# ECON220C Discussion Section 2

# From Pooled OLS to Fixed Effect Model

Lapo Bini

# Before Starting

## Review var/cov matrix

$$\sqrt{n}\left(\hat{\beta}_{OLS}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,E[x_ix_i^T]^{-1}E[u_i^2x_ix_i^T]E[x_ix_i^T]^{-1}\right)$$

# Roadmap

1. Some notation
2. Consistency of Pooled OLS
3. Fixed Effect model
4. Consistency of $\hat{\beta}_{FE}$
5. Exercise: IV and FE model.

# Static Panel Data Model

## New notation

$x_{it} = (x_{it1} \dots x_{itk})$ is a 1x$k$ vector, here k is the number of **covariates**.

$$X_i = \begin{bmatrix} \leftarrow & x_{i1} & \rightarrow \\ & \vdots & \\ \leftarrow & x_{iT} & \rightarrow \end{bmatrix}$$ is a $T$x$k$ matrix with all $i^{th}$**-individual observations**.

$$Y_i = \begin{bmatrix} x_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}$$ is a $T$x1 vector.

You can go further and stack all $X_i$ and $Y_i$ into $\mathbf{X}$ and $\mathbf{Y}$

# Pooled OLS

**Model** $y_{it} = x_{it}\beta + \varepsilon_{it}$, we can estimate $\beta$ by minimization of mean squared error as usual.

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^T x_{it}\right)^{-1} \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^T y_{it}\right)$$

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{n} X_i^T X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^T Y_{it}\right)$$

$$\hat{\beta}_{POLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \left(\mathbf{X}^T\mathbf{Y}\right)$$

# Pooled OLS

## Consistency

Given $y_{it} = x_{it}\beta + \varepsilon_{it}$, we estimate $\beta$ by Pooled OLS. Under what conditions on $x_{it}$ and $u_{it}$ is the estimator consistent? *Hint: write Pooled OLS formula and substitute* $Y_i$.

# Problem: Unobserved Heterogeneity

**Structure in the error term**

- Issue: we have two types of error: $\varepsilon_{it} = \alpha_i + u_{it}$.
- We will assume that $E[u_{it} x_{it}] = 0 \ \ \forall t$ in what follows.
- **Random effect model**: $\alpha_i | X_i \sim (0, \sigma_\alpha^2)$.

**Questions**

(i) Is Pooled OLS estimator consistent under the assumptions above?

(ii) Assume $Var(u_i | X_i) = \sigma_u^2$ and $E[u_i \alpha_i | X_i] = 0$, derive the conditional variance of $\varepsilon_i$ given $X_i$. Is the Pooled OLS estimator efficient?

# Problem: Unobserved Heterogeneity

**Answers**

# Fixed Effect Model

**Different structure**: now we assume that $x_i t$ and $\alpha_i$ are **correlated**. The model we consider is $y_{it} = x_{it}\beta + \alpha_i + u_{it}$

**Get rid of $\alpha_i$**: we want to rewrite the model as $\tilde{y}_{it} = \tilde{x}_{it}\beta + \tilde{u}_{it}$, where $\sim$ means transformation of the data, and then apply Pooled OLS. Different approaches:

(a) **Within** approach: just demean LHS and RHS $\implies$ $\tilde{y}_{it} \equiv y_{it} - \bar{y}_i$

(b) **First difference** estimator $\implies$ $\tilde{y}_{it} = \Delta y_{it} \equiv y_{it} - y_{it-1}$

(c) **Forward or backward** demeaning $\implies$ $\tilde{y}_{it} \equiv y_{it} - \vec{y}_i$

# Fixed Effect Model - Consistency

**Questions**: let's start from the pooled OLS formula:

$$\hat{\beta}_{POLS} = \left( \sum_i \sum_t \tilde{x}_{it}^T \tilde{x}_{it} \right)^{-1} \left( \sum_i \sum_t \tilde{x}_{it}^T \tilde{y}_{it} \right)$$

(i) Under what conditions is $\hat{\beta}_{POLS}$ consistent for $\beta$ for a fixed $T$ as $N \to \infty$ if we use the first difference or within approach?

(ii) Under what conditions is $\hat{\beta}_{POLS}$ unbiased for $\beta$ if we use the first difference or within approach?

# Fixed Effect Model - Consistency

**Answers**

# Exercise: IV & FE Model (Midterm 2024)

Consider a linear structural model:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

where $x_{it}$ may be correlated with $\alpha_i$ and $u_{is}$ $\forall s$. There is an instrumental variable $Z_{it}$ that is correlated with $\alpha_i$ but $E[u_{it}|z_{it}]$ for all $i$ and $t$. Consider the following estimator:

$$\hat{\beta} = \frac{\sum_i \sum_t (z_{it} - \vec{z}_i)(y_{it} - \vec{y}_i)}{\sum_i \sum_t (z_{it} - \vec{z}_i)(x_{it} - \vec{x}_i)}$$

# Exercise: IV & FE Model (Midterm 2024)

**Questions**

(i) Under what condition is $\hat{\beta}$ consistent?

(ii) Is the condition $E[u_{it}|z_{it}] = 0 \ \forall i, \ \forall t$ enough for consistency?

(iii) Is the condition $E[u_{it}|z_{is}] = 0 \ \forall i, \ \forall t, \ \forall s = 1, \dots, t$, enough for consistency?

(iv) Now suppose $\hat{\beta}$ is consistent, and we want to compute its standard error. Explain when you need to use the cluster-robust standard error.

(v) Explain how you would estimate the asymptotic variance of your estimator.

# Exercise: IV & FE Model (Midterm 2024)

**Answers**