

ECON220C Discussion Section 1

Introduction to Static Panel Data

Lapo Bini

UC San Diego

April 13, 2026

ROADMAP

1. **Data Structure & Algebra**
2. **Pooled OLS**
3. **Random Effect Model**

INTRODUCTION

► Panel Data Structure

We observe outcome variables y_{it} and covariates $\mathbf{x}_{it}$ for **individuals** $i = 1, \dots, N$ across multiple **time horizons** $t = 1, \dots, T$.

(i) $\mathbf{x}_{it} = (x_{it1} \dots x_{itk})$ is a $1 \times k$ vector of **covariates** for the scalar outcome y_{it}

(ii) $\mathbf{X}_i = \begin{bmatrix} \leftarrow & \mathbf{x}_{i1} & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{x}_{iT} & \rightarrow \end{bmatrix}$ is a $T \times k$ matrix with all i^{th} **observations** for $\mathbf{Y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}$

(iii) You can go further and stack all $\mathbf{X}_i$ and $\mathbf{Y}_i$ into $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix}$

► Minimization Problem

The model is $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$, we estimate the parameter vector by **minimizing** the sum of the **squared prediction errors**:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\mathbf{b}} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}_{it}\mathbf{b})^2$$

Remember

(i) $\frac{\partial \mathbf{x}_{it}\mathbf{b}}{\partial \mathbf{b}} =$

(ii) $\mathbf{X}'_i \mathbf{X}_i =$

(iii) $\mathbf{X}' \mathbf{X} =$

EXERCISE: POOLED OLS

(i) Consistency

Consider the model $y_{it} = x_{it}\beta + \varepsilon_{it}$ propose a condition on x_{it} and u_{it} and show the **consistency** of the pooled-OLS estimator.

(ii) Model Misspecification

Now suppose that the true model is $y_{it} = x_{it}^3 + u_{it}$ where $E[x_{it}] = E[u_{it}] = 0$. Suppose that $\{y_{it}, x_{it}, u_{it}\}$ is an *iid* sequence and $x_{it} \perp u_{it}$. You mistakenly run the regression in (i), find the **probability limit** of $\hat{\beta}$.

(iii) Heteroskedasticity Robust Estimator

Knowing the true model, suppose you want to compute the asymptotic variance of $\sqrt{N}(\hat{\beta} - \beta)$. Do you need a variance estimator robust to **conditional heteroskedasticity**?

RANDOM EFFECT MODEL

► Problem: Unobserved Heterogeneity

1. Issue: **individual-specific heterogeneity** in the error term: $\varepsilon_{it} = \alpha_i + u_{it}$.
2. We will assume that $E[u_{it}\mathbf{x}_{it}] = \mathbf{0} \quad \forall t$ in what follows.
3. **Random effect model**: $\alpha_i | \mathbf{X}_i \sim (0, \sigma_\alpha^2)$.

► Exercise

- (i) Is the Pooled OLS estimator **consistent** under the assumptions above?
- (ii) Assume $\text{Var}(\mathbf{u}_i | \mathbf{X}_i) = \sigma_u^2 \mathbb{I}_T$ and $E[\mathbf{u}_i \alpha_i | \mathbf{X}_i] = \mathbf{0}$, derive the conditional variance of $\boldsymbol{\varepsilon}_i$ given $\mathbf{X}_i$. Is the Pooled OLS estimator efficient?