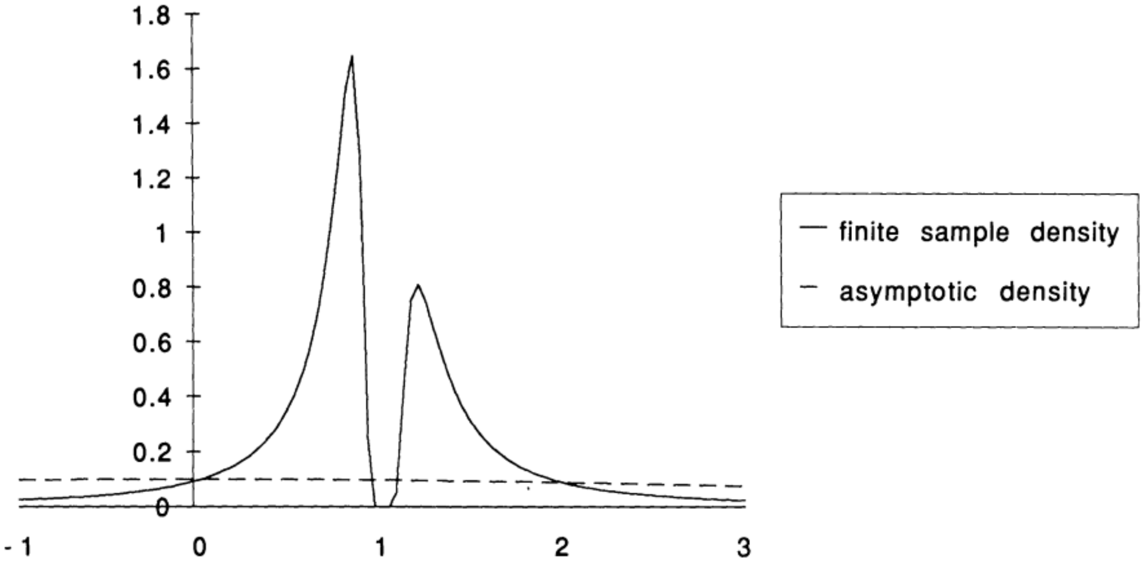
Relevance of Instruments, Weak Instruments

- We say that instruments are relevant if $\mathrm{rank}\mathbb{E}(z_i'x_i) = K$, or equivalently if $\mathrm{rank}(\Gamma) = K$. These conditions imply that the instruments z_i have explanatory power for all the regressors x_i.
- As $n \to \infty$ this is the "whole story", i.e. instruments are either relevant or not. However, at finite sample things are more complicated.
- We say that instruments are weak if they have little explanatory power for the endogenous regressors, e.g. for $K = L = 1$ the instrument is weak if Γ in $x_i = z_i\Gamma + \varepsilon_i$ is close to zero.
- If instruments are weak then the asymptotic theory above might give a bad approximation of the finite sample distribution.
- It is therefore important to test that instruments are sufficiently relevant.

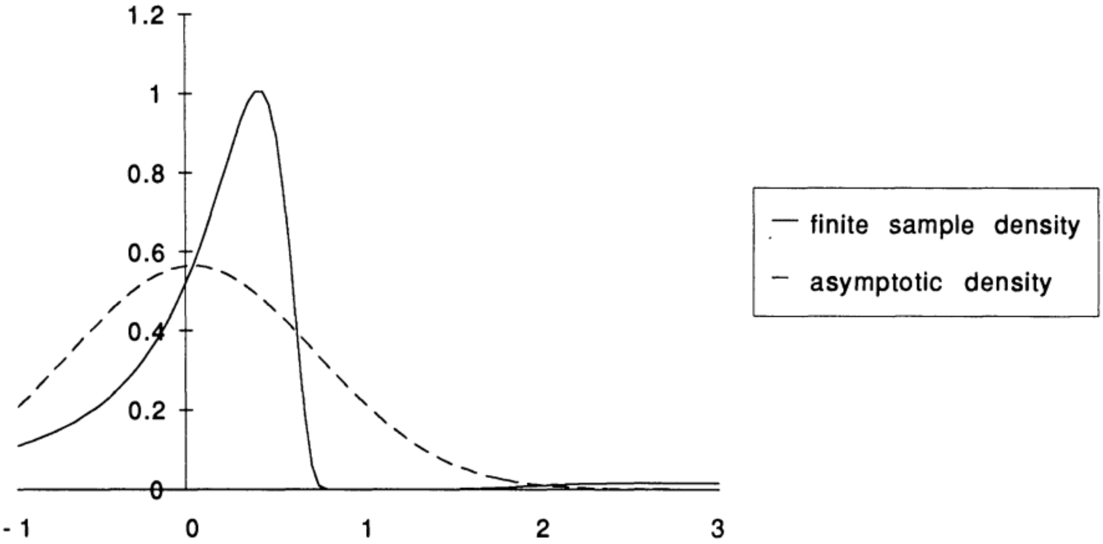
Relevance of Instruments, Weak Instruments (cont.)



Relevance of Instruments, Weak Instruments (cont.)



Relevance of Instruments, Weak Instruments (cont.)



Relevance of Instruments, Weak Instruments (cont.)

- It is standard practice to report the F-test (or t-test if only one excluded instruments) for testing significance of the excluded instruments in the first stage regression. This is a measure of the strength of the instruments.
- Example (as before): only endogenous regressors is x_{iK}, and one excluded instrument w_i, i.e. $z_i = (x_{i1}, \ldots, x_{i,K-1}, w_i)$. Then, first stage regression reads

  excluded IV

  $$x_{iK} = x_{i1}\gamma_1 + \ldots + x_{i,K-1}\gamma_{K-1} + w_i\gamma_K + \varepsilon_i.$$

  Null hypothesis of irrelevant instrument reads $H_0 : \gamma_K = 0$. Corresponding t-test statistics reads $t = \hat{\gamma}_K / \widehat{\mathrm{se}}(\hat{\gamma}_K)$.
- If the excluded instruments are not sufficiently significant in the first stage regression (i.e. if $|t|$ or F are not large enough), then one cannot expect the 2SLS estimator to have good properties.
- Rule of thumb: need $F \geq 10$ or $|t| \geq 3.2$ (p-value ≤ 0.0016)

Relevance of Instruments, Weak Instruments (cont.)

- While the rule $F \geq 10$ is widely used in practice, it is well known that such a pretest results in invalid inference
- A recent study suggests using conservative $F \geq 104$ instead
- In fact, one should use identification robust inference instead, which is valid regardless whether your IVs are weak or not
- One example of a valid test is the Anderson-Rubin (AR) test

AR test

- We want to test $H_0 : \beta = \beta_0$ vs. $H_a : \beta \neq \beta_0$
- Under the null, we can correctly compute the errors $u_i = y_i - x_i\beta = y_i - x_i\beta_0$

$$\mathbb{E}[z_i' u_i] = 0$$

- Hence, under the null $\mathbb{E}(z_i' u_i) = \mathbb{E}\left(z_i'(y_i - x_i\beta_0)\right) = 0$
- Effectively, we test whether a $L \times 1$ vector $v_i = z_i'(y_i - x_i\beta_0)$ has zero mean, i.e. we test $H_0 : \mathbb{E}v_i = 0$ vs. $H_a : \mathbb{E}v_i \neq 0$

$$H_0: \mathbb{E}[\underbrace{z_i'(y_i - x_i\beta_0)}_{v_i(\beta_0)}] = 0$$

$$\bar{V} = \frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta_0)$$

$$\frac{1}{\sqrt{n}} \sum v_i \overset{H_0}{\Rightarrow} N(0, \Sigma_v)$$

$$H_0: \mathbb{E}\underset{L\times 1}{v_i} = 0$$

$$n\,\bar{V}'\,\underset{\shortparallel}{\hat{\Sigma}_v^{-1}}\,\bar{V} \Rightarrow \chi^2_L$$

$$\hat{\Sigma}_v = \frac{1}{n}\sum v_i v_i'$$

$$v_i = z_i'(y_i - x_i\beta_0)$$

AR test (cont.)

- Under the null, $\mathbb{E}(v_i) = 0$ and we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_i \Rightarrow \mathcal{N}\left(0, \Sigma_v\right), \quad \Sigma_v = \mathbb{E}(v_i v_i'),$$

  which implies

$$n\bar{v}' \Sigma_v^{-1} \bar{v} \Rightarrow \chi_L^2, \quad \bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i$$

- Under the null, can consistently estimate Σ_v by

$$\hat{\Sigma}_v = \frac{1}{n} \sum_{i=1}^{n} v_i v_i',$$

  so

$$n\bar{v}' \hat{\Sigma}_v^{-1} \bar{v} \Rightarrow \chi_L^2$$

AR test (cont.)

- The Anderson-Rubin statistic for testing $H_0 : \beta = \beta_0$ is

$$\underline{AR(\beta_0)} = n\bar{v}(\beta_0)'\hat{\Sigma}_v(\beta_0)^{-1}\bar{v}(\beta_0) \overset{H_0}{\Rightarrow} \chi^2_L,$$

where

$$\bar{v}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta_0), \quad \hat{\Sigma}_v(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\beta_0)^2 z_i' z_i$$

- Reject $H_0 : \beta = \beta_0$ at significance level α if $AR(\beta_0) \underline{> \chi^2_{L,1-\alpha}}$
- Also can construct $1 - \alpha$ confidence sets for β as

$$CS_{1-\alpha} = \{\beta_0 : AR(\beta_0) \leqslant \chi^2_{L,1-\alpha}\}$$

Testing Overidentifying Restrictions

Testing Overidentifying Restrictions

- If $L > K$ we say that we are overidentified, because we have more exclusion restrictions than are needed to estimate β.
- In the overidentified case we can test if the exclusion restrictions $\mathbb{E}(z_i' u_i) = 0$ are valid.
- See **[Wooldridge, Ch.6.3.2]**

Testing Overidentifying Restrictions (cont.)

- If IVs are exogenous (and the model is correctly specified) then $\mathbb{E}\left(z_i'(y - x\beta_0)\right) = 0$ for some β_0

- This can be tested using J-statistic

$$J = \min_{\beta_0 \in \mathbb{R}^k} AR(\beta_0)$$

- If the instruments are strong, then, under H_0,

$$J \overset{H_0}{\Rightarrow} \boxed{\chi^2_{L-K}}$$

- The null hypothesis H_0 says that all instruments are exogenous. We reject this hypothesis at 5% significance level if J is larger than the 95% quantile of χ^2_{L-K}. If this is the case then this is a strong indication that at least some of the instruments are not exogenous.

- We will come back to that in the context of GMM framework