DegreeLive

Bachelor of Commerce (B2001) | Level 3 • Part B Major | Monash University

Principles of econometrics

Course Code: ETC3400 • Module 3 Notes

Module 3: Instrumental Variables, Two-Stage Least Squares & Simultaneous Equations

3.1 Endogeneity Sources and the Failure of OLS Consistency

In Module 2 we proved that OLS is consistent provided the weak orthogonality condition $E(x_i u_i) = 0$ holds. When at least one explanatory variable x_{ij} is correlated with the structural disturbance term u_i ($E(x_i u_i) \neq 0$), that regressor is **endogenous**, and the probability limit of the OLS estimator diverges from the true parameter vector β:

$$\operatorname{plim} \hat{\beta}_{OLS} = \beta + [E(x_i x_i')]^{-1} E(x_i u_i) = \beta + Q_{xx}^{-1} E(x_i u_i) \neq \beta$$

Where:

- $Q_{xx} = E(x_i x_i')$ is the $k \times k$ population second-moment matrix of the regressors.
- $Q_{xx}^{-1} E(x_i u_i)$ is the asymptotic inconsistency (asymptotic bias) vector of OLS, which does not vanish even as $n \to \infty$.

Four Structural Sources of Endogeneity in Economics

Endogeneity arises pervasively in non-experimental Australian economic and business datasets across four principal mechanisms:

1. **Omitted Variable Bias (Unobserved Confounding):** Suppose the true structural equation is $y_i = x_{1i}' \beta_1 + \gamma a_i + \varepsilon_i$, where a_i is an unobserved confounder (such as innate ability in a HILDA wage regression or managerial quality in an ASX firm productivity study). If a_i is omitted, it is absorbed into the composite error $u_i = \gamma a_i + \varepsilon_i$. Whenever $\gamma \neq 0$ and $\operatorname{Cov}(x_{1i}, a_i) \neq 0$, we obtain $E(x_{1i} u_i) = \gamma E(x_{1i} a_i) \neq 0$, biasing OLS.
2. **Classical Measurement Error (Errors-in-Variables):** Suppose $y_i = \beta_0 + \beta_1 x_i^* + \varepsilon_i$, where true x_i^* (e.g., permanent household income) is unobserved and the analyst observes a noisy proxy $x_i = x_i^* + e_i$ with classical measurement error e_i satisfying $E(e_i) = 0$, $\operatorname{Cov}(x_i^*, e_i) = 0$, and $\operatorname{Cov}(\varepsilon_i, e_i) = 0$. Substituting $x_i^* = x_i - e_i$ yields $y_i = \beta_0 + \beta_1 x_i + (\varepsilon_i - \beta_1 e_i)$. Because x_i is correlated with $u_i = \varepsilon_i - \beta_1 e_i$ via e_i, OLS suffers **attenuation bias** toward zero:

   $$\operatorname{plim} \hat{\beta}_{1,OLS} = \beta_1 \left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2} \right) = \lambda \beta_1, \qquad 0 < \lambda < 1$$

   Where $\lambda = \sigma_{x^*}^2 / (\sigma_{x^*}^2 + \sigma_e^2)$ is the reliability ratio of the measured regressor.
3. **Simultaneity (Reverse Causality):** In market equilibrium models—such as Australian housing supply and demand or Reserve Bank interest rate reaction functions—price p_i and quantity q_i are jointly determined by a system of structural equations. A positive demand shock shifts p_i and q_i simultaneously, inducing correlation between price and the supply error term.

DegreeLive

4. **Dynamic Specification with Serially Correlated Errors:** In a dynamic equation $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ where $u_t = \rho u_{t-1} + \varepsilon_t$ ($\rho \neq 0$), y_{t-1} depends on u_{t-1} and hence correlates with u_t, rendering OLS inconsistent.

Endogeneity Mechanism	Composite Error Structure	Direction of OLS Bias (Scalar Case)	Typical Australian Empirical Example
Omitted Variable	$u_i = \gamma a_i + \varepsilon_i$	Sign of $\gamma \cdot \mathrm{Cov}(x_i, a_i)$	Education returns omitting ability in HILDA
Classical Measurement Error	$u_i = \varepsilon_i - \beta_1 e_i$	Attenuation toward 0 (factor $\lambda < 1$)	Self-reported household wealth in ABS surveys
Simultaneity	Reduced-form shock passes into x_i	Mix of structural slopes and error variances	Electricity spot price & demand in NEM market
Lagged y_{t-1} + AR(1) Error	$u_t = \rho u_{t-1} + \varepsilon_t$	Upward bias if $\rho > 0$	Dynamic Australian consumption or inflation models

3.2 Instrumental Variables Identification and the Simple IV Estimator

Instrumental Variables (IV) estimation restores consistent parameter identification without observing the omitted confounder or eliminating simultaneity. Consider the linear structural equation:

$$y_i = x_i'\beta + u_i, \qquad E(x_i u_i) \neq 0$$

Suppose there exists an $\ell \times 1$ vector of instrumental variables z_i (which includes any exogenous regressors already in x_i, plus excluded instruments). For z_i to identify β cleanly, two foundational conditions must hold:

1. **Instrument Exogeneity (Exclusion Restriction / Orthogonality):**

   $$E(z_i u_i) = 0$$

   The instruments must be uncorrelated with the structural error term u_i. Equivalently, excluded instruments may affect y_i only indirectly through their causal effect on the endogenous regressors x_i, never directly or through omitted confounders. Because u_i is unobserved, exogeneity cannot be tested directly when $\ell = k$ and must be justified by institutional, geographic, or experimental design knowledge.

2. **Instrument Relevance (Rank Condition):**

   $$\mathrm{rank}\,(E(z_i x_i')) = k$$

DegreeLive

Where $E(z_i x_i^{'}) = Q_{zx}$ is an $\ell \times k$ matrix. Full column rank k requires $\ell \geq k$ (the **order condition**: at least as many instruments as regressors) and guarantees that the instruments exhibit genuine partial correlation with every endogenous regressor after conditioning on exogenous controls.

Just-Identified IV Estimator ($\ell = k$)

When $\ell = k$, the system of k population moment equations $E[z_i(y_i - x_i^{'}\beta)] = 0$ yields $Q_{zx}\beta = Q_{zy}$, where $Q_{zy} = E(z_i y_i)$. Because Q_{zx} is square ($k \times k$) and invertible under the rank condition, β is uniquely identified as $\beta = Q_{zx}^{-1} Q_{zy}$. Replacing population expectations by sample averages gives the simple IV estimator:

$$\hat{\beta}_{IV} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i^{'}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} z_i y_i\right) = (Z^{'}X)^{-1} Z^{'}y$$

Where Z is the $n \times k$ matrix of instruments. Substituting $y = X\beta + u$ gives $\hat{\beta}_{IV} = \beta + (\frac{1}{n}Z^{'}X)^{-1}(\frac{1}{n}Z^{'}u) \xrightarrow{p} \beta + Q_{zx}^{-1} \cdot 0 = \beta$, proving consistency.

The Wald Estimator and Local Average Treatment Effect (LATE)

In the scalar model $y_i = \beta_0 + \beta_1 x_i + u_i$ with a single binary instrument $z_i \in \{0, 1\}$ (such as random lottery eligibility for a training voucher or distance cutoff), the just-identified IV slope reduces to the **Wald estimator**:

$$\hat{\beta}_{1,IV} = \frac{\widehat{\mathrm{Cov}}(z_i, y_i)}{\widehat{\mathrm{Cov}}(z_i, x_i)} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

Where $\bar{y}_1$ and $\bar{x}_1$ are the sample means of y_i and x_i when $z_i = 1$, and $\bar{y}_0$ and $\bar{x}_0$ are the sample means when $z_i = 0$. Under heterogeneous treatment effects where β_{1i} varies across individuals, Imbens and Angrist (1994) demonstrated that if z_i is randomly assigned, satisfies the exclusion restriction, has non-zero first stage, and obeys **monotonicity** (no defiers—shifting z_i from 0 to 1 never reduces x_i), then the Wald estimator identifies the **Local Average Treatment Effect (LATE)**: the average causal effect strictly among *compliers* whose treatment status x_i is altered by the instrument.

Worked Example 1: Binary Instrument Wald Estimator and LATE

Suppose a policy researcher evaluates the effect of completing a tertiary university degree ($x_i \in \{0, 1\}$) on weekly log earnings (y_i) in a sample of $n = 1{,}000$ Australian adults. Because unobserved ability biases OLS upward to $\hat{\beta}_{1,OLS} = 0.55$, the analyst uses a binary indicator $z_i \in \{0, 1\}$ equal to 1 if the individual lived within 15 km of a university campus at age 17 ($n_1 = 500$) and 0 otherwise ($n_0 = 500$). Sample averages are $\bar{y}_1 = 7.12, \bar{x}_1 = 0.65$ for $z_i = 1$, and $\bar{y}_0 = 6.96, \bar{x}_0 = 0.25$ for $z_i = 0$.

Computing the reduced-form numerator $\bar{y}_1 - \bar{y}_0 = 7.12 - 6.96 = 0.16$ and the first-stage denominator $\bar{x}_1 - \bar{x}_0 = 0.65 - 0.25 = 0.40$ yields:

Degreelive

$$\hat{\beta}_{1,IV} = \frac{0.16}{0.40} = 0.40$$

Under the LATE assumptions, 40% of the population are compliers induced into university completion by campus proximity, and their average earnings return is 0.40 log points (approximately 49.2% in levels via $\exp(0.40) - 1$).

3.3 Two-Stage Least Squares (2SLS) in Matrix Notation

When the number of valid instruments exceeds the number of regressors ($\ell > k$, an **overidentified** model), $(Z'X)$ is $\ell \times k$ and non-square, so $(Z'X)^{-1}$ does not exist. Rather than discarding valuable instruments, Two-Stage Least Squares (2SLS) projects every column of X onto the column space $\mathrm{C}(Z)$ using the ℓ-dimensional projection matrix P_Z:

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

Where Z is the $n \times \ell$ matrix of all exogenous variables (both included exogenous regressors X_1 and excluded instruments Z_2). Partitioning $X = [X_1 \;\; X_2]$ (where X_1 is exogenous and X_2 is endogenous) and $Z = [X_1 \;\; Z_2]$, the two-stage algorithm proceeds algebraically as follows:

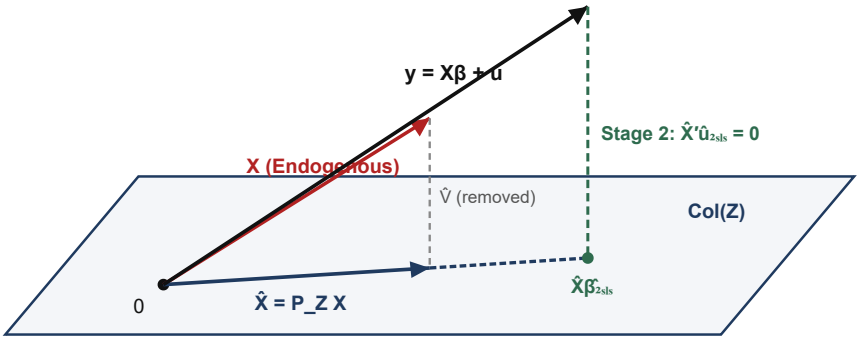
1. **Stage 1 (First-Stage Projection):** Regress each column of X on Z by OLS to form the matrix of first-stage fitted values $\hat{X} = P_Z X = Z\hat{\Pi}$, where $\hat{\Pi} = (Z'Z)^{-1}Z'X$. Notice that for any exogenous column $x_j \in \mathrm{C}(Z)$, $P_Z x_j = x_j$ exactly.
2. **Stage 2 (Second-Stage Regression):** Regress y on $\hat{X}$ by OLS:

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

   Where $P_Z'P_Z = P_Z$ holds because P_Z is symmetric and idempotent. Note carefully that Stage 2 standard errors must be computed using actual X in the residuals $\hat{u}_{2SLS} = y - X\hat{\beta}_{2SLS}$ rather than $y - \hat{X}\hat{\beta}_{2SLS}$, because the Stage 2 software output uses the wrong error variance $\hat{u}'_{Stage2}\hat{u}_{Stage2}/(n-k)$ instead of $\hat{u}'_{2SLS}\hat{u}_{2SLS}/(n-k)$.

Degree.live

Geometric Representation of 2SLS



Asymptotic Distribution of 2SLS

Let $Q_{zz} = E(z_i z_i')$ ($\ell \times \ell$, positive definite), $Q_{zx} = E(z_i x_i')$ ($\ell \times k$, full rank k), and $Q_{xz} = Q_{zx}'$. Substituting $y = X\beta + u$ into $\hat{\beta}_{2SLS}$ and scaling by $\sqrt{n}$ yields:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \left[\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\left(\frac{1}{n}Z'X\right)\right]^{-1}\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\left(\frac{1}{\sqrt{n}}Z'u\right)$$

Where:

- By the WLLN and CMT, $\frac{1}{n}Z'X \xrightarrow{p} Q_{zx}$ and $\frac{1}{n}Z'Z \xrightarrow{p} Q_{zz}$.
- By the Multivariate CLT, $\frac{1}{\sqrt{n}}Z'u \xrightarrow{d} N(0, \Omega_{zz})$, where $\Omega_{zz} = E(u_i^2 z_i z_i')$.

Applying Slutsky's Theorem gives the general heteroskedasticity-robust asymptotic distribution of 2SLS:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0,\ A^{-1}BA^{-1})$$

Where $A = Q_{xz}Q_{zz}^{-1}Q_{zx}$ and $B = Q_{xz}Q_{zz}^{-1}\Omega_{zz}Q_{zz}^{-1}Q_{zx}$. Under conditional homoskedasticity $E(u_i^2 \mid z_i) = \sigma^2$, $\Omega_{zz} = \sigma^2 Q_{zz}$ and the asymptotic covariance simplifies to:

$$\text{Avar}\left(\sqrt{n}(\hat{\beta}_{2SLS} - \beta)\right) = \sigma^2 \left(Q_{xz}Q_{zz}^{-1}Q_{zx}\right)^{-1}$$

In finite samples, the homoskedastic covariance matrix is estimated by $\hat{\sigma}^2(X'P_Z X)^{-1}$ with $\hat{\sigma}^2 = \hat{u}_{2SLS}'\hat{u}_{2SLS}/(n-k)$ (or $/n$), while the heteroskedasticity-robust sandwich estimator substitutes $\hat{\Omega}_{zz} = \sum_{i=1}^n \hat{u}_{i,2SLS}^2 z_i z_i'$ into $(X'P_Z X)^{-1}X'Z(Z'Z)^{-1}\hat{\Omega}_{zz}(Z'Z)^{-1}Z'X(X'P_Z X)^{-1}$. Note that when X is in fact exogenous, OLS has strictly smaller asymptotic variance than 2SLS because $Q_{xx} - Q_{xz}Q_{zz}^{-1}Q_{zx} \geq 0$: IV pays an efficiency cost in exchange for eliminating endogeneity bias.

Degree Live

3.4 Weak Instruments: Finite-Sample Bias and Diagnostic Rules

Standard asymptotic normality of 2SLS relies crucially on the denominator matrix $Q_{xz}Q_{zz}^{-1}Q_{zx}$ being bounded well away from zero. If the excluded instruments Z_2 are only weakly correlated with the endogenous regressor X_2—a condition known as **weak instruments**—the normal asymptotic approximation breaks down severely even in samples of several thousand observations.

Consequences of Weak Instruments

Consider a structural equation with one endogenous regressor x_2 and m_2 excluded instruments Z_2 after partialling out exogenous controls X_1, with first-stage equation $\widetilde{x}_2 = \widetilde{Z}_2\pi + v$. Define the dimensionless **concentration parameter** $\mu^2 = \pi'\widetilde{Z}_2'\widetilde{Z}_2\pi/\sigma_v^2$, which governs instrument strength. When μ^2 is small:

1. **Finite-Sample Bias Toward OLS:** Because v is correlated with the structural error u, the finite-sample expectation of $\hat{\beta}_{2SLS}$ is biased in the exact same direction as OLS, approximately satisfying $E(\hat{\beta}_{2SLS} - \beta) \approx \frac{\sigma_{uv}}{\sigma_v^2}\frac{m_2-2}{\mu^2+m_2}$. Notice that adding many weak instruments (m_2 large) worsens the finite-sample bias of 2SLS toward OLS!
2. **Non-Normal Distributions and Size Distortion:** The distribution of $\hat{\beta}_{2SLS}$ becomes skewed, multimodal, or Cauchy-like, while estimated standard errors severely understate uncertainty, causing Wald t-tests to reject true null hypotheses far more frequently than the nominal 5% level.

Testing for Weak Instruments and Robust Inference

- **First-Stage F-Statistic and Staiger-Stock Rule:** Test $H_0 : \pi = 0$ in the first-stage regression of the endogenous variable on all exogenous variables and excluded instruments. Under homoskedasticity with one endogenous regressor, Staiger and Stock (1997) established the benchmark rule of thumb $F > 10$. More formally, Stock and Yogo (2005) tabulated exact critical values: for example, with $m_2 = 1$ instrument, the critical value is 16.38 to ensure maximal 2SLS size distortion below 10%; with $m_2 = 3$ instruments, $F > 13.91$ bounds relative bias below 10%. Under heteroskedasticity or clustering, the Montiel Olea-Pflueger effective F-statistic should be used.
- **Limited Information Maximum Likelihood (LIML):** Unlike 2SLS, LIML is approximately median-unbiased even with many weak instruments, though it can exhibit heavy tails when m_2 is small.
- **Anderson-Rubin (AR) Test:** Provides exact weak-instrument-robust hypothesis testing for $H_0 : \beta_2 = \beta_{2,0}$ without requiring strong first-stage identification. Under H_0, $y - X_2\beta_{2,0} = X_1\beta_1 + u$ contains no endogenous variables; regressing $y - X_2\beta_{2,0}$ on X_1 and Z_2 and testing the joint significance of Z_2 via a standard F-test (or $\chi^2_{m_2}$ Wald test) maintains correct size regardless of how weak Z_2 is.

3.5 Specification Tests: Overidentifying Restrictions and Endogeneity

In an overidentified model ($\ell > k$), the analyst possesses $m = \ell - k > 0$ surplus moment conditions beyond those strictly required to identify β. These extra moment conditions permit testing whether the instruments are mutually coherent with $E(z_i u_i) = 0$ (assuming at least k instruments are valid), while comparing 2SLS and OLS evaluates whether X is truly endogenous.

DegreeLive

Sargan and Hansen J Tests of Overidentifying Restrictions

If all ℓ instruments are valid ($E(z_i u_i) = 0$), the 2SLS residual vector $\hat{u}_{2SLS} = y - X\hat{\beta}_{2SLS}$ should be approximately orthogonal to every column of Z. Under conditional homoskedasticity, the **Sargan statistic** measures the length of the projection of $\hat{u}_{2SLS}$ onto $\mathrm{C}(Z)$:

$$S = \frac{\hat{u}'_{2SLS} P_Z \hat{u}_{2SLS}}{\hat{\sigma}^2} = n R^2_{\hat{u}} \quad \overset{d}{} \quad \chi^2_{\ell - k}$$

Where:

- $\hat{\sigma}^2 = \hat{u}'_{2SLS}\hat{u}_{2SLS}/n$ is the consistent error variance estimate from 2SLS residuals.
- $R^2_{\hat{u}}$ is the uncentred R^2 from an auxiliary OLS regression of $\hat{u}_{2SLS}$ on all instruments Z.
- $\ell - k$ is the number of overidentifying restrictions (degrees of freedom).

Under heteroskedasticity or autocorrelation, the Sargan test is invalid and is replaced by **Hansen's J-statistic** (from GMM theory, Module 4), computed using a consistent meat matrix $\hat{S} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}^2_{i,2SLS} z_i z_i'$:

$$J = n\left(\frac{1}{n} Z' \hat{u}_{2SLS}\right)' \hat{S}^{-1} \left(\frac{1}{n} Z' \hat{u}_{2SLS}\right) \quad \overset{d}{} \quad \chi^2_{\ell - k}$$

Rejection of the null hypothesis indicates that at least one instrument is correlated with u_i (or the structural equation is misspecified). Failure to reject does not prove all instruments are exogenous—for instance, if all instruments share the exact same invalid correlation direction, the test lacks power.

Durbin-Wu-Hausman (DWH) Test for Endogeneity

Because OLS is asymptotically more efficient than 2SLS when all regressors are exogenous, we test $H_0 : E(x_i u_i) = 0$ (exogeneity) against $H_1 : E(x_i u_i) \neq 0$ (endogeneity) using two algebraically equivalent approaches:

1. **Hausman Quadratic Form Test:** Under H_0, OLS is consistent and efficient while 2SLS is consistent but inefficient, implying $\mathrm{Cov}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}, \hat{\beta}_{OLS}) = 0$ asymptotically and therefore $\mathrm{Avar}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) = \mathrm{Avar}(\hat{\beta}_{2SLS}) - \mathrm{Avar}(\hat{\beta}_{OLS})$. The Hausman statistic compares the difference vector:

   $$H = (\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})' \left[\underline{\mathrm{Var}}(\hat{\beta}_{2SLS}) - \underline{\mathrm{Var}}(\hat{\beta}_{OLS})\right]^{+} (\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) \quad \overset{d}{} \quad \chi^2_{k_2}$$

   Where k_2 is the number of suspected endogenous regressors and $^{+}$ denotes a generalised inverse (since exogenous regressors may leave the difference matrix singular). In practice, finite-sample differences between $\hat{\sigma}^2_{2SLS}$ and $\hat{\sigma}^2_{OLS}$ can make the covariance difference non-positive definite unless a common variance estimate is used.

DegreeLive

2. **Regression-Based Control Function Test (Always Positive Definite and Robust):** Obtain the first-stage residuals $\hat{V}_2 = X_2 - Z\hat{\Pi}_2$ from regressing the k_2 suspect endogenous variables on $Z = [X_1 \;\; Z_2]$. Next, augment the structural equation by OLS:

$$y = X_1\beta_1 + X_2\beta_2 + \hat{V}_2\delta + \text{error}$$

By the Frisch-Waugh-Lovell theorem, the OLS estimates of (β_1, β_2) in this augmented regression equal $\hat{\beta}_{2SLS}$ identically! Testing $H_0 : \delta = 0$ with a standard F-test (or heteroskedasticity-robust Wald test) directly tests whether X_2 is endogenous without any risk of negative variance differences.

Worked Example 2: Matrix 2SLS, Sargan Overidentification, and DWH Control Function Test

Consider a demeaned structural equation $y = x\beta + u$ with $n = 5$ observations, one endogenous regressor x, and $\ell = 2$ excluded instruments $Z = [z_1 \;\; z_2]$:

$$Z = \begin{pmatrix} -2 & 1 \\ -1 & -1 \\ 0 & 0 \\ 1 & -1 \\ 2 & 1 \end{pmatrix}, \qquad x = \begin{pmatrix} -0.5 \\ -1.5 \\ -2.0 \\ 0.5 \\ 3.5 \end{pmatrix}, \qquad y = \begin{pmatrix} -0.4 \\ -2.8 \\ -5.6 \\ 1.2 \\ 7.6 \end{pmatrix}$$

Step 1: Compute OLS and First-Stage Projection.

Here $x'x = 19.0$ and $x'y = 42.8$, so $\hat{\beta}_{OLS} = 42.8/19.0 = 2.2526$. Next, $Z'Z = \text{diag}(10, 4)$ (orthogonal instruments) and $Z'x = (10, 4)'$, yielding first-stage slopes $\hat{\pi} = (Z'Z)^{-1}Z'x = (1.0, 1.0)'$. The first-stage fitted vector is $\hat{x} = z_1 + z_2 = (-1, -2, 0, 0, 3)'$, with $\hat{x}'\hat{x} = 14$ and first-stage residual $\hat{v} = x - \hat{x} = (0.5, 0.5, -2.0, 0.5, 0.5)'$.

Step 2: Compute 2SLS Estimate and Residuals.

Computing $Z'y = (20.0, 8.8)'$ and $\hat{x}'y = \hat{\pi}'(Z'y) = 1.0(20.0) + 1.0(8.8) = 28.8$ gives:

$$\hat{\beta}_{2SLS} = \frac{\hat{x}'y}{\hat{x}'\hat{x}} = \frac{28.8}{14} \approx 2.0571$$

The 2SLS residuals $\hat{u}_{2SLS} = y - 2.0571x = (0.6286, 0.2857, -1.4857, 0.1714, 0.4000)'$ have $\hat{u}'_{2SLS}\hat{u}_{2SLS} = 2.8735$, so $\hat{\sigma}^2 = 2.8735/5 = 0.5747$.

Step 3: Perform the Sargan Overidentification Test ($df = \ell - k = 1$).

Computing $Z'\hat{u}_{2SLS} = Z'y - \hat{\beta}_{2SLS}Z'x = (-0.5714, 0.5714)'$ gives:

DegreeLive

$$S = \frac{(Z'\hat{u}_{2SLS})'(Z'Z)^{-1}(Z'\hat{u}_{2SLS})}{\hat{\sigma}^2} = \frac{(-0.5714)^2(0.10) + (0.5714)^2(0.25)}{0.5747} = \frac{0.1143}{0.5747} \approx 0.199$$

Comparing $S = 0.199$ against the 5% χ^2_1 critical value 3.84, we do not reject the overidentifying restriction.

Step 4: Perform the DWH Control Function Test for Endogeneity.

Regressing y on x and $\hat{v}$ by OLS yields $\hat{\beta}_{2SLS} = 2.0571$ on x and $\hat{\delta} = \frac{\hat{v}'y}{\hat{v}'\hat{v}} - \hat{\beta}_{2SLS} = \frac{14.0}{5.0} - 2.0571 = 0.7429$ on $\hat{v}$, with residual sum of squares $SSR_{CF} = 0.1143$ and $s^2_{CF} = 0.1143/(5-2) = 0.0381$. Since $x'\hat{v} = 0$, $\text{se}(\hat{\delta}) = \sqrt{0.0381(1/14 + 1/5)} = 0.1017$, yielding $t_{\hat{\delta}} = 0.7429/0.1017 = 7.31$. We strongly reject $H_0 : \delta = 0$, confirming that x is endogenous and OLS (2.2526) is biased upward.

3.6 Simultaneous Equations Models (SEMs) and System Estimation

Simultaneous Equations Models describe markets or macroeconomic systems where G endogenous variables $y_i = (y_{i1}, \ldots, y_{iG})'$ are determined jointly by G structural equations alongside K exogenous (predetermined) variables $x_i = (x_{i1}, \ldots, x_{iK})'$:

$$\Gamma' y_i + B' x_i = u_i \qquad \text{or in stacked } n\text{-observation form:} \qquad Y\Gamma + XB = U$$

Where:

- Y is the $n \times G$ matrix of endogenous variables and Γ is a $G \times G$ non-singular matrix of structural coefficients linking endogenous variables (with diagonal entries normalised to 1 or -1).
- X is the $n \times K$ matrix of exogenous variables ($E(X'U) = 0$) and B is a $K \times G$ matrix of structural coefficients on exogenous variables.
- U is the $n \times G$ matrix of structural disturbances with cross-equation covariance $E(u_i u_i') = \Sigma$.

Structural Form vs. Reduced Form and Identification Conditions

Post-multiplying $Y\Gamma + XB = U$ by Γ^{-1} solves for the endogenous variables solely as functions of exogenous variables and disturbances, yielding the **reduced form**:

$$Y = X\Pi + V, \qquad \Pi = -B\Gamma^{-1}, \qquad V = U\Gamma^{-1}$$

Where Π is the $K \times G$ matrix of reduced-form parameters (always consistently estimable by OLS) and V is the $n \times G$ matrix of reduced-form errors. The **identification problem** asks whether the structural parameters (Γ, B) of a specific equation j can be recovered uniquely from the reduced-form matrix Π.

Suppose equation j normalises its own dependent variable coefficient to 1 and imposes linear exclusion restrictions by omitting G_j^* endogenous variables and K_j^* exogenous variables (leaving $G_j = G - G_j^*$ included endogenous variables, i.e., $G_j - 1$ right-hand-side endogenous variables, and $K_j = K - K_j^*$

DegreeLive

included exogenous variables):

- **Order Condition (Necessary):** The number of excluded exogenous variables K_j^* must be at least as large as the number of included right-hand-side endogenous variables $(G_j - 1)$:

  $$K_j^* \geq G_j - 1$$

  Equivalently, the total number of excluded variables from equation j must be at least $G - 1$. If $K_j^* < G_j - 1$, equation j is underidentified; if $K_j^* = G_j - 1$, it is potentially just-identified; if $K_j^* > G_j - 1$, it is potentially overidentified.

- **Rank Condition (Necessary and Sufficient):** Let Φ_j be the submatrix formed from the columns of the full structural matrix $A = [\Gamma' \quad B']'$ corresponding to all variables excluded from equation j, keeping only the rows of the other $G - 1$ equations. Equation j is identified if and only if $\text{rank}(\Phi_j) = G - 1$. Equivalently, the matrix of reduced-form coefficients linking the included right-hand-side endogenous variables to the excluded exogenous variables K_j^* must have full column rank $G_j - 1$.

Single-Equation (2SLS) vs. System Estimation (3SLS)

Once equation j is identified, it can be estimated equation-by-equation using 2SLS, treating all K exogenous variables X in the system as the instrument set $Z = X$. Single-equation 2SLS is consistent even if other equations in the system are misspecified, but it ignores cross-equation error correlations $\Sigma_{jm} = \text{Cov}(u_{ij}, u_{im}) \neq 0$.

When all G equations are correctly specified and Σ is non-diagonal, **Three-Stage Least Squares (3SLS)** combines 2SLS with Seemingly Unrelated Regressions (GLS):

1. **Stage 1:** Estimate the reduced form $Y = X\Pi + V$ and obtain $\hat{Y} = P_X Y$.
2. **Stage 2:** Estimate each structural equation separately by 2SLS using $\hat{Y}$ to obtain consistent residual matrix $\hat{U}_{2SLS}$ and estimate $\hat{\Sigma} = \frac{1}{n}\hat{U}_{2SLS}'\hat{U}_{2SLS}$.
3. **Stage 3:** Stack all G structural equations into a single system and apply GLS using the system covariance matrix $\hat{\Sigma} \otimes I_n$ and system instruments $I_G \otimes X$:

   $$\hat{\delta}_{3SLS} = [\hat{W}'(\hat{\Sigma}^{-1} \otimes P_X)\hat{W}]^{-1}\hat{W}'(\hat{\Sigma}^{-1} \otimes P_X)y_{vec}$$

   Where $\hat{W} = \text{diag}(W_1, \ldots, W_G)$ stacks each equation's regressors $[Y_j \quad X_j]$ and $\otimes$ denotes the Kronecker product. Under homoskedastic errors, 3SLS achieves asymptotic efficiency superior to 2SLS whenever equations are overidentified and Σ is non-diagonal.

Worked Example 3: Identification and Structural Recovery in a Supply-Demand System

Consider a two-equation equilibrium model of the Australian regional grain market across n regions, with quantity q_i and price p_i jointly endogenous ($G = 2$), and regional disposable income y_i and growing-season rainfall r_i strictly exogenous ($K = 3$ including the intercept):

Degree Live

$$\text{Demand: } q_i = \alpha_0 + \alpha_1 p_i + \alpha_2 y_i + u_{i1}, \qquad \text{Supply: } q_i = \gamma_0 + \gamma_1 p_i + \gamma_2 r_i + u_{i2}$$

Where economic theory predicts $\alpha_1 < 0$ (downward-sloping demand), $\alpha_2 > 0$, $\gamma_1 > 0$ (upward-sloping supply), and $\gamma_2 > 0$.

Step 1: Verify Order and Rank Conditions.

The demand equation omits one exogenous variable (r_i, so $K_1^* = 1$) and includes one right-hand-side endogenous variable (p_i, so $G_1 - 1 = 1$); thus the order condition holds with equality (just-identified). The rank condition holds provided $\gamma_2 \neq 0$ (rainfall genuinely shifts supply). Symmetrically, the supply equation omits y_i ($K_2^* = 1 = G_2 - 1$) and satisfies the rank condition provided $\alpha_2 \neq 0$.

Step 2: Derive Reduced-Form Equations and Recover Structural Parameters.

Equating demand and supply and solving for p_i and q_i yields the reduced form $p_i = \pi_{p,0} + \pi_{p,y} y_i + \pi_{p,r} r_i + v_{ip}$ and $q_i = \pi_{q,0} + \pi_{q,y} y_i + \pi_{q,r} r_i + v_{iq}$, where the reduced-form slopes satisfy:

$$\pi_{p,y} = \frac{\alpha_2}{\gamma_1 - \alpha_1}, \qquad \pi_{p,r} = \frac{-\gamma_2}{\gamma_1 - \alpha_1}, \qquad \pi_{q,y} = \frac{\gamma_1 \alpha_2}{\gamma_1 - \alpha_1}, \qquad \pi_{q,r} = \frac{-\alpha_1 \gamma_2}{\gamma_1 - \alpha_1}$$

Suppose OLS estimation of the reduced form yields $\hat{\pi}_{p,y} = 0.40$, $\hat{\pi}_{p,r} = -0.50$, $\hat{\pi}_{q,y} = 0.60$, and $\hat{\pi}_{q,r} = 0.25$. By Indirect Least Squares (identical to 2SLS in just-identified systems), the structural slopes are uniquely recovered as:

$$\hat{\alpha}_1 = \frac{\hat{\pi}_{q,r}}{\hat{\pi}_{p,r}} = \frac{0.25}{-0.50} = -0.50, \qquad \hat{\gamma}_1 = \frac{\hat{\pi}_{q,y}}{\hat{\pi}_{p,y}} = \frac{0.60}{0.40} = 1.50$$

Substituting $\hat{\gamma}_1 - \hat{\alpha}_1 = 1.50 - (-0.50) = 2.00$ gives $\hat{\alpha}_2 = \hat{\pi}_{p,y}(\hat{\gamma}_1 - \hat{\alpha}_1) = 0.40(2.00) = 0.80$ and $\hat{\gamma}_2 = -\hat{\pi}_{p,r}(\hat{\gamma}_1 - \hat{\alpha}_1) = 0.50(2.00) = 1.00$. Intuitively, rainfall shocks trace out the demand curve while income shocks trace out the supply curve.

3.7 Comparative Summary of IV and System Estimators & Panel Extensions

In longitudinal datasets such as the HILDA Survey or Australian corporate panels, unobserved time-invariant heterogeneity α_i induces endogeneity when correlated with regressors. First-differencing or fixed-effects transformation removes α_i, yet dynamic specifications $y_{it} = \rho y_{i,t-1} + x_{it}'\beta + \alpha_i + \varepsilon_{it}$ still suffer Nickell bias because $\Delta y_{i,t-1}$ correlates with $\Delta \varepsilon_{it}$. The Arellano-Bond GMM/IV framework resolves this by using deeper lagged levels $y_{i,t-2}, y_{i,t-3}, \ldots$ as valid instruments for the first-differenced equation whenever ε_{it} is serially uncorrelated.

DegreeLive

Estimator	Consistency Condition	Efficiency & Finite-Sample Behaviour	Diagnostic Requirement
OLS	$E(x_i u_i) = 0$ (strict/weak exogeneity)	BLUE under A1–A4; lowest variance when exogenous	DWH test to verify exogeneity
2SLS	$E(z_i u_i) = 0$ and $\text{rank}(Q_{zx}) = k$	Biased toward OLS when instruments are weak or ℓ is large	First-stage $F > 10$; Sargan / Hansen J-test
LIML	Same as 2SLS (normal-likelihood ratio)	Approximately median-unbiased with many weak instruments	Anderson-Rubin robust confidence sets
3SLS	All G structural equations correctly specified	More efficient than 2SLS when Σ is non-diagonal; sensitive to misspecification in any equation	System overidentification & Hausman 2SLS vs. 3SLS check

DegreeLive