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

Block B. Two-stage least squares

Instructor: Henry Redondo

Dpto. de Fundamentos del Análisis Económico
Universidad de Alicante

Econometrics II · Academic Year 2026/2027

Outline of Block B

1. B. Two-stage least squares
    - The 2SLS estimator
    - Asymptotic properties of the 2SLS estimator
    - Potential pitfalls with IV estimation: weak instruments
    - Application: the programme had two arms

Several instruments for one regressor

- Consider the linear model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u \tag{16}$$

where

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

and x_k is potentially correlated with u.

- Assume now that we have more than one instrument for x_k. Let $z_1, z_2, .., z_M$ be variables that are not in (16) and satisfy

$$cov(z_h, u) = 0, \quad h = 1, .., M$$

- If each of the z_h is partially correlated with x_k once $x_1, .., x_{k-1}$ have been netted out, we could have M different IV estimators.
- Actually we could have an infinite number of different IV estimators because any linear combination of $x_1, .., x_{k-1}, z_1, .., z_M$ will be uncorrelated with u, and those linear combinations that are partially correlated with x_k (once $x_1, .., x_{k-1}$ have been netted out) are valid instruments.

Choosing the best instrument

- The question then is, how can we choose among all the potential instruments?
- We will first derive the $2SLS$ estimator intuitively and we will show later that under some assumptions the $2SLS$ estimator is somehow optimal.
- Let z be the $L \times 1$ ($L = k + M$) vector of all the exogenous variables

$$\mathbf{z} = (1, x_1, .., x_{k-1}, z_1, ..z_M)'$$

- The $2SLS$ method chooses to use as instrument for x_k the linear combination of the elements of $\mathbf{z}$ that is most highly correlated with x_k.
- The linear combination of the elements of $\mathbf{z}$ that is most highly correlated with x_k is the LP of x_k on $\mathbf{z}$

$$x_k^* = L(x_k \mid \mathbf{z}) = \delta_0 + \delta_1 x_1 + .. + \delta_{k-1} x_{k-1} + \theta_1 z_1 + .. + \theta_M z_M$$

The 2SLS estimator

- If x_k^* were observable we could use it as an instrument for x_k, but x_k^* is not observable because $\delta_0, \delta_1, .., \delta_{k-1}, \theta_1, .., \theta_M$ are population parameters that are unknown.
- However, these parameters can be consistently estimated by *OLS* in the regression of x_k on $1, x_1, .., x_{k-1}, z_1, .., z_M$.
- Let $\widehat{\delta}_0, \widehat{\delta}_1, .., \widehat{\delta}_{k-1}, \widehat{\theta}_1, .., \widehat{\theta}_M$ be these *OLS* estimates, then the 2*SLS* method uses as instrument for each observation on x_{ki}

$$\widehat{x}_{ki} = \widehat{\delta}_0 + \widehat{\delta}_1 x_{1i} + .. + \widehat{\delta}_{k-1} x_{k-1i} + \widehat{\theta}_1 z_{1i} + .. + \widehat{\theta}_M z_{Mi}$$

- and the 2*SLS* estimator is given by

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{x}}_i \mathbf{x}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{x}}_i y_i \right) = \left(\sum_{i=1}^{N} \widehat{\mathbf{x}}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^{N} \widehat{\mathbf{x}}_i y_i \right) \tag{17}$$

where $\widehat{\mathbf{x}}_i = (1, x_{1i}, ..x_{k-1i}, \widehat{x}_{ki})'$.

2SLS in matrix notation

- We can also write the $2SLS$ estimator in matrix notation as

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} \tag{18}$$

where

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

2SLS is an OLS estimator

- The *2SLS* estimator turns out to be an *OLS* estimator since

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

- *Proof*
  - Notice that

$$\widehat{\mathbf{x}}_k = \begin{pmatrix} \widehat{x}_{k1} \\ \vdots \\ \widehat{x}_{kN} \end{pmatrix} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}_k$$

  - Also, for the exogenous x_j (including the 1), if we run the regression of x_j on z, we can also write

$$\widehat{\mathbf{x}}_j = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}_j$$

    and $\widehat{\mathbf{x}}_j = \mathbf{x}_j$ because x_j is one of the elements of z.
  - Then, we can write $\widehat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$, and

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

  - and substituting in (18) we have that

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

2SLS in two steps

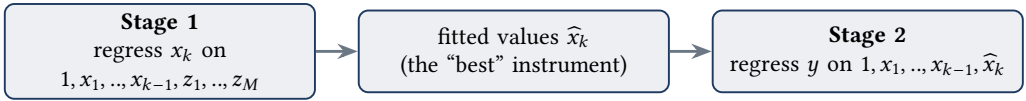
Remarks

- To summarize, the $2SLS$ estimator can be computed in two steps:
  - **1.** Obtain the fitted values $\widehat{x}_k$ from the regression of x_k on $1, x_1, .., x_{k-1}, z_1, .., z_M$. This is called the **first-stage regression.**
  - **2.** Run the OLS regression of y on $1, x_1, .., x_{k-1}, \widehat{x}_k$. This is called the **second-stage regression**, and it produces the $\widehat{\beta}_j$.
- Since $\widehat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$, substituting in (18), we have that the $2SLS$ estimator can also be computed as

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \tag{19}$$

- In practice it is better to use a statistical package that has a $2SLS$ command instead of using the 2-step procedure. The reason is that the usual standard errors in the second step OLS regression are not valid for the $2SLS$ estimator. We will show this later.

2SLS in two stages



Do not report the stage-2 standard errors. The stage-2 residuals are computed with $\widehat{x}_k$ instead of x_k, so $\widehat{\sigma}^2$ and hence the standard errors are wrong. The 2SLS residuals must use the **original** x_k: $\widehat{u}_i = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$.

Use a 2SLS command (e.g. `AER::ivreg` in R): same coefficients, correct standard errors.

2SLS and simple IV

- The $2SLS$ estimator and the simple IV estimator are identical when there is just one instrument for x_k.
- If we have J potentially endogenous explanatory variables in the structural equation and $M > J$ instruments, we can define the $2SLS$ estimator as in (17), where $\widehat{\mathbf{x}}_i$ will include all the exogenous explanatory variables and all the predictions (based on the instruments and the exogenous explanatory variables) for the endogenous explanatory variables.

2SLS in the general case

- We are now going to study the general case where we can have more than one potentially endogenous explanatory variable.
- Let's write the model in vector notation as

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

  where $\mathbf{x} = (1, x_2, .., x_k)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, .., \beta_k)'$.
- Assume that we have J potentially endogenous regressors and M instruments, and let $\mathbf{z}$ be the $L \times 1$ vector of all the exogenous explanatory variables and instruments.
- What are the analogous conditions to $IV1$ and $IV2$?.
  - The equivalent to $IV1$ is

$$2SLS1: \qquad E(\mathbf{z}u) = 0$$

  - The equivalent to $IV2$ is

$$2SLS2: \qquad a) \text{ rank } E(\mathbf{z}\mathbf{z}') = L \qquad b) \text{ rank } E(\mathbf{z}\mathbf{x}') = k$$

  - and we will show that under $2SLS1$ and $2SLS2$ the $2SLS$ estimator is a consistent estimator of $\boldsymbol{\beta}$.

Order and rank conditions

Remarks

- Assumption $2SLS1$ cannot be tested in general.
- Assumption $2SLS2a$) is technically very important but it will always be satisfied.
- Assumption $2SLS2b$) is called the **rank condition**.
  - A necessary condition is that $L \geq k$, that is, we have at least as many instruments as potentially endogenous regressors. This is called the **order condition**. When $L = k$ we say the model is **just identified** and when $L > k$ we say the model is **overidentified**.
  - Assumption $2SLS2b$) can be easily tested when we have just one potentially endogenous regressor. We have to test

$$H_0 : \theta_1 = \theta_2 = ... = \theta_M = 0$$

    where $\theta_1, \theta_2, ..., \theta_M$ are the coefficients of the instruments in the LP of x_k on $\mathbf{z}$.
  - If we have more than one potentially endogenous regressor, testing $2SLS2b$) is not so trivial. This assumption can be written in terms of the coefficients of the LP of all the endogenous variables on $\mathbf{z}$ but the test refers to the rank of a submatrix of those coefficients.

Asymptotic properties of the 2SLS estimator

- In this section we study the asymptotic properties of the 2SLS estimator.
- We first study under which assumptions the 2SLS estimator is consistent.
- Next, we obtain the asymptotic distribution of the 2SLS estimator.
- Finally, we find the estimated variance matrix of the 2SLS estimator.

$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is a product of sample moments ($\frac{1}{N}\sum \mathbf{x}_i\mathbf{z}_i'$, $\frac{1}{N}\sum \mathbf{z}_i\mathbf{z}_i'$, their inverses) times $\frac{1}{N}\sum \mathbf{z}_i u_i$. LLN + continuous mapping for the first, LLN (consistency) or CLT (normality) for the last, then Slutsky.

Consistency of 2SLS

Theorem (consistency of 2SLS). If assumptions $2SLS1$ and $2SLS2$ are satisfied, the $2SLS$ estimator based on a random sample, $\{(y_i, \mathbf{x}_i', \mathbf{z}_i')' : i = 1, .., N\}$, is a consistent estimator of $\boldsymbol{\beta}$.

- *Proof*
  - Substituting $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ in (19), we have that

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}\end{aligned}$$

  - and using vector notation

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right) \quad (20)$$

Proof of consistency of 2SLS, continued

- *Proof (cont.)*
  - $\mathbf{z}_i u_i$ are iid and $E(\mathbf{z}_i u_i) = 0$ (by $2SLS1$), then using the LLN

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i u_i \rightarrow_p 0 \tag{21}$$

  - $\mathbf{z}_i \mathbf{z}_i'$ are iid, then using the LLN

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

  - and since $E(\mathbf{z}\mathbf{z}')$ is non-singular (by $2SLS2a$), using the continuous function theorem, we have that

$$\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i'\right)^{-1} \rightarrow_p E(\mathbf{z}\mathbf{z}')^{-1} \tag{22}$$

Proof of consistency of 2SLS, continued

- *Proof (cont.)*
  - $\mathbf{x}_i\mathbf{z}_i'$ are iid, then, using the *LLN*,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i' \to_p E(\mathbf{x}\mathbf{z}') \tag{23}$$

  - We know that $E(\mathbf{x}\mathbf{z}')$ has full rank (by *2SLS2b*)) and therefore $E(\mathbf{x}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{x}')$ has also full rank. Then, using (22), (23) and the continuous function theorem, we have that

$$\begin{aligned}&\left(\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1} \\ &\to_p \quad \left(E(\mathbf{x}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{x}')\right)^{-1}\end{aligned} \tag{24}$$

  - Finally, using (21), (22), (23), (24) and the continuous function theorem, we have that

$$\widehat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$$

Asymptotic normality of 2SLS

Theorem (asymptotic normality of 2SLS). If assumptions $2SLS1$ and $2SLS2$ are satisfied and $E(u^2\mathbf{zz}')$ exists, the $2SLS$ estimator of $\boldsymbol{\beta}$ is asymptotically normal and the asymptotic distribution is

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

where

$$\begin{aligned} \mathbf{B} &= \mathbf{C}^{-1}E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(u^2\mathbf{zz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}')\mathbf{C}^{-1} \\ &\quad \text{and} \\ \mathbf{C}^{-1} &= (E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}'))^{-1} \end{aligned}$$

- *Proof*
  - Using (20), we have that

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

Proof of asymptotic normality of 2SLS, continued

- *Proof (cont.)*
  - $\mathbf{z}_i\mathbf{z}_i'$ are iid, then using the *LLN*

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

  - and since $E(\mathbf{z}\mathbf{z}')$ is non-singular (by $2SLS2a$), using the continuous function theorem, we have that

$$\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} \rightarrow_p E(\mathbf{z}\mathbf{z}')^{-1} \tag{25}$$

  - $\mathbf{x}_i\mathbf{z}_i'$ are iid, then, using the *LLN*,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i' \rightarrow_p E(\mathbf{x}\mathbf{z}') \tag{26}$$

Proof of asymptotic normality of 2SLS, continued

- *Proof (cont.)*
  - $E(\mathbf{xz}')$ has full rank (by $2SLS2b$)) and therefore $E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}')$ has also full rank. Then, using (25), (26) and the continuous function theorem, we have that

$$\left(\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)\right)^{-1}$$
$$\rightarrow_p \quad \left(E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}')\right)^{-1} = \mathbf{C}^{-1} \tag{27}$$

  - $\mathbf{z}_i u_i$ are iid, $E(\mathbf{z}_i u_i) = 0$ (by $2SLS1$) and $E(u^2\mathbf{zz}')$ exists, using the *CLT*

$$\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i \rightarrow_d N(0, E(u^2\mathbf{zz}')) \tag{28}$$

  - Then, using (28), (25), (26) and Slutsky's theorem, we have

$$\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\mathbf{z}_i u_i\right)$$
$$\rightarrow_d \quad N(0, E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(u^2\mathbf{zz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}')) \tag{29}$$

Proof of asymptotic normality of 2SLS, continued

- *Proof (cont.)*
  - Finally, using (29), (27) and Slutsky's theorem, we have that

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

where

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

2SLS under homoskedasticity

- If we now add a homoskedasticity assumption

$$2SLS3: \qquad E(u^2\mathbf{z}\mathbf{z}') = \sigma^2 E(\mathbf{z}\mathbf{z}')$$

  where $\sigma^2 = E(u^2)$.
- We have that

$$\begin{aligned} & E(\mathbf{x}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(u^2\mathbf{z}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{x}') \\ = \; & \sigma^2 E(\mathbf{x}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{x}') \\ = \; & \sigma^2 E(\mathbf{x}\mathbf{z}')E(\mathbf{z}\mathbf{z}')^{-1}E(\mathbf{z}\mathbf{x}') = \sigma^2\mathbf{C} \end{aligned}$$

- and the asymptotic variance simplifies to

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

- Then, we have that, under $2SLS1$, $2SLS2$ and $2SLS3$,

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

Variance matrix of 2SLS under homoskedasticity

- It can be shown that under $2SLS1$, $2SLS2$ and $2SLS3$

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

(or divided by N), where $\widehat{u}_i$ are the residuals based on the $2SLS$ estimator.

- Then, we have that

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

$$\to_p \sigma^2(E(\mathbf{xz}')E(\mathbf{zz}')^{-1}E(\mathbf{zx}'))^{-1}$$

Variance matrix of 2SLS under homoskedasticity (cont.)

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

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

Robust variance matrix of 2SLS

- When $2SLS3$ is not satisfied, the estimated variance of $\widehat{\boldsymbol{\beta}}$ given above is no longer valid.
- However, it can be shown that, under $2SLS1$ and $2SLS2$

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

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

$$\begin{aligned}\widehat{var(\widehat{\boldsymbol{\beta}})} &= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\left(\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'\right) \\ &\quad *(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\end{aligned}$$

Inference with 2SLS

- Once we have a consistent estimator of the asymptotic variance of the $2SLS$ estimator we can easily test any set of linear restrictions using a $Wald$ test.
- The W test statistic based on the $2SLS$ estimator is analogous to the W test statistic based on the IV estimator (or the OLS estimator). We just have to replace the IV estimator by the $2SLS$ estimator, and the estimated variance of the IV estimator by the estimated variance of the $2SLS$ estimator.
- The t statistic for testing one single restriction can also be used. We only need to replace the IV estimator by the $2SLS$ estimator, and the estimated variance of the IV estimator by the estimated variance of the $2SLS$ estimator.

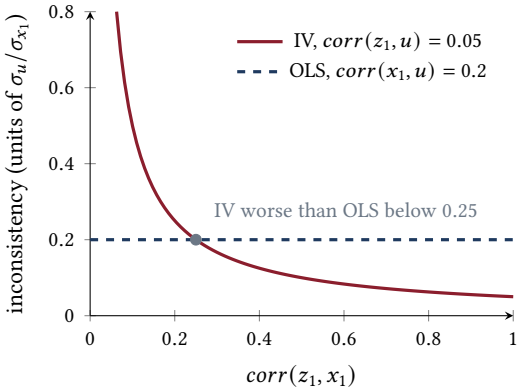
Bias with weak instruments

- When properly used *IV* methods can be a powerful tool for estimating structural equations. However there are some problems that we can find when applying *IV* in practice.
- It is worth mentioning that, unlike *OLS*, even under the zero conditional mean assumption, *IV* estimators are never unbiased. In fact under standard distributional assumptions the expected value of the *2SLS* estimator does not even exist. This is one of the reasons why we rely on large sample analysis to justify *GMM* estimation.
- Even in large samples, *IV* methods can be ill-behaved if we have **weak instruments**, that is if the instruments are only weakly partially correlated with the endogenous regressors.

Inconsistency with weak instruments

- Consider the simple model
$$y = \beta_0 + \beta_1 x_1 + u$$
and let $\widehat{\beta}_1$ be the simple IV that uses z_1 as an instrument for x_1.
- It can be shown that
$$plim\widehat{\beta}_1 = \beta_1 + \frac{cov(z_1, u)}{cov(z_1, x_1)} = \beta_1 + \frac{\sigma_u}{\sigma_{x_1}} \frac{corr(z_1, u)}{corr(z_1, x_1)}$$
- Then, when $cov(z_1, u) \neq 0$ the IV estimator is inconsistent and the size of the inconsistency bias can be arbitrarily large when $corr(z_1, x_1)$ gets close to zero.
- Thus, a small correlation between z_1 and u can cause a very large inconsistency bias if the instrument is weak and this bias can be larger than the bias of the OLS
$$plim\widehat{\beta}_1^{OLS} = \beta_1 + \frac{\sigma_u}{\sigma_{x_1}} corr(x_1, u)$$
- Unfortunately we cannot know the size of the inconsistency bias because we do not observe u, but we should be concerned if the correlation between x_1 and z_1 is small.

How weak instruments amplify small violations



Illustrative values.

$$plim\ \widehat{\beta}_1^{IV} - \beta_1 = \frac{\sigma_u}{\sigma_{x_1}} \frac{corr(z_1, u)}{corr(z_1, x_1)}$$

- Always report the first stage.
- A common rule of thumb flags instruments as weak when the first-stage F statistic is below 10.
- Large IV standard errors are a symptom of weak instruments.

Imprecision of IV

- Similar considerations arise when there are multiple variables and multiple instruments.
- Even when the instruments are truly exogenous in the population, if they are weak, *IV* estimators can be severely biased even for rather large sample sizes.
- Another potential problem when applying *IV* is that the standard errors can be quite large.
  - The magnitudes of the standard errors of the *IV* estimators depend on how good the instruments are.
  - When the instruments are weak (when the partial correlation between the endogenous explanatory variables and the instruments is very small) the standard errors of the *IV* estimators can be very large, even for large sample sizes.

The programme had two arms

- In Block A we pooled both arms into one dummy, `invited`.
- But a letter and a free audit tool are different nudges. We have **two** instruments for **one** endogenous regressor:

$$letter_i, \qquad software_i.$$

- Using only one of them throws information away. Using each separately gives two different IV estimates. How do we *combine* them?

Exercise. Which arm moves adoption more: the letter, or the letter plus free software? By roughly how much?

The programme had two arms

- In Block A we pooled both arms into one dummy, `invited`.
- But a letter and a free audit tool are different nudges. We have **two** instruments for **one** endogenous regressor:

$$letter_i, \qquad software_i.$$

- Using only one of them throws information away. Using each separately gives two different IV estimates. How do we *combine* them?

Exercise. Which arm moves adoption more: the letter, or the letter plus free software? By roughly how much?

With more instruments than endogenous regressors, the equation is **overidentified**. The answer to "how do we combine them" is two-stage least squares.

Analysis B1 · The first stage with both arms

First stage: `transp ~ letter + software +` *X, robust s.e.*

	Coef.	Robust s.e.	t
`letter`	0.0482	0.0153	3.2
`software`	0.2366	0.0165	14.3
Joint test: $F = 103.6$, $p < 10^{-40}$			

- A letter raises adoption by **5 pp**; the software by **24 pp**.
- This is the first-stage regression of 2SLS, and the joint F is the test of **relevance** ($H_0 : \theta_1 = \theta_2 = 0$).
- Remember the letter's $t = 3.2$. It will matter when we look at weak instruments.

Analysis B2 · 2SLS with both arms, and the two-step trap

Coefficient on `transp` *and two standard errors*

	$\widehat{\beta}_1$	Robust	Classical
`ivreg`	−0.0408	0.0098	0.0095
By hand (stage 2)	−0.0408	0.0105	0.0102

Residual standard deviation

With $transp$ (correct)	0.0647
With $\widehat{transp}$ (by hand)	0.0692

- 2SLS: −4.1 pp, 95% CI [−6.0, −2.2]. S.e. 0.0098 vs. 0.0132 with `invited`: **both arms buy back precision**.
- By hand: the *same* coefficient, different standard errors.
- Stage 2 forms residuals with $\widehat{transp}$ instead of $transp$, so $\widehat{\sigma}$ is wrong.

Here the error is 7%; elsewhere it can be large. Use a 2SLS command.

Is the equation identified?

Exercise. The equation has J endogenous regressors and M excluded instruments. What can you say?

(a) $J = 1, M = 0$: $transp$ with no programme at all.
(b) $J = 1, M = 3$: $transp$ instrumented by $letter$, $software$ and $peer_share$.
(c) $J = 2, M = 2$: $transp$ and $women_mgmt$ both endogenous, instrumented by $letter$ and $software$.

Is the equation identified?

Exercise. The equation has J endogenous regressors and M excluded instruments. What can you say?

(a) $J = 1, M = 0$: $transp$ with no programme at all.
(b) $J = 1, M = 3$: $transp$ instrumented by $letter$, $software$ and $peer_share$.
(c) $J = 2, M = 2$: $transp$ and $women_mgmt$ both endogenous, instrumented by $letter$ and $software$.

Answer.

(a) Not identified: the order condition $L \geq k$ fails.
(b) Overidentified: two overidentifying restrictions (identified if at least one $\theta_h \neq 0$).
(c) The order condition holds ($L = k$), but the **rank condition** almost surely fails: a lottery on transparency does not move women in management, so the 2×2 matrix of instrument coefficients has (close to) a row of zeros.

Analysis B3 · Who moves, men or women?

2SLS, both arms, controls X, robust s.e.

Outcome (2021–2025)	$\widehat{\beta}_1$	s.e.
Wage growth, men	−0.0294	0.0064
Wage growth, women	+0.0113	0.0066
Men − women = Δgap	−0.0408	

For comparison, OLS

men −0.036, women +0.020

Exercise. When a firm makes pay visible, does the gap close because women's wages rise, or because men's wages grow more slowly?

Analysis B3 · Who moves, men or women?

2SLS, both arms, controls X, robust s.e.

Outcome (2021–2025)	$\widehat{\beta}_1$	s.e.
Wage growth, men	−0.0294	0.0064
Wage growth, women	+0.0113	0.0066
Men − women = Δgap	−0.0408	

For comparison, OLS
men −0.036, women +0.020

Exercise. When a firm makes pay visible, does the gap close because women's wages rise, or because men's wages grow more slowly?

- Most of the effect comes from **slower wage growth for men** (−2.9 pp, $t = 4.6$); women gain 1.1 pp ($t = 1.7$).
- The two effects add up *exactly* to the effect on the gap: 2SLS is linear in y.
- OLS overstates both, because culture moves both.

Is a policy that narrows the gap by holding men's wages back a success? Who should decide?

Analysis B4 · What if we only had the letter?

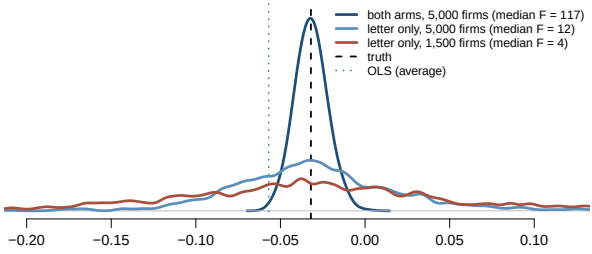
Firms in the control or letter arm (N = 3,813)

First stage: `letter`	0.048 (0.015)
First-stage F	**9.8**
2SLS $\widehat{\beta}_1$	−0.027
Robust s.e.	0.045
95% CI	[−0.116, 0.062]

Software arm only, for comparison
F = 208; $\widehat{\beta}_1$ = −0.041 (0.010)

- The letter moved adoption by 5 pp: significant, but F is just under the rule-of-thumb 10.
- The CI runs from "closes the gap by 12 pp" to "widens it by 6 pp": useless for the Ministry.
- Is −0.027 close to the true effect? One sample cannot tell. A simulation can.

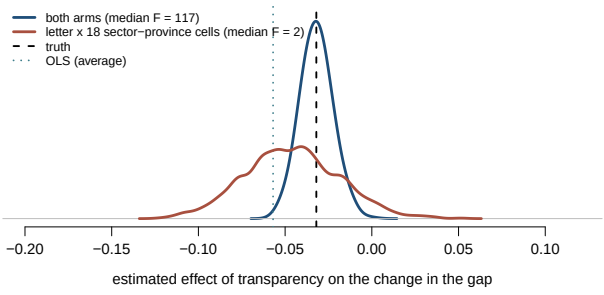
One weak instrument (Monte Carlo)



1,000 samples from the same simulated world. Share of estimates more than 5 pp from the truth: 0% ($F \approx 117$), 27% ($F \approx 12$), 54% ($F \approx 4$). `weak_iv_mc.R`

With one weak instrument the distribution stays roughly centred on the truth, but its tails explode. Any single estimate can land almost anywhere. The reported standard errors are also huge, so the intervals still cover the truth: the estimator is **useless** rather than misleading.

Many weak instruments (Monte Carlo)



- A data scientist's reflex: "use more information". Interact the letter with the 18 sector×province cells: 18 instruments, joint $F \approx 1.6$.
- The median estimate moves from −0.032 (truth) to **−0.046**, most of the way to OLS (−0.057).

Many weak instruments make 2SLS drift toward OLS, with standard errors that look reasonable. More features is not better when each one carries almost no signal: the first stage *overfits*, and the fitted $\widehat{x}$ starts to reproduce x itself, endogeneity included.

Block B in R

```
library(car)
d$letter   <- as.numeric(d$arm == 1)
d$software <- as.numeric(d$arm == 2)
Z <- "letter + software"
# B1: first stage and relevance
fs <- lm(paste("transp ~", Z, "+", X), d)
linearHypothesis(fs, c("letter=0", "software=0"),
                 vcov. = vcovHC(fs, "HC1"))
# B2: 2SLS, and the two-step trap
iv2 <- ivreg(paste("dgap ~ transp +", X, "|",
                   Z, "+", X), data = d)
rob(iv2, "HC0")
d$transp_hat <- fitted(fs)
rob(lm(paste("dgap ~ transp_hat +", X), d))
# B3: who moves?
d$growth_men <- log(d$wage_men_2025/d$wage_men_2021)
ivreg(paste("growth_men ~ transp +", X, "|",
            Z, "+", X), data = d)
# B4: the letter alone
ivreg(paste("dgap ~ transp +", X, "| letter +", X),
      data = d, subset = arm != 2)
```

1. Build `growth_wom`: do the two coefficients add up to the one for `dgap`?
2. Use the software arm only (`subset = arm != 1`). Compare F and s.e. with B4.
3. Why are the B2 by-hand s.e. wrong?

Summary of two-stage least squares

1. 2SLS uses the linear projection of the endogenous regressors on $\mathbf{z}$ as the instrument; with one instrument it is simple IV.
2. $\widehat{\boldsymbol{\beta}}_{2SLS} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$. Compute it with a 2SLS command: the two-step standard errors are wrong.
3. Identification needs the order condition ($L \geq k$) and the rank condition ($rank\, E(\mathbf{zx}') = k$).
4. Always report the first stage. One weak instrument gives a useless estimate; many weak instruments give a biased one that looks precise.

The case so far

OLS −5.6 pp; IV (invited) −3.8 pp; 2SLS (both arms) −4.1 pp (s.e. 1.0). Most of the effect runs through men's wages.