Advanced Econometrics: 2SLS and Instrumental Variables

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute

New Delhi

April 25, 2025

Contents

1 Introduction: Why Instrumental Variables? 2
- 1.1 The Problem of Endogeneity . . . 2
- 1.2 Instrumental Variables (IV): The Solution . . . 2

2 What is an Instrumental Variable? 2
- 2.1 Formal Definition . . . 2
- 2.2 Intuitive Example . . . 2

3 Two-Stage Least Squares (2SLS): Theory and Derivation 3
- 3.1 The Basic Model . . . 3
- 3.2 2SLS Procedure (Step-by-Step) . . . 3
- 3.3 Matrix Formulation . . . 3
- 3.4 Properties of 2SLS . . . 3

4 Conditions for Valid Instruments 3
- 4.1 Instrument Relevance . . . 3
- 4.2 Instrument Exogeneity . . . 4
- 4.3 Overidentification . . . 4

5 Instrumental Variables in Impact Assessment 4
- 5.1 Why IV for Impact Evaluation? . . . 4
- 5.2 Example: Training Program for Farmers . . . 4
- 5.3 Interpretation . . . 4

6 Testing Instrument Validity and Strength 4
- 6.1 Tests for Weak Instruments . . . 4
- 6.2 Overidentification Tests . . . 5
- 6.3 Testing for Endogeneity . . . 5
- 6.4 Testing Instrument Exogeneity . . . 5

7 Practical Examples from Agricultural Economics 5
- 7.1 Example 1: Impact of Microcredit on Farm Productivity . . . 5
- 7.2 Example 2: Fertilizer Use and Crop Yield . . . 5
- 7.3 Example 3: Extension Services and Technology Adoption . . . 6
- 7.4 Example 4: Irrigation and Farm Income . . . 6

8 Summary Table: IV and 2SLS in Practice 6

1 Introduction: Why Instrumental Variables?

1.1 The Problem of Endogeneity

In many econometric models, an explanatory variable may be correlated with the error term. This endogeneity can arise from:

- Simultaneity (joint determination of variables)
- Omitted variable bias
- Measurement error
- Reverse causality

Consequence: OLS estimates are biased and inconsistent.

1.2 Instrumental Variables (IV): The Solution

The **instrumental variables** method provides a way to obtain consistent estimates when regressors are endogenous. An **instrument** is a variable that is:

- **Relevant**: Correlated with the endogenous regressor.
- **Valid (Exogenous)**: Uncorrelated with the error term.

2 What is an Instrumental Variable?

2.1 Formal Definition

Given the model:

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

If x_i is endogenous ($\text{Cov}(x_i, u_i) \neq 0$), an instrument z_i satisfies:

1. **Relevance:** $\text{Cov}(z_i, x_i) \neq 0$
2. **Exogeneity:** $\text{Cov}(z_i, u_i) = 0$

2.2 Intuitive Example

Suppose x_i is fertilizer use, y_i is crop yield, and u_i includes unobserved soil quality. If farmers choose x_i based on soil quality, OLS is biased. Suppose z_i is a government program assignment (randomized), which affects x_i but not u_i directly. Then z_i is a valid instrument.

3 Two-Stage Least Squares (2SLS): Theory and Derivation

3.1 The Basic Model

Consider the structural equation:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

where x_{1i} is endogenous, x_{2i} is exogenous.

3.2 2SLS Procedure (Step-by-Step)

1. **First Stage:** Regress x_{1i} on all exogenous variables (instruments), obtain fitted values $\hat{x}_{1i}$.
$$x_{1i} = \pi_0 + \pi_1 x_{2i} + \pi_2 z_i + v_i$$
2. **Second Stage:** Regress y_i on $\hat{x}_{1i}$ and x_{2i}:
$$y_i = \beta_0 + \beta_1 \hat{x}_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

Result: The 2SLS estimator is consistent even if x_{1i} is endogenous.

3.3 Matrix Formulation

Let Y be $n \times 1$, X $(n \times k)$ contain both endogenous and exogenous regressors, and Z $(n \times l)$ the instruments (all exogenous variables).

$$\hat{\beta}_{2SLS} = (X_P' X_P)^{-1} X_P' Y$$

where $X_P = P_Z X$ and $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix onto the space spanned by Z.

3.4 Properties of 2SLS

- Consistent and asymptotically normal if instruments are valid.
- In exactly identified models, 2SLS and IV are equivalent.
- In overidentified models, 2SLS uses all available instruments.

4 Conditions for Valid Instruments

4.1 Instrument Relevance

- The instrument must be strongly correlated with the endogenous regressor.
- Weak instruments (low correlation) lead to bias and large standard errors.
- **Check:** First-stage F-statistic should typically exceed 10.

4.2 Instrument Exogeneity

- The instrument must be uncorrelated with the error term in the structural equation.
- This is a theoretical assumption, but can be tested (see below).

4.3 Overidentification

- If the number of instruments > number of endogenous regressors, the equation is overidentified.
- Overidentification allows for tests of instrument validity (see below).

5 Instrumental Variables in Impact Assessment

5.1 Why IV for Impact Evaluation?

In program evaluation and policy analysis, treatment or exposure is often endogenous (e.g., farmers self-select into a training program). IV methods allow us to estimate causal effects when randomization is not feasible.

5.2 Example: Training Program for Farmers

Suppose y_i is yield, D_i is participation in a training program (endogenous), Z_i is random assignment (instrument).

$$y_i = \alpha + \beta D_i + \gamma X_i + u_i$$

First stage: $D_i = \pi_0 + \pi_1 Z_i + \pi_2 X_i + v_i$ Second stage: $y_i = \alpha + \beta \hat{D}_i + \gamma X_i + \varepsilon_i$ β is the local average treatment effect (LATE) if Z_i is a valid instrument.

5.3 Interpretation

IV estimates the causal effect for those whose treatment status is affected by the instrument (compliers).

6 Testing Instrument Validity and Strength

6.1 Tests for Weak Instruments

- **First-stage F-statistic:** Should be > 10 for strong instruments (Staiger and Stock, 1997).
- **Partial R^2:** How much of the variation in the endogenous regressor is explained by the instruments.

6.2 Overidentification Tests

- **Sargan-Hansen Test:** Tests whether instruments are uncorrelated with the error term.
- **Procedure:** Estimate the model by 2SLS, regress residuals on all instruments, compute test statistic.

6.3 Testing for Endogeneity

- **Hausman Test:** Compares OLS and IV/2SLS estimates. Significant difference suggests endogeneity.

6.4 Testing Instrument Exogeneity

- If overidentified, use Sargan-Hansen test.
- If just-identified, exogeneity cannot be tested directly.

7 Practical Examples from Agricultural Economics

7.1 Example 1: Impact of Microcredit on Farm Productivity

Source: Diagne, A., Zeller, M. (2001). "Access to Credit and Its Impact on Welfare in Malawi." *Research Report 116, IFPRI.*

Endogeneity: Access to microcredit (D_i) is endogenous because more productive farmers may self-select into borrowing.

Instrument: Distance to the nearest microfinance institution (Z_i).

How IV Solves It: Z_i affects the likelihood of borrowing but is plausibly unrelated to unobserved farmer productivity.

Details:

- First stage: $D_i = \pi_0 + \pi_1 Z_i + \pi_2 X_i + v_i$
- Second stage: $Y_i = \alpha + \beta \hat{D}_i + \gamma X_i + u_i$
- IV estimate of β gives the causal effect of microcredit access on productivity.

7.2 Example 2: Fertilizer Use and Crop Yield

Source: Sheahan, M., Barrett, C. B. (2017). "Ten striking facts about agricultural input use in Sub-Saharan Africa." *Food Policy*, 67, 12–25.

Endogeneity: Fertilizer use is endogenous (farmers with better soils may use more fertilizer).

Instrument: Village-level fertilizer price or government subsidy eligibility.

How IV Solves It: Price/subsidy is correlated with fertilizer use but plausibly exogenous to plot-level unobserved soil quality.

Details:

- First stage: Fertilizer use regressed on price/subsidy and controls.

- Second stage: Yield regressed on predicted fertilizer use.
- IV corrects for bias in OLS due to omitted soil quality.

7.3 Example 3: Extension Services and Technology Adoption

Source: Kondylis, F., Mueller, V., Zhu, J. (2017). "Seeing is believing? Evidence from an extension network experiment." *Journal of Development Economics*, 125, 1–16.

Endogeneity: Participation in extension services is endogenous (motivated farmers more likely to participate).

Instrument: Randomized invitation to extension meetings.

How IV Solves It: Random assignment is exogenous and affects participation, allowing causal estimation of extension effect on technology adoption.

Details:

- First stage: Participation regressed on invitation status and controls.
- Second stage: Adoption regressed on predicted participation.
- IV estimate captures causal impact of extension services.

7.4 Example 4: Irrigation and Farm Income

Source: Dillon, B. (2017). "Selling crops early reduces storage and fertilizer use: Evidence from a randomized controlled trial in Kenya." *Journal of Development Economics*, 126, 172–189.

Endogeneity: Access to irrigation is endogenous (richer farmers may have more access).

Instrument: Randomized allocation of irrigation pumps.

How IV Solves It: Randomization ensures exogeneity, allowing estimation of causal effect of irrigation on income.

Details:

- First stage: Irrigation access regressed on random allocation.
- Second stage: Income regressed on predicted irrigation access.
- IV estimate gives causal effect.

8 Summary Table: IV and 2SLS in Practice

Step	Purpose	Key Output	Tests
1. Choose IVs	Solve endogeneity	Instrument variables	Relevance, exogeneity
2. First stage	Predict endogenous regressor	Fitted values, F-statistic	Weak IV test
3. Second stage	Estimate structural equation	Causal effect estimate	Hausman test
4. Overid. test	Test instrument validity	Sargan-Hansen statistic	Overid. p-value

Glossary

- **Instrumental Variable (IV)**: A variable correlated with the endogenous regressor but uncorrelated with the error.
- **2SLS**: Two-stage least squares, an estimation method using IVs.
- **Endogeneity**: Correlation between regressor and error term.
- **Weak Instrument**: Instrument only weakly correlated with endogenous regressor.
- **Overidentification**: More instruments than endogenous regressors.
- **Sargan-Hansen Test**: Test for instrument exogeneity.
- **Hausman Test**: Test for endogeneity.
- **Local Average Treatment Effect (LATE)**: Causal effect for compliers in IV settings.

Practice Questions

1. Explain the two conditions for a valid instrument. Give an example from agriculture.
2. Outline the steps of 2SLS estimation and explain why each is necessary.
3. What are the consequences of using a weak instrument?
4. How can you test for instrument relevance and exogeneity?
5. In an impact evaluation, why might IV be preferred over OLS?
6. For each of the agricultural economics examples above, explain what was endogenous and how IV solved the issue.
7. Design an IV strategy for evaluating the impact of a fertilizer subsidy program.
8. Discuss the interpretation of IV estimates when treatment effects are heterogeneous.

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.
- Diagne, A., & Zeller, M. (2001). "Access to Credit and Its Impact on Welfare in Malawi." IFPRI Research Report 116.
- Sheahan, M., & Barrett, C. B. (2017). "Ten striking facts about agricultural input use in Sub-Saharan Africa." *Food Policy*, 67, 12–25.

- Kondylis, F., Mueller, V., & Zhu, J. (2017). "Seeing is believing? Evidence from an extension network experiment." *Journal of Development Economics*, 125, 1–16.
- Dillon, B. (2017). "Selling crops early reduces storage and fertilizer use: Evidence from a randomized controlled trial in Kenya." *Journal of Development Economics*, 126, 172–189.
- Staiger, D., & Stock, J. H. (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3), 557–586.