# DSE - Data-Driven Economic Analysis
# Econometrics Module
# Lecture 9 - Additional Topics on IV

Michele De Nadai
michele.denadai@unimi.it

Trimester II, 2023

UNIVERSITÀ
DEGLI STUDI
DI MILANO

# Introduction

- We have introduced Instrumental Variables to handle endogenous control variables.
- Valid instruments are required to be **exogenous** and **relevant**.
- In the literature several tests have been considered to formally **test** these requirements.

# Endogeneity Tests

- Consider the simple model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i,$$

where $\mathbf{x}_i$ may be endogenous and a set of instruments $\mathbf{z}_i$ is available.

- Recall the control function approach:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{u}_i'\boldsymbol{\gamma} + \varepsilon_i, \qquad \mathrm{E}[\mathbf{x}_i\varepsilon_i] = 0 \qquad \mathrm{E}[\mathbf{u}_i\varepsilon_i] = 0,$$

where $\mathbf{u}_i = \mathbf{x}_i - \boldsymbol{\Gamma}'\mathbf{z}_i$ are the first stage errors.

- In this model $\mathbf{x}_i$ is endogenous iff $\boldsymbol{\gamma} \neq 0$.
- A test for $\mathrm{H}_0 : \mathrm{E}[\mathbf{x}_i e_i] = 0$ vs. $\mathrm{H}_1 : \mathrm{E}[\mathbf{x}_i e_i] \neq 0$ is then **equivalent** to testing for $\mathrm{H}_0 : \boldsymbol{\gamma} = 0$ vs $\mathrm{H}_1 : \boldsymbol{\gamma} \neq 0$.

# Endogeneity Tests

- We could test for $\mathrm{H}_0 : \boldsymbol{\gamma} = 0$ vs $\mathrm{H}_1 : \boldsymbol{\gamma} \neq 0$:
  1. Obtain $\widehat{\mathbf{u}}_i$ as the residuals from the linear regression of $\mathbf{x}_i$ on $\mathbf{z}_i$.
  2. Regress $y_i$ on $\mathbf{x_i}$ and $\widehat{\mathbf{u}}_i$.
  3. Perform a Wald test for the equality to zero of all the coefficients associated with $\widehat{\mathbf{u}}_i$ using a Wald test statistic:

     $$W = \widehat{\boldsymbol{\gamma}}' V_{\widehat{\boldsymbol{\gamma}}}^{-1} \widehat{\boldsymbol{\gamma}}$$

     Under $\mathrm{H}_0$ $W \sim \chi^2_q$.
- This class of tests is called **Durbin-Wu-Hausman Tests**, or just **Hausman Test**.

# Over-identification Tests

- When the number of instruments exceeds the number of endogenous variables the model is **overidentified**.
- We could in principle estimate the parameters of interest using a **smaller set of instruments**.
- **Over-identification tests** use these additional restrictions to test the validity of instruments.
- A set of $l$ instruments $\mathbf{z}_i$ is valid if $E[\mathbf{z}_i e_i] = 0$, i.e.

  $$E[\mathbf{z}_i(y_i - \mathbf{x}_i \boldsymbol{\beta})] = 0$$

  or equivalently

  $$E[\mathbf{z}_i y_i] = E[\mathbf{z}_i \mathbf{x}_i'] \boldsymbol{\beta}$$

  which define $l$ restrictions.

# Over-identification Tests (cont.)

- To simplify, suppose $k = 1$ and $l = 2$, i.e. $\mathbf{z}_i = (z_{1i}, z_{2i})'$:

$$\begin{aligned} E[z_{1i}y_i] &= E[z_{1i}x_i]\beta \\ E[z_{2i}y_i] &= E[z_{2i}x_i]\beta \end{aligned}$$

- We could use either of the two restrictions to **identify** $\beta$.
- In practice the corresponding estimators will be different in finite samples, but the difference should not be statistically significant, if both instruments are valid.
- Over-identification tests aim at comparing these two estimators.

# Sargan Test

- The null hypothesis of interest is

$$\begin{aligned} H_0 &: \mathrm{E}[\mathbf{z}_i e_i] = 0 \\ H_1 &: \mathrm{E}[\mathbf{z}_i e_i] \neq 0 \end{aligned}$$

- Consider the (unfeasible) regression of $e_i$ on the instruments:

$$e_i = \mathbf{z}_i' \boldsymbol{\alpha} + \varepsilon_i, \qquad \mathrm{E}[e_i^2 | \mathbf{z}_i] = \sigma^2$$

- The vector is then defined as

$$\boldsymbol{\alpha} = \mathrm{E}[\mathbf{z}_i \mathbf{z}_i']^{-1} \mathrm{E}[\mathbf{z}_i e_i]$$

  which is zero iff $\mathrm{E}[\mathbf{z}_i e_i] = 0$.

- We could then equivalently write $H_0 : \boldsymbol{\alpha} = 0$.

# Sargan Test (cont.)

- $e_i$ is not observed but, under $H_0$, they can be replaced by the residuals of a 2SLS regression of $y_i$ on $\mathbf{x}_i$ using $\mathbf{z}_i$ as instruments ($\widehat{e}_i$).
- Then
$$\widehat{\boldsymbol{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\mathbf{e}}$$
- **Sargan**'s test statistic is equivalent to the Wald test statistic for $H_0 : \boldsymbol{\alpha} = 0$
$$\begin{aligned} S &= (\widehat{\boldsymbol{\alpha}} - 0)'\widehat{Var}(\widehat{\boldsymbol{\alpha}})^{-1}(\widehat{\boldsymbol{\alpha}} - 0) \\ &= ((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\mathbf{e}})'(\widehat{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1})^{-1}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\mathbf{e}}) \\ &= \frac{\widehat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\mathbf{e}}}{\widehat{\sigma}^2} \\ &= \frac{\widehat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\mathbf{e}}}{\widehat{\sigma}^2} \qquad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 \end{aligned}$$
- Under $H_0$ we have $S \xrightarrow{d} \chi^2_{l-k}$.

# Caution

- Sargan test relies on the assumption that **at least one** IV is valid.
- Not-rejecting the null implies that estimates for $\boldsymbol{\beta}$ obtained with different sets of instruments are not statistically different.
- Rejecting the null implies that using different sets of instruments to estimate $\boldsymbol{\beta}$ give you different results.
- You still have no clue on which IV is valid and which is not.

# What does IV identify with Binary Treatments

- Consider a simple setup where an outcome of interest $(y_i)$ depends on an endogenous binary treatment $(d_i)$ and we have availability of valid and relevant binary instrument $(z_i)$.
- Typical example might be the return to college using proximity to the college as an instrument.
- We could estimate this effect by setting up a linear model

$$y_i = \beta_0 + \beta_1 d_i + u_i$$

and consider the IV estimator

$$\beta_{IV} = E[\mathbf{z}_i \mathbf{d}_i']^{-1} E[\mathbf{z}_i y_i] = \begin{pmatrix} \frac{Cov(d_i, z_i y_i) - Cov(d_i z_i, y_i)}{Cov(d_i, z_i)} \\ \frac{Cov(y_i, z_i)}{Cov(d_i, z_i)} \end{pmatrix},$$

where $\mathbf{d}_i = (1, d_i)'$ and $\mathbf{z}_i = (1, z_i)'$.

# Wald Ratio

- The IV estimator for the slope can be written as

$$\frac{Cov(y_i, z_i)}{Cov(d_i, z_i)} = \frac{Cov(y_i, z_i)}{Var(z_i)} \frac{Var(z_i)}{Cov(d_i, z_i)}$$

  which is the ratio of the slopes on the reduced form regression

$$y_i = \delta_0 + \delta_1 z_i + v_i$$

  and on the first stage regression

$$d_i = \gamma_0 + \gamma_1 z_i + e_i$$

- Since $z_i$ is a dummy this is equivalent to

$$\frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{E[d_i|z_i = 1] - E[d_i|z_i = 0]}$$

- This is often referred to as **Wald ratio**.

# Heterogeneous Treatment Effects

- Does the Wald ratio identify a causal parameter if we allow for **heteregeneous effects** of the treatment?
- Consider a generalization of the potential outcomes framework where treatment is also potential:
  - $Y_1$ is the value of the outcome with treatment
  - $Y_0$ is the value of the outcome without treatment
  - $D_1$ is the treatment individual would have chosen if $Z = 1$
  - $D_0$ is the treatment individual would have chosen if $Z = 0$
- Observed treatment is defined as $D = D_0 + (D_1 - D_0)Z$.
- Observed outcome is (as always) $Y = Y_0 + (Y_1 - Y_0)D$.
- Depending on the pair $(D_1, D_0)$ each individual will happen to belong to one of four **latent types**.

# Latent Types

- We have four different latent types:
  - **Compliers** ($\mathcal{C}$): those who always comply with the assignment outcome. They would get the treatment if $Z = 1$ and would not otherwise: $D_0 = 0$ and $D_1 = 1$.
  - **Defiers** ($\mathcal{D}$): those who never comply with the assignment outcome. They would not get the treatment if $Z = 1$ and would otherwise: $D_0 = 1$ and $D_1 = 0$.
  - **Always Takers** ($\mathcal{A}$): those who always get the treatment regardless of $Z$: $D_0 = 1$ and $D_1 = 1$.
  - **Never Takers** ($\mathcal{N}$): those who never get the treatment regardless of $Z$: $D_0 = 0$ and $D_1 = 0$.

| | $D_0 = 0$ | $D_0 = 1$ |
|---|---|---|
| $D_1 = 0$ | Never Takers | Defiers |
| $D_1 = 1$ | Compliers | Always Takers |

# Reduced Form

- Using Law of Iterated Expectations we could write:

$$\begin{aligned}
E[Y|Z=z] &= E[Y|\mathcal{C}, Z=z]Pr[\mathcal{C}|Z=z] \\
&+ E[Y|\mathcal{A}, Z=z]P[\mathcal{A}|Z=z] \\
&+ E[Y|\mathcal{N}, Z=z]P[\mathcal{N}|Z=z] \\
&+ E[Y|\mathcal{D}, Z=z]P[\mathcal{D}|Z=z] \\
&= E[Y_z|\mathcal{C}]\pi_{\mathcal{C}} \\
&+ E[Y_1|\mathcal{A}]\pi_{\mathcal{A}} \\
&+ E[Y_0|\mathcal{N}]\pi_{\mathcal{N}} \\
&+ E[Y_{1-z}|\mathcal{D}]\pi_{\mathcal{D}}
\end{aligned}$$

- Therefore the reduced form equation identifies

$$E[Y|Z=1] - E[Y|Z=0] = E[Y_1 - Y_0|\mathcal{C}]\pi_{\mathcal{C}} - E[Y_1 - Y_0|\mathcal{D}]\pi_{\mathcal{D}}$$

# First Stage

- Similarly:

$$\begin{aligned} E[D|Z=z] &= E[D|\mathcal{C}, Z=z]Pr[\mathcal{C}|Z=z] \\ &+ E[D|\mathcal{A}, Z=z]P[\mathcal{A}|Z=z] \\ &+ E[D|\mathcal{N}, Z=z]P[\mathcal{N}|Z=z] \\ &+ E[D|\mathcal{D}, Z=z]P[\mathcal{D}|Z=z] \\ &= z\pi_{\mathcal{C}} \\ &+ \pi_{\mathcal{A}} \\ &+ \pi_{\mathcal{N}} \\ &+ (1-z)\pi_{\mathcal{D}} \end{aligned}$$

- Therefore the first stage equation identifies

$$E[D|Z=1] - E[D|Z=0] = \pi_{\mathcal{C}} - \pi_{\mathcal{D}}$$

# Local Average Treatment Effect

- In general we have

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = \frac{E[Y_1 - Y_0|\mathcal{C}]\pi_{\mathcal{C}} - E[Y_1 - Y_0|\mathcal{D}]\pi_{\mathcal{D}}}{\pi_{\mathcal{C}} - \pi_{\mathcal{D}}}$$

- In general this is not a causal parameter of interest
- Typical assumption is: **no defiers** - $\pi_{\mathcal{D}} = 0$, which would imply

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = \frac{E[Y_1 - Y_0|\mathcal{C}]\pi_{\mathcal{C}}}{\pi_{\mathcal{C}}} = E[Y_1 - Y_0|\mathcal{C}]$$

# Local Average Treatment Effect

- The Wald Ratio identifies $E[Y_1 - Y_0|\mathcal{C}]$ the Average Treatment Effect for the sub-population of compliers. This is the **Local Average Treatment Effect** parameter.
- Compliers are individuals who are induced into treatment by being assigned $Z = 1$.
- The population of compliers might not be the population of interest and the ATE for them might be very different from the ATE for other latent types.
- Still this is the **only** sub-group for which we can identify a causal parameter.
- When can we identify $ATE = E[Y_1 - Y_0]$?

# LATE and Latent Index Models

- The **monotonicity assumption**:

$$D_1 \geqslant D_0$$

  is required to interpret IV estimates when treatment effects are heterogeneous.

- Suppose we model observed treatment $D$ as

$$D = \begin{cases} 1 & \text{if } \gamma_0 + \gamma_1 Z - \nu > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Potential treatments are now defined as

$$D_0 = \mathbf{1}[\gamma_0 > \nu] \qquad D_1 = \mathbf{1}[\gamma_0 + \gamma_1 > \nu]$$

- Monotonicity is here satisfied by the condition $\gamma_1 \geqslant 0$.

# LATE and Latent Index Models

- In this setting $\nu$ indexes heterogeneity between individuals
- **Latent types** are determined by the (unobserved) value of $\nu$

$$\text{if } \nu < \gamma_0 \text{ then } D_0 = D_1 = 1$$

$$\text{if } \gamma_0 < \nu \leqslant \gamma_0 + \gamma_1 \text{ then } D_0 = 0, D_1 = 1$$

$$\text{if } \nu > \gamma_0 + \gamma_1 \text{ then } D_0 = D_1 = 0$$

- Individual treatment effects can also be indexed by the unobserved $\nu$:

$$E(Y_1 - Y_0 | V = \nu)$$

- LATE identifies

$$E(Y_1 - Y_0 | \gamma_0 < \nu \leqslant \gamma_0 + \gamma_1)$$

# Multivalued Treatments

- LATE parameter can be generalized to the case of multivalued treatments
- Let $S_1$ and $S_0$ be treatment indicators for the amount of treatment received with and without $Z = 1$
- $S = S_0 + (S_1 - S_0)Z$ is the observed amount of treatment
- Suppose $S_0$ and $S_1$ take values in $[0, 1, 2]$
- The potential outcomes here are $Y_0$, $Y_1$ and $Y_2$
- Which kind of **causal parameter** are we interested in?

# Multivalued Treatments

- IV regression of $Y$ on $D$ using a binary $Z$ as an instrument still identifies the Wald ratio:

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[S|Z=1]-E[S|Z=0]}$$

- Let us assume **monotonicity** of treatment as in the binary case, that is

$$S_1 \geqslant S_0$$

- The Wald estimator identifies a **weighted average** of treatment effects for different groups of compliers

# Numerator

We have

$$\begin{aligned} E[Y|Z=1] &= E[Y|S_0=0,S_1=0,Z=1]Pr(S_0=0,S_1=0|Z=1) \\ &+ E[Y|S_0=0,S_1=1,Z=1]Pr(S_0=0,S_1=1|Z=1) \\ &+ E[Y|S_0=0,S_1=2,Z=1]Pr(S_0=0,S_1=2|Z=1) \\ &+ E[Y|S_0=1,S_1=1,Z=1]Pr(S_0=1,S_1=1|Z=1) \\ &+ E[Y|S_0=1,S_1=2,Z=1]Pr(S_0=1,S_1=2|Z=1) \\ &+ E[Y|S_0=2,S_1=2,Z=1]Pr(S_0=2,S_1=2|Z=1) \end{aligned}$$

# Numerator

We have

$$
\begin{aligned}
E[Y|Z=1] &= E[Y_0|S_0=0,S_1=0,Z=1]\Pr(S_0=0,S_1=0)\\
&+ E[Y_1|S_0=0,S_1=1,Z=1]\Pr(S_0=0,S_1=1)\\
&+ E[Y_2|S_0=0,S_1=2,Z=1]\Pr(S_0=0,S_1=2)\\
&+ E[Y_1|S_0=1,S_1=1,Z=1]\Pr(S_0=1,S_1=1)\\
&+ E[Y_2|S_0=1,S_1=2,Z=1]\Pr(S_0=1,S_1=2)\\
&+ E[Y_2|S_0=2,S_1=2,Z=1]\Pr(S_0=2,S_1=2)
\end{aligned}
$$

# Numerator

We have

$$
\begin{aligned}
\mathrm{E}[Y|Z=1] &= \mathrm{E}[Y_0|S_0=0, S_1=0]\pi_{00} \\
&+ \mathrm{E}[Y_1|S_0=0, S_1=1]\pi_{01} \\
&+ \mathrm{E}[Y_2|S_0=0, S_1=2]\pi_{02} \\
&+ \mathrm{E}[Y_1|S_0=1, S_1=1]\pi_{11} \\
&+ \mathrm{E}[Y_2|S_0=1, S_1=2]\pi_{12} \\
&+ \mathrm{E}[Y_2|S_0=2, S_1=2]\pi_{22}
\end{aligned}
$$

# Numerator

Similarly for $Z = 0$ we obtain

$$\begin{aligned}
E[Y|Z=0] &= E[Y_0|S_0=0, S_1=0]\pi_{00} \\
&+ E[Y_0|S_0=0, S_1=1]\pi_{01} \\
&+ E[Y_0|S_0=0, S_1=2]\pi_{02} \\
&+ E[Y_1|S_0=1, S_1=1]\pi_{11} \\
&+ E[Y_1|S_0=1, S_1=2]\pi_{12} \\
&+ E[Y_2|S_0=2, S_1=2]\pi_{22}
\end{aligned}$$

# Numerator

Taking the difference we get

$$
\begin{aligned}
E[Y|Z=1] - E[Y|Z=0] &= E[Y_1 - Y_0|S_0 = 0, S_1 = 1]\pi_{01} \\
&+ E[Y_2 - Y_0|S_0 = 0, S_1 = 2]\pi_{02} \\
&+ E[Y_2 - Y_1|S_0 = 1, S_1 = 2]\pi_{12} \\
&= E[Y_1 - Y_0|S_0 = 0, S_1 = 1]\pi_{01} \\
&+ E[Y_1 - Y_0|S_0 = 0, S_1 = 2]\pi_{02} \\
&+ E[Y_2 - Y_1|S_0 = 0, S_1 = 2]\pi_{02} \\
&+ E[Y_2 - Y_1|S_0 = 1, S_1 = 2]\pi_{12} \\
&= E[Y_1 - Y_0|S_0 < 1 \leqslant S_1]\Pr(S_0 < 1 \leqslant S_1) \\
&+ E[Y_2 - Y_1|S_0 < 2 \leqslant S_1]\Pr(S_0 < 2 \leqslant S_1)
\end{aligned}
$$

This is a **weighted average** of treatment effects for different groups of compliers

# Denominator

At the denominator of the Wald Ratio we obtain:

$$
\begin{aligned}
E[S|Z=1] - E[S|Z=0] &= E[S_1|Z=1] - E[S_0|Z=0] \\
&= E[S_1 - S_0] \\
&= E[S_1 - S_0|S_0 = 0, S_1 = 0]\pi_{00} \\
&+ E[S_1 - S_0|S_0 = 0, S_1 = 1]\pi_{01} \\
&+ E[S_1 - S_0|S_0 = 0, S_1 = 2]\pi_{02} \\
&+ E[S_1 - S_0|S_0 = 1, S_1 = 1]\pi_{11} \\
&+ E[S_1 - S_0|S_0 = 1, S_1 = 2]\pi_{12} \\
&+ E[S_1 - S_0|S_0 = 2, S_1 = 2]\pi_{22} \\
&= \pi_{01} + 2\pi_{02} + \pi_{12} \\
&= \Pr(S_0 < 1 \leqslant S_1) + \Pr(S_0 < 2 \leqslant S_1)
\end{aligned}
$$

# Wald Ratio with Multivalued Treatments

- In general if $S$ takes values in $\{0, 1, \ldots, K\}$ then the Wald ratio identifies

$$\sum_{j=1}^{K} \omega_j E[Y_j - Y_{j-1} | S_0 < j \leqslant S_1]$$

where

$$\omega_j = \frac{\Pr(S_0 < j \leqslant S_1)}{\sum_{i=1}^{K} \Pr(S_0 < i \leqslant S_1)}$$

- Without monotonicity we are not able to interpret IV estimates causally.
- Still, monotonicity has some **testable implications**. In fact it must be

$$\Pr(S \geqslant j | Z = 1) \geqslant \Pr(S \geqslant j | Z = 0)$$

# References

- ”Mostly Harmless Econometrics”, Angrist J. and Pischke J. (2009) **Chapter 4**