Econometrics and Applications

Kirby CHEN

Academic Year 2024-2025

Contents

1 Lecture 3: Endogeneity and Instrumental Variables

1.1 Motivation and Overlook

Example:

- Omitted variables bias
- Measurement error
- Simultaneous equations bias (reverse causality)

Our Goal

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

The endogenous variable x has a real impact on Y, and we aim to find the true value of β_1.

1. **Using an Instrumental Variable to Derive the Model's Covariance**

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Taking the covariance of both sides with the instrumental variable z:

$$\text{cov}(Y, z) = \text{cov}(\beta_0 + \beta_1 X + \varepsilon, z)$$

Expanding the covariance expression:

$$\text{cov}(Y, z) = \text{cov}(\beta_0, z) + \beta_1 \times \text{cov}(X, z) + \text{cov}(\varepsilon, z)$$

Since the instrumental variable z is uncorrelated with both β_0 and the error term ε, these covariance terms disappear:

$$\text{cov}(Y, z) = \beta_1 \times \text{cov}(X, z)$$

Solving for β_1:

$$\beta_1 = \frac{\text{cov}(Y, z)}{\text{cov}(X, z)}$$

Instrumental Variables (IV) estimator of β_1, β_{IV}.

2. **Reduced-form Equation: Indirect Least Square, ILS**

$$x = \delta_0 + \delta_1 \times z + u$$

$$Y = \pi_0 + \pi_1 \times z + v$$

Reduced-form equation: Writing an endogenous variable in terms of exogenous variables.

$$x = \delta_0 + \delta_1 \times z + u$$
$$Y = \pi_0 + \pi_1 \times z + v$$

$$\delta_1 = \frac{\text{cov}(x, z)}{\text{var}(z)}$$
$$\pi_1 = \frac{\text{cov}(Y, z)}{\text{var}(z)}$$

We know:

$$Y = \beta_0 + \beta_1 \times x + \varepsilon$$

Regression coefficient:

$$\beta_1 = \frac{\text{cov}(Y, x)}{\text{var}(x)}$$

Using the instrumental variable:

$$\frac{\pi_1}{\delta_1} = \frac{\frac{\text{cov}(Y,z)}{\text{var}(z)}}{\frac{\text{cov}(x,z)}{\text{var}(z)}} = \frac{\text{cov}(Y,z)}{\text{cov}(x,z)} = \beta_{IV} = \beta_1$$

$$x = \delta_0 + \delta_1 \times z + u$$

$$Y = \pi_0 + \pi_1 \times z + v$$

$$\delta_1 = \frac{\text{cov}(x,z)}{\text{var}(z)}$$

$$\pi_1 = \frac{\text{cov}(Y,z)}{\text{var}(z)}$$

*Reduced-form Equation

$$x = \delta_0 + \delta_1 \times z + u$$

$$Y = \pi_0 + \pi_1 \times z + v$$

$$\delta_1 = \frac{\text{cov}(x,z)}{\text{var}(z)}$$

$$\pi_1 = \frac{\text{cov}(Y,z)}{\text{var}(z)}$$

*Indirect Least Squares (ILS) Method

$$Y = \beta_0 + \beta_1 \times x + \varepsilon$$

$$= \beta_0 + \beta_1 \times (\delta_0 + \delta_1 \times z + u) + \varepsilon$$

$$= \beta_0 + \beta_1 \times \delta_0 + \beta_1 \times \delta_1 \times z + \beta_1 \times u + \varepsilon$$

$$= (\beta_0 + \beta_1 \times \delta_0) + \beta_1 \times \delta_1 \times z + (\beta_1 \times u + \varepsilon)$$

$$\pi_0 = \beta_0 + \beta_1 \times \delta_0, \quad \pi_1 = \beta_1 \times \delta_1, \quad v = \beta_1 \times u + \varepsilon$$

Question: when IVs more than endogenous variables, the above two method fails.

3. **Two Stage Least Squares (2SLS/TSLS)**

   *First Stage

$$x = \delta_0 + \delta_1 \times z + u$$
$$x = \hat{\delta_0} + \hat{\delta_1} \times z + \hat{u}$$
$$\hat{x} = \delta_0 + \delta_1 \times z$$

   *Second Stage

$$Y = \beta_{0,2SLS} + \beta_{1,2SLS} \times \hat{x} + \varepsilon_{2SLS}$$

   *Does the Model Have Endogeneity?

$$Y = \beta_0 + \beta_1 \times x + \varepsilon$$
$$= \beta_0 + \beta_1 \times (\hat{x} + \hat{u}) + \varepsilon$$
$$= \beta_0 + \beta_1 \times \hat{x} + \beta_1 \times \hat{u} + \varepsilon$$

$$\text{cov}(\hat{x}, \varepsilon_{2SLS}) = \text{cov}(\hat{x}, \beta_1 \times \hat{u} + \varepsilon)$$

$$= \beta_1 \times \text{cov}(\hat{x}, \hat{u}) + \text{cov}(\hat{x}, \varepsilon) = 0$$

   When there exists many IVs:

   *First Stage

$$x = \delta_0 + \delta_1 \times z_1 + \delta_2 \times z_2 + u$$
$$\hat{x} = \hat{\delta_0} + \hat{\delta_1} \times z_1 + \hat{\delta_2} \times z_2$$

   *Second Stage

$$Y = \beta_{0,2SLS} + \beta_{1,2SLS} \times \hat{x} + \varepsilon_{2SLS}$$

1.2 Math Section

1.2.1 Assumption

1. **Linearity**: $Y = X\beta + \epsilon$.
2. **Full rank**: $\text{rank}(X) = k$.
3. **Exogeneity**: $\mathbb{E}[\epsilon|X] = 0$.

   Law of iterated expectations:

$$\mathbb{E}[\epsilon] = \mathbb{E}[\mathbb{E}[\epsilon|X]] = \mathbb{E}[0] = 0.$$

4. **Homoscedasticity and nonautocorrelation**:

$$\text{Var}(\epsilon_i|X) = \sigma^2, \quad i = 1, 2, \ldots, n.$$

$$\text{Var}(\epsilon_i, \epsilon_j|X) = 0, \quad i \neq j, \quad \text{Var}(\epsilon_i \epsilon) = \sigma^2 I.$$

5. X may be fixed and random.

We assume that there is an additional vector of variables z_i, with $L \geq k$.

(1) **Exogeneity**: z_i is uncorrelated with disturbance ϵ_i.

(2) **Relevance**: z_i is correlated with explanatory variable x_i.

(3) **Homoscedasticity**: $\mathbb{E}[\epsilon_i^2|z_i] = \sigma^2$.

(4) **Random Sampling** $(x_i, z_i, \epsilon_i) \overset{iid}{\sim}$.

(5) **Moments of x_i and z_i**:

$$\mathbb{E}[x_i x_i'] = Q_{XX} < \infty, \quad \text{rank}(Q_{XX}) = k.$$

$$\mathbb{E}[z_i z_i'] = Q_{ZZ} < \infty, \quad \text{rank}(Q_{ZZ}) = L.$$

$$\mathbb{E}[z_i x_i'] = Q_{ZX} < \infty, \quad \text{rank}(Q_{ZX}) = k.$$

$$(L \times k) \quad (\text{since } L \geq k).$$

(6) **Exogeneity of Instruments**:

$$\mathbb{E}[\epsilon_i|b_i] = 0.$$

1.2.2 Property of OLS

1. **OLS is biased**.

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon.$$

$$\mathbb{E}[\hat{\beta}|X] = \beta + \mathbb{E}[(X'X)^{-1}X'\epsilon|X].$$

$$= \beta + (X'X)^{-1}X'\mathbb{E}[\epsilon|X].$$

$$= \beta + (X'X)^{-1}X'\eta \neq \beta$$

(biased).

2. **OLS is inconsistent in big sample**.

Recall: $\mathbb{E}[\epsilon|X] = 0, \quad \mathbb{E}[\epsilon_i x_i]$

$$= \mathbb{E}\left[\mathbb{E}[\epsilon_i x_i|X]\right] = \mathbb{E}\left[x_i \mathbb{E}[\epsilon_i|X]\right] = 0.$$

2. **OLS is inconsistent**.

$$\mathbb{E}[x_i\epsilon_i] = \mathbb{E}[x_i\eta] \neq 0.$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\right).$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} Q_{XX}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i \xrightarrow{p} \eta \neq 0.$$

$$\Rightarrow \hat{\beta} \xrightarrow{p}\neq \beta.$$

moment non.

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

OLS? .

3. **A method of moment estimator β_{mom} sets the sample analogue to 0**:

$$\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\beta_{\text{mom}}) = 0.$$

$$\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i x_i'\right)\beta_{\text{mom}} = 0.$$

.

$$\left(\sum_{i=1}^{n} x_i x_i'\right)\beta_{\text{mom}} = \sum_{i=1}^{n} x_i y_i.$$

$$\beta_{\text{mom}} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right).$$

$$= (X'X)^{-1}X'y = \beta_{\text{ols}}.$$

IV Model Assumptions

- (1), (2), (3) were replaced with (7).

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

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

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

(In sample),

$$\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i'\beta_{IV}) = 0.$$

$$\sum_{i=1}^{n} z_i y_i - \left(\sum_{i=1}^{n} z_i x_i'\right)\beta_{IV} = 0.$$

$$\left[\sum_{i=1}^{n} z_i x_i'\right]\beta_{IV} = \sum_{i=1}^{n} z_i y_i.$$

If $L = k$, then

$$\beta_{IV} = \left(\sum_{i=1}^{n} z_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i y_i\right).$$

$$\beta_{IV} = (Z'X)^{-1}Z'y.$$

$$\beta_{OLS} = (X'X)^{-1}X'y.$$

1.2.3 WTS: Consistency

When $L = k$, $\mathbb{E}[z_i x_i'] = Q_{ZX}$, and:

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

$$= (Z'X)^{-1}Z'(X\beta + \epsilon).$$

$$= \beta + (Z'X)^{-1}Z'\epsilon.$$

$$= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i\epsilon_i\right).$$

$$\xrightarrow{p} \mathbb{E}[z_i x_i'] = Q_{ZX}, \quad \text{for using WLLN}.$$

$$\Rightarrow \hat{\beta}_{IV} \xrightarrow{p} \beta + (\mathbb{E}[z_i x_i'])^{-1}\mathbb{E}[z_i\epsilon_i].$$

$$\mathbb{E}[z_i\epsilon_i] = \mathbb{E}[\mathbb{E}[z_i\epsilon_i|z_i]] = \mathbb{E}[z_i\mathbb{E}[\epsilon_i|z_i]].$$

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

$$\Rightarrow \hat{\beta}_{IV} \xrightarrow{p} \beta.$$

IV estimator is consistent.
WTS: Asymptotic normality proof

$$\hat{\beta}_{IV} - \beta = \left(\frac{1}{n} \sum_{i=1}^{n} z_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} z_i \epsilon_i \right).$$

By CLT,

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \left[\frac{1}{n} \sum_{i=1}^{n} z_i x_i' \right]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \epsilon_i \right).$$

$$\xrightarrow{p} Q_{ZX}$$

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} z_i \epsilon_i \right) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} z_i \epsilon_i - \mathbb{E}[z_i \epsilon_i] \right).$$

$$\xrightarrow{d} N(0, \sigma^2 Q_{ZZ}).$$

$$\text{Var}(z_i \epsilon_i) = \mathbb{E}[z_i \epsilon_i - 0](z_i \epsilon_i - 0)'.$$

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

$$= \mathbb{E}[\mathbb{E}[\epsilon_i^2 | z_i] z_i z_i'].$$

$$= \sigma^2 \mathbb{E}[z_i z_i'] = \sigma^2 Q_{ZZ}.$$

By Slutsky's theorem,

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \to dN(0, \sigma^2 Q_{ZX}^{-1} Q_{ZZ} Q_{ZX}^{-1}).$$

Consistency.
But IV is biased:

$$\hat{\beta}_{IV} = \beta + (Z'X)^{-1} Z' \epsilon.$$

$$\mathbb{E}[\hat{\beta}_{IV} | X, Z] = \beta + (Z'X)^{-1} Z' \mathbb{E}[\epsilon | X, Z] \neq \beta.$$

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

Matrix dimensions:

$$Z : n \times L, \quad Z' : L \times n, \quad X : n \times k.$$

$$L > k.$$

When $L > k$:

$$X \to Z \quad \text{projection.}$$

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

$$= ZCZ'Z'.$$

L L Z Z . .

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

$$\hat{X} = Z(Z'Z)^{-1}Z'X.$$

$$L \times L, \quad L \times n, \quad L \times k.$$

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

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

Replaced the Z.

2SLS

When the number of instrumental variables (m) exceeds the number of endogenous regressors (k), the usual inverse $(Z'X)^{-1}$ does not exist because $Z'X$ is not square or may not be full rank. To address this issue, we use the **Two-Stage Least Squares (2SLS) approach** to estimate the regression coefficients.

Steps of 2SLS *Step 1: First Stage Regression To address endogeneity in X, we first express X in terms of the instrumental variables Z:

$$X = ZC + V$$

where:
- Z is the matrix of instrumental variables ($n \times m$).
- C is the coefficient matrix to be estimated.
- V is the error term.

Since $m > k$, the equation for C is obtained using the **Ordinary Least Squares (OLS) estimator**:

$$\hat{C} = (Z'Z)^{-1}Z'X.$$

Thus, we obtain the predicted values of X:

$$\hat{X} = Z\hat{C} = Z(Z'Z)^{-1}Z'X.$$

Since $\hat{X}$ is the part of X that is explained by Z, we can decompose:

$$X = \hat{X} + \hat{V},$$

where $\hat{V}$ represents the residuals.

Step 2: Second Stage Regression

Now, instead of using the original X (which is endogenous), we use the predicted values $\hat{X}$ to estimate the relationship between Y and X:

$$Y = X\tilde{\beta}^{2SLS} + \tilde{u}^{2SLS}.$$

Since X contains endogenous variables, we use $\hat{X}$ as an instrument:

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

Expanding $\tilde{\beta}^{2SLS}$:

$$\tilde{\beta}^{2SLS} = [(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}(X'Z)(Z'Z)^{-1}Z'Y.$$

Consistency of the Estimator

To show that $\tilde{\beta}^{2SLS}$ is a **consistent estimator**, we take the probability limit:

$$plim\,\tilde{\beta}^{2SLS} = \beta + plim\left[(X'Z)(Z'Z)^{-1}(Z'Z)\right]^{-1} \cdot plim X'Z(Z'Z)^{-1}Z'u.$$

Since $plim X'Z(Z'Z)^{-1}Z'u = 0$ under exogeneity conditions, we obtain:

$$plim\,\tilde{\beta}^{2SLS} = \beta.$$

Thus, the estimator is **consistent**.

*Variance of $\tilde{\beta}^{2SLS}$ The variance of $\tilde{\beta}^{2SLS}$ is given by:

$$\widehat{\mathrm{Var}}(\tilde{\beta}^{2SLS}) = \hat{\sigma}_u^2(X'Z(Z'Z)^{-1}Z'X).$$

where the estimated error variance is:

$$\hat{\sigma}_u^2 = \frac{\tilde{u}'\tilde{u}}{n}.$$

Important Note: The residuals are computed as:

$$\tilde{u} = Y - X\tilde{\beta}^{2SLS}, \quad \text{not as} \quad Y = \hat{X}\tilde{\beta}^{2SLS}.$$

Conclusion The **2SLS method** ensures that the estimator is **consistent** when X is endogenous. The key intuition is:

1. The **first stage** removes endogeneity by regressing X on the instruments Z, isolating the exogenous variation.

2. The **second stage** uses this exogenous variation to estimate β, ensuring that the regression is not biased by endogeneity.

Thus, 2SLS provides an effective way to obtain **unbiased and consistent estimates** in the presence of endogeneity.

1.2.4 The Property of 2SLS

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

$$= (X'P_ZX)^{-1}X'P_Z(X\beta + \varepsilon)$$

$$= \beta + (X'P_ZX)^{-1}X'P_Z\varepsilon$$

We want to show that $\hat{\beta}_{2SLS}$ is a consistent estimator, which requires proving that:

$$(X'P_ZX)^{-1}X'P_Z\varepsilon \xrightarrow{p} 0.$$

*Step-by-Step Derivation

$$(X'P_ZX)^{-1}X'P_Z\varepsilon$$

$$= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon$$

$$= \left[\left(\frac{X'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'X}{n}\right)\right]^{-1}\left(\frac{X'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'\varepsilon}{n}\right).$$

By the Weak Law of Large Numbers (WLLN):

$$\frac{1}{n}\sum_{i=1}^{n} X_i z_i' \xrightarrow{p} E[X_i z_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \xrightarrow{p} E[z_i z_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} z_i X_i' \xrightarrow{p} E[z_i X_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} z_i \varepsilon_i \xrightarrow{p} E[z_i \varepsilon_i] = 0.$$

Thus,

$$(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}(Q_{XZ}Q_{ZZ}^{-1}0) = 0.$$

This shows that:

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

*Asymptotic Normality

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta)$$

$$= \left[\frac{X'Z}{n}\frac{Z'Z}{n}^{-1}\frac{Z'X}{n}\right]^{-1}\frac{X'Z}{n}\frac{Z'Z}{n}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i\varepsilon_i.$$

By the Central Limit Theorem (CLT):

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i\varepsilon_i \xrightarrow{d} N(0, \sigma^2 Q_{ZZ}).$$

Since

$$\mathrm{Var}(Z'\varepsilon) = E[z_i\varepsilon_i\varepsilon_i z_i'] = E[\varepsilon_i^2 z_i z_i'] = \sigma^2 E[z_i z_i'] = \sigma^2 Q_{ZZ},$$

we obtain:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, \sigma^2(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}).$$

Thus,

$$\hat{\beta}_{2SLS} \sim N(\beta, \frac{\sigma^2}{n}(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}).$$

1.2.5 Efficiency

*Variance Comparison and Positive Semi-Definiteness

Statement: If β_{OLS} variance is smaller than β_{IV},

$$A - B > 0 \quad \text{(positive semi-definite)}$$

then $B^{-1} - A^{-1}$ is also positive semi-definite.

*Derivation

$$Q_{XX} - (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

$$= Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}$$

$$= \operatorname*{plim}_{n\to\infty}\frac{X'X}{n} - \operatorname*{plim}_{n\to\infty}\frac{X'Z}{n}\left(\operatorname*{plim}_{n\to\infty}\frac{Z'Z}{n}\right)^{-1}\operatorname*{plim}_{n\to\infty}\frac{Z'X}{n}$$

$$= \operatorname*{plim}_{n\to\infty}\left[\frac{X'X}{n} - \frac{X'Z}{n}(Z'Z/n)^{-1}Z'X/n\right]$$

$$= \operatorname*{plim}_{n\to\infty}\left[\frac{X'(I-P_Z)X}{n}\right] = \operatorname*{plim}_{n\to\infty}\frac{X'M_ZX}{n}$$

where $M_Z = I - P_Z$ and $P_Z = Z(Z'Z)^{-1}Z'$.

If A is positive semi-definite, then A is positive semi-definite, where $A_n \xrightarrow{p} A$.

*Important Observation

$$X'M_ZX = X'M_ZM_ZX = X'M_Z(X'M_Z)$$

For any $r \neq 0$, let $V = r'(X'M_Z)$,

$$r'X'M_ZXr' = \gamma'X'M_ZXM_ZXr' = \gamma'\mathbb{D}\gamma = \sum_{i=1}^{p} v_i^2$$

$$\sigma_{IV}^2 \geq \sigma_{OLS}^2.$$

Thus, the asymptotic variance satisfies:

$$\text{Asy. Var}(\beta_{OLS}) \leq \text{Asy. Var}(\beta_{IV}).$$

*Conclusion:
However, note that σ^2 is still useful, so further testing is needed.

1.2.6 Test

1. **Hausman Test**

   - **Null Hypothesis**
     $$H_0 : E[\varepsilon_i|x_i] = 0 \quad \Rightarrow \quad \text{Exogeneity}$$
     - Under H_0, IV and OLS are consistent.
   - Define the difference:
     $$d = \hat{\beta}_{IV} - \hat{\beta}_{OLS}$$
     (similar to a linear restriction). Under H_0,
     $$d \xrightarrow{p} 0.$$
   - **Test Statistic** If we can derive:
     $$\sqrt{n}d \xrightarrow{d} N(0, V),$$
     and estimate V by $\hat{V}$, then we can test H_0 using the Wald statistic:
     $$W = \sqrt{n}d'\hat{V}^{-1}\sqrt{n}d = nd'\hat{V}^{-1}d \xrightarrow{d} \chi^2(r).$$
   - **Variance of d**
     $$\text{Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS})$$
     $$= \text{Var}(\hat{\beta}_{IV}) + \text{Var}(\hat{\beta}_{OLS}) - 2\text{Cov}(\hat{\beta}_{IV}, \hat{\beta}_{OLS}).$$
   - **Hausman's Principle**
     Let b_E be an estimator of β such that:
     $$\sqrt{n}(b_E - \beta) \xrightarrow{d} N(0, V_E).$$
     Suppose b_E is efficient in the sense that for any other estimator b of β such that:

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, V),$$

we have:

$$V \geq V_E.$$

Let b_I be an inefficient estimator of β, namely:

$$\sqrt{n}(b_I - \beta) \xrightarrow{d} N(0, \Sigma), \quad \text{where } \Sigma \geq V_E.$$

Then the asymptotic variance satisfies:

$$\text{Asy. Var}(b_E, b_I) = \text{Asy. Var}(b_E).$$

- **Proof of a Scalar Case**
  Let β be a scalar.
  Consider an estimator:

$$\hat{\beta} = \alpha b_I + (1 - \alpha) b_E = b_E + d(b_I - b_E)$$

for a constant α.
Then,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Omega).$$

Asymptotic Variance

$$\Omega = \text{Asy. Var}[b_E + d(b_I - b_E)]$$

$$= \text{Asy. Var}[b_E] + d^2 \text{Asy. Var}[b_I - b_E] + 2d \text{Asy. Cov}(b_E, b_I - b_E)$$

$$= \text{Asy. Var}[b_E] + 2d \text{Asy. Cov}(b_E, b_I - b_E) + d^2 \text{Asy. Var}(b_I - b_E).$$

Minimization Condition

$$\Omega \text{ is minimized when } d = -\frac{\text{Asy. Cov}(b_E, b_I - b_E)}{\text{Asy. Var}(b_I - b_E)}.$$

Efficiency Argument If $\alpha^* \neq 0$, then $\hat{\beta}$ with $\alpha = \alpha^*$ will have a smaller asymptotic variance than $\hat{\beta}$ with $\alpha = 0$, which contradicts the efficiency of b_E.
Thus, we conclude:

$$\alpha^* = 0 \quad \Rightarrow \quad \text{Asy. Cov}(b_E, b_I - b_E) = 0.$$

Final Covariance Expression Using the identity:

$$\text{Cov}(A + B, C) = \text{Cov}(A, C) + \text{Cov}(B, C),$$

we obtain:

$$\text{Asy. Cov}(b_I, b_E) - \underbrace{\text{Asy. Cov}(b_E, b_E)}_{\text{Asy. Var}(b_E)} = 0.$$

- **Final Test Statistic**

$$\sqrt{n}d \xrightarrow{d} N(0, V),$$

where:

$$V = \text{Asy. Var}(\hat{\beta}_{IV} - \hat{\beta}_{OLS})$$

$$= \text{Asy. Var}(\hat{\beta}_{IV}) - \text{Asy. Var}(\hat{\beta}_{OLS}) - 2\text{Asy. Cov}(\hat{\beta}_{IV}, \hat{\beta}_{OLS})$$

$$= \text{Asy. Var}(\hat{\beta}_{IV}) - \text{Asy. Var}(\hat{\beta}_{OLS}).$$

Let:

$$\hat{V}_{IV} \xrightarrow{p} \text{Asy. Var}(\hat{\beta}_{IV}),$$

$$\hat{V}_{OLS} \xrightarrow{p} \text{Asy. Var}(\hat{\beta}_{OLS}).$$

Then the final test statistic is:

$$W = nd'(\hat{V}_{IV} - \hat{V}_{OLS})^{-1} d \xrightarrow{d} \chi^2(r).$$

Question: Does the instrumental variable z need to be uncorrelated with the dependent variable y?

No!

- The instrumental variable z affects the dependent variable y through the endogenous variable x:

$$z \to x \to y$$

- The instrumental variable z does not directly affect the dependent variable y:

$$\text{cov}(z, y|x) = 0$$

- The instrumental variable z **can and must** influence the dependent variable y **only through** the endogenous variable x.

Suppose that there is a set of instrumental variables $Z = (Z_0 \quad Z_1 \quad \ldots Z_K)$ that meet the following condition:

1. plim $n^{-1}Z'X = Q_{ZX}$ (non-singular)
2. plim $n^{-1}Z'Z = Q_{ZZ}$ (positive definite)

3. plim $n^{-1}Z'u = 0$

$$Y = X\beta + u \Rightarrow Z'Y = Z'X\beta + Z'u$$

Let $\tilde{\beta}$ be an estimator of β. Then we have:

$$Z'Y = Z'X\tilde{\beta} + Z'\tilde{u} \Rightarrow Z\tilde{U} =$$

$$Z'(Y - X\tilde{\beta}) \Rightarrow \tilde{u} = Y - X\tilde{\beta}$$

$$(Z'\tilde{u})(Z'\tilde{u}) = (Z'Y - Z'X\tilde{\beta})'(Z'Y - Z'X\tilde{\beta})$$

$$= Y'Z'ZY - 2\tilde{\beta}'X'Z'ZY + \tilde{\beta}'X'Z'Z'X\tilde{\beta}$$

$$\frac{\partial(Z'\tilde{u})(Z'\tilde{u})}{\partial\tilde{\beta}} = -2X'Z'ZY + 2X'Z'Z'X\tilde{\beta} = 0$$

hence $X'Z'ZY = X'Z'Z'X\tilde{\beta}$. Then premultiplying by $(X'Z)^{-1}$ leads to

$$\tilde{\beta}^{IV} = (Z'X)^{-1}Z'Y$$

We further have:

$$\tilde{\beta}^{IV} = (Z'X)^{-1}Z'(X\beta + u)$$

$$= \beta + (Z'X)^{-1}Z'u$$

$$\text{plim } \tilde{\beta}^{IV} = \beta + \left[\text{plim}\left(\frac{Z'X}{n}\right)\right]^{-1} \cdot \text{plim}\frac{Z'u}{n}$$

$$= \beta + Q_{ZX}^{-1} \cdot 0 = \beta$$

Therefore $\tilde{\beta}^{IV}$ is consistent.

1.3 Problem Set

*Problem 2 Derive the limiting distribution of the two-stage least squares estimator (2SLS) and consistency of the estimator for the variance-covariance matrix. For each step make exactly clear which assumptions are needed. You may assume homoskedasticity of the errors, or not, but if so state it as an assumption.

(a). Verify that

$$\hat{\beta}_{2SLS} - \beta = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'\varepsilon.$$

1. **Solution**

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

$$= \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'X}{n}\right) \right]^{-1} \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'\varepsilon}{n}\right) \right].$$

To use the Weak Law of Large Numbers (WLLN) in Hansen chapter 6, P164, the following assumptions are needed:

2. *Assumptions

   - **A1:** (y_i, x_i, z_i) are i.i.d.
   - **A2:** $E|y_i|^2 < \infty$, $E||x_i||^2 < \infty$, $E||z_i||^2 < \infty$.

3. *Detour:

   - The WLLN in Hansen only needs the first moment, as in A2: $E|y_i| < \infty$, $E||x_i|| < \infty$, $E||z_i|| < \infty$; but in A2, we ask for the second moment to exist. The reason is that the cross product behaves like a degree-2 term. By the **Cauchy-Schwarz inequality**, one can prove that the expectation of the cross product exists and is finite using A2.
   - For example, using the inequality:

   $$E(|x_{ik} z_{i\ell}|) \leq \sqrt{E|x_{ik}^2| E|z_{i\ell}^2|}$$

   where x_{ik} is the k-th element, and $z_{i\ell}$ is the ℓ-th element. Since A2 ensures the second moment of x_i and z_i exists and is finite, it follows that $E[x_i z_i']$ exists and is finite.

4. *By the WLLN, we obtain:

$$\frac{X'Z}{n} \xrightarrow{p} Q_{XZ}, \quad \frac{Z'Z}{n} \xrightarrow{p} Q_{ZZ}, \quad \frac{Z'X}{n} \xrightarrow{p} Q_{ZX}.$$

5. *By the Continuous Mapping Theorem, and the additional assumptions:

   - **A3:** $E[z_i \varepsilon_i] = 0$ (the **exogeneity condition**).
   - **A4:** $E[z_i z_i'] = Q_{ZZ}$ is full rank/invertible/positive definite.
   - **A5:** $E[z_i x_i']$ has full column rank K (the **relevance condition**).

6. *Then, the 2SLS estimator is consistent as:

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{P} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1} \underbrace{\cancel{E[z_i \varepsilon_i]}}_{0} = 0$$

(Finite matrix).

b. Rescale the equation to converge to a random variable and establish the asymptotic distribution

*Solution

1. Use the usual scaling, multiply by $\sqrt{n}$, and

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

$$= \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'X}{n}\right) \right]^{-1} \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'\varepsilon}{\sqrt{n}}\right) \right].$$

2. **Weak Law of Large Numbers (WLLN) and Central Limit Theorem (CLT)** WLLN and CLT are needed to obtain the distribution. To use the CLT as in Hansen chapter 6, P164, we need assumptions **A1**, **A2**, and:

   - **A6:** $E||z_i z_i' \varepsilon_i^2|| < \infty$, since to use CLT for $z_i \varepsilon_i$, we need $z_i \varepsilon_i$ to have a **finite second moment**.
   - **A7:** $\Omega = E[z_i z_i' \varepsilon_i^2]$ is positive definite, so it is a valid asymptotic variance matrix.

   (Can have a different **A6'** as $E|y_i|^4 < \infty$, $E||z_i||^4 < \infty$, $E||x_i||^4 < \infty$, and then use the **Cauchy-Schwarz inequality** to prove $E||z_i z_i' \varepsilon_i^2|| < \infty$. Assumption **A6'** can replace both **A6** and **A2**, since a higher moment exists means a lower moment also exists.)

3. **Application of the Central Limit Theorem** By the CLT, we have:

$$\sqrt{n}\frac{Z'\varepsilon}{n} = \sqrt{n}\frac{1}{n}\sum_i z_i \varepsilon_i \xrightarrow{d} N(0, \Omega)$$

4. **Combining with WLLN**

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

$$\xrightarrow{d} (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1} Q_{XZ}Q_{ZZ}^{-1} N(0, \Omega) = N(0, V)$$

   where

$$V = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1} (Q_{XZ}Q_{ZZ}^{-1}\boldsymbol{\Omega}Q_{ZZ}^{-1}Q_{ZX})(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

(c). Estimator $\hat{V}$ for the Variance-Covariance Matrix
Solution: Detour

1. - This V is the variance-covariance matrix in $\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, V)$, not the asymptotic variance of $\hat{\beta}_{2SLS}$.
   - The asymptotic variance of $\hat{\beta}_{2SLS}$ is $\frac{V}{n}$.

2. **Under A8: Homoskedasticity,** $E[\varepsilon_i^2] = \sigma^2 < \infty$

$$\Omega = E[z_i z_i' \varepsilon_i^2] = \sigma^2 E[z_i z_i'] = \sigma^2 Q_{ZZ}$$

   Thus, V can be reduced to:

$$V = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1} (Q_{XZ}Q_{ZZ}^{-1}\sigma^2 Q_{ZZ}Q_{ZZ}^{-1}Q_{ZX})(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

$$= \sigma^2 (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

3. **Sample Analog $\hat{V}$**

$$\hat{V} = \hat{\sigma}^2 (\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1}$$

where

$$\hat{Q}_{ZZ} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' = \frac{1}{n} Z'Z$$

$$\hat{Q}_{XZ} = \frac{1}{n} \sum_{i=1}^{n} x_i z_i' = \frac{1}{n} X'Z$$

$$\hat{Q}_{ZX} = \frac{1}{n} \sum_{i=1}^{n} z_i x_i' = \frac{1}{n} Z'X$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i' \hat{\beta}_{2SLS})^2$$

4. **Heteroskedasticity Case**

If heteroskedasticity is present, then V cannot be simplified. With the Q items the same as above, the Ω matrix can be estimated by:

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' \hat{\varepsilon}_i^2 = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' (y_i - x_i' \hat{\beta}_{2SLS})^2$$

d. Establish consistency of $\hat{V}$
Solution

1. **Under A8: Homoskedasticity**,

$$\hat{V} = \hat{\sigma}^2 (\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1}$$

The convergence in probability of $(\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1}$ has been proven when establishing consistency, so the key is to show $\hat{\sigma}^2$ is a consistent estimator of σ^2.

To show this, write:

$$\hat{\varepsilon}_i = y_i - x_i' \hat{\beta} = x_i' \beta + \varepsilon_i - x_i' \hat{\beta} = x_i' (\beta - \hat{\beta}) + \varepsilon_i$$

$$\hat{\varepsilon}_i^2 = \varepsilon_i^2 + 2(\beta - \hat{\beta})' x_i \varepsilon_i + (\beta - \hat{\beta})' x_i x_i' (\beta - \hat{\beta})$$

Summing up:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 + 2(\beta-\hat{\beta})'\frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i + (\beta-\hat{\beta})'\frac{1}{n}\sum_{i=1}^n x_i x_i'(\beta-\hat{\beta})$$

- (1) By WLLN, $\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \xrightarrow{p} E[\varepsilon_i^2] = \sigma^2$.
- (3) By A2, $E[x_i x_i'] < \infty$, and $\hat{\beta}_{2SLS}$ is a consistent estimator of β, using WLLN that $\frac{1}{n}\sum_{i=1}^n x_i x_i' \xrightarrow{p} E[x_i x_i'] < \infty$, thus part (3) vanishes as $n \to \infty$.
- (2) Under A2, both $E[x_{ik}^2] < \infty$ and $E[\varepsilon_i^2] < \infty$, and by the **Cauchy-Schwarz inequality**:

  $$E(|x_{ik}\varepsilon_i|) \leq \sqrt{E[x_{ik}^2]E[\varepsilon_i^2]} < \infty.$$

  Using WLLN that $\frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i \xrightarrow{p} E[x_i\varepsilon_i] < \infty$, and again $\hat{\beta}_{2SLS}$ is a consistent estimator of β, part (2) vanishes as $n \to \infty$.

Hence, we obtain:

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2, \quad \text{and} \quad \hat{V} \xrightarrow{p} V.$$

2. **Heteroskedasticity Case**

$$V = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}(Q_{XZ}Q_{ZZ}^{-1}\mathbf{\Omega}Q_{ZZ}^{-1}Q_{ZX})(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

One needs to prove that:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n z_i z_i'(y_i - x_i'\hat{\beta}_{2SLS})^2 \xrightarrow{p} E[z_i z_i'\varepsilon_i^2].$$

Inserting $\hat{\varepsilon}_i$ back into $\hat{\Omega}$:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n z_i z_i'\varepsilon_i^2 + 2\frac{1}{n}\sum_{i=1}^n z_i z_i'[(\beta-\hat{\beta})'x_i\varepsilon_i] + \frac{1}{n}\sum_{i=1}^n z_i z_i'[(\beta-\hat{\beta})'x_i x_i'(\beta-\hat{\beta})]$$

- (1) By WLLN, $\frac{1}{n}\sum_{i=1}^n z_i z_i'\varepsilon_i^2 \xrightarrow{p} E[z_i z_i'\varepsilon_i^2]$.
- (2) In homoskedasticity, we could take $(\beta-\hat{\beta})$ out of summation, but here we cannot directly because:

  $$\frac{1}{n}\sum_{i=1}^n z_i z_i'\left[(\beta-\hat{\beta})'x_i\varepsilon_i\right]$$

Instead, consider the $k-\ell$ element in $\hat{\Omega}$:

$$\hat{\Omega}_{k\ell} = \frac{1}{n}\sum_{i=1}^n z_{ik}z_{i\ell}[(\beta-\hat{\beta})'x_i\varepsilon_i] = (\beta-\hat{\beta})'\frac{1}{n}\sum_{i=1}^n z_{ik}z_{i\ell}x_i\varepsilon_i$$

Then follow similar logic as in homoskedasticity and show that:

$$\hat{\Omega}_{k\ell} \xrightarrow{p} E[z_{ik} z_{i\ell} x_i \varepsilon_i] < \infty.$$

2 Pandel Data & Model

1. Panel Data System

For m individual units observed over T time periods, consider the system:

$$Y_i = X_i \beta_i + \varepsilon_i, \quad i = 1, \ldots, m$$

We can stack the system as:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Assuming:

$$E[\varepsilon\varepsilon' | X] = \Omega$$

If the errors are uncorrelated across time but possibly correlated across units, then:

$$\Omega = \Sigma \otimes I_T$$

Where $\Sigma \in \mathbb{R}^{m \times m}$ captures contemporaneous correlation across units, and I_T is a $T \times T$ identity matrix.

GLS Estimator

$$\hat{\beta}_{\text{GLS}} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y$$

If $\Omega = \Sigma \otimes I$, then:

$$\hat{\beta}_{\text{GLS}} = \left(X'(\Sigma^{-1} \otimes I) X \right)^{-1} X'(\Sigma^{-1} \otimes I) Y$$

This is useful in the context of Seemingly Unrelated Regressions (SUR).

2. Panel Data with Unobserved Heterogeneity

A more general model includes observed and unobserved heterogeneity:

$$y_{it} = X_{it}\beta + Z_i'\theta + C_i + \nu_{it}$$

Where:

- X_{it}: time-varying regressors

- Z_i: time-invariant observed regressors
- C_i: unobserved individual effect
- ν_{it}: idiosyncratic error

Let $V_{it} = C_i + \nu_{it}$, then:

$$y_{it} = X_{it}\beta + Z_i'\theta + V_{it}$$

2.1 Random Effects Model

Assume:

$$C_i \perp X_{it},\ Z_i$$

Then, define:

$$Z_i'\theta + C_i = \alpha_i, \quad E[C_i|X_{it}] = 0$$

So the model becomes:

$$y_{it} = X_{it}\beta + \alpha + u_i + \nu_{it}$$

Estimation: feasible GLS or MLE.

2.2 Fixed Effects Model

Assume:

$$C_i \text{ is correlated with } X_{it},\ Z_i$$

Then C_i cannot be treated as part of the error. Instead, eliminate C_i using the within transformation (demeaning over time) or using dummy variables:

$$y_{it} = X_{it}\beta + Z_i'\theta + C_i + \nu_{it} \Rightarrow y_{it} = X_{it}\beta + \alpha_i + \nu_{it}$$

Estimation: fixed effects (within estimator or LSDV method).

Summary

	Random Effects	Fixed Effects
Assumption	$C_i \perp X_{it}, Z_i$	C_i correlated with X_{it}, Z_i
Estimator	GLS / MLE	Within Estimator / Dummy Variables
Efficiency	More efficient if assumption holds	Robust to correlation
Consistency	Only if uncorrelated	Always consistent

Model Setup

Consider the fixed effects panel data model:

$$y_{it} = X_{it}\beta + \alpha_i + \nu_{it}$$

where:

- y_{it}: dependent variable
- X_{it}: time-varying regressors
- α_i: unobserved time-invariant individual-specific effect
- ν_{it}: idiosyncratic error

To consistently estimate β, we must eliminate α_i, which may be correlated with X_{it}.

1. Dummy Variables (LSDV Method)

Introduce a set of $N-1$ individual-specific dummy variables:

$$y_{it} = X_{it}\beta + \sum_{j=1}^{N-1} d_j\delta_j + \nu_{it}$$

where:

- $d_j = 1$ if the observation belongs to individual j, 0 otherwise.
- δ_j captures the effect α_j.

This formulation allows us to estimate β while absorbing the α_i via dummies.

2. Within Transformation (Time-Demeaning)

Take the average over time for each individual i:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$$

Subtract individual means from each observation:

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + (\alpha_i - \alpha_i) + (\nu_{it} - \bar{\nu}_i)$$

$$\Rightarrow \tilde{y}_{it} = \tilde{X}_{it}\beta + \tilde{\nu}_{it}$$

This transformation removes α_i, and OLS on the transformed variables yields the fixed effects estimator.

Summary

- **Dummy variables** absorb α_i by explicitly including it in the regression.
- **Within transformation** eliminates α_i by demeaning, leading to the "within" estimator.

Both approaches yield consistent estimates of β under the fixed effects assumption.

1. Model Setup

The fixed effects model with individual-specific effects is:

$$y_{it} = X_{it}'\beta + \alpha_i + \nu_{it}$$

We can write this in matrix form using dummy variables $D \in \mathbb{R}^{nT \times n}$, where each column of D corresponds to one individual:

$$Y = X\beta + D\alpha + \nu$$

Here:

- $Y \in \mathbb{R}^{nT \times 1}$: stacked vector of outcomes
- $X \in \mathbb{R}^{nT \times k}$: stacked covariates
- $D \in \mathbb{R}^{nT \times n}$: individual dummy matrix (one column per individual)
- $\alpha \in \mathbb{R}^{n \times 1}$: individual effects
- $\nu \in \mathbb{R}^{nT \times 1}$: error term

2. OLS with Dummy Variables (LSDV)

We can estimate β and α using OLS on:

$$Y = X\beta + D\alpha + \nu$$

However, when n is large, this becomes computationally inefficient due to the large number of dummies.

3. Within Transformation (Projection)

To eliminate α, we use the projection matrix:

$$M = I - D(D'D)^{-1}D'$$

This matrix projects any vector onto the orthogonal complement of the column space of D, i.e., it removes the individual-specific means.

Multiply both sides of the model by M:

$$MY = MX\beta + MD\alpha + M\nu$$

Since $MD = 0$, we get:

$$MY = MX\beta + M\nu$$

Thus, the **within estimator** is:

$$\hat{\beta}_{FE} = (X'MX)^{-1}X'MY$$

This estimator is numerically equivalent to the LSDV estimator for β.

4. Summary

- LSDV: Estimate β and α using dummy variables.
- Within Estimator: Use matrix M to remove α and estimate β directly.
- Both methods give the same estimate for β, but the within estimator is computationally efficient for large n.