

Econometrics and Applications

Kirby CHEN

Academic Year 2024-2025

Contents

1	Lecture 3: Endogeneity and Instrumental Variables	1
1.1	Motivation and Overlook	1
1.2	Math Section	4
1.2.1	Assumption	4
1.2.2	Property of OLS	5
1.2.3	WTS: Consistency	7
1.2.4	The Property of 2SLS	11
1.2.5	Efficiency	12
1.2.6	Test	13
1.3	Problem Set	16
2	Pandel Data & Model	21
3	Stationarity	47

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:

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

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

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** $(x_i, z_i, \epsilon_i) \stackrel{iid}{\sim}$.
- (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: $\mathbb{E}[\epsilon|X] = 0, \quad \mathbb{E}[\epsilon_i x_i]$

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

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,

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{IV} - \beta) &= \left[\frac{1}{n} \sum_{i=1}^n z_i x_i' \right]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \epsilon_i \right) \\ &\xrightarrow{p} Q_{ZX} \end{aligned}$$

$$\begin{aligned} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n z_i \epsilon_i \right) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n z_i \epsilon_i - \mathbb{E}[z_i \epsilon_i] \right) \\ &\xrightarrow{d} N(0, \sigma^2 Q_{ZZ}). \end{aligned}$$

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

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

.

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

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

By Slutsky's theorem,

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

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 $L > k$:

$X \rightarrow Z$ projection.

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

$$= ZCZ'Z'.$$

$$\begin{matrix} L & L & Z & & Z & . & . \\ & & & & & & \end{matrix}$$

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

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

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

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

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

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:

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

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:

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

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

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'P_ZX)^{-1}X'P_ZY \\ &= (X'P_ZX)^{-1}X'P_Z(X\beta + \varepsilon) \\ &= \beta + (X'P_ZX)^{-1}X'P_Z\varepsilon\end{aligned}$$

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

$$\begin{aligned}& (X'P_ZX)^{-1}X'P_Z\varepsilon \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon \\ &= \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'X}{n} \right) \right]^{-1} \left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'\varepsilon}{n} \right).\end{aligned}$$

By the Weak Law of Large Numbers (WLLN):

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n X_i z_i' &\xrightarrow{p} E[X_i z_i'] \\ \frac{1}{n} \sum_{i=1}^n z_i z_i' &\xrightarrow{p} E[z_i z_i'] \\ \frac{1}{n} \sum_{i=1}^n z_i X_i' &\xrightarrow{p} E[z_i X_i'] \\ \frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i &\xrightarrow{p} E[z_i \varepsilon_i] = 0.\end{aligned}$$

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

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

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

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

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

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

$$\begin{aligned} & Q_{XX} - (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} \\ &= Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX} \\ &= \text{plim}_{n \rightarrow \infty} \frac{X'X}{n} - \text{plim}_{n \rightarrow \infty} \frac{X'Z}{n} \left(\text{plim}_{n \rightarrow \infty} \frac{Z'Z}{n} \right)^{-1} \text{plim}_{n \rightarrow \infty} \frac{Z'X}{n} \\ &= \text{plim}_{n \rightarrow \infty} \left[\frac{X'X}{n} - \frac{X'Z}{n} (Z'Z/n)^{-1} Z'X/n \right] \\ &= \text{plim}_{n \rightarrow \infty} \left[\frac{X'(I - P_Z)X}{n} \right] = \text{plim}_{n \rightarrow \infty} \frac{X'M_Z X}{n} \end{aligned}$$

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

$$X'M_Z X = X'M_Z M_Z X = X'M_Z (X'M_Z)$$

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:

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

and estimate V by \hat{V} , then we can test H_0 using the Wald statistic:

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

- **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**

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

where:

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

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 :

$$z \rightarrow x \rightarrow y$$

- 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 $Z = (Z_0 \ Z_1 \ \dots Z_K)$ that meet the following condition:

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

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

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

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**

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

1. **Solution**

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

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

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:

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

(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**

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{2SLS} - \beta) &= \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'X}{n} \right) \right]^{-1} \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'\varepsilon}{\sqrt{n}} \right) \right] \\ &\xrightarrow{d} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1} N(0, \Omega) = N(0, V) \end{aligned}$$

where

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

(c). **Estimator \hat{V} for the Variance-Covariance Matrix**

Solution: Detour

- 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 $n \rightarrow \infty$.
- (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 $n \rightarrow \infty$.

Hence, we obtain:

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

2. Heteroskedasticity Case

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

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:

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

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 Panel Data & Model

1. Panel Data System

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

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

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:

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

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:

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

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 .

1. What is the Matrix M ?

In fixed effects models, we want to eliminate unobserved individual-specific effects α_i . To do this, we use the matrix:

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

This is a **projection matrix**:

- D is the dummy variable matrix for individuals.
- $D(D'D)^{-1}D'$ is the projection onto the column space of D , i.e., the space spanned by individual means.
- M projects onto the *orthogonal complement* of that space, removing variation due to individual means.

2. Why Does M Remove Individual Means?

Suppose we have panel data with n individuals and T time periods per individual. Each variable $y \in \mathbb{R}^{nT}$ is stacked as:

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{nT} \end{bmatrix}$$

Let $D \in \mathbb{R}^{nT \times n}$ be the dummy matrix for individuals. Then:

$$D'D = T \cdot I_n, \quad (D'D)^{-1} = \frac{1}{T} I_n$$

$$D(D'D)^{-1}D' = \frac{1}{T} DD'$$

This means for any vector y , we have:

$$P = D(D'D)^{-1}D', \quad Py = \text{individual mean vector}$$

So:

$$My = (I - P)y = y - \bar{y}_i$$

where each block of T observations is demeaned within individual i . Hence, M performs the **within transformation**.

3. Numerical Example (2 Individuals, 2 Time Periods)

Let:

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Then:

$$D'D = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad (D'D)^{-1} = \frac{1}{2} I$$

$$P = D(D'D)^{-1}D' = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$Py = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_2 \end{bmatrix}, \quad My = y - Py = \begin{bmatrix} y_{11} - \bar{y}_1 \\ y_{12} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ y_{22} - \bar{y}_2 \end{bmatrix}$$

This is the vector of **demeaned values**.

4. Why Use This in Estimation?

In the fixed effects model:

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

Multiplying both sides by M :

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

Since $MD = 0$, the individual effects α drop out:

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

So we can estimate β using OLS on the **demeaned model**:

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

5. Summary

- $D(D'D)^{-1}D'$ projects a variable onto the space of individual means.
- $M = I - D(D'D)^{-1}D'$ subtracts those means — it de-means each individual's observations.
- This transformation is the core of the **fixed effects (within)** estimator.

Model Recap

The estimated pooled regression model is:

$$\ln(\text{WAGE}_{it}) = \beta_0 + \delta_0 Y85_{it} + \beta_1 \text{EDUC}_{it} + \delta_1 (Y85_{it} \cdot \text{EDUC}_{it}) + \cdots + \varepsilon_{it}$$

We focus on two main effects:

- Return to education over time
- Gender wage gap over time

1. Returns to Education (EDUC)

From Output (Pooled Regression):

- Coefficient on **educ**: $\beta_1 = 0.0747209$
- Coefficient on **y85Educ**: $\delta_1 = 0.0184605$

Interpretation

- **1978:** Return to education is simply $\hat{\beta}_1 = 0.0747$, or a 7.47% increase in wages per additional year of education.
- **1985:** Return includes the interaction:

$$\hat{\beta}_1 + \hat{\delta}_1 = 0.0747 + 0.0185 = 0.0932 \Rightarrow 9.32\% \text{ increase}$$

Statistical Significance

- The p-value for y85Educ is 0.049, which is marginally below 0.05.
- Thus, the increase is **statistically significant at the 5% level**.

Conclusion:

- In 1978, the return to education was 7.5%.
- In 1985, it increased by 1.85 percentage points to 9.3%.
- The difference is marginally statistically significant.

2. Gender Wage Gap (FEMALE)

From Output (Pooled Regression):

- Coefficient on female: $\beta_5 = -0.3167$
- Coefficient on y85fem: $\delta_5 = 0.0859$

Interpretation

- **1978:** Women earned $\exp(-0.3167) - 1 \approx -31.67\%$ less than men.
- **1985:** Gap reduced to $-0.3167 + 0.0859 = -0.2308 \Rightarrow \exp(-0.2308) - 1 \approx -23.08\%$
- Difference = 8.6 percentage points
- But: $p = 0.117 \Rightarrow$ **not statistically significant**.

Conclusion:

- In 1978, women earned 31.7% less than men.
- In 1985, the gap closed to 23.1%.
- The change is not statistically significant, so we cannot conclude the gap truly changed.

3. Mathematical Summary

Term	Coefficient
educ	0.0747 (EDU effect in 1978)
y85Educ	0.0185 (Extra EDU return in 1985)
female	-0.3167 (Wage penalty for women in 1978)
y85fem	0.0859 (Reduction in penalty in 1985)

4. Summary Table

Year	Return to EDU	Gender Gap	Significance EDU	Significance Gender
1978	$\beta_1 = 0.0747$	$\beta_5 = -0.3167$	—	—
1985	$\beta_1 + \delta_1 = 0.0932$	$\beta_5 + \delta_5 = -0.2308$	Yes (5%)	No (p = 0.117)

Why Interpret Dummy and Continuous Variables Differently in Log-Wage Models?

In models where the dependent variable is in logarithmic form (e.g., $\ln(\text{WAGE})$), the interpretation of regression coefficients depends on whether the independent variable is continuous or binary.

1. Continuous Variables (e.g., EDUC, EXPER)

For a continuous variable X , the model takes the form:

$$\ln(\text{WAGE}) = \beta_0 + \beta_1 X + \varepsilon$$

Then:

$$\frac{d \ln(\text{WAGE})}{dX} = \beta_1 \Rightarrow \text{WAGE changes by approximately } 100 \cdot \beta_1 \%$$

Interpretation: Each one-unit increase in X is associated with an approximate $\beta_1 \cdot 100\%$ change in wages.

Example: If $\beta_1 = 0.076$, then each additional year of education increases wages by approximately 7.6%.

2. Dummy Variables (e.g., FEMALE, UNION)

For a binary variable $D \in \{0, 1\}$, the model is:

$$\ln(\text{WAGE}) = \beta_0 + \beta_1 D + \varepsilon$$

Then:

$$\begin{aligned} \text{WAGE}_1 &= \exp(\beta_0 + \beta_1), \quad \text{WAGE}_0 = \exp(\beta_0) \\ \Rightarrow \frac{\text{WAGE}_1}{\text{WAGE}_0} &= \exp(\beta_1) \Rightarrow \% \text{ difference} = \exp(\beta_1) - 1 \end{aligned}$$

Interpretation: Changing from 0 to 1 in the dummy variable is associated with a $(\exp(\beta_1) - 1) \cdot 100\%$ difference in wages.

Example: If $\beta_1 = -0.230$, then:

$$\exp(-0.230) - 1 \approx -0.206 \quad \Rightarrow \quad \text{Wages are 20.6\% lower}$$

3. Summary Table

Variable Type	Interpretation	Formula
Continuous (EDUC, EXPER)	Approx. % change per unit increase	$\beta \cdot 100$
Dummy (FEMALE, UNION)	% difference from base group	$(\exp(\beta) - 1) \cdot 100$

1. Contemporaneous Exogeneity (Weakest)

$$\mathbb{E}[\varepsilon_{it} \mid X_{it}] = 0 \quad \text{for } t = 1, \dots, T$$

- Error term is uncorrelated with the explanatory variables from the same time period.
- Assumption used in standard cross-sectional regressions.
- Allows for correlation with past or future values of the regressors.

2. Sequential Exogeneity (Intermediate)

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, X_{i2}, \dots, X_{it}] = 0 \quad \text{for } t = 1, \dots, T$$

- Error term is uncorrelated with current and all past regressors.
- Stronger than contemporaneous exogeneity.
- Appropriate when past X 's influence current outcomes, but future values do not respond to current shocks.
- Implies contemporaneous exogeneity.

3. Strict Exogeneity (Strongest)

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, X_{i2}, \dots, X_{iT}] = 0 \quad \text{for } t = 1, \dots, T$$

- Error term is uncorrelated with regressors from all time periods (past, present, and future).
- Strongest assumption; it rules out feedback from shocks to future values of X .
- Implies both sequential and contemporaneous exogeneity.
- Often unrealistic in economic applications with dynamic behavior.

4. Interpretation Under Sequential Exogeneity

Assume the model is:

$$Y_{it} = X'_{it}\beta + \varepsilon_{it}$$

If we assume sequential exogeneity, then:

$$\mathbb{E}[Y_{it} \mid X_{i1}, \dots, X_{it}] = X'_{it}\beta$$

That is, current and past values of X are sufficient to predict Y_{it} ; future values are not needed.

5. Summary Comparison

Assumption	Mathematical Form	Future X ?	Lagged X ?	Strength
Contemporaneous	$\mathbb{E}[\varepsilon_{it} \mid X_{it}] = 0$	Yes	No	Weakest
Sequential	$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{it}] = 0$	Yes	Yes	Intermediate
Strict	$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}] = 0$	No	Yes	Strongest

1. What Is Pooled OLS?

Pooled OLS treats panel data as if it is a simple cross-sectional regression by *stacking* all observations across time and individuals and estimating one large OLS regression:

$$Y_{it} = X'_{it}\beta + \varepsilon_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, n$$

The same β is assumed to apply across all individuals and time periods.

2. Example Setup

Suppose we have:

- $T = 3$ years: 1990, 1991, 1992
- Regressors: COMPUTER use, EDUCation, EXperience, FEMALE dummy
- Year dummies: D91, D92 to control for year effects

The regression model is:

$$\ln(\text{WAGE}_{it}) = \delta_0 + \delta_1 D91_t + \delta_2 D92_t + \beta_1 \text{COMPUTER}_{it} + \beta_2 \text{EDUC}_{it} + \beta_3 \text{EXPER}_{it} + \beta_4 \text{FEMALE}_{it} + \varepsilon_{it}$$

This is a **static model** — no lagged variables, no dynamics.

3. Estimating β : Pooled OLS

The OLS estimator is:

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

Where:

- X : stacked regressor matrix
- Y : stacked outcome variable

Alternatively, in panel structure:

$$\hat{\beta}_{\text{pooled}} = \left(\frac{1}{n} \sum_{i=1}^n X_i' X_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i' Y_i \right)$$

Where X_i is a $T \times K$ matrix and Y_i a $T \times 1$ vector for individual i .

4. Key Assumption

To be valid, pooled OLS requires:

$$\mathbb{E}[X_{it}\varepsilon_{it}] = 0$$

This is known as **contemporaneous exogeneity**. It rules out correlation between the regressor and the error term in the same period.

5. Limitations of Pooled OLS

- Ignores individual-specific unobserved heterogeneity
- Assumes no correlation between regressors and individual effects
- OLS standard errors can be misleading due to serial correlation within individuals

6. Alternatives

To address these issues, use panel-specific methods:

1. First-differencing
2. Fixed effects (within estimator)
3. Random effects (GLS)

7. Summary

Concept	Explanation
Pooled OLS	Runs OLS on stacked panel data
Key Assumption	$\mathbb{E}[X_{it}\varepsilon_{it}] = 0$
Limitation	Ignores time correlation within individuals
Usefulness	Simple, best when heterogeneity is small or controlled
Alternatives	Fixed effects, random effects, first differencing

The regression model shown is:

$$\ln(\text{WAGE}_{it}) = \delta_0 + \delta_1 D91_t + \delta_2 D92_t + \beta_1 \text{COMPUTER}_{it} + \beta_2 \text{EDUC}_{it} + \beta_3 \text{EXPER}_{it} + \beta_4 \text{FEMALE}_{it} + \varepsilon_{it}$$

This specification includes year dummies for 1991 and 1992, and assumes the slope coefficients β_1, β_2, \dots are constant over time.

Why Are There No Interaction Terms?

1. Homogeneous Effect Assumption

The model assumes that the effects of explanatory variables are the same in each year:

$$\frac{\partial \ln(\text{WAGE}_{it})}{\partial \text{EDUC}_{it}} = \beta_2 \quad \text{for all } t$$

Including interaction terms like $D91_t \cdot \text{EDUC}_{it}$ would allow slope coefficients to vary by year, but this is not necessary unless we have evidence that such effects differ across time.

2. Time Effects Are Captured by Intercepts

The dummies $D91_t$ and $D92_t$ control for time-specific shifts in wages, such as macroeconomic conditions or inflation, that affect all individuals equally. These shifts are captured by differences in the intercept:

Year-specific effects: δ_1 and δ_2

3. Parsimony and Precision

Including unnecessary interaction terms increases model complexity and reduces estimation precision. If the coefficients are truly constant across time, omitting the interactions yields more efficient estimates.

4. No Theoretical Justification for Varying Slopes

Unless theory or prior evidence suggests that the effect of a regressor (e.g., education) varies across years, we do not include interaction terms.

Summary Table

Reason	Explanation
Homogeneous effects	Assumes same slope β across all years
Time controls via dummies	Year-specific intercepts absorb common shocks
Model simplicity	Avoids unnecessary complexity and multicollinearity
No strong theory for interactions	No need to test for year-specific slopes

Unobserved Heterogeneity

In panel data models, a central concern is **unobserved heterogeneity**, which refers to unmeasured individual-specific characteristics that may bias estimation results.

Let C_i denote unobserved characteristics for individual i , such as ability or motivation. If C_i is related to regressors, omitting it leads to bias.

Population Objective

We are interested in the conditional expectation:

$$\mathbb{E}[Y_{it} \mid X_{i1}, X_{i2}, \dots, X_{iT}, C_i]$$

but unfortunately, C_i is not observed.

Linear Unobserved Effects Model

We often specify the model as:

$$Y_{it} = \beta_0 + X'_{it}\beta + C_i + \varepsilon_{it}$$

where:

- X_{it} : a $K \times 1$ vector of explanatory variables
- C_i : time-invariant, individual-specific unobservable effect
- ε_{it} : time-varying idiosyncratic error

We assume:

$$\mathbb{E}[\varepsilon_{it} \mid X_{it}] = 0 \quad \text{for } t = 1, \dots, T$$

However, if:

$$\mathbb{E}[X_{it}C_i] \neq 0$$

then X_{it} is endogenous, and pooled OLS will be **inconsistent** due to omitted variable bias.

Solutions to Unobserved Heterogeneity

Proxy Variable Use an observable variable (e.g., IQ) to capture the influence of C_i . This is rarely practical.

Instrumental Variables Use instruments that are correlated with X_{it} but uncorrelated with C_i . Validity of instruments is crucial.

Panel Data Solutions Panel data methods offer powerful tools to eliminate or control for C_i :

- **First Differences:**

$$\Delta Y_{it} = \Delta X'_{it}\beta + \Delta \varepsilon_{it}$$

Since C_i is time-invariant, it disappears under differencing.

- **Fixed Effects (Within Estimator):**

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

This removes C_i by demeaning each individual's observations.

- **Random Effects:** Assumes $C_i \sim \text{iid}$ and uncorrelated with X_{it} , i.e.:

$$\mathbb{E}[X_{it}C_i] = 0$$

Provides more efficient estimates than FE under stronger assumptions.

First Differencing

Suppose we have a panel with two periods $T = 2$, and we consider the linear unobserved effects model:

$$Y_{it} = \beta_0 + X'_{it}\beta + C_i + \varepsilon_{it}$$

Define the first differences:

$$\Delta Y_i = Y_{i2} - Y_{i1}$$

$$\Delta X_i = X_{i2} - X_{i1}$$

$$\Delta \varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1}$$

Now if we difference the model:

$$\begin{aligned} \Delta Y_i &= Y_{i2} - Y_{i1} \\ &= (\beta_0 + X'_{i2}\beta + C_i + \varepsilon_{i2}) - (\beta_0 + X'_{i1}\beta + C_i + \varepsilon_{i1}) \\ &= (X_{i2} - X_{i1})'\beta + (\varepsilon_{i2} - \varepsilon_{i1}) \\ &= \Delta X'_i\beta + \Delta \varepsilon_i \end{aligned}$$

Key insight: The time-invariant component C_i is removed through differencing.

When is First Differencing Valid?

To estimate β consistently, the following conditions must hold:

1. **Random sampling across individuals i .**
2. **Orthogonality (Strict Exogeneity):**

$$\mathbb{E}[\Delta X'_i \Delta \varepsilon_i] = 0$$

That is,

$$\mathbb{E}[(X_{i2} - X_{i1})'(\varepsilon_{i2} - \varepsilon_{i1})] = 0$$

Expanding:

$$\mathbb{E}[X'_{i2}\varepsilon_{i2}] + \mathbb{E}[X'_{i1}\varepsilon_{i1}] - \mathbb{E}[X'_{i2}\varepsilon_{i1}] - \mathbb{E}[X'_{i1}\varepsilon_{i2}] = 0$$

To ensure this, we require:

- **Contemporaneous exogeneity:** $\mathbb{E}[X'_{it}\varepsilon_{it}] = 0$

- **No serial correlation between X and ε :** $\mathbb{E}[X'_{i2}\varepsilon_{i1}] = \mathbb{E}[X'_{i1}\varepsilon_{i2}] = 0$

Combined, this implies **strict exogeneity**:

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, X_{i2}] = 0 \quad \text{for } t = 1, 2$$

3. Full rank condition:

$$\text{Rank}(\mathbb{E}[\Delta X_i \Delta X'_i]) = K$$

This ensures no perfect multicollinearity after differencing.

First-Differencing and Intercepts

Suppose we have a model with time-specific intercepts ($T = 2$) where $D2_t$ is an indicator variable for period $t = 2$:

$$Y_{it} = \beta_0 + \beta_1 D2_t + \beta_2 x_{it} + C_i + \varepsilon_{it}$$

After differencing, the second-period intercept becomes the new intercept in the differenced model:

$$\Delta Y_i = \beta_1 + \beta_2 \Delta x_i + \Delta \varepsilon_i$$

Why? Because:

$$\Delta D2_t = D2_2 - D2_1 = 1 - 0 = 1$$

So the difference contributes $\beta_1 \cdot 1 = \beta_1$, which appears as the new intercept.

If we are differencing with $T > 2$, it is generally easier to add the $T - 1$ period-specific intercepts (time dummies) *after* differencing, since the differencing transformation eliminates one time period.

Fixed Effects vs. Random Effects

We begin with the linear unobserved effects model:

$$Y_{it} = X'_{it}\beta + C_i + \varepsilon_{it}, \quad t = 1, 2, \dots, T$$

where:

- X_{it} is a $K \times 1$ vector of observed variables
- C_i is a scalar individual-specific unobserved effect
- ε_{it} is a time-varying error term

Interpretation of C_i

The term C_i is also referred to as:

- Latent variable

- Fixed effect
- Unobserved heterogeneity
- Individual effect

Traditional Definitions

- **Fixed Effects:** C_i is treated as a parameter to estimate for each i . It may be correlated with X_{it} .
- **Random Effects:** C_i is modeled as a random variable and assumed to be uncorrelated with X_{it} :

$$\text{Cov}(X_{it}, C_i) = 0 \quad \text{for all } t, \quad \text{or } \mathbb{E}[C_i | X_{i1}, \dots, X_{iT}] = \mathbb{E}[C_i]$$

Modern Clarification

- Random effects assume no correlation between C_i and the regressors.
- Fixed effects *do not* require C_i to be fixed (non-random), but allow arbitrary correlation between C_i and X_{it} .

Strict Exogeneity (Conditional)

When C_i is included in the model, we redefine strict exogeneity conditional on C_i :

$$\mathbb{E}[Y_{it} | X_{i1}, \dots, X_{iT}, C_i] = X'_{it}\beta + C_i$$

This implies:

“Regressors X_{it} are strictly exogenous conditional on C_i ”

Compare to Previous Assumption (Without C_i)

Previously, strict exogeneity was:

$$\mathbb{E}[Y_{it} | X_{i1}, \dots, X_{iT}] = X'_{it}\beta$$

But when C_i is correlated with X_{it} , this breaks down:

$$\mathbb{E}[Y_{it} | X_{i1}, \dots, X_{iT}] = X'_{it}\beta + \mathbb{E}[C_i | X_{i1}, \dots, X_{iT}]$$

Thus, we must condition on C_i for strict exogeneity to hold.

Pooled OLS Revisited

Under certain assumptions, we can still use pooled OLS to estimate the linear model with unobserved effects. Rewrite the unobserved effects model as:

$$Y_{it} = \beta_0 + X'_{it}\beta + C_i + \varepsilon_{it}$$

Define the composite error term:

$$\nu_{it} = C_i + \varepsilon_{it} \quad \Rightarrow \quad Y_{it} = \beta_0 + X'_{it}\beta + \nu_{it}$$

When is Pooled OLS Consistent?

Pooled OLS will yield consistent estimates of β if:

$$\mathbb{E}[X_{it}\nu_{it}] = 0 \quad \text{for all } t = 1, \dots, T$$

This requires two conditions:

$$\mathbb{E}[X_{it}\varepsilon_{it}] = 0 \tag{1}$$

$$\tag{1}$$

$$\mathbb{E}[X_{it}C_i] = 0 \tag{2}$$

$$\tag{2}$$

Equation (1) requires no contemporaneous correlation between regressors and idiosyncratic shocks. Equation (2) rules out omitted variable bias due to correlation between regressors and unobserved effects.

Interpretation and Limitations

- Assumption (2) is quite strong and often unrealistic.
- For example, in a model where Y_{it-1} is included as a regressor, it will likely be correlated with C_i , violating assumption (2).
- Therefore, pooled OLS would be inappropriate in such dynamic models.

Serial Correlation of Errors

Even when assumptions (1) and (2) hold, the composite error ν_{it} contains C_i , which is constant over time. Thus, for any $t \neq s$:

$$\text{Cov}(\nu_{it}, \nu_{is}) \neq 0$$

This means that ν_{it} is serially correlated. To address this, we recommend using cluster-robust standard errors when estimating with pooled OLS.

Asymptotics Matter

Consider two asymptotic frameworks:

- **Fixed T , large N :** Traditional panel data asymptotics. Pooled OLS may be consistent under assumptions (1) and (2).
- **Large T , fixed N :** Time-series asymptotics. Serial correlation in ν_{it} may persist as $T \rightarrow \infty$, violating assumptions of weak dependence.

Conclusion

Pooled OLS can be used under:

- $\mathbb{E}[X_{it}\varepsilon_{it}] = 0$
- $\mathbb{E}[X_{it}C_i] = 0$

However:

- Serial correlation in ν_{it} requires robust standard errors.
- Dynamic models often violate these assumptions.
- Consider asymptotic properties carefully when interpreting pooled OLS estimates.

Random Effects Estimation

We consider the standard panel data model with unobserved individual effects:

$$Y_{it} = X'_{it}\beta + C_i + \varepsilon_{it}$$

Define the composite error term:

$$\nu_{it} = C_i + \varepsilon_{it} \quad \Rightarrow \quad Y_{it} = X'_{it}\beta + \nu_{it}$$

Random effects estimation treats C_i as part of the error term but imposes additional assumptions to allow for efficient estimation via Generalized Least Squares (GLS).

Assumption RE.1

(a) **Strict Exogeneity (Conditional on C_i):**

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}, C_i] = 0 \quad \text{for } t = 1, \dots, T$$

(b) **Independence Between C_i and X_{it} :**

$$\mathbb{E}[C_i \mid X_{i1}, \dots, X_{iT}] = \mathbb{E}[C_i]$$

Assumption (a) is equivalent to strict exogeneity used in fixed effects models. Assumption (b) is stronger and assumes that C_i is mean independent of all regressors across time. It implies that knowing any X_{it} does not help predict C_i .

Why Random Effects?

RE estimation is more efficient than Fixed Effects (FE) when assumptions hold. It explicitly models the serial correlation in the composite error caused by the individual-specific effect C_i .

Stacked Formulation (over T periods)

Define:

$$Y_i = X_i\beta + \nu_i, \quad \text{where} \quad \nu_i = C_i \cdot \mathbf{1}_T + \varepsilon_i$$

Here:

- $\mathbf{1}_T$ is a $T \times 1$ vector of ones
- ε_i is a $T \times 1$ vector of idiosyncratic errors

Then the unconditional variance of the composite error is:

$$\Omega = \mathbb{E}[\nu_i \nu_i']$$

Properties of Ω :

- Ω is a $T \times T$ positive definite matrix
- It is the same for all individuals i , assuming iid sampling across i
- It reflects the structure induced by the presence of C_i

Covariance Matrix for Random Effects

Consider the random effects model:

$$Y_{it} = X'_{it}\beta + C_i + \varepsilon_{it}$$

where:

- $C_i \sim \text{IID}(0, \sigma_C^2)$ is the unobserved individual-specific effect,
- $\varepsilon_{it} \sim \text{IID}(0, \sigma_\varepsilon^2)$ is the idiosyncratic error,
- and C_i is independent of ε_{it} .

Stacking the model over time for individual i :

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \boldsymbol{\nu}_i, \quad \text{where} \quad \boldsymbol{\nu}_i = C_i \cdot \mathbf{1}_T + \boldsymbol{\varepsilon}_i$$

We are interested in the variance of $\boldsymbol{\nu}_i$, the composite error term. Define:

$$\Omega = \mathbb{E}[\boldsymbol{\nu}_i \boldsymbol{\nu}_i']$$

Under the assumptions:

$$\begin{aligned} \mathbb{E}[\varepsilon_{it}] &= 0, & \mathbb{E}[\varepsilon_{it}^2] &= \sigma_\varepsilon^2, \\ \mathbb{E}[C_i] &= 0, & \mathbb{E}[C_i^2] &= \sigma_C^2, \\ \mathbb{E}[\varepsilon_{it}\varepsilon_{is}] &= 0 \text{ for } t \neq s, \\ \mathbb{E}[C_i\varepsilon_{it}] &= 0 \text{ (independence),} \end{aligned}$$

Then for the composite error:

$$\nu_{it} = C_i + \varepsilon_{it}$$

We derive the elements of the covariance matrix:

- Diagonal elements (variance):

$$\mathbb{E}[\nu_{it}^2] = \mathbb{E}[(C_i + \varepsilon_{it})^2] = \sigma_C^2 + \sigma_\varepsilon^2$$

- Off-diagonal elements (covariance across time):

$$\mathbb{E}[\nu_{it}\nu_{is}] = \mathbb{E}[(C_i + \varepsilon_{it})(C_i + \varepsilon_{is})] = \sigma_C^2 \quad \text{for } t \neq s$$

Thus, the full $T \times T$ covariance matrix of $\boldsymbol{\nu}_i$ is:

$$\boxed{\Omega = \sigma_\varepsilon^2 I_T + \sigma_C^2 \mathbf{1}_T \mathbf{1}_T'}$$

Derivation of the Random Effects Covariance Matrix

We begin with the random effects model:

$$\nu_{it} = C_i + \varepsilon_{it}$$

Where:

- $C_i \sim \text{iid}(0, \sigma_C^2)$: individual-specific (time-invariant) random effect
- $\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2)$: idiosyncratic error
- $C_i \perp \varepsilon_{it}$ for all t

Stack over time $t = 1, \dots, T$ for individual i :

$$\boldsymbol{\nu}_i = \begin{bmatrix} \nu_{i1} \\ \nu_{i2} \\ \vdots \\ \nu_{iT} \end{bmatrix} = C_i \cdot \mathbf{1}_T + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} = C_i \cdot \mathbf{1}_T + \boldsymbol{\varepsilon}_i$$

We want to derive the covariance matrix:

$$\Omega = \mathbb{E}[\boldsymbol{\nu}_i \boldsymbol{\nu}_i']$$

Substitute the composite error:

$$\Omega = \mathbb{E}[(C_i \mathbf{1}_T + \boldsymbol{\varepsilon}_i)(C_i \mathbf{1}_T + \boldsymbol{\varepsilon}_i)']$$

Using the identity $(a + b)(a + b)' = aa' + ab' + ba' + bb'$:

$$\Omega = \mathbb{E}[C_i^2 \mathbf{1}_T \mathbf{1}_T'] + \mathbb{E}[C_i \mathbf{1}_T \boldsymbol{\varepsilon}_i'] + \mathbb{E}[\boldsymbol{\varepsilon}_i C_i \mathbf{1}_T'] + \mathbb{E}[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i']$$

Apply the assumptions:

$$\begin{aligned}\mathbb{E}[C_i^2] &= \sigma_C^2 \\ \mathbb{E}[C_i \boldsymbol{\varepsilon}'_i] &= \mathbf{0} \\ \mathbb{E}[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i] &= \sigma_\varepsilon^2 I_T\end{aligned}$$

So:

$$\Omega = \sigma_C^2 \mathbf{1}_T \mathbf{1}'_T + \sigma_\varepsilon^2 I_T$$

$$\Omega = \sigma_\varepsilon^2 I_T + \sigma_C^2 \mathbf{1}_T \mathbf{1}'_T$$

Interpretation

- $\sigma_\varepsilon^2 I_T$: reflects uncorrelated individual-time shocks.
- $\sigma_C^2 \mathbf{1}_T \mathbf{1}'_T$: adds equal covariance across time for the same individual, due to C_i .

This structure captures the serial correlation of errors in panel data induced by unobserved individual heterogeneity C_i . It is central to the random effects model and allows feasible GLS estimation.

Feasible GLS Estimation of the Random Effects Model

We begin with the random effects model:

$$Y_{it} = X'_{it} \beta + C_i + \varepsilon_{it}$$

Define the composite error term:

$$\nu_{it} = C_i + \varepsilon_{it}$$

Stacked form:

$$Y_i = X_i \beta + \nu_i, \quad \text{with } \nu_i = C_i \cdot \mathbf{1}_T + \varepsilon_i$$

Goal: Feasible GLS Estimator

The Generalized Least Squares (GLS) estimator is:

$$\hat{\beta}_{\text{GLS}} = \left(\sum_{i=1}^n X'_i \Omega^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X'_i \Omega^{-1} Y_i \right)$$

This is infeasible because $\Omega = \mathbb{E}[\nu_i \nu'_i]$ is unknown.

Assumptions (RE.3: Homoskedasticity)

- $\mathbb{E}[\varepsilon_i \varepsilon'_i \mid X_i, C_i] = \sigma_\varepsilon^2 I_T$
- $\mathbb{E}[C_i^2 \mid X_i] = \sigma_C^2$

Under these assumptions, the covariance matrix has the form:

$$\Omega = \sigma_\varepsilon^2 I_T + \sigma_C^2 \mathbf{1}_T \mathbf{1}_T'$$

Step-by-Step Estimation (Feasible GLS)

1. Estimate β using pooled OLS:

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

2. Compute residuals:

$$\hat{\nu}_{it} = Y_{it} - X'_{it}\tilde{\beta}$$

3. Estimate total variance of residuals (within + between):

$$\hat{\sigma}_\nu^2 = \frac{1}{nT - K} \sum_{i=1}^n \sum_{t=1}^T \hat{\nu}_{it}^2$$

4. Estimate cross-time component ($\hat{\sigma}_C^2$):

$$\hat{\sigma}_C^2 = \frac{1}{nT(T-1)/2 - K} \sum_{i=1}^n \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{\nu}_{it} \hat{\nu}_{is}$$

5. Solve for $\hat{\sigma}_\varepsilon^2$:

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\nu^2 - \hat{\sigma}_C^2$$

6. Form estimated $\hat{\Omega}$:

$$\hat{\Omega} = \hat{\sigma}_\varepsilon^2 I_T + \hat{\sigma}_C^2 \mathbf{1}_T \mathbf{1}_T'$$

7. Estimate β using FGLS:

$$\hat{\beta}_{\text{RE}} = \left(\sum_{i=1}^n X'_i \hat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^n X'_i \hat{\Omega}^{-1} Y_i \right)$$

This yields a consistent and efficient estimate of β under the random effects assumptions.

Within Transformation and Fixed Effects Estimator

To eliminate the unobserved individual effect C_i , the Fixed Effects (FE) estimator relies on the idea of differencing out C_i within an individual over time.

Step 1: Time Averaging (Over T periods)

We start with the panel data model:

$$Y_{it} = X'_{it}\beta + C_i + \varepsilon_{it}$$

Average over time (for each individual i):

$$\begin{aligned}\bar{Y}_i &= \frac{1}{T} \sum_{t=1}^T Y_{it} = \frac{1}{T} \sum_{t=1}^T (X'_{it}\beta + C_i + \varepsilon_{it}) \\ \Rightarrow \bar{Y}_i &= \bar{X}'_i\beta + C_i + \bar{\varepsilon}_i\end{aligned}$$

This is the ****averaged model****:

$$\bar{Y}_i = \bar{X}'_i\beta + C_i + \bar{\varepsilon}_i \quad (1.2)$$

Step 2: Subtracting (Demeaning)

Now subtract the averaged model from the original model:

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

This gives the ****within-transformed model****:

$$\tilde{Y}_{it} = \tilde{X}'_{it}\beta + \tilde{\varepsilon}_{it} \quad (1.3)$$

where:

$$\begin{aligned}\tilde{Y}_{it} &= Y_{it} - \bar{Y}_i \\ \tilde{X}_{it} &= X_{it} - \bar{X}_i \\ \tilde{\varepsilon}_{it} &= \varepsilon_{it} - \bar{\varepsilon}_i\end{aligned}$$

Interpretation

- C_i drops out because it does not vary over time.
- The estimator uses only ****within-individual variation**** in X_{it} .
- Time-invariant regressors are eliminated.

Estimation

We can now run OLS on the transformed equation:

$$\tilde{Y}_{it} = \tilde{X}'_{it}\beta + \tilde{\varepsilon}_{it}$$

This is called the ****Fixed Effects (within) estimator****. It provides a consistent estimate of β even when C_i is correlated with X_{it} .

Fixed Effects Estimator via Within Transformation

We begin with the panel data model:

$$Y_{it} = X'_{it}\beta + C_i + \varepsilon_{it}, \quad t = 1, \dots, T$$

where:

- X_{it} : time-varying regressors,
- C_i : unobserved individual-specific effect,
- ε_{it} : idiosyncratic error.

Step 1: Averaging Over Time

We average the model over time for each individual:

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}, \quad \bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}, \quad \bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}$$

Substituting into the model:

$$\bar{Y}_i = \bar{X}_i' \beta + C_i + \bar{\varepsilon}_i \quad (1.2)$$

Step 2: Subtracting (Within Transformation)

Now subtract equation (1.2) from the original model:

$$\begin{aligned} Y_{it} - \bar{Y}_i &= (X_{it} - \bar{X}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \\ \Rightarrow \tilde{Y}_{it} &= \tilde{X}_{it}' \beta + \tilde{\varepsilon}_{it} \end{aligned} \quad (1.3)$$

where:

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i, \quad \tilde{X}_{it} = X_{it} - \bar{X}_i, \quad \tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$$

This transformation eliminates the unobserved effect C_i from the model.

Implications

- The fixed effects estimator uses only within-individual variation.
- It allows for C_i to be arbitrarily correlated with X_{it} .
- It removes time-invariant variables from the model since their deviations are always zero.

Estimation

We can now estimate β by applying OLS to the within-transformed model:

$$\tilde{Y}_{it} = \tilde{X}_{it}' \beta + \tilde{\varepsilon}_{it}$$

This is the **Fixed Effects (Within) Estimator**.

Consistency of the Fixed Effects Estimator

After performing the within transformation, we arrive at the demeaned model:

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \ddot{\varepsilon}_{it}, \quad t = 1, \dots, T$$

where:

$$\begin{aligned}\ddot{Y}_{it} &= Y_{it} - \bar{Y}_i \\ \ddot{X}_{it} &= X_{it} - \bar{X}_i \\ \ddot{\varepsilon}_{it} &= \varepsilon_{it} - \bar{\varepsilon}_i\end{aligned}$$

This transformation removes the individual fixed effect C_i , allowing us to estimate β using OLS.

Key Condition for Consistency

To ensure consistency of the fixed effects estimator, we require the orthogonality condition:

$$\mathbb{E}[\ddot{X}_{it}\ddot{\varepsilon}_{it}] = 0$$

which expands to:

$$\mathbb{E}[(X_{it} - \bar{X}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0$$

Why This Holds

Under Assumption FE.1 (Strict Exogeneity):

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}, C_i] = 0 \quad \text{for all } t$$

This implies:

$$\mathbb{E}[\varepsilon_{it} \mid X_{i1}, \dots, X_{iT}] = 0 \quad \text{and} \quad \mathbb{E}[\bar{\varepsilon}_i \mid X_{i1}, \dots, X_{iT}] = 0$$

Thus:

$$\mathbb{E}[\ddot{\varepsilon}_{it} \mid \ddot{X}_{i1}, \dots, \ddot{X}_{iT}] = 0 \quad \Rightarrow \quad \mathbb{E}[\ddot{X}_{it}\ddot{\varepsilon}_{it}] = 0$$

Conclusion

This orthogonality ensures that:

- OLS on the within-transformed model yields a consistent estimator of β .
- The Fixed Effects estimator remains unbiased under strict exogeneity and appropriate sample conditions.

Identification and Efficiency of the Fixed Effects Estimator

Assumption FE.2 (Rank Condition).

$$\text{rank} \left(\sum_{t=1}^T \mathbb{E}[\ddot{X}_{it}\ddot{X}_{it}'] \right) = \text{rank} \left(\mathbb{E}[\ddot{X}_i'\ddot{X}_i] \right) = K$$

This assumption ensures that the regressors \ddot{X}_{it} (demeaned regressors) have sufficient variation over time. If any component of X_{it} is time-invariant across all t , then:

$$\ddot{X}_{it} = X_{it} - \bar{X}_i = 0$$

for all i , and such regressors cannot be identified in the model.

FE Estimator via OLS on Demeaned Data. The Fixed Effects estimator can now be written as:

$$\hat{\beta}_{\text{FE}} = \left(\sum_{i=1}^n \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^n \ddot{X}_i' \ddot{Y}_i \right)$$

This is the OLS estimator using the within-transformed (demeaned) data, where:

$$\ddot{X}_i = \text{stack of } \ddot{X}_{it} = X_{it} - \bar{X}_i \text{ for all } t$$

$$\ddot{Y}_i = \text{stack of } \ddot{Y}_{it} = Y_{it} - \bar{Y}_i \text{ for all } t$$

Even though C_i is removed by demeaning, the interpretation of β is still consistent with the original structural model.

Assumption FE.3 (Homoskedasticity and No Serial Correlation).

$$\mathbb{E}[\varepsilon_i \varepsilon_i' \mid X_i, C_i] = \sigma_\varepsilon^2 I_T$$

Under this assumption:

- The error terms are homoskedastic (constant variance over time and individuals).
- The errors are serially uncorrelated.
- The variance does not depend on X_i or C_i .

If FE.3 holds, then $\hat{\beta}_{\text{FE}}$ is not only consistent, but also **efficient**.

3 Stationarity

The sample $\{Y_1, \dots, Y_T\}$ is a realization of a dependent stochastic process. To analyze this process, we often consider the marginal distributions of Y_t and the joint distributions of $(Y_t, Y_{t+1}, \dots, Y_{t+l})$.

To make valid statistical inferences over time, we typically require some form of constancy in these distributions. This leads to the concept of **stationarity**.

Covariance Stationarity (Weak Stationarity)

Definition. A stochastic process $\{Y_t\}$ is said to be *covariance stationary* if:

1. $\mu = \mathbb{E}[Y_t]$ is constant over time.
2. $\sigma^2 = \text{Var}(Y_t) = \mathbb{E}[(Y_t - \mu)^2]$ is constant over time.
3. The autocovariance $\gamma(k) = \text{Cov}(Y_t, Y_{t-k})$ depends only on the lag k , not on t .

This type of stationarity is also known as *weak stationarity*.
Here:

- $\gamma(k)$: k -th order autocovariance
- $\rho(k) = \text{Corr}(Y_t, Y_{t-k})$: k -th order autocorrelation

Strict Stationarity

A stochastic process $\{Y_t\}$ is *strictly stationary* if the joint distribution of (Y_t, \dots, Y_{t+l}) is identical for all t and any l .

Definition. A process $\{Z_t\}$ is strictly stationary if, for any finite integer r , and any set of subscript indices t_1, t_2, \dots, t_r , the joint distribution of:

$$(Z_{t_1}, Z_{t_2}, \dots, Z_{t_r})$$

depends only on $t_1 - t, t_2 - t, \dots, t_r - t$, not on the time index t itself.

Comparison

Property	Covariance Stationarity	Strict Stationarity
Mean Constant?	Yes	Yes
Variance Constant?	Yes	Yes
Autocovariance Depends Only on Lag?	Yes	Yes
Joint Distribution Time-Invariant?	No	Yes
Implied by Strict Stationarity?	Yes	—
Implies Strict Stationarity?	Not necessarily	—

Further Concepts in Stationarity

Relative Time vs Absolute Time

In a strictly stationary process, only the *relative time* or spacing between observations matters:

$$(t_1 - t, t_2 - t, \dots, t_r - t)$$

not the absolute time values t .

Example:

$$\begin{aligned} r = 1, \quad t_1 = 5 &\Rightarrow (Z_1, Z_5) \\ t = 12 &\Rightarrow (Z_{12}, Z_{16}) \end{aligned}$$

Both pairs have the same joint distribution.

Theorem: Stationarity Under Transformation

Theorem 11. If Y_t is strictly stationary and

$$X_t = \phi(Y_t, Y_{t-1}, Y_{t-2}, \dots) \in \mathbb{R}^q$$

is a random vector, then X_t is strictly stationary.

Interpretation: Stationarity is preserved under deterministic transformations involving lags of the process. This is particularly useful for linear filters, moving averages, and constructing estimators.

Theoretical Examples: Stationary or Not?

1. $Y_t \sim \text{iid}$: **Stationary (strict)**
2. $Y_t = e_t + \theta e_{t-1}$, with $e_t \sim \text{iid}$, $\mathbb{E}[e_t] = 0$: **Stationary** (MA(1) process)
3. $Y_t = Z$, for some random variable Z : **Strictly stationary**, trivially constant
4. $Y_t = t$: **Non-stationary** (mean grows over time)
5. $Y_t = e_t + \frac{1}{t}e_{t-1}$: **Non-stationary** (coefficient changes with time)
6. $Y_t = Y_{t-1} + e_t$, $Y_0 = 0$: **Non-stationary** (random walk, variance increases with time)
7. $Y_t = (-1)^t$: **Strictly stationary** (periodic process with constant distribution)

Important Clarification from Hansen (2021)

Stationarity means that the distribution is constant over time. It does not, however, imply that the process has no dependence structure or that it lacks periodic patterns.

Interpretation: This time series has a deterministic upward trend. Although the fluctuations appear regular, the mean changes over time, so the series is **non-stationary**.

Ergodicity

Intuition: Ergodicity can be thought of as *asymptotic independence*. It's a minimal assumption needed to guarantee that a law of large numbers (LLN) holds for dependent data.

Heuristic (Hayashi, 2000, p.101):

A stationary process is ergodic if it is asymptotically independent, that is, if any two random variables positioned far apart in the sequence are almost independently distributed.

Recall that if two random variables X and Y are independent, then:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Definition (Hayashi, 2000)

A stationary process $\{Z_t\}$ is **ergodic** if for any two bounded functions

$$f : \mathbb{R}^{k+1} \rightarrow \mathbb{R}, \quad g : \mathbb{R}^{l+1} \rightarrow \mathbb{R},$$

we have:

$$\lim_{s \rightarrow \infty} |\mathbb{E}[f(Z_t, \dots, Z_{t+k}) \cdot g(Z_{t+s}, \dots, Z_{t+s+l})] - \mathbb{E}[f(Z_t, \dots, Z_{t+k})] \cdot \mathbb{E}[g(Z_t, \dots, Z_{t+l})]| = 0.$$

Interpretation: Two variables in the sequence that are far apart behave as if they are independent.

Stationary Ergodic Process

Definition: A process that is both stationary and ergodic is called an **ergodic stationary process** or a **stationary ergodic process**.

Theorem 12 (Ergodic Theorem)

Let $\{Z_t\}$ be a stationary and ergodic process with $\mathbb{E}[Z_1] = \mu$. Then:

$$\bar{Z}_T = \frac{1}{T} \sum_{t=1}^T Z_t \xrightarrow{\text{a.s.}} \mu.$$

Reference: See Theorem 9.5.5 of Karlin and Taylor (1975).

Why It Matters

- Kolmogorov's LLN requires i.i.d. observations.
- The ergodic theorem allows for **serial dependence**, as long as the dependence *dies out over time*.
- This makes it applicable to most stationary time series models used in econometrics.