Econometrics

Practical Session 22

Instrumental Variables: The General IV Regression Model

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

The General IV Regression Model

The general IV model extends the single-instrument case to multiple regressors:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \beta_{k+1} w_{1i} + \ldots + \beta_{k+r} w_{ri} + u_i$$

x's → endogenous

- k variables correlated with u_i
- Each requires at least one instrument

w's → exogenous included

- r variables uncorrelated with u_i
- Control variables that appear in the structural equation

z's → exogenous instruments

- m variables excluded from the structural equation
- Correlated with x's but not with u_i

Two Conditions for Valid Instruments

1. Relevance

$$\exists s \in 1, \ldots, k, \forall j \in 1, \ldots, k:$$

$$\text{corr}\left(z_{ji}, x_{si} \mid w_{1i}, \ldots, w_{ri}\right) \neq 0$$

→ Each instrument must be correlated with at least one endogenous regressor, **after partialling out the effect of w's**

2. Exogeneity

$$\text{corr}\left(z_{ji}, u_i\right) = 0$$

→ The same condition as before: the whole point of using an instrument is its exogeneity

Identification

A model is **identified** if there is enough exogenous variation to estimate all the coefficients on the endogenous regressors:

Exactly identified

- $m = k$ (instruments = endogenous regressors)
- Can **estimate β's**
- **Cannot test instrument exogeneity**

Overidentified

- $m > k$ (more instruments than needed)
- Can **estimate β's**
- Can **test overidentification** (J-statistic)

Underidentified

- $m < k$ (too few instruments)
- **Estimation is impossible**

TSLS/2SLS with Multiple Instruments and Controls

Stage 1: Regress **each endogenous x_k** on **all z's and all w's**:

$$x_{ki} = \pi_0 + \pi_1 z_{1i} + \ldots + \pi_m z_{mi} + \pi_{m+1} w_{1i} + \ldots + \pi_{m+r} w_{ri} + v_i$$

Stage 2: Regresses **y_i on $\hat{x}'$s and all w's**:

$$y_i = \beta_0 + \beta_1 \hat{x}_{1i} + \ldots + \beta_k \hat{x}_{ki} + \beta_{k+1} w_{1i} + \ldots + \beta_{k+r} w_{ri} + u_i$$

CRITICAL REQUIREMENT

All w's must be used in both stages. Otherwise, either first stage or second stage estimates are biased and inconsistent.

What If We Drop w From the First Stage?

- Suppose the true model is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + u_i$$

- Suppose also that $\text{cov}(z_i, u_i) = 0$, but $\text{cov}(z_i, w_i) \neq 0$
- For the **first stage**, you run:

$$x_i = \pi_0 + \pi_1 z_i + \tilde{v}_i, \quad \tilde{v}_i = \pi_2 w_i + v_i$$

- So, the first stage **estimates $\hat{\pi}_0$ and $\hat{\pi}_1$ are biased and inconsistent**

What If We Drop w_i From the Model?

- Suppose the true model is the same as before, but **you estimate this one instead (without w_i)**:

$$y_i = \beta_0 + \beta_1 x_i + \tilde{u}_i, \quad \tilde{u}_i = \beta_2 w_i + u_i$$

- Suppose also that $\text{cov}(z_i, u_i) = 0$, but $\text{cov}(z_i, w_i) \neq 0$
- ***z* is no longer exogenous** in the main second stage!

$$y_i = \beta_0 + \underbrace{\beta_1 \hat{\pi}_0 + \beta_1 \hat{\pi}_1 z_i}_{\beta_1 \hat{x}_i} + \tilde{u}_i$$

$$\text{cov}(z_i, \tilde{u}_i) = \beta_2 \,\text{cov}(z_i, w_i) + \text{cov}(z_i, u_i) = \beta_2 \,\text{cov}(z_i, w_i) \neq 0$$

- **Relevance** is now a **conditional** statement: $\text{cov}(z_i, x_i \mid w_i) \neq 0$ → ***z* must predict *x* after partialling out *w***

Finding Valid Instruments

1. Equation-specific variables

Variables that belong to one equation (supply shifters, demand shifters) **but not the other** serve as natural instruments

Example: rainfall shifts supply but not demand for agricultural goods

2. Quasi-random variation

Look for **factors that affect** x ***as if* randomly**, without directly affecting y

Examples: lottery drafts, date of birth, natural disasters, geographic discontinuities, policy cutoffs

Remember: relevance can be tested (first-stage F). Exogeneity cannot → it requires a compelling economic argument. Both matter equally.

Market Analysis: Graddy (1995), Fulton Fish Market

Data and Variables

- **Key question:** Is the fish market competitive? What is the price elasticity of demand?
- Graddy, K. (RAND-JE, 1995). “Testing for Imperfect Competition at the Fulton Fish Market.”
- **Daily price and quantity data for whiting** (a white fish) at New York’s Fulton Fish Market
- $n = 97$ **observations**, Dec 1991 to May 1992
- The market is an auction: fishermen sell to dealers each morning before retail

The elasticity of demand argument

- High elasticity
  → high sensitivity to price
  → low markups
  → close to perfect competition
- Low elasticity
  → low sensitivity to price
  → high markups
  → market power

Data and Variables

Variable	Role	Description
`ltotqty`	Outcome y	Log total quantity sold (lbs)
`lavgprc`	Endogenous x	Log average price per pound

ENDOGENEITY CONCERNS

Simultaneous causality → price and quantity are simultaneously determined → observed (P, Q) pairs are the market equilibrium → a mixture of both demand and supply shifts.

Data and Variables

Variable	Role	Description
`mon–thurs`	Instrument z_1	Weekday dummies. Quantity **demanded** depends on price and weekday patterns (Friday effect for Catholics, Monday/ weekend preferences)
`wave2`	Instrument z_2	Average wave height (ft) in the prior 2 days. Quantity **supplied** depends on price and sea conditions.
`speed2`	Instrument z_3	Average wind speed (knots) in the prior 2 days. Quantity **supplied** depends on price and sea conditions.

The Economic Model

Demand equation

$$\ln Q_i = \alpha_d + \beta_d \cdot \ln P_i + \delta_d \text{weekday}_i + \varepsilon_i^d$$

- $\beta_d < 0$ expected (downward-sloping demand)

Supply equation

$$\ln Q_i = \alpha_s + \beta_s \cdot \ln P_i + \gamma_s \text{weather}_i + \varepsilon_i^s$$

- $\beta_s > 0$ expected (upward-sloping supply)

THE PROBLEM OF RAW OLS

OLS of Q on P gives a **weighted average** of the demand and supply slopes. In a perfectly symmetric case: $\hat{\beta}^{\text{OLS}} \approx (\beta_d + \beta_s)/2$. In this case: $\hat{\beta}^{\text{OLS}} = -0.493$

Simultaneous Causality: OLS Is Not Structural

We need to **separately identify** the demand and supply equations. The strategy:

Equation to identify	Type of IV needed	Variable used
Demand curve	Supply shifter (affects cost but not preferences)	`wave2`, `speed2` (weather)
Supply curve	Demand shifter (affects preferences but not cost)	`mon–thurs` (weekday)

Step 1: Demand Equation – First Stage

$$\widehat{\text{lavgprc}}_i = \hat{\pi}_0 + \hat{\pi}_1 \text{wave2}_i + \hat{\pi}_2 \text{speed2}_i$$

Coefficient	Estimate	SE (HC1)	*t*-statistic
Intercept	-0.754	(0.105)	-7.15***
`wave2` (wave height)	+0.116	(0.023)	+5.02***
`speed2` (wind speed)	-0.007	(0.010)	-0.69

First-stage ***F*-statistic: 13.9 > 10** → instruments **are relevant** ✓

Rougher seas (higher waves) significantly raise prices → consistent with supply reduction. Wind speed alone is not significant.

Step 2: Supply Equation – First Stage

$$\widehat{\text{lavgprc}}_i = \hat{\pi}_0 + \hat{\pi}_1 \text{mon}_i + \hat{\pi}_2 \text{tues}_i + \hat{\pi}_3 \text{wed}_i + \hat{\pi}_4 \text{thurs}_i$$

First-stage $F = 0.2 < 10$ ⚠

Weak instruments → Weekday dummies explain essentially none of the variation in log price → the supply **2SLS estimate will be biased and unreliable**

Why might this be?

- Weekday effects primarily shift **quantity demanded**, not **price**
- If supply is relatively inelastic price is not driven mainly by predictable demand patterns

Step 3: 2SLS Results

CATÓLICA LISBON BUSINESS & ECONOMICS

Equation	OLS $\hat{\beta}$	2SLS $\hat{\beta}$	IVs (first-stage F)
Demand $\ln Q_i = \alpha_d + \beta_d \cdot \ln P_i + \delta_d \text{weekday}_i + \varepsilon_i^d$	−0.493	**−0.812**	weather ✓ ($F = 13.9$)
Supply $\ln Q_i = \alpha_s + \beta_s \cdot \ln P_i + \gamma_s \text{weather}_i + \varepsilon_i^s)$	−0.493	**0.53**	weekday ⚠ ($F = 0.2$)

- A **1% price increase reduces quantity demanded by 0.81 %** (elastic demand)
- The **demand curve slopes downward** → consistent with theory
- Supply equation (unreliable): the positive sign but this estimate is heavily biased

Discussion

1. If a bad storm reduced fish supply and raised prices by 10%, by how much would quantity demanded fall?

KEY TAKEAWAYS

- **Elasticity ≈ −0.812** (not perfectly elastic, not perfectly inelastic)
- A 10% price increase → quantity demanded falls by 0.81 × 10% = 8.1%
- **Consumers substitute away from fish when it gets expensive**, but not completely
- The 2SLS estimate of −0.812 is larger in magnitude than OLS (−0.493), suggesting that OLS underestimates the demand elasticity → likely because OLS confounds demand and supply movements

Discussion

2. Why is the supply equation unidentified in practice, and what would a researcher need to find a valid set of instruments for it?

KEY TAKEAWAYS

- Weekday dummies fail the **relevance condition** ($F = 0.2$) → they provide no useful exogenous variation to trace out the supply curve
- **Root cause:** price in the fish market is determined primarily by daily supply conditions (weather), not by predictable weekly demand patterns
- To identify the supply equation, a researcher needs **variables that shift demand** (and thus price) without shifting supply (income shocks, consumer preference changes, nearby competitor prices on specific days)
- Angrist, Graddy & Imbens (2000) revisit this market using different demand-shifter strategies and find the supply curve is indeed hard to identify cleanly

Discussion

3. In the demand equation, weekday dummies appear as **control variables (w's)**, not as instruments (z's). Why is this distinction important?

KEY TAKEAWAYS

- In the demand equation:
  - **w's (weekday dummies):** they affect quantity demanded **directly** and must be in the structural equation
  - **z's (weather):** excluded instruments → they affect price but only through supply
- If weekday dummies were omitted from the demand equation and weekday effects partly correlated with weather (e.g., storms happen more on certain days), the weather IVs would no longer be exogenous