Econometrics II

Lecture 2 - Instrumental Variables (Part II)

Lasha Chochua

2025

ISET

Notation Reference: Scalars

Symbol	Type	Definition
Y_1	scalar	Outcome variable (e.g. log wage)
e	scalar	Structural error term
u_1	scalar	Reduced form error for Y_1; equals $u_2'\beta_2 + e$
k_1	scalar	Number of exogenous regressors
k_2	scalar	Number of endogenous regressors
k	scalar	Total regressors; $k = k_1 + k_2$
ℓ	scalar	Total number of instruments
ℓ_2	scalar	Number of excluded instruments; $\ell_2 = \ell - k_1$
n	scalar	Sample size

Notation Reference: Vectors

Symbol	Dimension	Definition
X	$k \times 1$	Full regressor vector; $X = (X_1', X_2')'$
X_1	$k_1 \times 1$	Exogenous regressors
$X_2 = Y_2$	$k_2 \times 1$	Endogenous regressors
Z	$\ell \times 1$	Full instrument vector; $Z = (Z_1', Z_2')'$
Z_1	$k_1 \times 1$	Included exogenous instruments
Z_2	$\ell_2 \times 1$	Excluded instruments (the "true" IVs)
β	$k \times 1$	Structural parameter vector; $\beta = (\beta_1', \beta_2')'$
λ	$\ell \times 1$	Reduced form coefficients for Y_1 on Z
λ_1	$k_1 \times 1$	Reduced form coefficients on Z_1; $\lambda_1 = \beta_1 + \Gamma_{12}\beta_2$
λ_2	$\ell_2 \times 1$	Reduced form coefficients on Z_2; $\lambda_2 = \Gamma_{22}\beta_2$
u_2	$k_2 \times 1$	Reduced form error for Y_2
$\vec{Y}$	$(1 + k_2) \times 1$	Stacked endogenous variables; $\vec{Y} = (Y_1, Y_2')'$

Notation Reference: Matrices

Symbol	Dimension	Definition
Γ	$\ell \times k_2$	Reduced form coefficients for Y_2 on Z
Γ_{12}	$k_1 \times k_2$	Coefficients on Z_1 in reduced form for Y_2
Γ_{22}	$\ell_2 \times k_2$	Coefficients on Z_2 in reduced form for Y_2; rank$(\Gamma_{22}) = k_2$ for identification
$\bar{\Gamma}$	$\ell \times k$	Stacked matrix; $\bar{\Gamma} = \begin{bmatrix} I_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix}$
$\mathbf{P}_Z$	$n \times n$	Projection matrix; $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$
$\mathbf{Z}'\mathbf{Z}$	$\ell \times \ell$	Sample moment matrix of instruments
$\mathbf{Z}'\mathbf{X}$	$\ell \times k$	Sample cross-moment of instruments and regressors
$\mathbf{Z}'\mathbf{Y}_1$	$\ell \times 1$	Sample cross-moment of instruments and outcome

Consistency of 2SLS

- We now demonstrate the consistency of the 2SLS estimator for the structural parameter.
- The following is a set of regularity conditions:

Assumption 1

1. The variables (Y_{1i}, X_i, Z_i), $i = 1, \ldots, n$, are i.i.d.
2. $\mathbb{E}[Y_1^2] < \infty$.
3. $\mathbb{E}\|X\|^2 < \infty$.
4. $\mathbb{E}\|Z\|^2 < \infty$.
5. $\mathbb{E}[ZZ']$ is positive definite.
6. $\mathbb{E}[ZX']$ has full rank k.
7. $\mathbb{E}[Ze] = 0$.

Interpreting Assumptions

- Assumptions 1.2–1.4: all variables have finite variances
- Assumption 1.5: Z has an invertible design matrix (non-redundant instruments)
- Assumption 1.6: $\mathbb{E}[ZX']$ has full rank k (relevance condition)
- Assumption 1.7: $\mathbb{E}[Ze] = 0$ (instrument exogeneity)

These conditions mirror those for OLS but are applied to instruments.

Theorem: Consistency of 2SLS

Theorem 1 (Consistency of 2SLS)

Under Assumption 1,

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta \quad \text{as } n \to \infty.$$

Intuition and Proof Strategy

- 2SLS is consistent under the same general conditions as OLS
- The key differences:
  - Instruments replace regressors in the moment conditions
  - Require relevance ($\text{rank}(\mathbb{E}[ZX']) = k$)
  - Require exogeneity ($\mathbb{E}[Ze] = 0$)
- We begin with the structural model:

$$\mathbf{Y} = \mathbf{X}\beta + e$$

Consistency of 2SLS

- We plug into the 2SLS formula:

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

- Substitute $\mathbf{Y} = \mathbf{X}\beta + e$:

$$\hat{\beta}_{2SLS} = \beta + (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'e$$

Applying LLN and CMT

- Separate out the stochastic terms:

$$\hat{\beta}_{2SLS} - \beta = \left(\left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{X}\right)\right)^{-1}\left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{n}\mathbf{Z}'e\right)$$

- Under the assumptions:

$$\xrightarrow{p} (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\underbrace{\mathbb{E}[Ze]}_{=0} = 0$$

Definitions of Moment Matrices

- Define the following population moments:

$$\mathbf{Q}_{XZ} = \mathbb{E}[\mathbf{X}\mathbf{Z}']$$
$$\mathbf{Q}_{ZZ} = \mathbb{E}[\mathbf{Z}\mathbf{Z}']$$
$$\mathbf{Q}_{ZX} = \mathbb{E}[\mathbf{Z}\mathbf{X}']$$

- Law of Large Numbers applies under Assumptions 1.1, 1.2–1.4
- Continuous Mapping Theorem justifies convergence
- Invertibility from Assumptions 1.5 and 1.6
- Final step uses $\mathbb{E}[Ze] = 0$

Asymptotic Distribution of 2SLS

- We now show that the 2SLS estimator satisfies a central limit theorem.
- We first state a set of sufficient regularity conditions:

Assumption 2 *(in addition to Assumption 1)*

1. $\mathbb{E}[Y_1^4] < \infty$
2. $\mathbb{E}\|X\|^4 < \infty$
3. $\mathbb{E}\|Z\|^4 < \infty$
4. $\Omega = \mathbb{E}[ZZ'e^2]$ is positive definite

Theorem: Asymptotic Normality of 2SLS

Theorem 2 (Asymptotic Distribution of 2SLS)
Under Assumption 2, as $n \to \infty$,

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

where

$$V_\beta = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\Omega\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}$$

Interpretation and Intuition

- 2SLS converges at $\sqrt{n}$ rate to a normal vector
- The asymptotic variance V_β is more complex than in OLS
- In special cases with homoskedasticity ($\mathbb{E}[e^2 \mid Z] = \sigma^2$), we get:

$$\Omega = \mathbf{Q}_{ZZ}\sigma^2, \quad V_\beta = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\sigma^2$$

Proof Outline

- We start from the consistency expression and multiply by $\sqrt{n}$:

$$\sqrt{n}(\widehat{\beta}_{2SLS} - \beta) = \left(\left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{X}\right)\right)^{-1}\left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}\right)$$

- Apply the CLT for i.i.d. vectors:

$$\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i e_i \xrightarrow{d} \mathcal{N}(0, \Omega)$$

- In the end

$$\sqrt{n}(\widehat{\beta}_{2SLS} - \beta) = \left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\tfrac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\tfrac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}\right)$$
$$\xrightarrow{d} (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathcal{N}(0, \Omega) = \mathcal{N}(0, \mathbf{V}_{\beta})$$

Finite Second Moment of Ze

- To complete the proof we demonstrate that Ze has a finite second moment under Assumption 2
- By Minkowski's inequality:

$$\left(\mathbb{E}\left[e^4\right]\right)^{1/4} = \left(\mathbb{E}\left[\left(Y_1 - X'\beta\right)^4\right]\right)^{1/4} \leq \left(\mathbb{E}\left[Y_1^4\right]\right)^{1/4} + \|\beta\| \left(\mathbb{E}\|X\|^4\right)^{1/4} < \infty$$

  under Assumptions 2.1 and 2.2
- Then by the Cauchy–Schwarz inequality:

$$\mathbb{E}\|Ze\|^2 \leq \left(\mathbb{E}\|Z\|^4\right)^{1/2} \left(\mathbb{E}\left[e^4\right]\right)^{1/2} < \infty$$

  using Assumption 2.3

Determinants of 2SLS Variance

- It is instructive to examine the asymptotic variance of the 2SLS estimator to understand the factors which determine its precision. Under homoskedasticity, the asymptotic variance simplifies to:

$$\mathbf{V}_\beta^0 = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\sigma^2 = \left(\mathbb{E}[XZ'](\mathbb{E}[ZZ'])^{-1}\mathbb{E}[ZX']\right)^{-1}\mathbb{E}[e^2]$$

- Variance increases with $\mathbb{E}[e^2]$
- Variance decreases with correlation between X and Z
- The structure of Z does not affect the variance — $\mathbf{V}_\beta^0$ is invariant to linear transformations $Z \mapsto \mathbf{C}Z$.

Effect of Adding Instruments

- Suppose we partition $Z = (Z_a, Z_b)$ where $\dim(Z_a) \geq k$.
- Let $\hat{\beta}_a$ and $\hat{\beta}$ denote the 2SLS estimators using instruments Z_a and (Z_a, Z_b) respectively.
- Assume $Z_a \perp Z_b$ (or pre-project Z_b orthogonal to Z_a). Then:

$$\text{avar}[\hat{\beta}] = \left(\mathbb{E}[XZ'](\mathbb{E}[ZZ'])^{-1}\mathbb{E}[ZX']\right)^{-1}\sigma^2$$

$$= \left(\mathbb{E}[XZ_a'](\mathbb{E}[Z_aZ_a'])^{-1}\mathbb{E}[Z_aX'] + \mathbb{E}[XZ_b'](\mathbb{E}[Z_bZ_b'])^{-1}\mathbb{E}[Z_bX']\right)^{-1}\sigma^2$$

$$\leq \left(\mathbb{E}[XZ_a'](\mathbb{E}[Z_aZ_a'])^{-1}\mathbb{E}[Z_aX']\right)^{-1}\sigma^2 = \text{avar}[\hat{\beta}_a]$$

Bias–Variance Trade-Off

- Adding instruments $\Rightarrow$ asymptotic variance of $\hat{\beta}_{2SLS}$ decreases
- However, finite-sample bias generally **increases** with more instruments
- This introduces a **bias–variance trade-off** in instrument selection

Hence, while asymptotic theory supports using all valid instruments, in practice:

- Too many instruments $\Rightarrow$ potential finite-sample distortion
- Instrument choice should balance **precision** and **bias**

Covariance Matrix Estimation

- Estimation of the asymptotic covariance matrix V_β is done by replacing population moments with sample analogues:

$$\hat{\mathbf{V}}_\beta = (\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1}(\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\Omega}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})(\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1}$$

- where

$$\hat{\mathbf{Q}}_{ZZ} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' = \frac{1}{n}\mathbf{Z}'\mathbf{Z}, \quad \hat{\mathbf{Q}}_{XZ} = \frac{1}{n}\sum_{i=1}^{n} X_i Z_i' = \frac{1}{n}\mathbf{X}'\mathbf{Z}$$

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{e}_i^2, \quad \hat{e}_i = Y_i - X_i'\hat{\beta}_{2SLS}$$

Warning: Correct Residuals for 2SLS Standard Errors

- The covariance matrix **must** be constructed using the correct 2SLS residual:

$$\hat{e}_i = Y_i - X_i'\hat{\beta}_{\text{2sls}}$$

- This is **different** from the residual obtained by naively running the second stage OLS
- In the two-stage computation:
  - **Stage 1:** Regress X on Z $\to$ obtain $\hat{X}_i = \hat{\Gamma}' Z_i$
  - **Stage 2:** Regress Y on $\hat{X}$ $\to$ obtain $\hat{\beta}_{\text{2sls}}$, with residual:

$$Y_i = \hat{X}_i'\hat{\beta}_{\text{2sls}} + \hat{v}_i \tag{12.41}$$

Warning: Correct Residuals for 2SLS Standard Errors (Cont.)

- The second-stage residual $\hat{v}_i$ satisfies:

$$\hat{v}_i = Y_i - X_i'\hat{\beta}_{\text{2sls}} + (X_i - \hat{X}_i)'\hat{\beta}_{\text{2sls}} = \hat{e}_i + \hat{u}_i'\hat{\beta}_{\text{2sls}} \neq \hat{e}_i$$

- So the standard errors reported by the second-stage regression use $\hat{\sigma}_v^2$ instead of $\hat{\sigma}^2$:

$$\hat{V}_\beta = \left(\frac{1}{n}\hat{X}'\hat{X}\right)^{-1}\hat{\sigma}_v^2 = \left(\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}\right)^{-1}\hat{\sigma}_v^2, \quad \hat{\sigma}_v^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i^2$$

- Since $\hat{v}_i \neq \hat{e}_i$, these standard errors are **incorrect**
- **Solution**: construct $\hat{\beta}_{\text{2sls}}$ directly and compute standard errors using $\hat{e}_i = Y_i - X_i'\hat{\beta}_{\text{2sls}}$

Homoskedastic Case and Theorem

- Under homoskedasticity:

$$\hat{\mathbf{V}}_\beta^0 = (\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1}\hat{\sigma}^2, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2$$

Theorem 3 (Consistency of the Covariance Estimator)
Under Assumption 2, as $n \to \infty$,

$$\hat{V}_\beta^0 \xrightarrow{p} V_\beta^0 \quad \text{and} \quad \hat{V}_\beta \xrightarrow{p} V_\beta$$

Is Homoskedasticity in 2SLS realistic?

- As in OLS, the asymptotic variance simplifies under **conditional homoskedasticity**:

$$\mathbb{E}[e^2 \mid Z] = \sigma^2$$

- This holds when Z and e are **independent**
- It is reasonable to assume $e \perp Z_2$ (excluded instruments):
  - By the exclusion restriction, Z_2 affects Y **only through** X
  - So Z_2 carries no direct information about e
- But there is **no reason** to expect $e \perp X_1$ (included exogenous variables):
  - X_1 enters the structural equation directly
  - Heteroskedasticity with respect to X_1 is entirely plausible
- Hence **heteroskedasticity should be equally expected** in 2SLS and OLS

Functions of Parameters

- Given the distribution theory in Theorems 2 and 3, we can derive the asymptotic distribution of smooth nonlinear functions of the coefficient estimators.
- Given a function $r(\beta) : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q$, define the parameter $\theta = r(\beta)$.
- Given $\hat{\beta}_{2SLS}$, a natural estimator of θ is $\hat{\theta}_{2SLS} = r(\hat{\beta}_{2SLS})$.

Theorem 4

Under Assumptions 1 and **7.3 in the book**, as $n \to \infty$,

$$\hat{\theta}_{2SLS} \xrightarrow{p} \theta.$$

Covariance Matrix for Transformed Parameters

- If $r(\beta)$ is differentiable, the asymptotic covariance matrix of $\hat{\theta}_{2SLS}$ is:

$$\hat{\mathbf{V}}_\theta = \hat{\mathbf{R}}\hat{\mathbf{V}}_\beta\hat{\mathbf{R}}', \quad \hat{\mathbf{R}} = \frac{\partial}{\partial\beta} r(\hat{\beta}_{2SLS})'$$

- We similarly define the homoskedastic version:

$$\hat{\mathbf{V}}_\theta^0 = \hat{\mathbf{R}}\hat{\mathbf{V}}_\beta^0\hat{\mathbf{R}}'$$

Asymptotic Distribution of Transformed Estimator

Theorem 5

Under Assumptions 2 and **7.3 in the book**, as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_{2SLS} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_\theta)$$

and

$$\hat{\mathbf{V}}_\theta \xrightarrow{p} \mathbf{V}_\theta, \quad \text{where} \quad \mathbf{V}_\theta = \mathbf{R}'\mathbf{V}_\beta\mathbf{R}, \quad \mathbf{R} = \frac{\partial}{\partial\beta} r(\beta)'.$$

Recall Card (1995)

	OLS	IV(a)	IV(b)	2SLS(a)	2SLS(b)	LIML
education	0.074	0.132	0.133	0.161	0.160	0.164
	(0.004)	(0.049)	(0.051)	(0.040)	(0.041)	(0.042)
experience	0.084	0.107	0.056	0.119	0.047	0.120
	(0.007)	(0.021)	(0.026)	(0.018)	(0.025)	(0.019)
$\text{experience}^2/100$	−0.224	−0.228	−0.080	−0.231	−0.032	−0.231
	(0.032)	(0.035)	(0.133)	(0.037)	(0.127)	(0.037)
Black	−0.190	−0.131	−0.103	−0.102	−0.064	−0.099
	(0.017)	(0.051)	(0.075)	(0.044)	(0.061)	(0.045)
south	−0.125	−0.105	−0.098	−0.095	−0.086	−0.094
	(0.015)	(0.023)	(0.0284)	(0.022)	(0.026)	(0.022)
urban	0.161	0.131	0.108	0.116	0.083	0.115
	(0.015)	(0.030)	(0.049)	(0.026)	(0.041)	(0.027)
Sargan				0.82	0.52	0.82
p-value				0.37	0.47	0.37

Example: Standard Error for Scalar Transformation

- When $q = 1$, a standard error for $\hat{\theta}_{2SLS}$ is given by:

$$s(\hat{\theta}_{2SLS}) = \sqrt{n^{-1}\hat{V}_{\theta}}$$

- Let's consider parameter estimates from the **fifth** column of Table from Card (1995) (2SLS with three endogenous regressors and four excluded instruments). Suppose we are interested in the return to experience, which depends on the level of experience.

Example: Return to Experience at 10 Years

- We consider a nonlinear function:

$$\theta = r(\beta) = \beta_2 + 2 \cdot 10 \cdot \frac{\beta_3}{100}$$

- This computes the marginal effect of experience when experience $= 10$.
- From the 2SLS(b) column:
  - $\hat{\beta}_2 = 0.047$ (experience)
  - $\hat{\beta}_3 = -0.032$ (experience^2/100)
- So:

$$\hat{\theta} = 0.047 + 2 \cdot 10 \cdot \frac{-0.032}{100} = 0.047 - 0.0064 = 0.0406 \approx 0.041$$

Delta Method: Standard Error for $\hat{\theta}$

- Let $r(\beta) = \beta_2 + 2c \cdot \beta_3$ with $c = \frac{10}{100} = 0.1$
- Then:

$$\mathbf{R} = \begin{bmatrix} 1 & 2c \end{bmatrix} = \begin{bmatrix} 1 & 0.2 \end{bmatrix}$$

- Assuming $\hat{\beta}_2$ and $\hat{\beta}_3$ are uncorrelated:
  - $\text{se}(\hat{\beta}_2) = 0.025$
  - $\text{se}(\hat{\beta}_3) = 0.127$
- Variance of $\hat{\theta}$:

$$\hat{V}_\theta = 1^2 \cdot (0.025)^2 + 0.2^2 \cdot (0.127)^2 = 0.000625 + 0.000645 \approx 0.00127$$

- Standard error:

$$s(\hat{\theta}) = \sqrt{0.00127} \approx 0.0356$$

Using the Full Covariance Matrix

- In practice, $\hat{\beta}_2$ and $\hat{\beta}_3$ are **negatively correlated**, which lowers the standard error.
- Reported standard error in the **book** is **0.003**
- So $\hat{V}_\beta$ must have a large negative covariance term:

$$\hat{V}_\theta = \mathbf{R}\hat{V}_\beta\mathbf{R}'$$

- Hence:

$$s(\hat{\theta}) = \sqrt{\hat{V}_\theta} = 0.003$$

- Accurate standard errors require the **full estimated covariance matrix** $\hat{V}_\beta$, not just the standard errors.

Hypothesis Tests

- For a differentiable function $r(\beta) : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q$, define $\theta = r(\beta)$.
- We test:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

- The **Wald statistic** is:

$$W = n(\hat{\theta} - \theta_0)' \hat{V}_\theta^{-1} (\hat{\theta} - \theta_0)$$

Asymptotic Wald Distribution

- From Theorem 5, W has a chi-squared limit distribution under H_0:

Theorem 6

Under Assumption 2, **Assumption 7.3** in the book, and H_0, as $n \to \infty$,

$$W \xrightarrow{d} \chi_q^2$$

Let $G_q(u)$ be the CDF of χ_q^2. For c satisfying $\alpha = 1 - G_q(c)$, the test:

$$\text{Reject } H_0 \text{ if } W > c$$

has asymptotic size α.

Application: Return to Experience

- Return to experience is governed by:
  - β_2 (experience)
  - β_3 (experience^2/100)
- From 2SLS(b), neither is individually significant at the 5% level:
  - se$(\hat{\beta}_2) = 0.025,\ p > 0.05$
  - se$(\hat{\beta}_3) = 0.127,\ p > 0.05$
- We test the joint hypothesis:

$$H_0 : \beta_2 = \beta_3 = 0$$

- The Wald statistic is:

$$W = 244$$

- This yields an **asymptotic p-value** **0.0000** strongly reject H_0.
- The joint test shows that experience does have a statistically significant effect, despite individual insignificance.

Control Function Regression

- We present an alternative way of computing the 2SLS estimator via least squares, often useful in nonlinear contexts or for **testing endogeneity**.
- Start with the structural and reduced form equations:

$$Y = X_1'\beta_1 + X_2'\beta_2 + e$$
$$X_2 = \Gamma_{12}'Z_1 + \Gamma_{22}'Z_2 + u_2$$

- Since X_2 is endogenous, u_2 and e are correlated. We project e on u_2:

$$e = u_2'\alpha + \nu, \quad \alpha = \left(\mathbb{E}[u_2u_2']\right)^{-1}\mathbb{E}[u_2e], \quad \mathbb{E}[u_2\nu] = 0$$

Substituting the Projection

- Substitute into the structural equation:

$$Y = X_1'\beta_1 + X_2'\beta_2 + u_2'\alpha + \nu \tag{*}$$

- This yields the moment conditions:
  - $\mathbb{E}[X_1\nu] = 0$
  - $\mathbb{E}[X_2\nu] = 0$
  - $\mathbb{E}[u_2\nu] = 0$
- So X_2 is uncorrelated with ν. The endogeneity of X_2 is accounted for via u_2.

Estimation Strategy

- If u_2 were observed, estimate $(*)$ by regressing Y on (X_1, X_2, u_2).
- Since u_2 is unobserved, estimate it from the reduced form:

$$\hat{u}_{2i} = X_{2i} - \hat{\Gamma}'_{12} Z_{1i} - \hat{\Gamma}'_{22} Z_{2i}$$

- Then run OLS of Y on $(X_1, X_2, \hat{u}_2)$ and get:

$$Y_i = X_i' \hat{\beta}_1 + \hat{u}'_{2i} \hat{\alpha} + \hat{\nu}_i \qquad (**)$$

Matrix Notation and Equivalence to 2SLS (Optional)

- In matrix form:

$$\mathbf{Y} = \mathbf{X}\hat{\beta} + \hat{\mathbf{U}}_2\hat{\alpha} + \hat{\nu}$$

- This gives an alternative expression for the 2SLS estimator.
- Reduced form residual can be written:

$$\hat{\mathbf{U}}_2 = (\mathbf{I}_n - \mathbf{P}_Z)\mathbf{X}_2$$

- Then, by the FWL (Frisch–Waugh–Lovell) theorem:

$$\hat{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y} \qquad (***)$$

- where:

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2], \quad \tilde{\mathbf{X}}_1 = \mathbf{X}_1 - \hat{\mathbf{U}}_2(\hat{\mathbf{U}}_2'\hat{\mathbf{U}}_2)^{-1}\hat{\mathbf{U}}_2'\mathbf{X}_1 = \mathbf{X}_1$$

Matrix Notation and Equivalence to 2SLS (Optional)

- From earlier, we had:

$$\tilde{\mathbf{X}}_2 = \mathbf{X}_2 - \hat{\mathbf{U}}_2(\hat{\mathbf{U}}_2'\hat{\mathbf{U}}_2)^{-1}\hat{\mathbf{U}}_2'\mathbf{X}_2 = \mathbf{P}_Z\mathbf{X}_2$$

- So:

$$\tilde{\mathbf{X}} = [\mathbf{X}_1, \mathbf{P}_Z\mathbf{X}_2] = \mathbf{P}_Z\mathbf{X}$$

- Substituting into the estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y} = \hat{\beta}_{2SLS}$$

- This confirms that OLS on the control function regression gives the 2SLS estimator.

Endogeneity Tests

- The 2SLS estimator allows X_2 to be endogenous, meaning it may be correlated with the structural error e.
- We can test this restriction. The null hypothesis is:

$$H_0 : \mathbb{E}[X_2 e] = 0$$

- with the alternative:

$$H_1 : \mathbb{E}[X_2 e] \neq 0$$

- Since the maintained assumption is $\mathbb{E}[Ze] = 0$ and $X_1 \subseteq Z$, we could equivalently write:

$$H_0 : \mathbb{E}[Xe] = 0$$

Endogeneity Tests (Cont.)

- Recall the control function regression (*):

$$Y = X_1'\beta_1 + X_2'\beta_2 + u_2'\alpha + \nu$$

$$\alpha = (\mathbb{E}[u_2 u_2'])^{-1}\mathbb{E}[u_2 e]$$

Testing Endogeneity via $\alpha = 0$

- We have:

$$\mathbb{E}[X_2 e] = 0 \iff \mathbb{E}[u_2 e] = 0 \iff \alpha = 0$$

- So, test:

$$H_0 : \alpha = 0 \quad \text{vs} \quad H_1 : \alpha \neq 0$$

- Use the Wald statistic W for $\alpha = 0$ in the control function regression.
- Under H_0:

$$W \xrightarrow{d} \chi^2_{k_2}$$

- Alternatively, under normality of error:

$$F = \frac{W^0}{k_2} \sim F(k_2, n - k_1 - 2k_2)$$

Estimation Strategy

1. Estimate the reduced form:

$$X_{2i} = \hat{\Gamma}'_{12} Z_{1i} + \hat{\Gamma}'_{22} Z_{2i} + \hat{u}_{2i}$$

2. Estimate the control function regression:

$$Y_i = X_i' \hat{\beta} + \hat{u}_{2i}' \hat{\alpha} + \hat{\nu}_i \qquad (****)$$

- Let W, W^0, and $F = W^0/k_2$ denote:
  - Robust Wald
  - Homoskedastic Wald
  - F statistic

Distribution Results

Theorem 7

Under H_0,

$$W \xrightarrow{d} \chi^2_{k_2}$$

Let $c_{1-\alpha}$ solve $\mathbb{P}[\chi^2_{k_2} \leq c_{1-\alpha}] = 1 - \alpha$. The test "Reject H_0 if $W > c_{1-\alpha}$" has asymptotic size α.

F Test under Normality

Theorem 8

Suppose $e \mid X, Z \sim \mathcal{N}(0, \sigma^2)$. Under H_0:

$$F \sim F(k_2, n - k_1 - 2k_2)$$

Let $c_{1-\alpha}$ solve $\mathbb{P}[F(k_2, n - k_1 - 2k_2) \leq c_{1-\alpha}] = 1 - \alpha$.

The test "Reject H_0 if $F > c_{1-\alpha}$" has exact size α.

Durbin-Wu-Hausman (DWH) Test for Endogeneity

- There is a classic alternative way to test for endogeneity:
- Under H_0, both OLS and 2SLS are consistent.
- Under H_1, only 2SLS is consistent.
- Thus, the **difference** between OLS and 2SLS is a valid test statistic:

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2) \quad \text{(OLS)}, \quad \tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2) \quad \text{(2SLS)}$$

Test Statistic and Interpretation

- Under homoskedasticity and H_0, OLS is efficient:

$$\mathrm{Var}[\hat{\beta}_2 - \tilde{\beta}_2] = \mathrm{Var}[\tilde{\beta}_2] - \mathrm{Var}[\hat{\beta}_2]$$

$$= \left((\mathbf{X}_2'(\mathbf{P}_Z - \mathbf{P}_1)\mathbf{X}_2)^{-1} - (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \right) \sigma^2$$

  Where $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{P}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{P}_1$

- The test statistic is:

$$T = \frac{(\hat{\beta}_2 - \tilde{\beta}_2)' \left[(\mathbf{X}_2'(\mathbf{P}_Z - \mathbf{P}_1)\mathbf{X}_2)^{-1} - (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \right]^{-1} (\hat{\beta}_2 - \tilde{\beta}_2)}{\hat{\sigma}^2}$$

- This T-statistic has an asymptotic $\chi^2_{k_2}$ distribution under H_0, or an F version under normality.

Equivalence Across Methods

- Durbin (1954), Wu (1973), and Hausman (1978) proposed equivalent tests:
- If $\hat{\sigma}^2$ is the variance from the regression in (****), then $T = W^0$
- These methods coincide with the control function Wald test
- In practice, they yield identical statistics up to how $\hat{\sigma}^2$ is estimated
- The class of tests includes:
  - **Durbin-Wu-Hausman tests**
  - **Wu-Hausman tests**
  - **Hausman tests**
  - **Robust DWH test** (Wooldridge (1995))

Overidentification Tests

- When $\ell > k$, the model is **overidentified**, meaning there are more moment conditions than free parameters.
  - This is a restriction, and it is testable.
  - The corresponding tests are called **overidentification tests**.
- The IV model assumes:

$$\mathbb{E}[\mathbf{Z}e] = 0$$

- Equivalently, since $e = Y - \mathbf{X}'\beta$:

$$\mathbb{E}[\mathbf{Z}Y] - \mathbb{E}[\mathbf{Z}\mathbf{X}']\beta = 0$$

- This is an $\ell \times 1$ system of moment restrictions. But since β is $k \times 1$ and $\ell > k$, it's not guaranteed that such a β exists — and we can test that!

A Simple Example

- Suppose:
  - One endogenous regressor X_2
  - No X_1
  - Two instruments: Z_1, Z_2
- Then valid instruments imply:

$$\mathbb{E}[Z_1 Y] = \mathbb{E}[Z_1 X_2]\beta$$
$$\mathbb{E}[Z_2 Y] = \mathbb{E}[Z_2 X_2]\beta$$

- Both equations must hold for the same β.
- This is a strong requirement — it's what the test is checking.
- If both instruments are valid, both IV estimates (using Z_1 or Z_2) target the same β.
- If one is invalid, the estimates diverge in large samples.

Interpretation and Practical Illustration

- Overidentification testing is equivalent to:
  - "Are all available instruments consistent with one another?"
- In terms of estimation, you can:
  - Use Z_1 alone to estimate β
  - Or use Z_2 alone
  - Or both (overidentified)
- If overidentification holds, all estimates should agree.
- For example, in Table showing Card (1995) results:
  - The sixth column uses both `public` and `private` as instruments for education.
  - If we instead use just `public`, we get a different estimate.
- If this difference is large, the overidentification test may reject the null that both instruments are valid.

Deriving the Sargan Overidentification Test

- To test:

$$H_0 : \mathbb{E}[\mathbf{Z}e] = 0 \quad \text{vs} \quad H_1 : \mathbb{E}[\mathbf{Z}e] \neq 0$$

- we assume homoskedasticity:

$$\mathbb{E}[e^2 \mid \mathbf{Z}] = \sigma^2$$

- To avoid homoskedasticity assumption, a GMM approach is best. But under homoskedasticity, we can proceed with the classical test.

Deriving the Sargan Overidentification Test (Cont.)

- To implement a test of H_0, consider the regression:

$$e = \mathbf{Z}'\alpha + \nu$$

with $\alpha = (\mathbb{E}[\mathbf{Z}\mathbf{Z}'])^{-1}\mathbb{E}[\mathbf{Z}e]$.

- We can rewrite H_0 as $\alpha = 0$. Since e is unobserved, we use:

$$\hat{e} = Y - \mathbf{X}\hat{\beta}_{2SLS}$$

- Then estimate:

$$\hat{\alpha} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}$$

Sargan's Score Test Statistic

- The test statistic is:

$$S = \hat{\alpha}' \left(\mathrm{Var}(\hat{\alpha})\right)^{-1} \hat{\alpha} = \frac{\hat{\mathbf{e}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}}{\hat{\sigma}^2}$$

  Where:

$$\hat{\sigma}^2 = \frac{1}{n}\hat{\mathbf{e}}'\hat{\mathbf{e}}$$

- Basmann (1960) proposed a version where:

$$\hat{\sigma}^2 = \frac{1}{n}\hat{v}'\hat{v}, \quad \text{with } \hat{v} = \hat{e} - \mathbf{Z}\hat{\alpha}$$

- Both tests are asymptotically equivalent under homoskedasticity.
- Sargan's version is more commonly used.

Sargan's Score Test

Theorem 9 Under Assumption 2 and $\mathbb{E}[e^2 \mid \mathbf{Z}] = \sigma^2$, then as $n \to \infty$,

$$S \xrightarrow{d} \chi^2_{\ell-k}.$$

For c satisfying $\alpha = 1 - G_{\ell-k}(c)$, $\mathbb{P}[S > c \mid H_0] \to \alpha$, so the test "Reject H_0 if $S > c$" has asymptotic size α.

Reading

- **Hansen (2022), Econometrics.**
  - Sections: 12.15-12.18, 12.20, 12.21, 12.28, 12.29, 12.31