Instrumental Variables

(IV)

by

Luis Chancí[1]

1 Introduction

In the previous notes, we studied OLS as the benchmark estimator for linear models. Under the standard exogeneity condition,

$$\mathbb{E}[x_i u_i] = 0,$$

OLS is consistent, asymptotically normal, and easy to interpret. But that conclusion depends critically on the regressors being orthogonal to the structural error term. Once that condition fails, the logic of OLS breaks down.

This is the setting in which Instrumental Variables (IV) enters. IV is designed for situations in which one or more regressors are endogenous, that is, correlated with the error term. In such cases, OLS generally converges to the wrong object, even in large samples. The role of IV is to recover exogenous variation in the problematic regressor by using another variable, called an *instrument*, that shifts the endogenous regressor but is otherwise unrelated to the structural error.

The connection with OLS is useful to keep in mind from the start. OLS works by projecting y onto the regressor space spanned by X. IV modifies that logic: instead of using the full variation in X, it uses only the variation in X that is induced by a valid instrument Z. In that sense, IV can be understood as replacing contaminated variation with exogenous variation. This makes IV a natural extension of OLS rather than a completely separate technique.

[1]Contact me at: luis.chanci@usach.cl, luischanci@santotomas.cl, or lchanci1@binghamton.edu. These notes mirror my natural thought process for teaching this material, blending the accessible style of Wooldridge or Angrist with the formal rigor of Hansen. In the spirit of transparency, I used AI help (like Gemini) to check spelling and grammar. This version: 2026.

These notes are organized around four main ideas. First, IV is motivated by the failure of OLS under endogeneity. Second, the validity of IV rests on a small number of core conditions, especially instrument relevance and instrument exogeneity. Third, the resulting estimator can be written in a very transparent matrix form, known as Two-Stage Least Squares (2SLS). Fourth, the asymptotic theory of IV follows directly from the moment-condition logic that will later be generalized in GMM.

The structure of these notes is as follows. Section 2 introduces the endogeneity problem and the intuition behind instruments. Section 3 presents the main conditions a valid instrument must satisfy. Section 4 studies the exactly identified scalar case, where the basic IV logic is most transparent. Section 5 develops the general matrix formulation of 2SLS and shows how it relates back to OLS. Section 6 derives consistency, asymptotic normality, and feasible variance estimation step by step. Section 7 discusses practical issues such as weak instruments and overidentification. Finally, Section 8 shows how IV fits naturally into the broader GMM framework that will be developed in the next set of notes.

1.1 Directed Acyclic Graphs (DAGs) and the intuition of IV

Before introducing IV formally, it is useful to visualize the endogeneity problem with a *Directed Acyclic Graph* (DAG). A DAG is a graphical representation of causal or dependence relationships among variables. Each node represents a variable, and each arrow represents a directed relationship. The term *acyclic* means that the graph does not contain feedback loops: following the direction of the arrows, one cannot return to the same node.

In econometrics, DAGs are useful because they help clarify which correlations are harmless and which ones threaten identification. In the OLS framework, the key requirement is that the regressor of interest be orthogonal to the structural error. In DAG language, this means that there should be no open path connecting the regressor to the unobserved determinants of the outcome. Endogeneity arises precisely when such a path exists.

Figure 1 illustrates this idea. On the left, the regressor X is exogenous: the unobserved component U does not affect X, so OLS is conceptually appropriate. In the middle, endogeneity appears because U affects both X and Y, creating correlation between the regressor and the structural error. On the right, an instrumental variable Z provides exogenous variation in X: it is correlated with X, but orthogonal to U. This is the variation that IV exploits for identification.

Thus, the role of an instrument can be summarized graphically as follows: it must

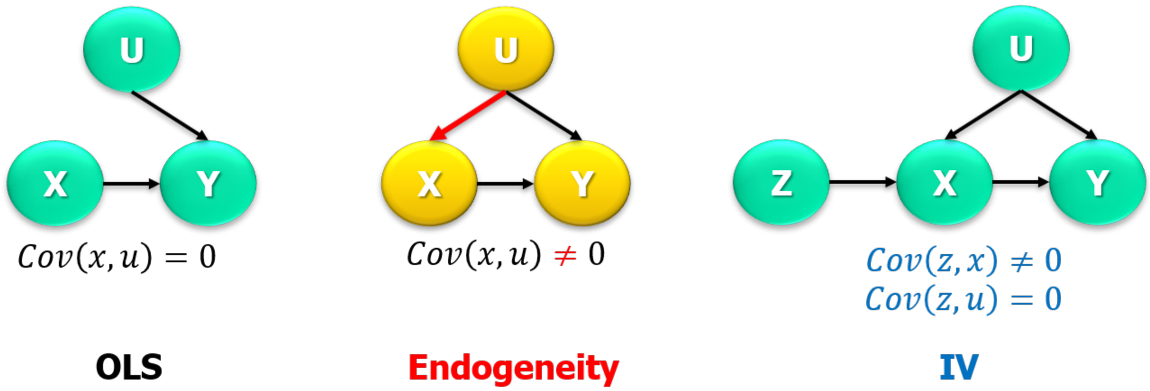

Figure 1: DAG intuition for exogeneity, endogeneity, and instrumental variables.

shift the endogenous regressor, but it must do so without being connected to the unobserved determinants of the outcome. This graphical intuition is useful because it anticipates the two core IV requirements that we formalize next: relevance and exogeneity.

2 Endogeneity and the Intuition of IV

Consider the linear model

$$y_i = x_i'\beta + u_i.$$

If

$$\mathbb{E}[x_i u_i] = 0,$$

then OLS is consistent. If instead

$$\mathbb{E}[x_i u_i] \neq 0,$$

then OLS is generally inconsistent.

This failure is usually called *endogeneity*. Conceptually, it means that the regressor contains variation that is systematically related to the unobserved determinants of the outcome. Once that happens, the regressor is no longer cleanly separated from the error term, so the OLS coefficient mixes the causal or structural effect of interest with spurious correlation.

2.1 Sources of endogeneity

Several mechanisms can generate endogeneity:

- **Omitted variables.** A relevant variable that affects y is excluded from the model and is correlated with x.
- **Simultaneity.** The explanatory variable and the dependent variable are jointly determined.
- **Measurement error.** The regressor is measured with error, contaminating the relationship between x and u.

The omitted-variable case is the simplest way to see the problem. Suppose the true population model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad \mathbb{E}[u_i \mid x_{1i}, x_{2i}] = 0,$$

but we estimate the short regression, omitting x_2:

$$y_i = \alpha_0 + \beta_1 x_{1i} + v_i,$$

where the composite error is

$$v_i = \beta_2 x_{2i} + u_i.$$

Because OLS estimates the linear projection of y_i onto x_{1i}, the probability limit of the OLS estimator is

$$\text{plim}\, \hat{\beta}_1^{OLS} = \frac{\text{Cov}(x_{1i}, y_i)}{\text{Var}(x_{1i})} = \beta_1 + \beta_2 \frac{\text{Cov}(x_{1i}, x_{2i})}{\text{Var}(x_{1i})}.$$

The second term is the omitted-variable bias. It shows that, even if u_i is orthogonal to the regressors in the true model, omitting x_2 generally induces correlation between x_{1i} and the composite error v_i. Thus, OLS is inconsistent unless either $\beta_2 = 0$ or $\text{Cov}(x_{1i}, x_{2i}) = 0$.

2.2 The basic IV idea

Suppose there exists a variable z_i such that:

(i) it is correlated with the endogenous regressor x_i, and

(ii) it is uncorrelated with the structural error u_i.

Then z_i provides an exogenous source of variation in x_i. The idea of IV is to isolate the component of x_i that is predictable from z_i and use only that component to estimate the structural effect.

This is why instruments are often described informally as creating a "natural experiment." They move the endogenous regressor in a way that is not contaminated by the omitted factors collected in the error term.

Remark 1. The intuition is easy to summarize: OLS uses all variation in x, while IV uses only the variation in x that comes from z. If that induced variation is exogenous, then IV can recover the structural parameter even when OLS cannot.

3 Conditions for a Valid Instrument

The usefulness of IV depends entirely on the validity of the instrument. In practice, this is the central challenge of IV applications.

3.1 Informal conditions

An instrument must satisfy three intuitive requirements:

- **Relevance:** the instrument must be related to the endogenous regressor.
- **Exogeneity:** the instrument must be orthogonal to the structural error.
- **Exclusion:** the instrument must affect the dependent variable only through the endogenous regressor.

The exclusion restriction is often the key economic assumption. It is what rules out direct effects of the instrument on the outcome.

3.2 Formal conditions

To formalize the setup, consider the structural equation

$$y_i = x_i'\beta + u_i,$$

and let z_i denote an $r \times 1$ vector of instruments.

Assumption 1 (Instrument validity)**.** Assume:

(i) **Instrument exogeneity:**

$$\mathbb{E}[z_i u_i] = 0.$$

(ii) **Instrument relevance:**

$$\operatorname{rank}\big(\mathbb{E}[z_i x_i']\big) = k,$$

where k is the number of regressors to be instrumented.

The first condition gives the orthogonality restrictions that identify the parameter. The second ensures that the instruments contain enough variation to pin down the coefficients. If relevance fails, the instruments are weak or irrelevant, and IV breaks down.

Remark 2 (Exclusion versus exogeneity). In structural applications, exclusion and exogeneity are conceptually distinct. Exclusion means that the instrument does not enter the structural equation directly. Exogeneity means that it is uncorrelated with the structural error. In reduced-form notation, the key formal condition for consistency is $\mathbb{E}[z_i u_i] = 0$, but credible IV practice usually requires an economic argument for exclusion as well.

4 The Exactly Identified Scalar Case

The simplest IV case is the one-regressor, one-instrument model:

$$y_i = \beta x_i + u_i,$$

with scalar instrument z_i.

This case is especially useful because the estimator can be derived directly from covariance restrictions.

4.1 Covariance derivation

Assume

$$\mathbb{E}[z_i u_i] = 0.$$

Taking covariance with z_i on both sides of the structural equation,

$$\begin{aligned} \operatorname{Cov}(z_i, y_i) &= \operatorname{Cov}(z_i, \beta x_i + u_i) \\ &= \beta \operatorname{Cov}(z_i, x_i) + \operatorname{Cov}(z_i, u_i). \end{aligned}$$

Using the orthogonality condition,

$$\operatorname{Cov}(z_i, u_i) = 0,$$

we obtain
$$\beta = \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)}.$$

Replacing population covariances with their sample analogues yields the IV estimator:
$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}.$$

This formula makes the logic of IV very transparent: the coefficient is identified by how the instrument co-moves with the outcome relative to how it co-moves with the endogenous regressor.

4.2 Moment-condition derivation

The same estimator can also be derived as a method-of-moments estimator. Since

$$u_i = y_i - x_i\beta,$$

the orthogonality condition implies

$$\mathbb{E}[z_i(y_i - x_i\beta)] = 0.$$

The sample analogue is
$$\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i\beta) = 0.$$

Solving for β gives
$$\hat{\beta}_{IV} = \left(\sum_{i=1}^{n} z_i x_i\right)^{-1}\left(\sum_{i=1}^{n} z_i y_i\right).$$

This expression is algebraically equivalent to the covariance formula above after centering when an intercept is included. More importantly, it reveals the logic that will later become central in GMM: IV is fundamentally an estimator based on orthogonality conditions.

Remark 3. In the exactly identified scalar case, the IV estimator solves one sample moment condition with one unknown parameter. This is the cleanest setting in which the connection between IV and GMM becomes visible.

5 Two-Stage Least Squares and the Matrix Formulation

The scalar case is useful for intuition, but most applications involve multiple regressors and often multiple instruments. The standard estimator in that setting is Two-Stage Least Squares (2SLS).

5.1 Structural model with endogenous and exogenous regressors

Partition the regressors as

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

where:

- x_{1i} is an $m \times 1$ vector of endogenous regressors,
- x_{2i} is a $p \times 1$ vector of included exogenous regressors.

Let the instrument vector be

$$z_i = \begin{bmatrix} z_{1i} \\ x_{2i} \end{bmatrix},$$

where z_{1i} are excluded instruments and the included exogenous regressors instrument themselves.

Stacking the sample, write

$$y = X\beta + u,$$

where $X = [X_1 \; X_2]$, and let Z denote the $n \times r$ matrix of instruments.

5.2 The 2SLS estimator

Define the projection matrix onto the column space of Z:

$$P_Z = Z(Z'Z)^{-1}Z'.$$

The Two-Stage Least Squares estimator is

$$\hat{\beta}_{2SLS} = (X'P_ZX)^{-1}X'P_Zy.$$

This formula can be understood from the two-stage algorithm:

First stage. Project the endogenous regressors onto the instrument space:

$$\hat{X} = P_ZX.$$

Second stage. Regress y on the fitted regressors:

$$\hat{\beta}_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'y.$$

Since $\hat{X} = P_Z X$, this gives exactly

$$\hat{\beta}_{2SLS} = (X'P_Z X)^{-1}X'P_Z y.$$

Remark 4 (Why are naive second-stage standard errors wrong?)*.* In the manual two-stage description, the fitted regressors from the first stage are estimated objects, not fixed data. Therefore, the standard OLS standard errors from the second-stage regression do not account correctly for first-stage estimation. Proper 2SLS variance formulas must incorporate the full sampling variation of the estimator.

5.3 Connection with OLS

The matrix expression for 2SLS makes the link with OLS immediate. If the regressors are all exogenous, then we may choose

$$Z = X.$$

In that case,

$$P_Z = P_X = X(X'X)^{-1}X'.$$

Substituting into the 2SLS formula:

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'P_X X)^{-1}X'P_X y \\ &= (X'X)^{-1}X'y \\ &= \hat{\beta}_{OLS}.\end{aligned}$$

So OLS is a special case of IV in which the regressor matrix instruments itself.

Remark 5*.* This identity is useful conceptually. IV does not replace OLS in general; rather, it extends OLS to situations where some regressors are endogenous. When no endogeneity is present, the two estimators coincide.

6 Asymptotic Theory of IV

The statistical justification for IV is mainly asymptotic. In finite samples, IV estimators can be biased and imprecise, especially when instruments are weak. But under standard conditions, 2SLS is consistent and asymptotically normal.

6.1 Regularity conditions

Let

$$Q_{ZX} = \mathbb{E}[z_i x_i'], \qquad Q_{ZZ} = \mathbb{E}[z_i z_i'].$$

Assumption 2 (Regularity conditions for IV)**.** Assume:

(i) $\{(y_i, x_i, z_i)\}_{i=1}^n$ is i.i.d.

(ii) $\mathbb{E}[z_i u_i] = 0$.

(iii) Q_{ZZ} exists and is positive definite.

(iv) $\text{rank}(Q_{ZX}) = k$.

(v) Suitable moment conditions hold so that LLN and CLT apply to the relevant sample moments.

The key identifying ingredients are still exogeneity and relevance. The remaining conditions ensure the sample analogues converge to their population counterparts.

6.2 Consistency: step-by-step derivation

Start from the 2SLS estimator:

$$\hat{\beta}_{2SLS} = (X'P_Z X)^{-1} X' P_Z y.$$

Substitute the structural model $y = X\beta_0 + u$:

$$\begin{aligned} \hat{\beta}_{2SLS} &= (X'P_Z X)^{-1} X' P_Z (X\beta_0 + u) \\ &= \beta_0 + (X'P_Z X)^{-1} X' P_Z u. \end{aligned}$$

Using $P_Z = Z(Z'Z)^{-1}Z'$, rewrite as

$$\hat{\beta}_{2SLS} = \beta_0 + \left(\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'X}{n} \right)^{-1} \left(\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'u}{n} \right).$$

Now apply the Law of Large Numbers:

$$\frac{X'Z}{n} \xrightarrow{p} Q_{XZ}, \qquad \frac{Z'Z}{n} \xrightarrow{p} Q_{ZZ}, \qquad \frac{Z'u}{n} \xrightarrow{p} \mathbb{E}[z_i u_i] = 0.$$

Therefore,

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta_0.$$

Proposition 6.1 (Consistency of 2SLS)**.** *Under the regularity conditions above,*

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta_0.$$

The result is intuitive. The instruments isolate a component of X that remains orthogonal to the structural error. As the sample grows, the estimator uses that exogenous component to recover the true parameter vector.

6.3 Asymptotic normality: step-by-step derivation

To obtain the limiting distribution, start again from

$$\hat{\beta}_{2SLS} - \beta_0 = (X'P_Z X)^{-1} X'P_Z u.$$

Multiply by $\sqrt{n}$:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) = A_n^{-1} C_n,$$

where

$$A_n = \frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n},$$

and

$$C_n = \frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'u}{\sqrt{n}}.$$

Now analyze each piece separately.

Step 1: the deterministic part. By LLN,

$$A_n \xrightarrow{p} A = Q_{XZ} Q_{ZZ}^{-1} Q_{ZX}.$$

Step 2: the stochastic part. By CLT,

$$\frac{Z'u}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where

$$\Omega = \mathbb{E}[z_i z_i' u_i^2].$$

Therefore,

$$C_n \xrightarrow{d} \mathcal{N}(0,\, Q_{XZ}Q_{ZZ}^{-1}\Omega Q_{ZZ}^{-1}Q_{ZX}).$$

Step 3: combine the pieces. By Slutsky's theorem,

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V_{IV}),$$

with asymptotic variance

$$V_{IV} = A^{-1}(Q_{XZ}Q_{ZZ}^{-1}\Omega Q_{ZZ}^{-1}Q_{ZX})A^{-1}.$$

Theorem 6.2 (Asymptotic normality of 2SLS)**.** *Under the regularity conditions above,*

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V_{IV}),$$

where

$$V_{IV} = A^{-1}(Q_{XZ}Q_{ZZ}^{-1}\Omega Q_{ZZ}^{-1}Q_{ZX})A^{-1}, \qquad A = Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}.$$

Under homoskedasticity, this reduces to

$$\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \frac{\hat{\sigma}^2}{n}\left(\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}\right)^{-1}.$$

Using the sample-average definitions

$$\hat{Q}_{XZ} = \frac{X'Z}{n}, \qquad \hat{Q}_{ZZ} = \frac{Z'Z}{n}, \qquad \hat{Q}_{ZX} = \frac{Z'X}{n},$$

we obtain

$$\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \frac{\hat{\sigma}^2}{n}\left(\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n}\right)^{-1} = \hat{\sigma}^2\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}.$$

Since

$$P_Z = Z(Z'Z)^{-1}Z',$$

this can be written equivalently as

$$\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \hat{\sigma}^2(X'P_ZX)^{-1}.$$

This is the familiar homoskedastic 2SLS variance formula commonly reported by statistical software, and it makes explicit the connection between the asymptotic derivation and the projection matrix P_Z.

6.4 Feasible variance estimation

In practice, we replace the population matrices by sample analogues. Define

$$\hat{Q}_{XZ} = \frac{X'Z}{n}, \qquad \hat{Q}_{ZZ} = \frac{Z'Z}{n},$$

and let

$$\hat{u} = y - X\hat{\beta}_{2SLS}.$$

The heteroskedasticity-robust variance estimator is

$$\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \frac{1}{n}\hat{A}^{-1}\hat{B}\hat{A}^{-1},$$

where

$$\hat{A} = \hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX},$$

and

$$\hat{B} = \hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{\Omega}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}, \qquad \hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 z_i z_i'.$$

Under homoskedasticity, this reduces to

$$\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \frac{\hat{\sigma}^2}{n}\left(\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}\right)^{-1}.$$

This is the correct large-sample variance for the IV estimator itself, not for the scaled object $\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0)$.

7 Inference and Diagnostics

7.1 Testing hypotheses on the coefficients

Once a consistent variance estimator is available, inference follows the same large-sample logic as in OLS and MLE. For a single coefficient,

$$t_j = \frac{\hat{\beta}_j - \beta_{j,0}}{\widehat{se}(\hat{\beta}_j)} \xrightarrow{d} \mathcal{N}(0,1).$$

For joint linear restrictions

$$H_0 : R\beta = r,$$

the Wald statistic is

$$W = (R\hat{\beta} - r)' \left[R\widehat{\text{Var}}(\hat{\beta})R' \right]^{-1} (R\hat{\beta} - r) \xrightarrow{d} \chi^2_q.$$

7.2 Weak instruments

The asymptotic theory above assumes the instruments are relevant enough for the first-stage relationship to be informative. In practice, instruments may be only weakly correlated with the endogenous regressors. When that happens, 2SLS can behave poorly in finite samples: it becomes biased, imprecise, and inference based on standard asymptotic approximations can be severely distorted.

A simple diagnostic is the first-stage regression. In the single endogenous regressor case, one usually examines the first-stage F-statistic on the excluded instruments. A very small first-stage F is a warning sign that the instrument may be weak.

Remark 6. Weak instruments are one of the main practical concerns in IV estimation. They do not merely reduce efficiency; they can also invalidate standard inference. This is one reason why assessing instrument relevance is as important as arguing for instrument exogeneity.

7.3 Overidentification

If the number of instruments exceeds the number of endogenous regressors, the model is *overidentified*. In that case, the orthogonality conditions

$$\mathbb{E}[z_i u_i] = 0$$

provide more moment restrictions than unknown parameters.

This creates an opportunity for testing the joint validity of the overidentifying restrictions. The basic idea is simple: if all instruments are valid, then the sample moment conditions evaluated at the IV estimate should be close to zero. A standard overidentification test is the Hansen J-statistic:

$$J = n\, g_n(\hat{\beta}_{2SLS})'\hat{\Omega}^{-1} g_n(\hat{\beta}_{2SLS}),$$

where

$$g_n(\beta) = \frac{1}{n} Z'(y - X\beta),$$

and

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z_i z_i'$$

is the consistent estimator of the covariance matrix of the sample moments, with $\hat{u} = y - X\hat{\beta}_{2SLS}$.

Under the null hypothesis that all instruments are valid,

$$J \xrightarrow{d} \chi^2_{r-k}.$$

Remark 7 (The Sargan statistic). Under homoskedasticity, the weighting matrix simplifies, and the overidentification test can be written as

$$\text{Sargan} = \frac{\hat{u}' P_Z \hat{u}}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{r-k}.$$

Equivalently, it is often computed as $n \times R_u^2$, where R_u^2 is the uncentered R^2 from a regression of the 2SLS residuals $\hat{u}$ on the full set of instruments Z.

This statistic is especially important because it anticipates the general GMM logic: testing whether the empirical moment conditions are jointly distinguishable from zero.

7.4 Testing exogeneity

Another common question is whether OLS and IV differ systematically enough to conclude that endogeneity is present. This leads to exogeneity tests such as the Durbin–Wu–Hausman test. The basic idea is to compare an estimator that is efficient under exogeneity (OLS) with one that remains consistent under endogeneity (IV). A large difference between the two suggests that the OLS exogeneity condition fails.

Remark 8. In practice, however, exogeneity tests should be interpreted with care. They depend on the quality of the instruments: a weak or invalid instrument can make the comparison between OLS and IV misleading.

8 Bridge to GMM

The IV estimator is not only a solution to an endogeneity problem; it is also one of the cleanest examples of a moment-based estimator. This is precisely what makes it the natural bridge to GMM.

Recall the IV moment condition:

$$\mathbb{E}[z_i(y_i - x_i'\beta)] = 0.$$

Define the sample moment vector

$$g_n(\beta) = \frac{1}{n}Z'(y - X\beta).$$

8.1 Exact identification

If the number of instruments equals the number of parameters, say $r = k$, then the IV estimator solves

$$g_n(\hat{\beta}) = 0.$$

In this case, the estimator is exactly identified, and there is no need to choose a weighting matrix. The moment equations can be solved directly.

8.2 Overidentification and the GMM objective

If instead $r > k$, then the system

$$g_n(\beta) = 0$$

contains more equations than unknowns and generally cannot be solved exactly. The natural strategy is therefore to choose β so that the sample moments are as close to zero as possible. This leads to the quadratic criterion

$$Q_n(\beta) = g_n(\beta)'W_n g_n(\beta),$$

where W_n is a positive definite weighting matrix.

Minimizing this objective yields the GMM estimator:

$$\hat{\beta}_{GMM} = \arg\min_{\beta} Q_n(\beta).$$

The corresponding first-order conditions imply

$$\hat{\beta}_{GMM} = \left(X'ZW_nZ'X\right)^{-1} X'ZW_nZ'y.$$

If we choose

$$W_n = (Z'Z)^{-1},$$

then this becomes

$$\hat{\beta} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'y,$$

which is exactly the usual IV/2SLS estimator.

Remark 9 (Why IV is the natural bridge to GMM). IV already contains all the essential ingredients of GMM:

- a vector of orthogonality conditions,
- a sample moment function,
- and, in the overidentified case, a quadratic criterion used to make the moments as close to zero as possible.

The main conceptual step from IV to GMM is therefore not a new estimation philosophy, but a generalization of the same logic to richer sets of moment conditions and more flexible weighting matrices.

9 Summary

Instrumental Variables estimation is designed for situations in which OLS fails because one or more regressors are endogenous. The core idea is to replace contaminated variation in the regressors with exogenous variation induced by valid instruments. We saw how the IV logic is most transparent in the exactly identified scalar case, how it generalizes to the matrix 2SLS estimator, why OLS appears as a special case when the instruments coincide with the regressors, and how the asymptotic theory of IV follows from the LLN and CLT applied to moment conditions. We also discussed practical issues such as weak instruments, overidentification, and robust variance estimation. In the next set of notes, this moment-condition perspective will be generalized formally through GMM, where IV will reappear as a central special case.