Unit 3. Instrumental variables estimation of single-equation linear models

Block A. The simple IV estimator

Instructor: Henry Redondo

Dpto. de Fundamentos del Análisis Económico
Universidad de Alicante

Econometrics II · Academic Year 2026/2027

Motivation

Every EU country must now require firms to report their gender pay gap (Directive 2023/970). Does making pay visible actually close the gap?

- Firms that publish their pay *choose* to do so. Comparing them with the rest mixes the policy with the kind of firm that adopts it.
- Unit 2 called this **endogeneity**. This unit is about how to get around it: find something that moves the regressor but is unrelated to the error.

Four blocks.

A. **The simple IV estimator.** One instrument, one endogenous regressor.
B. **Two-stage least squares.** Several instruments; weak instruments.
C. **GMM.** More moments than parameters; the optimal weighting.
D. **Multiple indicators.** Two noisy measures of the omitted variable.

Outline of Block A

1. A. The simple IV estimator
    - The exercise: Programa Transparenta
    - Identification and the IV estimator
    - Asymptotic properties of the IV estimator
    - Application: identification and estimation in our case
    - Asymptotic normality and the variance matrix of IV
    - Inference
    - Application: precision and inference in our case

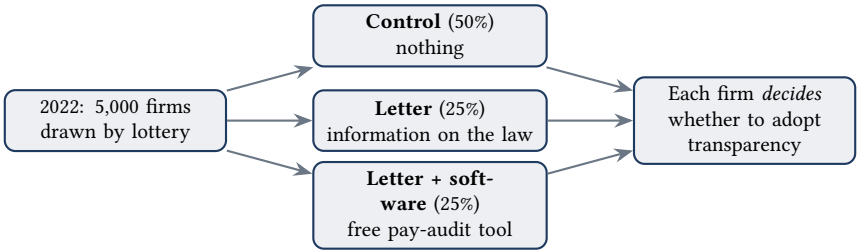
5,000 firms in the Comunitat Valenciana

Variable	Meaning
`wage_men_2021`, …	mean monthly wage, men and women, 2021 and 2025 (€)
`transp`	adopted pay transparency by 2025
`arm`	programme arm: control, letter, software
`size`, `n_women`, `n_men`	employees
`sector`, `province`	6 sectors, 3 provinces
`parent_abroad`	parent company abroad
`peer_share`	share of other firms in the same sector and province that adopted
`fair_score`	employees' pay-fairness rating (0–10)
`women_mgmt`	% of managers who are women

What the data look like

- Median firm: 44 employees. 32.2% of firms adopted transparency by 2025.
- Mean wage in 2021: men 2,150 €, women 1,821 €. Mean log gap **0.163**.
- The outcome is the change in the gap, $\Delta gap = gap_{2025} - gap_{2021}$, where $gap = \log(\bar{w}_{men}) - \log(\bar{w}_{women})$.

Programa Transparenta, an instrument we design



Nobody can *force* a firm to adopt transparency, but an agency can randomly choose whom to *encourage*: an A/B test where not every treated unit takes the treatment.

Exercise. Adopters' gap fell 4.8 pp; non-adopters' rose 1.2 pp. Is the effect about −6 pp? More? Less?

The rules of the game

These data are simulated. The firms, the programme and every number are invented for teaching. The design is inspired by real pay-transparency laws (Denmark, UK, Canada); the real evidence is at the end of the unit.

`transparency_firms.dta`: wages by sex, adoption, programme arm and firm characteristics.

The oracle. `transparency_oracle.dta`: each firm's egalitarian **culture**. It is what a real analyst never sees. We open it once, in Block A, to measure the bias.

Because we built the world, we know the true effect. We reveal it only when a simulation needs it (Block B), and we compare every estimator with it at the end.

The problem in our case

- The equation we want, with exogenous controls $\mathbf{w}_i$ (log size, share of women, sector, province):

$$\Delta gap_i = \beta_0 + \beta_1\, transp_i + \mathbf{w}_i'\boldsymbol{\delta} + u_i$$

- β_1: how much adopting transparency changes the gap (log points ≈ pp).
- u_i contains the firm's **culture**: egalitarian firms close their gap faster *and* adopt transparency more often.

$cov(transp, u) \neq 0$: `transp` is endogenous and OLS is inconsistent (Unit 2, omitted variables).

Exercise. Which way does the bias go? Culture lowers Δgap and raises adoption.

The next slides develop the theory of instrumental variables. We come back to this exercise once we know that the IV estimator is consistent.

An endogenous regressor

- Consider the linear model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u \tag{1}$$

where

$$E(u) = 0 \quad \text{and} \quad cov(x_j, u) = 0, \quad \text{for all } j = 1, .., k-1 \tag{2}$$

and x_k might be correlated with u

- We are then assuming that x_k is potentially endogenous in equation (1), whereas assumption (2) says that $x_1, .., x_{k-1}$ are exogenous in equation (1).
- The endogeneity might come from several reasons such as measurement errors, omitted variables, simultaneity, etc.
- As we already know, in general, if one of the regressors is endogenous, the *OLS* estimator of (1) is not a consistent estimator of any of the β_j. Moreover, without more information we cannot consistently estimate any of the parameters in equation (1).

Instrumental variables

- The method of instrumental variables (IV) provides a general solution to having an endogenous regressor. To use the IV approach we need a variable z_1 that is not in equation (1) and satisfies the following two conditions
  - **a**) z_1 is uncorrelated with u

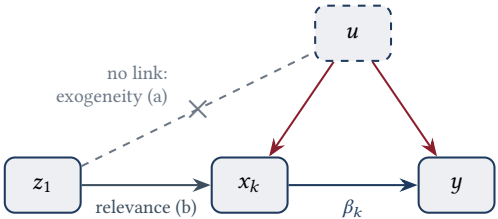
$$Cov(z_1, u) = 0$$

  - **b**) The coefficient of z_1 in the LP of x_k on $1, x_1, .., x_{k-1}, z_1$ is non zero, that is, if

$$L(x_k \mid 1, x_1, .., x_{k-1}, z_1) = \delta_0 + \delta_1 x_1 + .. + \delta_{k-1} x_{k-1} + \theta_1 z_1 \tag{3}$$

    is the LP of x_k on $1, x_1, .., x_{k-1}$ and z_1, assumption **b** is that $\theta_1 \neq 0$.
- When assumptions **a** and **b** are satisfied, z_1 is said to be **an instrumental variable** for x_k or **an instrument** for x_k.
- Notice that we have not imposed any restriction on the distribution of x_k or z_1. They can be both continuous or discrete, or one of them continuous and the other discrete.

Two conditions, one picture



(a) Exogeneity. $Cov(z_1, u) = 0$: the instrument affects y *only* through x_k. It **cannot be tested**.

(b) Relevance. $\theta_1 \neq 0$: the instrument moves x_k once the other exogenous regressors are netted out. It **can be tested**.

IV uses only the part of the variation in x_k that comes from z_1, which by (a) is unrelated to the unobservables in u.

Relevance and the reduced form

- Condition **b** is loosely described as "z_1 is correlated with x_k" but this statement is not correct in general. Condition **b** means that z_1 is partially correlated with x_k once the other exogenous variables $x_1, .., x_{k-1}$ have been netted out.
- Notice however that if x_k were the only explanatory variable in equation (1), the *LP* of x_k on the exogenous variables will just be

$$L(x_k \mid 1, z_1) = \delta_0 + \theta_1 z_1$$

where

$$\theta_1 = \frac{cov(x_k, z_1)}{var(z_1)}$$

- Then, in this case, condition **b** is equivalent to "z_1 is correlated with x_k" since

$$\theta_1 \neq 0 \Leftrightarrow cov(x_k, z_1) \neq 0$$

- The *LP* in (3) can be written in error form as:

$$x_k = \delta_0 + \delta_1 x_1 + .. + \delta_{k-1} x_{k-1} + \theta_1 z_1 + r_k \qquad (4)$$

and this equation is called the **reduced form equation** for the potentially endogenous explanatory variable x_k.

Identification

- We are now going to show that assumptions **a** and **b,** together with assumption (2), solve the **identification problem** in equation (1).
- By identification we mean that, under assumptions **a, b** and (2)**,** we can write the vector of structural parameters in equation (1) in terms of population moments of observable variables, and therefore, since the population moments can be consistently estimated by the corresponding sample moments, we can obtain a consistent estimator of the structural parameters.
- We can write equation (1) in vector notation as

$$y = \mathbf{x}'\boldsymbol{\beta} + u$$

where $\mathbf{x} = (1, x_2, .., x_k)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, .., \beta_k)'$, and let $\mathbf{z} = (1, x_2, .., x_{k-1}, z_1)'$ be the vector of all exogenous variables (the exogenous regressors and the instrument).

Orthogonality and rank conditions

- Assumptions **a** and (2) imply the following k orthogonality conditions.

$$E(\mathbf{z}u) = 0 \qquad IV1 \tag{5}$$

- and substituting $u = y - \mathbf{x}'\boldsymbol{\beta}$

$$E(\mathbf{z}y) = E(\mathbf{z}\mathbf{x}')\boldsymbol{\beta} \tag{6}$$

  where $E(\mathbf{z}\mathbf{x}')$ is a $k \times k$ matrix and $E(\mathbf{z}y)$ is a $k \times 1$ vector.
- Then, if $E(\mathbf{z}\mathbf{x}')$ has full rank, that is

$$\text{rank of } E(\mathbf{z}\mathbf{x}') = k \qquad IV2 \tag{7}$$

  the system of linear equations (6) has a unique solution that is

$$\boldsymbol{\beta} = E(\mathbf{z}\mathbf{x}')^{-1}E(\mathbf{z}y) \tag{8}$$

- It is clear that assumption **a** is needed for $IV1$, but, where have we used assumption **b**?

Where is relevance used?

- Assumption **b** is needed for $IV2$, since, if the exogenous variables are not perfectly collinear, that is $E(\mathbf{zz}')$ has full rank, $E(\mathbf{zx}')$ has full rank if and only if assumption **b** is satisfied.
- *Proof*
  - We can write

$$\mathbf{x} = \mathbf{\Pi z} + \mathbf{v}$$

    where

$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{I}_{k-1} & \mathbf{0}_{k-1,1} \\ \boldsymbol{\delta}' & \theta_1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} \mathbf{0}_{k-1,1} \\ r_k \end{bmatrix}$$

  - Then

$$E(\mathbf{zx}') = E(\mathbf{zz}')\mathbf{\Pi}' + E(\mathbf{zv}')$$

    and since r_k is the error of the LP of x_k on $\mathbf{z}$, $E(\mathbf{z}r_k) = 0$ and

$$E(\mathbf{zx}') = E(\mathbf{zz}')\mathbf{\Pi}'$$

    Since $E(\mathbf{zz}')$ has full rank, $E(\mathbf{zx}')$ has full rank if and only if $\mathbf{\Pi}$ has full rank, that is if and only if $\theta_1 \neq 0$

The IV estimator

- Given a random sample, $\{(y_i, \mathbf{x}_i', z_{1i})' : i = 1, .., N\}$, from the population, the **instrumental variables (IV) estimator** of $\boldsymbol{\beta}$ is the sample counterpart of (8)

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i y_i\right) = \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{z}_i y_i\right) \tag{9}$$

- $\widehat{\boldsymbol{\beta}}$ can also be written using matrix notation as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_N' \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix}, \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Remarks on the IV estimator

Remarks

- The *OLS* estimator of β is an *IV* estimator that uses x_k as an instrument for itself.
- When searching for an instrument for an endogenous explanatory variable both conditions **a** and **b** are equally important.
- However, there is a practically important difference between them:
  - Condition **b** can be tested (it is just to test the significance of z_1 in the regression of x_k on $1, x_2, .., x_{k-1}$ and z_1)
  - Condition **a** has to be maintained and cannot be tested because it involves the unobservable u.

Where do instruments come from?

- **Randomised encouragement.** Participation is voluntary and endogenous, but eligibility or encouragement is assigned at random: job-training offers, school vouchers, our letters. In tech it is an A/B test with non-compliance: you randomise the *invitation* to a new feature, not its use.
- **Natural experiments and policy rules.** Some feature of the setting, often unintended, produces variation in the endogenous regressor that is unrelated to u. Size thresholds in labour law are a classic source: Denmark's 2006 pay-statistics law applied to firms with at least 35 employees; the UK's reporting rule (first reports in 2018) applies from 250 employees.
- The price: we must **argue** exogeneity for each of them, as we do when we grade candidate instruments in the application.

Where do instruments come from? (cont.)

- Regional variation in prices or taxes are sometimes used as instruments for endogenous explanatory variables appearing in individual-level equations.
  - For example, in estimating the effects of alcohol consumption on performance in college, the local price of alcohol can be used as an instrument for alcohol consumption, provided other regional factors that affect college performance have been appropriately controlled for.
  - The idea is that the price of alcohol, including any taxes, can be assumed to be exogenous to each individual.

More than one endogenous regressor

- IV estimation can also be applied when there is more than one potentially endogenous explanatory variable in the structural equation.
  - The only thing is that we need at least as many instruments as potentially endogenous regressors.
- Assume that we have J potentially endogenous explanatory variables and J instruments $z_1, z_2, ..z_J$, and let $\mathbf{z}$ be the vector of all the exogenous x_j and all the instruments.
- If $\mathbf{z}$ satisfies $IV1$ and $IV2$, the vector $\boldsymbol{\beta}$ is also identified and the IV estimator is defined by (9).
- The only problem is that if $J \geq 2$, assumption $IV2$ cannot be so easily tested.
  - Assumption $IV2$ can always be written in terms of the coefficients of the LP of all the endogenous regressors on all the exogenous variables (including the instruments), but the condition is on the rank of a matrix and therefore is more difficult to test.

The rank condition with two endogenous regressors

- For instance if $J = 2$, and x_{k-1} and x_k are the potentially endogenous regressors and z_1 and z_2 are the instruments, we can write the LP of x_{k-1} and the LP of x_k on all the exogenous variables as

$$\begin{aligned} x_{k-1} &= \delta_{k-1,0} + \delta_{k-1,1}x_1 + .. + \delta_{k-1,k-2}x_{k-2} \\ &\quad +\theta_{k-1,1}z_1 + \theta_{k-1,2}z_2 + r_{k-1} \\ x_k &= \delta_{k,0} + \delta_{k,1}x_1 + .. + \delta_{k,k-2}x_{k-2} \\ &\quad +\theta_{k,1}z_1 + \theta_{k,2}z_2 + r_k \end{aligned}$$

- Then, condition $IV2$ is satisfied if and only if the matrix of coefficients

$$\begin{pmatrix} \theta_{k-1,1} & \theta_{k-1,2} \\ \theta_{k,1} & \theta_{k,2} \end{pmatrix}$$

has full rank.

Asymptotic properties of the IV estimator

- In this section we study the asymptotic properties of the IV estimator.
- We first study under which assumptions the IV estimator is consistent.
- Next, we obtain the asymptotic distribution of the IV estimator.
- Finally, we find the estimated variance matrix of the IV estimator.

Consistency of IV

Theorem (consistency of IV). Under assumptions $IV1$ and $IV2$, the IV estimator is a **consistent estimator** of $\boldsymbol{\beta}$.

- *Proof*
  - Substituting $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ in (9), we can write the IV estimator $\widehat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{z}_i y_i\right) = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i'\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i u_i\right)$$

  - Given that we have a random sample, $\mathbf{z}_i u_i$ are iid, using the law of large numbers we have that

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i u_i \rightarrow_p E(\mathbf{z}u) \underset{\text{by } IV1}{=} 0 \tag{10}$$

Proof of consistency of IV, continued

- *Proof (cont.)*
  - $\mathbf{z}_i\mathbf{x}_i'$ are also *iid*, and using the law of large numbers we have that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i' \rightarrow_p E(\mathbf{z}\mathbf{x}')$$

  - and since $E(\mathbf{z}\mathbf{x}')$ is non-singular (by $IV2$), using the continuous function theorem, we have that

$$\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1} \rightarrow_p E(\mathbf{z}\mathbf{x}')^{-1} \tag{11}$$

  - Then, using (10) and (11) we have that

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right) \rightarrow_p \boldsymbol{\beta}$$

Analysis A0 · Was the lottery fair?

Means before the programme, by arm

	Control	Letter	Software
Firms	2,550	1,263	1,187
Gap in 2021	0.164	0.163	0.159
Employees	62.0	62.7	62.4
Share of women	0.509	0.504	0.510
Parent abroad	0.098	0.109	0.113
Fairness score	5.90	5.94	5.85
Adopted by 2025	**0.253**	**0.305**	**0.487**

- Joint test that the 2021 gap is equal across arms: $F = 0.74$, $p = 0.48$. Same for firm size: $p = 0.57$.
- Before the programme the three groups look alike: this is an **A/A test**.
- After it, adoption differs a lot. The lottery *moved* adoption without being related to anything else about the firm: exactly what an instrument must do.

Analysis A1 · The naive answer, and what the oracle says

OLS with and without the oracle (controls X, robust s.e.)

	Coef. on `transp`	Robust s.e.
OLS: `dgap ~ transp + X`	−0.0561	0.0019
With the oracle: `dgap ~ transp + culture + X`	−0.0310	0.0020

- OLS: transparency closes the gap by 5.6 pp, with a tiny standard error.
- Controlling for culture (coefficient −0.0300, s.e. 0.0010), the effect falls to **3.1 pp**. Almost half of the OLS number was culture.
- Adopters have mean culture +0.54; non-adopters −0.37 (culture has s.d. ≈ 1).

In real life there is no oracle. We need a way to get −0.031 *without* observing culture.

Is the invitation an instrument?

Let $invited_i = 1$ if the firm was drawn into either arm of Programa Transparenta.

(a) Exogeneity: $Cov(invited, u) = 0$**.**

- The lottery decided who was invited, so *invited* is independent of culture and of everything else in u.
- **Exclusion**: a letter should change the gap *only* by changing whether the firm adopts transparency.

(b) Relevance: $\theta_1 \neq 0$**.**

- Invited firms should adopt more often.
- This one we can **check** in the data: regress *transp* on *invited* and the controls.

Exercise. Randomisation guarantees independence. Does it also guarantee exclusion? Give a story in which the letter affects the gap directly.

Is the invitation an instrument?

Let $invited_i = 1$ if the firm was drawn into either arm of Programa Transparenta.

(a) Exogeneity: $Cov(invited, u) = 0$**.**

- The lottery decided who was invited, so *invited* is independent of culture and of everything else in u.
- **Exclusion**: a letter should change the gap *only* by changing whether the firm adopts transparency.

(b) Relevance: $\theta_1 \neq 0$**.**

- Invited firms should adopt more often.
- This one we can **check** in the data: regress $transp$ on *invited* and the controls.

Exercise. Randomisation guarantees independence. Does it also guarantee exclusion? Give a story in which the letter affects the gap directly.

Answer. The letter reminds managers of the gap and they raise women's pay *without* publishing anything. Then *invited* enters u, and exclusion fails even though the lottery was random.

Analysis A2 · The IV estimator, by hand

Three regressions (controls X, robust s.e.)

	Coef.	s.e.
First stage: *transp* on *invited*	0.1395	0.0129
Reduced form: Δgap on *invited*	−0.0052	0.0020
IV (`ivreg`)	**−0.0375**	0.0132

The Wald estimator.

$$\widehat{\beta}_1^{IV} = \frac{\text{reduced form}}{\text{first stage}} = \frac{-0.0052}{0.1395} = -0.0375$$

- Invited firms were **14 pp** more likely to adopt ($F = 118$).
- Their gap fell by **0.5 pp** more: the *intention-to-treat* effect, adopters or not.
- If the invitation works *only* through adoption, the effect of adopting is $0.52/0.14 \approx 3.7$ pp.

IV rescales the effect of the lottery by how much the lottery moved the regressor.

Grading the candidate instruments

Instrument for *transp*	(a) Exogenous?	(b) Relevant?	The story that breaks it
`invited` (lottery)	✓	✓	a letter that changes pay directly
`parent_abroad`	doubtful	✓	multinationals differ in culture
`peer_share`	×	✓	same sector and town, same culture
Last digit of the tax ID	✓	×	unrelated to adoption
Size ≥ 250 (EU rule)	near the cutoff	✓	size affects pay; firms stop at 249

The most *relevant* candidates tend to have the weakest exclusion arguments. Relevance can be tested; exogeneity has to be argued.

Analysis A3 · What bad instruments do

`ivreg(dgap ~ transp + X | z + X)`, robust s.e.

z	First-stage F	$\widehat{\beta}_1$	s.e.
`invited`	118	−0.038	0.013
`parent_abroad`	72	−0.111	0.015
`peer_share`	93	−0.188	0.021
OLS		−0.056	0.002

- All three instruments are strongly relevant.
- The two invalid ones land **further from the oracle (−0.031) than OLS does**, with reassuring standard errors.
- $corr(z, culture)$: 0.001 (`invited`), 0.14 (`parent_abroad`), 0.42 (`peer_share`).

A strong first stage does not make an instrument valid. An invalid IV can be worse than doing nothing about endogeneity.

Asymptotic normality of IV

Theorem (asymptotic normality of IV). Under $IV1$ and $IV2$, if $E(u^2\mathbf{zz}')$ exists, the IV estimator is asymptotically normal and the asymptotic distribution is

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d N(0, E(\mathbf{zx}')^{-1}E(u^2\mathbf{zz}')E(\mathbf{xz}')^{-1})$$

- *Proof*
  - We have seen that
  $$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right)$$
  - Multiplying by $\sqrt{N}$ we have
  $$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right)$$
  - Given that we have a random sample, $\mathbf{z}_i u_i$ are iid, and since $E(\mathbf{z}_i u_i) = 0$ (by $IV1$) and $E(u_i^2\mathbf{z}_i\mathbf{z}_i')$ exists, using the central limit theorem we have that
  $$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right) \to_d N(0, E(u^2\mathbf{zz}')) \tag{12}$$

Proof of asymptotic normality of IV, continued

- *Proof (cont.)*
  - $\mathbf{z}_i\mathbf{x}_i'$ are also *iid*, then, using the law of large numbers

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i' \to_p E(\mathbf{z}\mathbf{x}')$$

  - and since $E(\mathbf{z}\mathbf{x}')$ is non-singular (by $IV2$), using the continuous function theorem, we have that

$$\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1} \to_p E(\mathbf{z}\mathbf{x}')^{-1} \tag{13}$$

  - Then, using (12), (13) and Slutsky's theorem, we have that

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \left(\tfrac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\sqrt{N}\left(\tfrac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right)\\ &\to_d N(0, E(\mathbf{z}\mathbf{x}')^{-1}E(u^2\mathbf{z}\mathbf{z}')E(\mathbf{x}\mathbf{z}')^{-1})\end{aligned}$$

IV under homoskedasticity

- Analogously as for the *OLS* estimator we can make some homoskedasticity assumption. The weakest **homoskedasticity assumption** is

$$E(u^2\mathbf{z}\mathbf{z}') = \sigma^2 E(\mathbf{z}\mathbf{z}') \qquad IV3$$

  where

$$\sigma^2 = E(u^2)$$

- Then, under $IV1$, $IV2$ and $IV3$,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d N(0, \sigma^2 E(\mathbf{z}\mathbf{x}')^{-1} E(\mathbf{z}\mathbf{z}') E(\mathbf{x}\mathbf{z}')^{-1})$$

Variance matrix of IV under homoskedasticity

- Under assumptions $IV1$-$IV3$, it can be shown that

$$\widehat{\sigma}^2 = \frac{1}{N-k}\sum_{i=1}^{N}\widehat{u}_i^2 \rightarrow_p \sigma^2$$

  where $\widehat{u}_i$ are the residuals based on the IV estimator.
- Then, we have that

$$\widehat{\sigma}^2\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1} \rightarrow_p \sigma^2 E(\mathbf{z}\mathbf{x}')^{-1}E(\mathbf{z}\mathbf{z}')E(\mathbf{x}\mathbf{z}')^{-1}$$

- and the estimated variance matrix of $\widehat{\boldsymbol{\beta}}$ that is appropriate when the homoskedasticity assumption is satisfied is given by

$$\begin{aligned}\widehat{var(\widehat{\boldsymbol{\beta}})} &= \frac{1}{N}\widehat{\sigma}^2\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1}\\ &= \widehat{\sigma}^2\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1}\end{aligned}$$

Robust variance matrix of IV

- When $IV3$ is not satisfied, the estimated variance of $\widehat{\boldsymbol{\beta}}$ given above is no longer valid.
- However, it can be shown that, under $IV1$ and $IV2$

$$\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'\right) \rightarrow_p E(u^2\mathbf{z}\mathbf{z}')$$

- Then

$$\left(\tfrac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\tfrac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'\right)\left(\tfrac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1}$$

$$\rightarrow_p E(\mathbf{z}\mathbf{x}')^{-1}E(u^2\mathbf{z}\mathbf{z}')E(\mathbf{x}\mathbf{z}')^{-1}$$

Robust variance matrix of IV (cont.)

- and the heteroskedasticity-robust variance matrix estimator of $\widehat{\boldsymbol{\beta}}$ is given by

$$\begin{aligned}\widehat{var(\widehat{\boldsymbol{\beta}})} &= \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1} \\ &= \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'\right)\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)^{-1} \qquad (14)\end{aligned}$$

Wald tests with IV

- Let's consider that we want to test a set of J linear restrictions

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$
$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

  where $\mathbf{R}$ is a $J \times k$ matrix and $\mathbf{r}$ is a $J \times 1$ vector.

- We know that under $IV1$ and $IV2$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow_d N(0, \mathbf{B}) \tag{15}$$

  where

$$\mathbf{B} = E(\mathbf{z}\mathbf{x}')^{-1} E(u^2 \mathbf{z}\mathbf{z}') E(\mathbf{x}\mathbf{z}')^{-1}$$

- Furthermore if $IV3$ is satisfied the expression for $\mathbf{B}$ simplifies to

$$\mathbf{B} = \sigma^2 E(\mathbf{z}\mathbf{x}')^{-1} E(\mathbf{z}\mathbf{z}') E(\mathbf{x}\mathbf{z}')^{-1}$$

Wald tests with IV, continued

- Using (15), we have that

$$\sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \to_d N(0, \mathbf{RBR}')$$

- We know that

$$N\widehat{var(\widehat{\boldsymbol{\beta}})} \to_p \mathbf{B}$$

- Then

$$N\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}' \to_p \mathbf{RBR}'$$

- and the $Wald$ test statistic is

$$\begin{aligned} W &= \sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(N\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}\sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \\ &= (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \to_d \chi^2_J \quad \text{under } H_0 \end{aligned}$$

- If $IV3$ is satisfied $\widehat{var(\widehat{\boldsymbol{\beta}})}$ can be replaced by $\widehat{var(\widehat{\boldsymbol{\beta}})}$.
- If the significance level is α, the rejection region is $\{W > \chi^2_{J,\alpha}\}$.

Remarks on inference with IV

Remarks

- Notice that the W statistic based on the IV estimator has the same expression as the W statistic based on the OLS estimator.
- Analogously, if we want to test one single restriction using the IV estimator, we can use the t statistic proposed for OLS replacing the OLS estimator by the IV estimator and the estimated variance of the OLS estimator by the estimated variance of the IV estimator.
- The asymptotic distribution of the t statistic will be standard normal under the null and the rejection regions will be the same as the rejection regions for the t statistic based on OLS.

Analysis A4 · The price of identification

OLS vs. IV with `invited`

	$\widehat{\beta}_1$	Robust s.e.	95% CI
OLS	−0.056	0.0019	[−0.060, −0.052]
IV	−0.038	0.0132	[−0.063, −0.012]



- The IV standard error is **6.8 times** the OLS one.
- Why: IV uses only the variation in *transp* that comes from the lottery, which is 14 pp out of a regressor that varies between 0 and 1.
- OLS is precise and *wrong*; IV is right on average and *noisy*.

The same trade-off as in model selection: IV buys consistency with variance. Block B shows how to buy some of it back.

Block A in R

```
library(AER); library(sandwich); library(lmtest)
d <- read.csv("data/transparency_firms.csv")
d$sector <- factor(d$sector); d$province <- factor(d$province)
d$gap2021 <- log(d$wage_men_2021/d$wage_women_2021)
d$dgap <- log(d$wage_men_2025/d$wage_women_2025) - d$gap2021
d$lsize <- log(d$size); d$share_women <- d$n_women/d$size
d$invited <- as.numeric(d$arm >= 1)
X <- "lsize + share_women + sector + province"
rob <- function(m, t = "HC1") coeftest(m, vcovHC(m, t))

# A0: was the lottery fair?
rob(lm(gap2021 ~ factor(arm), d))
# A1: naive OLS
ols <- lm(paste("dgap ~ transp +", X), d); rob(ols)
# A2: first stage, reduced form, IV
lm(paste("transp ~ invited +", X), d)
lm(paste("dgap ~ invited +", X), d)
iv <- ivreg(paste("dgap ~ transp +", X, "| invited +", X), data = d)
rob(iv, "HC0")
# A3: bad instruments
ivreg(paste("dgap ~ transp +", X, "| peer_share +", X), data = d)
```

Full script: `01_class_analysis.R`. Arms are coded 0 = control, 1 = letter, 2 = letter + software.

Block A in R: try it yourself

1. Divide the reduced-form coefficient by the first-stage one. Do you get the `ivreg` estimate exactly?
2. Drop `X` from all three regressions. How much does the IV estimate move? Why so little?
3. Compare the IV standard error with `vcov(iv)` and with `vcovHC(iv, "HC0")`.

Summary of the simple IV estimator

1. An instrument must be **exogenous** ($E(\mathbf{z}u) = 0$, not testable) and **relevant** (partial correlation with x_k, testable).
2. Under $IV1$ and $IV2$: $\boldsymbol{\beta} = E(\mathbf{z}\mathbf{x}')^{-1}E(\mathbf{z}y)$ and $\widehat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$. With one instrument: reduced form ÷ first stage.
3. IV is consistent and asymptotically normal; use robust standard errors unless $IV3$ is credible.
4. Wald and t tests work exactly as with OLS, replacing the estimator and its variance.
5. Consistency has a price: much larger standard errors.

The case so far

OLS says −5.6 pp; the lottery says −3.8 pp (s.e. 1.3 pp).