Panel Data models

Prepared & presented by Prof. Guechari Yasmina

1. Introduction to Panel Data

- **Definition:**
  - Panel data consists of observations on **multiple entities** (individuals, firms, countries) **over time**.
  - It combines elements of both **cross-sectional** and **time-series** data.
- **Example Structure:**
  - A dataset that tracks **GDP growth** of multiple countries (5) over years (20years).
  - Follows the same individuals' **income and spending habits** for five years.

2. Key Characteristics of Panel Data

- Observations are collected for **the same units** over multiple time periods.
  ✓ Allows tracking **changes over time** within each entity.
  ✓ Two types:
- **Balanced Panel:** Data available for all entities across all time periods.
- **Unbalanced Panel:** Some entities have missing observations in certain periods.

3. Types of Data in Econometrics

Type	Description	Example
Cross-Sectional Data	Data collected at a single point in time	Income levels of 1,000 individuals in 2024
Time-Series Data	Data collected over time for one entity	Inflation rate of a country from 2000 to 2023
Panel Data	Data collected over time for multiple entities	Income levels of 1,000 individuals from 2010-2024

4. Examples of Panel Data in Economic Research

- **Labor Economics:** Wage of individuals over time.
- **Macroeconomics:** Economic growth across countries for several decades.
- **Corporate Finance:** Financial performance of firms across different years.

5. Panel data analysis

- There are two main models
  - Fixed Effects Model (FEM)
  - Random Effects Model (REM)
- Estimate the model using software like Eviews

6. Fixed Effects Model (FEM)

- **What is Fixed Effects Model (FEM)?**
- FEM is a way to analyze data where we observe the **same individuals, companies, or countries over time**.
- The goal is to **measure the effect of some variables** (like education, experience, investment, etc.) on an outcome (like productivity, salary, GDP, profit, etc.) **while accounting for things that don’t change over time** for each individual or entity.

6. Fixed Effects Model (FEM)

- **Fixed Effects Model controls for all characteristics that do not change over time** (for each individual, country, company...) — even if we don’t observe or measure them. But we handle this by giving **each entity its own "starting point" (intercept)** in the model.

6. Fixed Effects Model (FEM)

- **Mathematical Representation**
- $\boldsymbol{Y_{it} = \alpha_i + \beta X_{it} + \mu_{it}}$ where:
- α_i = **individual-specific intercept**, capturing unobserved characteristics.
- **Key Feature:** FEM allows **each entity to have a unique intercept**.
- Each cross-section effect (intercept) shows how much higher or lower the entity's base outcome is, compared to the average intercept, after controlling for the independent variables.

6. Fixed Effects Model (FEM)

- **Assumptions:**

1. **Unobserved characteristics don’t change over time**
   Example: A person’s natural talent, or a country's location — these don’t change every year.
2. **These unobserved characteristics are related to the variables we’re studying**
   Example: A very talented person (unobserved) may also choose to get more education (observed), so their talent affects education and productivity. We need a model that can **handle that link** — that’s why FEM is used.

6. Fixed Effects Model (FEM)

- **Why FEM is Useful:**
- It helps us **focus only on changes over time** (like how education and experience affects productivity) by removing the **effect of fixed traits** (like personality, family background, or culture) that stay the same over the years.

6. Fixed Effects Model (FEM)

- **Example:** Suppose we follow **5 individuals** over **10 years**, and we collect data on:
- Their **years of education**
- Their **experience**,
- Their **productivity score** (from job evaluations),
- But we **cannot measure** their **personality**, **family background**, or **cultural values**.
- These **unmeasured traits** may affect productivity, **but they don't change over time.**

6. Fixed Effects Model (FEM)

- We suspect:
- People with strong personalities or from supportive families tend to be more productive **regardless** of their education/experience.
- And these traits **also affect their access to better education or job experience.**
- So, to **avoid bias**, we use a Fixed Effects Model to control for these **unchanging personal traits** by giving each person their own **intercept**.

6. Fixed Effects Model (FEM)

- **The FEM Equation (simplified):**
- $Productivity_{it} = \alpha_i + \beta Educ_{it} + \gamma Exper_{it} + \varepsilon_{it}$
- α_i = the fixed effect (the unique intercept for each person — controls for personality, family background, culture)
- β , γ = how much education and experience affect productivity
- *i* = the individual
- *t* = the year
- Assume after estimation, we've got this result:
- $Prod_{it} = 12.2 + 1.8Educ_{it} + 0.6Exper_{it}$

6. Fixed Effects Model (FEM)

- **Interpretation**
- **Intercept (12.2):** is the estimated productivity ($Prod_{it}$) when both education ($Educ_{it}$) and experience ($Exper_{it}$) are zero.

➢ In fixed effects models, we don't focus too much on this number because we care more about how things change over time for the same person in the sample.

- **Coefficient on Education (1.8):** Holding experience constant, a **one-unit increase in education** is associated with an **increase of 1.8 units in productivity** for the same individual in sample over time.
- **Coefficient on Experience (0.6):** Holding education constant, a **one-unit increase in experience** is associated with an **increase of 0.6 units in productivity** for the same individual in the sample over time.

6. Fixed Effects Model (FEM)

- **Interpreting Individual Fixed Effects:** These are the **individual-specific effects** (the α_i in our model) — i.e., the impact of **unmeasured, time-invariant traits** like personality or family background.

Individual	Effect	Interpretation
Individual 1	+2.50	Has unobserved traits that increase productivity by 2.5 points above average.
Individual 2	-1.00	Has traits that reduce productivity by 1 point below average, all else equal.
Individual 5	-1.60	Likely has unfavorable unmeasured traits (e.g., lack of motivation), reducing productivity by 1.6 points.

6. Fixed Effects Model (FEM)

- In a **Fixed Effects Model**, the intercept for each entity (here: individual) is calculated as:

$$Entity - specific\ intercept = average\ intercept + enttity\ effect$$

Example: individual1 intercept=12.2+2.5=14.7

Individual 3 intercept = 12.2– 1.6 =10.6

The interpretation: Holding all else constant, Individual 1 tends to have a value of the dependent variable that is2.5 **units higher than the sample average intercept of 12.5**, resulting in an entity-specific intercept of **14.7, this indicate that individual 1 has productivity equal to 14.5 point when educ and exper both are zero.**

7. Random Effects Model (REM)

- **Definition & Assumptions**
- REM assumes that individual-specific effects (αi) are **randomly distributed** and uncorrelated with explanatory variables. This allows us to estimate **time-invariant variables** (such as gender, region…..), unlike FEM.
- **Mathematical Representation**
- $\boldsymbol{Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \varepsilon_{it}}$
- where:
- α_i is a random variable ε_{it} is the error term.

7. Random Effects Model (REM)

- In the **Random Effects Model (REM)**:
  - The individual-specific effect α_i is treated as:
  - ➢ a **random variable**, not a fixed number.
  - ➢ It is assumed to be **uncorrelated** with the independent variables X_{it}.Because of this assumption, REM allows you to **include and estimate the effect of time-invariant variables** (like gender, region, culture, etc.).

7. Random Effects Model (REM)

- **Example:** You collect data on **100 manufacturing firms** over **5 years** (panel data); and you investigate What factors affect **firm productivity** over time?

- **Dependent variable (Y):** Productivity (measured as output per worker)
- **Independent variables (X):**
  - **Capital investment** (changes over time)
  - **Training hours** (changes over time)
  - **Firm size** (number of employees, changes over time)
  - **Industry type** (e.g., food, textile, electronic)

7. Random Effects Model (REM)

- $Productivity = \beta_0 + \beta_1 cap_{it} + \beta_2 train_{it} + \beta_3 size_{it} + \beta_4 indust_{it} + \alpha_i + \varepsilon_{it}$
- α_i : firm-specific random effect (not observed)
- ε_{it} : regular error term

Since Industry Type is a categorical variable, EViews cannot read it directly. You must create dummy variables:

•Food_dummy = 1 if Industry = Food, 0 otherwise
•Textile_dummy = 1 if Industry = Textile, 0 otherwise
•Electronics_dummy = 1 if Industry = Electronics, 0 otherwise
👉 How to create them:
•Go to Quick → Generate Series

7. Random Effects Model (REM)

Variable	Coefficient (β)	Std. Error	p-value
Intercept (β0)	10.5	1.2	0.000
Capital (β1)	0.8	0.2	0.001
Training (β2)	1.1	0.3	0.000
Firm Size (β3)	0.05	0.01	0.000

Industry Type (β4) Food = base category, Textile = -0.6, Electronics = 1.2 —

7. Random Effects Model (REM)

- **Intercept (10.5):** If a firm has 0 capital investment, 0 training, 0 employees, and is in the food industry, then its expected productivity is 10.5 units.
- **Capital Investment (0.8):** For every 1 unit increase in capital investment, a firm's productivity increases by 0.8 units, holding other factors constant.

✓This means that firms investing more in machinery or tech tend to become more productive.

- **Training Hours (1.1):** For every 1 extra hour of employee training, productivity increases by 1.1 units.

✓Training employees leads to a strong boost in output

7. Random Effects Model (REM)

- **Firm Size (0.05):** Each additional employee is associated with a 0.05 unit increase in productivity.

✓Larger firms are slightly more productive, possibly due to economies of scale.

- **Industry Type:** Textile industry firms produce 0.6 units less than food industry firms, holding everything else constant.

✓Electronics industry firms produce 1.2 units more than food industry firms.

✓Different industries have different productivity levels, likely due to differences in technology or processes.

7. Random Effects Model (REM)

- After finishing the estimation and interpretation, we are going to estimate the random effects for the cross-section in Eviews
- **View → Fixed/Random Effects → Random Effects → Cross-section effects**
- EViews will display the **estimated random effects** α_i for **each cross-section unit** (e.g., each firm).
- If **Firm A** has α=1.2, this means its productivity is, on average, **1.2 units higher** than the average firm **due to unobserved firm characteristics** (e.g., better management, culture, tech).
- If **Firm B** has α=−0.5, it's **0.5 units lower** than average — possibly due to unobserved inefficiencies.
- So these α_i values reflect the **"individual effects"** captured by the model.

8. Choosing between FEM & REM

- The **Hausman test** helps determine whether to use **FEM or REM**:
  - **H0 (Null Hypothesis):** No correlation between individual effects and explanatory variables → **Use REM**.
  - **H1 (Alternative Hypothesis):** Correlation exists → **Use FEM**.
- ✅ **Null hypothesis**: Random Effects is appropriate.
- **If p-value < 0.05** → Reject RE → Use **Fixed Effects**.
- **If p-value > 0.05** → Fail to reject RE → **Random Effects is OK**.

8. Choosing between FEM & REM

Correlated Random Effects - Hausman Test
Equation: Untitled
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	6.849448	3	0.0769

Based on the table above, the p-value is greater than 5%, so we fail to reject the null hypothesis. This means that the Random Effects Model is appropriate for estimating the underlying relationship.

9. Model Diagnostic

- Check for

i. Serial correlation
ii. Heteroskedasticity
iii. Cross-sectional dependence (large panel)
iv. Stationarity (use panel unit root test)

9. Diagnostic check

Test 1: Serial Correlation Test

- In panel data, especially with time series, **autocorrelation** can bias standard errors.
- **Common test**:
  - **Wooldridge test** for autocorrelation in panel data. This test is not available in Eviews.
- **In EViews**: you can check serial correlation After estimation panel data model, follow this steps

Step 1: Click Proc → Make Residual Series → call it, for example, resid1.

- This saves the residuals from the first model.

9. Diagnostic check

- Now you need to **create a lag** of the residuals:
- Go to **Quick → Generate Series**, and enter this code:
- resid1_lag = resid1(-1)
- Quick → Estimate Equation → Type:
- resid1 c resid1_lag
- Look at the **coefficient** on resid1_lag.

9. Diagnostic check

Dependent Variable: RESID01
Method: Panel Least Squares
Date: 04/28/25 Time: 18:31
Sample (adjusted): 2021 2024
Periods included: 4
Cross-sections included: 5
Total panel (balanced) observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.199196	0.244302	0.815368	0.4255
RESID01_LAG	0.663242	0.245473	2.701898	0.0146

R-squared	0.288545	Mean dependent var	0.084574
Adjusted R-squared	0.249019	S.D. dependent var	1.241593
S.E. of regression	1.075954	Akaike info criterion	3.078932
Sum squared resid	20.83819	Schwarz criterion	3.178505
Log likelihood	-28.78932	Hannan-Quinn criter.	3.098370
F-statistic	7.300252	Durbin-Watson stat	1.275467
Prob(F-statistic)	0.014593		

9. Diagnostic check

- If the coefficient of lag resid is **significantly different from zero**, it suggests **serial correlation**.
- You can also check the **F-statistic** and its **p-value**:
  - **Null Hypothesis (H_0)**: No serial correlation (coefficient = 0).
  - **Alternative Hypothesis (H_1)**: Serial correlation exists (coefficient ≠ 0).
- ✅ If p-value < 0.05 → Reject H_0 → **Serial correlation detected.**

Based on the results in the table, If the coefficient of lag resid is **significantly different from zero, and the p-value of f-stat is less than 5% so, we reject the null hypothesis which means there is a serial correlation**

9. Diagnostic check

Test 2: Cross-sectional Dependence Test in Panel Data

- In **panel data**, we often assume that the units (like countries, firms,...) are **independent of each other**.
  But in reality, **they can influence one another**.
- Example: If oil prices rise, it might affect the productivity of **many countries** or **firms** at the same time.
- If companies are part of the same supply chain, what happens to one might affect others.
- 👉 This shared influence leads to **cross-sectional dependence (correlation)** in the error terms.

9. Diagnostic check

- **How to test it?**
- 🧷 **1. Pesaran's CD test**
- Most commonly used.
- Works well **even when T is small** (short time periods).
- **Null hypothesis (H_0)**: No cross-sectional dependence.
- **Alternative (H_1)**: There **is** cross-sectional dependence.
- 🧷 **2. Breusch-Pagan LM test**
- Better when **T is large and N is small**.

9. Diagnostic check

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: Untitled
Periods included: 4
Cross-sections included: 5
Total panel observations: 20
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	27.21593	10	0.0024
Pesaran scaled LM	3.849599		0.0001
Pesaran CD	1.168892		0.2424

Based on the results of PersanCD test in the table above, p-value > 5%, so we fail to reject the null hypothesis.

Fail to reject H_0 → No significant cross-sectional dependence. Model is OK.

9. Diagnostic check

- **Test 3: Heteroskedasticity Test**
- In panel data, heteroskedasticity can occur within cross-sections, or across cross-sections, or even both. Common Tests for Heteroskedasticity in Panel Data (EViews-compatible):after FE regression):
- Click Proc → Make Residual Series → name it: resid1
- Now generate a new series: residuals squared:
- Quick → Generate Series → Type:
- Entering thid code: resid1_sq = resid1^2
- Quick → Estimate Equation → Type: resid1_sq c
- Under Panel Options → Select: Effects specification → Cross-section Fixed Effects.

9. Diagnostic check

- Look at the **F-statistic** for cross-section fixed effects.
- **Null Hypothesis (H_0)**: Homoskedasticity (no variance difference across entities).
- **Alternative Hypothesis (H_1)**: Heteroskedasticity (variance differs across entities).
- ✅ If **p-value < 0.05** → Reject H_0 → **Evidence of heteroskedasticity** across entities.

9. Diagnostic check

Dependent Variable: RESID01_SQ
Method: Panel Least Squares
Date: 04/28/25 Time: 19:47
Sample: 2020 2024
Periods included: 5
Cross-sections included: 5
Total panel (balanced) observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.342835	0.503300	2.668061	0.0148

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.337836	Mean dependent var	1.342835
Adjusted R-squared	0.205403	S.D. dependent var	2.823082
S.E. of regression	2.516499	Akaike info criterion	4.860471
Sum squared resid	126.6554	Schwarz criterion	5.104246
Log likelihood	-55.75589	Hannan-Quinn criter.	4.928084
F-statistic	2.551000	Durbin-Watson stat	1.518488
Prob(F-statistic)	0.070952		

Based on the result in the table, p-value is greater than the 5% so we fail to reject the null hypothesis this means that there is no evidence of heteroskedasticity

9. Diagnostic check

- Test 4: **Unit Root/Stationarity Tests (in panel data)**
- To check if variables are stationary. Here's a full, clear guide on **how to perform a panel unit root test** in **EViews**, step by step.

1. In the **Import Wizard**:
    1. Select the **ID series** (cross-section ID) — e.g., country
    2. Select the **date or time series** — e.g., year
    3. Choose **Panel structured workfile** in the last step
    4. Click **Finish**

- Now your panel is ready to test.
- Suppose you want to test GDP.
- In the **Workfile window**, double-click GDP.

9. Diagnostic check

- With GDP open, go to:
  View > Unit Root Test > Panel Unit Root Test
- Choose the test you want:
- **step 4: Interpret the Output**
- **Look at the p-value** of the test statistic.
- If **p-value < 0.05** → Reject the null = **Stationary**
- If **p-value ≥ 0.05** → Cannot reject null = **Non-stationary**

9. Diagnostic check

Statistic	Prob.
PP-Fisher Chi-square	0.0234

PP-Fisher Test:

- **Null Hypothesis (H_0)**: Unit root exists (non-stationary).
- **Result**: p-value = **0.0234** (< 0.05)
- ➔ **Reject the null**: **GDP is stionary** according to PP-Fisher test.

9. Diagnostic check

Test Name	Null Hypothesis	Use When:
Levin-Lin-Chu (LLC)	All panels have unit root	Balanced panel, longer time
Im-Pesaran-Shin (IPS)	Some panels have unit root	Unbalanced panels allowed
ADF-Fisher	Unit root in all panels	Works with fewer obs
PP-Fisher	Unit root in all panels	Robust to serial correlation
Hadri	**Stationarity** (inverse logic)	Small samples