Handout 1: Overview of the Course. Introduction to Panel data

Master in Data Science for Decision Making
Barcelona School of Economics

Laura Mayoral

IAE and BSE

Barcelona, Winter 2026

Welcome!

This course

- Two sections

- Part I: Introduction to Panel data and Nonparametric econometrics (10h)

- Part II 2: DAGs, causality, etc. (10h, Prof: Aleix Ruiz de Villa)

- From now on, we will focus on Part I: Introduction to Panel data and Nonparametric econometrics.

Today's Goal

1. Description of the overall logistics of the course.
2. 9 Questions to review basic econometrics
3. Motivation and Overview.
4. The course itself: Introduction to panel data.

1. About logistics

- This course will review to main topics
  - Estimation of panel data models
  - Introduction to non-parametric and semiparemetric
- Both are very broad topics in econometrics: this course will be a short introduction, focused on explaining the main ideas and how the models are estimated rather than the technical details.
- 10 hours with me+ 2 hours with the RA
- Website of the course: click here
  - Check the syllabus for information about grading, references, etc.
  - please check it regularly for updates
  - Main course materials can also be found in classroom.

2. 9 Questions to review basic econometrics

9 Questions to review basic econometrics

1. What's the role of Econometrics?

(Order doesn't imply relevance)

9 Questions to review basic econometrics

1. What's the role of Econometrics?

(Order doesn't imply relevance)

1. Prediction: Uses data on a number of variables to predict another one.

2. Estimation and Inference: It develops and applies statistical methods to quantify and test causal relationships between economic variables.

9 Questions to review basic econometrics

1. What's the role of Econometrics?

(Order doesn't imply relevance)

1. Prediction: Uses data on a number of variables to predict another one.

2. Estimation and Inference: It develops and applies statistical methods to quantify and test causal relationships between economic variables.

This course will focus on (2).

2. Let y be a dependent variable of interest and let X be an independent variable.

- Does correlation between y and X imply causation?

2. Let y be a dependent variable of interest and let X be an independent variable.

- Does correlation between y and X imply causation?
- and viceversa?
- can you provide examples?

2. Let y be a dependent variable of interest and let X be an independent variable.

- Does correlation between y and X imply causation?
- and viceversa?
- can you provide examples?

2. Let y be a dependent variable of interest and let X be an independent variable.

- Does correlation between y and X imply causation?
- and viceversa?
- can you provide examples?

- Since

1) correlation and causation are two different concepts and

2) because we're interested in causal relationships,

$\Rightarrow$ the goal is to obtain estimates that can be interpreted causally.

3.a What type of function relating y and X is typically estimated by econometricians?

3.a What type of function relating y and X is typically estimated by econometricians?

$$E(y|X): \text{conditional expectation}$$

(There are exceptions: quantile regression).

3.a What type of function relating y and X is typically estimated by econometricians?

$$E(y|X) : \text{conditional expectation}$$

(There are exceptions: quantile regression).

3.b Why so much interest placed on estimating the conditional mean $E(y|X)$?

3.a What type of function relating y and X is typically estimated by econometricians?

$$E(y|X): \text{conditional expectation}$$

(There are exceptions: quantile regression).

3.b Why so much interest placed on estimating the conditional mean $E(y|X)$?

- Response: Conditional expectation $E(y|X)$ is the optimal.* predictor of y given X.

- Meaning:

- Consider the problem: what's the best way of combining information on X to produce the best predictor for y, best=" lowest mean squared error (MSE)"

- Answer: $E(y|X)$

4. What is the simplest model you can postulate for the conditional expectation?

4. What is the simplest model you can postulate for the conditional expectation?

$$E(y|X) = X\beta$$

- But, does the conditional expectation need to be linear?

4. What is the simplest model you can postulate for the conditional expectation?

$$E(y|X) = X\beta$$

- But, does the conditional expectation need to be linear?
  - No! in general $E(y|X) = g(X)$ can be highly non linear

4. What is the simplest model you can postulate for the conditional expectation?

$$E(y|X) = X\beta$$

- But, does the conditional expectation need to be linear?
  - No! in general $E(y|X) = g(X)$ can be highly non linear
- A typical assumption in the first econometric courses is linearity:

Assumption 1: the conditional expectation of y given X is linear.

- Why do we do this?

4. What is the simplest model you can postulate for the conditional expectation?

$$E(y|X) = X\beta$$

- But, does the conditional expectation need to be linear?
  - No! in general $E(y|X) = g(X)$ can be highly non linear
- A typical assumption in the first econometric courses is linearity:

Assumption 1: the conditional expectation of y given X is linear.

- Why do we do this?
  - Simplicity

4. What is the simplest model you can postulate for the conditional expectation?

$$E(y|X) = X\beta$$

- But, does the conditional expectation need to be linear?
  - No! in general $E(y|X) = g(X)$ can be highly non linear
- A typical assumption in the first econometric courses is linearity:

Assumption 1: the conditional expectation of y given X is linear.

- Why do we do this?
  - Simplicity
  - There's one case where we know that the conditional expectation is linear, which one?

5. For a linear conditional expectation, what is the model we take to the data?

$$y = X\beta + \epsilon \quad (1)$$

ϵ: random noise

“Chance is only the measure of our ignorance.” (Henry Poincaré, French mathematician).

A key assumption: recall that we want $E(y|X) = X\beta$, therefore Assumption 2: $E(\epsilon|X) = 0$

Assumption 2 demands that other variables we ignore that can have an impact on y (which are assembled in ϵ) should be uncorrelated with X.

Under Assumption 2, taking expectations in (1).

$$E(y|X) = E(X\beta|X) + E(\epsilon|X) = X\beta$$

6. Under Assumptions 1 and 2, how would you estimate (1)?

6. Under Assumptions 1 and 2, how would you estimate (1)?

OLS.

6. Under Assumptions 1 and 2, how would you estimate (1)?

OLS.

(Bottom line: Econometrics would be very simple if Assumptions 1 and 2 were always true!)

7. Under Assumptions 1 and 2, what would be the (asymptotic) properties of your estimator?

7. Under Assumptions 1 and 2, what would be the (asymptotic) properties of your estimator?

- $\hat{\beta}$ is consistent, i.e., as the sample size $N \to \infty$

7. Under Assumptions 1 and 2, what would be the (asymptotic) properties of your estimator?

- $\hat{\beta}$ is consistent, i.e., as the sample size $N \to \infty$

$$\hat{\beta}_N \xrightarrow{p} \beta$$

- $\hat{\beta}_N$ is asymptotically normally distributed

  - Allows very easy inference (confidence intervals, testing hypotheses...).

8. Under Assumptions 1 and 2, does $\hat{\beta}$ have a “causal” interpretation?

8. Under Assumptions 1 and 2, does $\hat{\beta}$ have a “causal” interpretation?

YES!

8. Under Assumptions 1 and 2, does $\hat{\beta}$ have a “causal” interpretation?

YES!

9. And what if either Assumption 1 or Assumption 2 holds?

- NO

- If Assumptions 1 or 2 fail, $\hat{\beta}$ will be a measure of the linear association between y and X.

- Why?

9. And what if either Assumption 1 or Assumption 2 holds?

- NO

- If Assumptions 1 or 2 fail, $\hat{\beta}$ will be a measure of the linear association between y and X.

- Why?
  - If Assumption 2 fails: $\hat{\beta} \overset{p}{\nrightarrow} \beta$
  - If Assumption 1 fails: what does β even mean, if the relationship between y and X is not linear?

Key Takeaways

1. One of the main goals of econometrics is the estimation of causal relationships

2. Correlation doesn't imply causation (and viceversa!)

3. Estimating the conditional expectation is typically our main goal (i.e., given the value of the covariates, what's the average value of y).

4. Typical assumptions in elementary econometric courses: linearity of the conditional expectation and exogeneity of the regressors.

5. These are strong assumptions that we will try to relax in this course.

2. Motivation and Overview of the course

This course

- In this (first half) of the course we're going to estimate models where assumptions (1) and/or (2) might not hold.

- We will discuss

1) whether these are assumptions are reasonable or are too demanding,

2) what are the consequences of their violation and, most importantly,

3) we will review some methods that will allow us to obtain consistent estimators when these assumptions are violated.

Overview of the course

- We will depart from the above-outlined framework in two directions:

Direction I

- Interest in estimation methods that are valid (in certain cases) when Assumption 2 is violated.
  - One of the reasons why Assumption 2 is violated is due to omitted variables that are in the residual term and are correlated with the X.
  - We will analyse how and under what circumstances the use of panel data models solves this problem.

Overview of the course, II

Direction II

- Interest in estimation methods that are valid under mild assumptions on the functional form: $E(y|X) = f(X)$
  - Imposing linearity and/or a specific distribution on the data are strong assumptions
- Tradeoff between efficiency and validity:
  - Imposing assumptions that are correct leads to more efficient estimators
  - Imposing assumptions that are not true leads to inconsistent estimators

Non parametric estimation

- Departure point: in the vast majority of cases we don't know the "true" model or the "true" distribution of the data.

- Approach: We will look at methods that are valid under mild assumptions about the DGP (we won't impose restrictions about the DGP)

→ Non-parametric (or semi-parametric) estimators.

  - Note: a parametric model is known up to some parameters, for instance: $E(y|X) = X\beta$

  - A nonparametric model is one in which the function itself is unknown: $E(y|X) = g(X)$

4. Introduction to Panel Data Models

Introduction to panel data models: Roadmap

1. Basic questions: What is panel data? what is it useful?

2. Review of Omitted variable bias.

3. Types of Panel Data: Balanced vs Unbalanced; Micro vs. Macro panel data.

4. Types of Panel Data Models: Linear vs. Nonlinear; Static vs. Dynamic.

5. Estimation of Panel data models

   5.1. Fixed Effect Models: estimation and inference

   5.2. Other estimators: Random Effects Models, Pooled OLS, Between estimator

1. Basic questions

What is panel data?

- Data where the same individual/unit of observation is observed several times (more than 1).

- Many consecutive cross sections, where we can link units over time.

  - N: the number of units (the cross-sectional dimension of the data)

  - T: number of time periods (the time or longitudinal dimension of the data).

- However, panel data refers to all data sets that span (at least) two dimensions:

- Example 1: Individuals observed every year for a number of years.

- Example 2: Firms, each having a number of establishments.

- Example 3: Schools, each having a number of students

Why is panel data useful?

- Two main advantages:

- Recall that omitted variables are a common cause of violation of Assumption 2, i.e., they often cause violation of the exogeneity assumption.

1. The use of panel data helps avoiding the omitted variables bias

- Why? it allows to control for unobserved characteristics that are constant over the time dimension.

- Unobserved characteristics are accounted for, not left in the residual term (therefore, avoiding the correlation between the regressors and the residual term).

A quick example

- Context: you want to study whether studying more years leads to a better salary.

- You have a sample of N individuals observed at a point in time and you estimate:

$$salary_i = \beta_0 + \beta_1 yearseduc_i + \epsilon_i$$

- Problem: individuals are heterogeneous as they differ (among other things) in their innate ability

A quick example

- Context: you want to study whether studying more years leads to a better salary.

- You have a sample of N individuals observed at a point in time and you estimate:

$$salary_i = \beta_0 + \beta_1 yearseduc_i + \epsilon_i$$

- Problem: individuals are heterogeneous as they differ (among other things) in their innate ability

- More ability will lead to more years of education AND to have a higher salary (for reasons different from education)⇒ omitted variable

- Panel data will help us solve this problem:

  - Having repeated observations for these individuals will allow us to “control” for all the individual unobserved heterogeneity

2. Panel data also helps studying dynamics:

Example: Habit Formation in Consumption

Suppose you want to study habit formation in consumption. Does last period's consumption affect this period's consumption, beyond what's explained by current income?

With cross-sectional data, you can't observe how an individual's consumption evolves over time. But with panel data, you can estimate:

$$C_{it} = \beta_0 + \beta_1 C_{it-1} + \beta_2 \text{Income}_{it} + \alpha_i + \varepsilon_{it} \tag{1}$$

Here you're directly testing whether lagged consumption (C_{it-1}) matters, controlling for individual fixed effects (α_i). If $\beta_1 > 0$ and significant, it suggests consumption habits persist over time.

- Other dynamic examples:
  - **Labor market dynamics**: Does being unemployed today affect your probability of being unemployed next period (state dependence)?
  - **Firm investment**: Do current profits affect next period's investment, or do firms smooth investment over time?
  - **Health dynamics**: Does being sick this year affect health outcomes next year?

2. Motivation for Panel data Models:
(Review of) Omitted variable Bias

- Consider a ("true") model that verifies Assumptions 1 and 2.

$$y = \alpha + \beta X + \gamma \eta + \epsilon$$

- Assume however that the following model is estimated:

$$y = \alpha + \beta X + u$$

with $u = \gamma \eta + \epsilon$.

- It follows that:

$$\hat{\beta} \xrightarrow{p} \frac{\mathrm{Cov}(y, X)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(\alpha + \beta X + \gamma \eta + \epsilon, X)}{\mathrm{Var}(X)} = \beta + \gamma \frac{\mathrm{Cov}(\eta, X)}{\mathrm{Var}(X)}$$

- Omitted variable bias (OVB):

- If Cov (η, X) $= 0$, then the estimate of $\hat{\beta}$ is consistent.
- If Cov (η, X) $\neq 0$, then the estimate of $\hat{\beta}$ is not consistent.

- First important fact to remember:

omitting variables that are uncorrelated with the regressors doesn't lead to bias.

- If $Cov(\eta, X) \neq 0$, the bias ($\hat{\beta} - \beta$) is

$$\hat{\beta} - \beta = \gamma \frac{\mathrm{Cov}(\eta, X)}{\mathrm{Var}(X)} + o_p(1)$$

- Omitted variable bias (OVB):

- If Cov $(\eta,$ X$) = 0$, then the estimate of $\hat{\beta}$ is consistent.
- If Cov $(\eta,$ X$) \neq 0$, then the estimate of $\hat{\beta}$ is not consistent.

- First important fact to remember:

omitting variables that are uncorrelated with the regressors doesn't lead to bias.

- If $Cov(\eta, X) \neq 0$, the bias $(\hat{\beta} - \beta)$ is

$$\hat{\beta} - \beta = \gamma \frac{\mathrm{Cov}(\eta, X)}{\mathrm{Var}(X)} + o_p(1)$$

- Second important fact to remember: Omitted variable bias formula

$$\hat{\beta} - \beta = \gamma \frac{\mathrm{Cov}(\eta, X)}{\mathrm{Var}(X)} + o_p(1)$$

- The sign of the bias depends on the product of two terms:
  - the correlation of X and the omitted variable
  - the coefficient γ of the omitted variable, η

- If this product is positive, the bias is positive: $\hat{\beta}$ will tend to be larger than the true β

- If this correlation is negative, the bias is negative: $\hat{\beta}$ will tend to be smaller than the true β

- Understanding this formula well is important: it will allow you to predict the direction of the bias of your estimates!

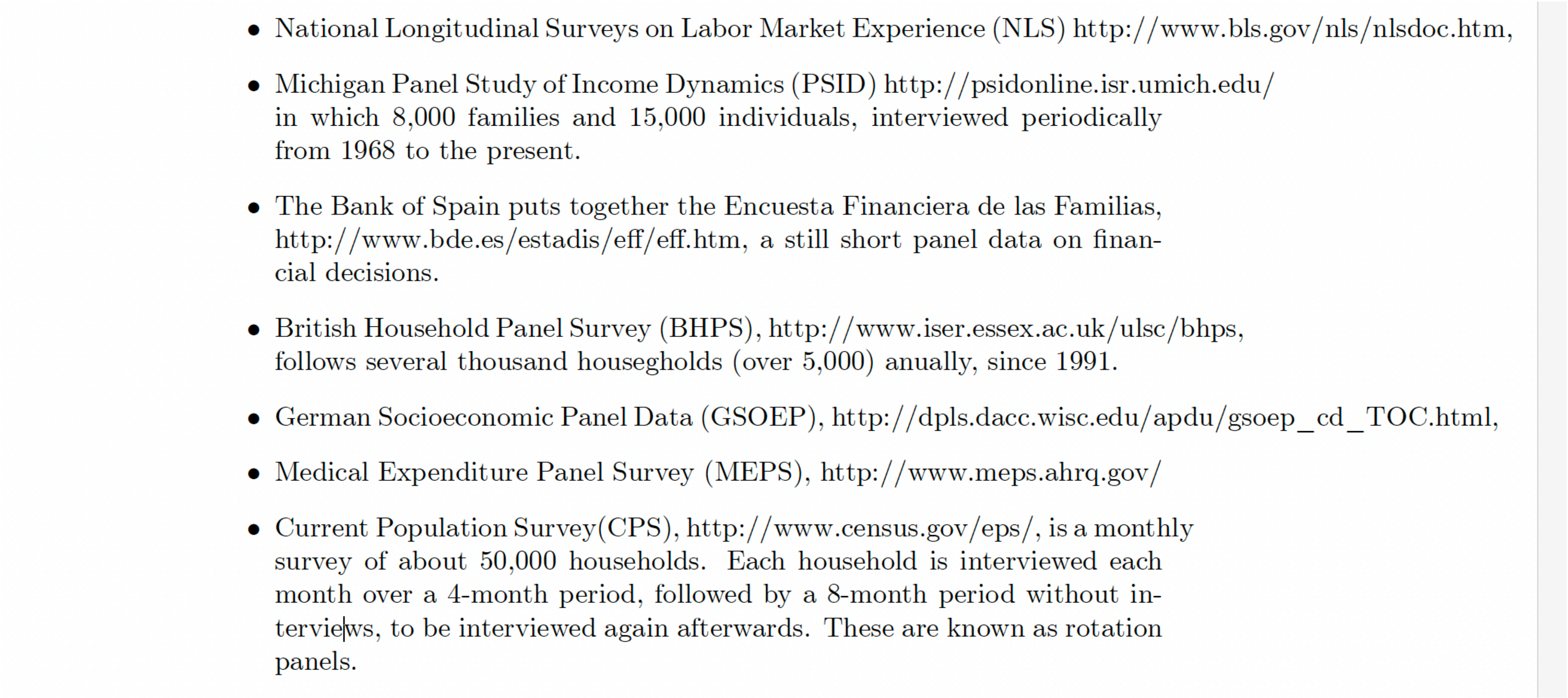
An Example

- Consider this example:

You want to estimate the impact of studying a master in data science on wages and you have data on both variables for a representative sample of people in their 30's. If you regress wages on 'master':

  - What omitted variables could be in this regression?
  - Is it reasonable to expect that these variables are uncorrelated with the variable "master"?
  - Can you anticipate the direction of the bias?

Some examples of datasets with panel structure



- National Longitudinal Surveys on Labor Market Experience (NLS) http://www.bls.gov/nls/nlsdoc.htm,
- Michigan Panel Study of Income Dynamics (PSID) http://psidonline.isr.umich.edu/ in which 8,000 families and 15,000 individuals, interviewed periodically from 1968 to the present.
- The Bank of Spain puts together the Encuesta Financiera de las Familias, http://www.bde.es/estadis/eff/eff.htm, a still short panel data on financial decisions.
- British Household Panel Survey (BHPS), http://www.iser.essex.ac.uk/ulsc/bhps, follows several thousand housegholds (over 5,000) anually, since 1991.
- German Socioeconomic Panel Data (GSOEP), http://dpls.dacc.wisc.edu/apdu/gsoep_cd_TOC.html,
- Medical Expenditure Panel Survey (MEPS), http://www.meps.ahrq.gov/
- Current Population Survey(CPS), http://www.census.gov/eps/, is a monthly survey of about 50,000 households. Each household is interviewed each month over a 4-month period, followed by a 8-month period without interviews, to be interviewed again afterwards. These are known as rotation panels.

3. Types of panel data: A First Classifications of Panels

1. Balanced and Unbalanced panels

2. Short and Long panels (or micro and macro panels)

Balanced vs. Unbalanced panels.

- Balanced panel: every $i \in N$ has T observations.
- Unbalanced panel: if the above is not true.

- Example: consider a panel of countries observed over time, developed countries tend to have all observations available, developing ones typically have some missing values for some time periods.

- For simplicity, we will typically consider balanced panels in the following.

- Methods that allow for unbalancedness are not complicated, see Chapter 17 in Wooldridge (you will learn about sample selection issues and attrition).

Short vs Long panels

- Short panels (micro panels): Large N, short T. Example: A sample of workers observed three time periods.
- Long panels (macro panels): Large T (N can be smaller or comparable in size). Example: OECD countries observed at a monthly frequency for 30 years.

Short vs Long panels

- Short panels (micro panels): Large N, short T. Example: A sample of workers observed three time periods.

- Long panels (macro panels): Large T (N can be smaller or comparable in size). Example: OECD countries observed at a monthly frequency for 30 years.

- The techniques needed to deal with these type of datasets may differ.

  - If N is the dominant dimension (short panels), asymptotics are computed considering $N \to \infty$, similar to cross-sectional data

  - But if T is the dominant dimension (long panels), asymptotics are computed considering $T \to \infty$ or $T, N \to \infty$, more similar to time-series data

- In this course we will consider $N >> T$

4. Types of Panel Data Models

- Let's write now the panel data models that then we will take to the data. First distinction: linear vs. nonlinear models.

- Linear panel data model, e.g.:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ denote the first and second dimensions of the data. For instance, i can denote, individuals, firms, countries, etc. and t, time (or space, or other dimensions that the data might have).

- Non linear panel data model:

$$y_{it} = g(c_i, X_{it}, \epsilon_{it})$$

where g is a nonlinear function.

- Estimation of nonlinear panel data models presents additional complications (due to the incidental parameters problem) and requires alternative estimation approaches.

- We will start by considering linear models.

Second distinction: Static vs. Dinamic panels

- Static panel data models: no lagged dependent variable in the regression. E.g.,

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

- Dynamic panel data models: lag(s) of the dependent variable are included in the model:

$$y_{it} = c_i + X_{it}\beta + \gamma y_{it-1} + \varepsilon_{it}$$

- Introducing dynamics in the regression complicates estimation because y_{it-1} is endogeneous.

- Different estimation methods: GMM.

First models we will take to the data:

Static and Linear Panel data Models

- We will begin by considering linear and static panel data models (i.e., do not include lags of the dependent variable).

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

- Let's focus on c_i, the "novelty" in this model.

- c contains all the characteristics of individuals that are constant over time (no t subindex!).
- It's typically non-observable, so c is often called "unobserved individual heterogeneity".
- Despite being non-observable, having panel data allows controlling for this term

More on c

- Because c is constant across individuals, it's called "fixed";
- . . . but it changes across individuals, so it's considered to be a random variable (don't be fooled by the name!).

- Note 1: The term "Fixed effect" is also typically employed in a different context: models where c and X are allowed to be correlated. We will go back to this below.

- Note 2: Obviously it's also possible to estimate "conventional" models with panel data, i.e, $y_{it} = c + X_{it}\beta + \varepsilon_{it}$. This is typically not a good idea unless one wants to estimate the impact of a variable that doesn't vary over time.

5. Estimation of panel data models

- We're interested in estimating this model.

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

with $i = 1, \ldots, N$, $t = 1, \ldots, T$.

- X_{it} is a $1 \times K$ vector of regressors. In general, it can contain variables that only vary over t, or vary over the two dimensions.

- Depending on the estimator employed, it can contain (or not) variables that only vary over i.

Types of Estimators for linear and static panel data models

- The assumptions placed on c_i will determine the estimator that should be employed.

- Assumptions on c_i: alternative scenarios

1. First case: $c_i = c$ is constant and non-observable.

Types of Estimators for linear and static panel data models

- The assumptions placed on c_i will determine the estimator that should be employed.

- Assumptions on c_i: alternative scenarios

1. First case: $c_i = c$ is constant and non-observable.

- Then, use “pooled OLS”, i.e., estimate everything with OLS.

- This is just the type of regressions you’re used to.

- No omitted variable bias (despite c being non-observable). (Why?)

- Second case: c_i is random but observable.

- Then, include these variables in the regression, estimate by OLS.

- For instance: age, gender, education of the parents, etc. etc.

- No omitted variable bias in this case (assuming that all the relevant characteristics are observed!). (Why?)

- Third case: c_i non-constant (i.e. random) and non-observable: this is the interesting case!

- Two types of assumptions on c_i: Fixed or Random Effects

- Fixed Effects models: allow for arbitrary correlation between c and X. (Implications for OLS?)

- Random Effects models: assume that the correlation between c and X is zero. (Implications for OLS?)

Exercise

- Which approach do you think is more general/less problematic?

- Why?

Fixed or Random Effects?

- FE estimators: valid under any value of $corr(X, c)$, including zero.

- RE estimators: only valid if $corr(X, c) = 0$

  - In theory: It's possible to test for random or fixed effects (Hausman tests).

  - In practice: it's complicated. The test itself relies on stringent assumptions.

- Always try to use estimators that are valid under general assumptions!$\Rightarrow$ Fixed effects are much safer.

Lecture 1: Key Takeaways

- In econometric models, a key threat to identification is not including all relevant variables in the model

- In most cases, this will give rise to biased estimators: omitted variable bias

- Having panel data allows us to control for all the unobserved and time-invariant individual heterogeneity

- Panel data: repeated observations⇒

- Data with two subindices, $i = 1, \ldots, N$ and $t = 1, \ldots, T$
- Typically: individuals over time, but not necessarily

Lecture 1: Key Takeaways, II

- In this course: N is large and T is small (micro-panel). (Other cases also possible)

- We're interested in estimating this model.

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

with $i = 1, \ldots, N$, $t = 1, \ldots, T$.

- c_i: unobserved, captures individual heterogeneity that is invariant over time

Lecture 1: Key Takeaways, III

How to estimate this model?

- Different assumptions of $c_i \Rightarrow$ different estimators.

1. First case: $c_i = c$ is constant: “pooled OLS”, i.e., estimate everything with OLS.

2. Second case: c_i is random. Two cases

   - $cov(c_i, X_i)$ is unrestricted (interesting case!): Fixed effects estimator

   - $cov(c_i, X_i) = 0$ (very limited use!): Randoms effects estimator

5.1. Fixed Effects Estimator

Roadmap

I. Fixed Effects Estimator

- FE estimation: Within transformation
- FE estimation:Dummy variable estimator
- FE estimation: First difference transformation

2. Trade-offs

3. Two-way fixed effects; The relation with DiD models

4. Inference in Fixed Effects models: robust versus clustered-robust standard errors.

5.1. Fixed Effects Estimator

- Recall the main framework:

  - c random, nonobservable, c and X are allowed to be correlated
  - Because this is much more general, this should be your first choice!

- Model to be estimated:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

where $cov(c_i, X_{it})$ can take any value (including zero).

Main identification assumption (FE1): Strict exogeneity

- Strict Exogeneity:

$$\text{FE1} : E(\varepsilon_{it}|X_i, c_i) = 0$$

- Meaning of strict: the cond. expectation needs to be zero for all values of t, past, contemporaneous and future values. This implies $cov(\varepsilon_{it}, X_{it+h}) = 0$ for all h.

- An additional condition: X_{it} cannot contain time-invariant variables, we need to drop those from the equation (we'll see why).

- Important: FE estimator only allows to estimate the impact of time-varying explanatory variables

Exercise: Identification in Panel Data Models

- Consider the following panel data model with time period dummies $d_{2t}, \ldots, d_{Tt}$, time-constant observables z_i, and time-varying variables w_{it}:

$$y_{it} = \theta_1 + \theta_2 d_{2t} + \cdots + \theta_T d_{Tt} + z_i\gamma_1 + d_{2t}z_i\gamma_2 + \cdots + d_{Tt}z_i\gamma_T + w_{it}\delta + c_i + u_{it}$$

with $E(u_{it}|z_i, w_{i1}, \ldots, w_{iT}, c_i) = 0$ for $t = 1, 2, \ldots, T$

- Questions:

- (a) Explain why θ_1 and γ_1 cannot be separately identified from c_i.

- (b) Which parameters involving z_i *can* be identified? Provide intuition.

- (c) Suppose $y_{it} = \log(\text{wage}_{it})$ and z_i includes a female indicator. What can we estimate about the gender wage gap, and what can we not estimate?

Solution

- (a) Identification problem with θ_1 and γ_1:

- The term $\theta_1 + z_i\gamma_1$ cannot be distinguished from c_i because both are time-constant.

- Any value attributed to the intercept or to z_i's effect in period 1 could equivalently be absorbed into the unobserved individual effect.

- (b) Identifiable parameters:

- The vectors $\gamma_2, \gamma_3, \ldots, \gamma_T$ are identified.

- These measure *differences* in the partial effects of time-constant variables relative to the base period $(t = 1)$.

- We can test whether effects of time-constant variables have changed over time.

- (c) Gender wage gap application:

- We *cannot* estimate the gender gap in any particular time period. We *can* estimate how the gender gap has changed over time.

How to estimate fixed effects models?

- In a nutshell: transform the model, get rid of c_i, then estimate!

- The idea is simple:

  - Linear panel data models allow for transformations that get rid of c_i from the model.

  - Since c_i disappears from the model, we can use OLS on the transformed model

  - There are different types of transformations/estimators: within transformation, first differences transformation, dummy variables estimator.

  - First transformation: within transformation or fixed effects transformation

FE estimation: Within transformation

- Step 1: Consider the FE model and average each variable over $t = 1, \dots, T$ to get:

$$\bar{y}_{it} = c_i + \bar{X}_i \beta + \bar{\varepsilon}_i$$

where $\bar{y}_{it} = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{X}_i = T^{-1} \sum_{t=1}^{T} X_{it}$, $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$

- Step 2: Compute the difference $y_{it} - \bar{y}_{it}$:

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + \varepsilon_{it} - \bar{\varepsilon}_i$$

Notice that c_i disappears in the transformation!

- Step 3: Estimate the resulting model by (pooled) OLS: consistent, as there are not omitted variables!!

- In practice:

- Use software to do the two steps (don't do them yourself)

- Why? There's an adjustment in the degrees of freedom that affects the computation of the residual variance, the software will do it automatically:

$$\widehat{\sigma}_u^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\tilde{u}}_{it}^{\,2}}{NT - N - K}.$$

Interpretation

- In a nutshell: the fixed effect estimator is a pooled OLS estimator applied on a model where all the variables have been demeaned.

- Technical Note I: To achieve consistency strict exogeneity is key! Why?

- The "transformed" (demeaned) variables contain all values of the variables for $t = 1, ...T$, not only the contemporaneous ones.
- For these variables to be exogeneous (in the usual sense) then:

$$E\big[(x_{it} - \bar{x}_i)'(u_{it} - \bar{u}_i)\big] = 0.$$

Under Assumption FE.1, u_{it} is uncorrelated with x_{is} for all $s, t = 1, 2, \ldots, T$. It follows that u_{it} and $\bar{u}_i$ are uncorrelated with x_{it} and $\bar{x}_i$ for all $t = 1, 2, \ldots, T$.

- **Technical Note II**: X cannot include time-invariant variables. Why?

- Technical Note II: X cannot include time-invariant variables. Why?
  - After demeaning, time invariant variables will become a vector of zeros in the matrix X. Then, that matrix will become non-invertible!
- This is in fact the second identification condition:

$$\text{FE2} : rank((X - \bar{X})'(X - \bar{X})) = K$$

Fixed effect estimator=Within estimator

Interpretation of the coefficients estimated using the within estimator

- The within estimator only exploits the within-variation for identification

  - the within transformation removes all differences across the units: all of them have the same mean, equal to zero.

  - Therefore, all the variation employed for identification comes from within-units.

Example

- How does joining a union affect a worker's wage?
- Setup:
  - We have panel data on workers over multiple years.
  - Each worker i is observed for $t = 1, 2, \ldots, T$ time periods.
  - Let w_{it} be the *log wage* of individual i at time t.
  - Let union_{it} be an indicator that is 1 if worker i is a union member at time t, and 0 otherwise.
  - There are unobserved, time-invariant characteristics (e.g. innate ability, ambition) that might affect wage levels.

- A naive "pooled OLS" model could be:

$$w_{it} = \beta_0 + \beta_1 \, \text{union}_{it} + u_{it}.$$

- But: only consistent if 1) individuals are identical (so β_0 can capture unobserved effects) OR if the unobserved effects are uncorrelated with joining an union.

- Fixed Effects model:

$$w_{it} = \underbrace{c_i}_{\text{time-invariant FE}} + \beta_1 \, \text{union}_{it} + \varepsilon_{it},$$

where c_i is a worker-specific intercept capturing all time-invariant traits of individual i.

- The Within Transformation ("De-meaning")

$$(w_{it} - \overline{w}_i) \; = \; \beta_1 \, (\text{union}_{it} - \overline{\text{union}_i}) \; + \; (\varepsilon_{it} - \overline{\varepsilon}_i).$$

- Interpretation of the Within-Estimator Coefficients:

  - We can only estimate this equation if there are workers that switch from being a union member to a non-member (and viceversa). Why?
  - β_1 measures how *log wage* changes for the **same individual** when that individual **switches** from being non-union to union.
  - β_1 is identified by those individuals who *change* their union status at least once during the panel. Individuals who are always union or never union provide no within variation to identify β_1.

Asymptotic Properties of the FE estimator

- Recall $\hat{\beta}_{FE} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$, where the "$\sim$" denotes that the data has been demeaned.

- Under FE1 and FE2, as $N \to \infty$ and fixed T

- $\hat{\beta}_{FE}$ is consistent and asymptotically normal:

$$N^{-1/2}(\hat{\beta}_{FE} - \beta) \xrightarrow{d} N(0, Avar\hat{\beta}_{FE})$$

where $Avar\hat{\beta}_{FE}$ denotes the asymptotic variance of $\hat{\beta}_{FE}$

- The specific shape of $Avar\hat{\beta}_{FE}$ will depend on the specific assumptions about heteroskedasticity and serial correlation).

Key Takeaways

- Fixed effects estimator: can deal with unobserved heterogeneity, correlated with the X's.

- Modus operandi: 1) demean all variables, 2) apply OLS

- Interpretation: We exploit only **within-individual (or within-unit)** variation over time.

- **Time-Invariant Factors Are Removed**: Any unchanging traits such as innate skill or background are taken out by "de-meaning" each person's data.

Alternative Approaches for estimating models with Fixed Effects

- As mentioned earlier, the basic idea for estimating linear panel data models with fixed effects consists of transforming the data to get rid of c_i

- The within transformation is one way of doing so (=demean all the variables).

- Two additional (and equivalent) approaches:

- First differencing estimator: transforms the model to get rid of c_i by taking first differences

- Dummy-variable estimator: estimates c_i by including individual-level dummy variables.

Alternative approach I: Dummy-variable estimator

- Key idea: This estimator treats c_i as parameters to be estimated.

- How? Include dummy variables D_i in the model so that for each i, D_i is 1 for the T values of i and zero otherwise (exclude the constant of the model or omit one D_i to avoid perfect multicollinearity).

- It turns out that is estimator is identical to the fixed effects one (numerically identical!).

- Why?

Dummy-variable estimator, II

- Recall the **Frisch–Waugh–Lovell theorem**: the coefficient on X_{it} obtained from a regression that includes individual dummies is identical to the coefficient obtained by:

  1. partialling out the individual dummies from y_{it} and X_{it}, and
  2. regressing the resulting residuals on each other.

- Key fact: Demeaning within individuals is exactly the residualization step implied by FWL.

Conclusion: the DV estimator and the fixed effects estimator yield identical estimates of β.

Dummy-variable estimator, III

- The DV estimator provides estimates for the c_i parameters, in contrast to the FE estimator.

- Key fact: However, in short panels these parameters ARE NOT estimated consistently. Why?

- The Incidental Parameter Problem: arises when the number of nuisance parameters increases with the sample size.

- In panel data models with individual fixed effects, with large N but fixed T: the number of nuisance parameters (the fixed effects) increases with the sample size, N.

- Intuition: New data doesn't help to "learn"/accumulate knowledge on the c', because the number of these parameters keeps growing as new data arrives!

- Therefore:
  - In short panels ($N \to \infty$, fixed T)
    - $\hat{c}_i$ are inconsistent
  - If T is sufficiently large, we can obtain the estimated c's, plot the distribution and have a relatively precise idea of the degree of heterogeneity in the distribution.
  - If T also tends to infinity: c_i will be consistently estimated.

- How to obtain estimates for the $c's$?

- Most statistical packages don't report directly the "c"'s

- Alternative: run an OLS regression with dummies as explained above.

- Or, if you've used the within estimator, you could also obtain these values by computing:

$$\hat{c}_i = \bar{y_i} - \bar{X}_i \hat{\beta}$$

(same problems apply of course!)

Alternative approach 2: First differencing methods

- Recall that the key idea for estimating models with FE is to transform the model so that we can get rid of c_i.

- First difference transformation: get rid of c_i by taking first differences in the model, i.e, $\Delta y_{it} = y_{it} - y_{it-1}$

- Recall the model:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it} \quad (1)$$

$$y_{it-1} = c_i + X_{it-1}\beta + \varepsilon_{it-1} \quad (2)$$

- Compute (1)-(2) to obtain:

$$\Delta y_{it} = \Delta X_{it}\beta + \Delta\varepsilon_{it}$$

- c_i has disappeared!

Comparison FE and FD estimators

- If T=2, both yield the same $\hat{\beta}$

- If $T > 2$, then they can be different.

- Choosing one or the other hinges on assumptions on the persistence of the serial correlation of the error term. (See Wooldridge, 10.7.1)

- Under (very unrealistic) assumptions of $i.i.d.$ residuals, FE estimator is more efficient.

- In applied work, the FE is typically more applied/reported.

2. Pros and cons: Tradeoffs of using FE models

- FE models are great to avoid OVB (if OV are time invariant)

- But there are some cons:

1. If there's measurement error in the data, it can become worse (therefore, it can have a larger impact on the estimates)

2. By demeaning the variables, we can also eliminate variation in the data that is "good" and therefore, estimates can be much less precisely estimated.

- Why is this?

- Fixed effects estimation can exacerbate measurement error because demeaning removes most of the true signal when regressors vary little over time, while the measurement error remains, lowering the signal-to-noise ratio and increasing attenuation bias.
- This problem is more severe when T is small and the regressor has little variation.

- We will see these points in two examples

Example

Studying the Effects of Unions on Wages

- Freeman (1984) studies the effect of unions on wages.

- Identification is tricky in this problem due to many potential omitted variables.

- He provides a comparison of estimates using OLS in cross section and FE.

Studying the Effects of Unions on Wages

- Cross sectional and FE estimates:

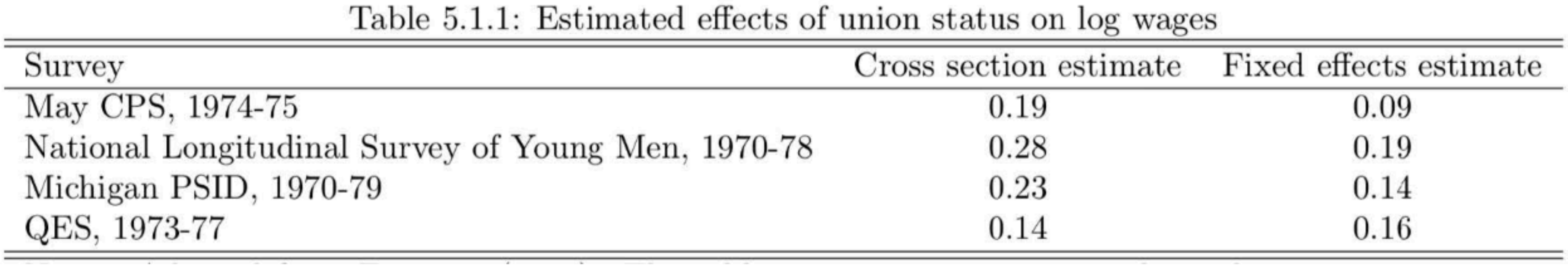


Table 5.1.1: Estimated effects of union status on log wages

Survey	Cross section estimate	Fixed effects estimate
May CPS, 1974-75	0.19	0.09
National Longitudinal Survey of Young Men, 1970-78	0.28	0.19
Michigan PSID, 1970-79	0.23	0.14
QES, 1973-77	0.14	0.16

- Cross sectional analysis delivers higher coefficients, why can this be?

Comparing Cross-Section to Panel Results

- A potencial explanation: OVB is positive, i.e., either: $Cov(\varepsilon_{it}, c_i) \geq 0$ AND $\gamma > 0$ OR both are negative.

- Can you think of omitted variables that could create this correlation?

- However, there is another suspect: Measurement Error

Comparing CS to Panel: Measurement Error

- The use of FE models can typically worsen measurement error. Why?

- The variation in the data is typically due to two terms: the "true" variation and potentially, the variation induced by noise or measurement error.

- When transforming the data to get rid of c, the "true" variation in the data decreases (we're removing all the between variation!). However, the within transformation doesn't get rid of the noise.

- As a result, the measurement error becomes relatively larger: the signal-to-noise ratio decreases.
- Recall that (classical) measurement error leads to biased coefficients. The bias is always towards zero (attenuation bias).
- As the measurement error is larger, attenuation due to it can also be larger.

Second tradeoff: FE can eliminate “good” variation in the data

Example: Class Size and Test Scores

- **Research Question:** Does smaller class size improve test scores?
- **Cross-Section OLS:**

$$\text{TestScore}_s = \alpha + \beta\,\text{ClassSize}_s + u_s$$

  - Uses variation across schools (some large, some small).
  - Often finds a **negative** relationship:

$$\widehat{\beta} < 0 \quad \text{(larger classes} \rightarrow \text{lower test scores).}$$

Consider now panel data on schools and introduce school FE:

$$\text{TestScore}_{s,t} = \alpha_s + \beta\,\text{ClassSize}_{s,t} + u_{s,t}$$

- Controls for time-invariant differences across schools (good to control for school-level omitted variables!).

- Identification now relies on **within-school** fluctuations over time.

- **Outcome:**

  - Variation in class size within each school (e.g., 25.5 to 24.8) may be small.
  - This can lead to a **smaller** (or less precise) estimate of β.
  - Large cross-sectional differences are no longer exploited, hence “chewed up” by fixed effects.

3. Two-way fixed effects

■ The two-way fixed effects model extends the standard FE approach by controlling for unobserved heterogeneity across two dimensions: individuals (or entities) and time periods.

$$y_{it} = c_i + \lambda_t + X_{it}\beta + \varepsilon_{it},$$

where

- c_i: Individual-specific fixed effect.
- λ_t: Time-specific fixed effect.

■ λ_t captures shocks or trends common to all individuals in period t (e.g., economic changes, changes in policies, etc).

Estimation Methods for TWFE models

- **Dummy Variable Approach:** Include dummy variables for each individual and each time period.

- **Within Transformation:** Demean the data by subtracting individual and time averages to remove fixed effects, avoiding a large number of dummy variables.

Example

■ Consider analyzing the impact of job training programs on wages:

$$\text{Wage}_{it} = \alpha_i + \lambda_t + \beta_1 \text{Training}_{it} + \varepsilon_{it}.$$

- α_i: Controls for innate ability/other individual-specific factors.

- λ_t: Controls for year-specific economic conditions.

- This specification isolates the effect of Training_{it} on Wage_{it} by accounting for unobserved individual and time-specific influences.

- More generally:

- You can construct models with a lot of different types of FE.

- An example: you have panel data on conflict at the country level over a number of months and want to study the impact of a country-level variable that varies over time. In addition to country FE, you can write models that contain

1) month FE: control for global trends

2) region-specific month FE: you let the month FE to change across regions/continents (because the trends can differ across regions)

3) country-specific decade -FE: you allow for unobserved factors that create slowly moving trends that are country-specific.

. . .

The relation between TWFE and Difference-in-Differences

- Recall the two-period diff-in-diff setup:

$$y_{it} = \alpha + \gamma \cdot \text{Treat}_i + \lambda \cdot \text{Post}_t + \delta \cdot (\text{Treat}_i \times \text{Post}_t) + \varepsilon_{it}$$

- With individual and time fixed effects, TWFE generalizes this:

$$y_{it} = c_i + \lambda_t + \delta \cdot D_{it} + \varepsilon_{it}$$

where $D_{it} = 1$ if unit i is treated at time t.

- In the **canonical 2x2 case** (2 groups, 2 periods), TWFE recovers the standard diff-in-diff estimator.

- Intuition: c_i absorbs group differences, λ_t absorbs time trends, δ captures the treatment effect.

Outside of the canonical case: apply with caution!

- TWFE models allow to estimate a diff-in-diff set up if $T = 2$, but what if $T > 2$?

- The analogy can break!

- This realization is relatively new in the diff-in-diff literature, you will see a lot of papers that apply (naive) TWFE when they want to estimate diff-in diff models when $T > 2$.

- When does the analogy break? Staggered treatment timing

The Problem: Staggered Treatment Timing

- In practice, units often receive treatment at **different times** (staggered adoption).

- The natural approach: run TWFE with $D_{it} = 1$ if unit i is treated by time t.

- **Problem:** TWFE estimates a weighted average of many 2x2 diff-in-diff comparisons, including:

- "Good" comparisons: treated vs. not-yet-treated

- "Bad" comparisons: late-treated vs. already-treated (using treated units as controls!)

- If treatment effects are **heterogeneous over time**, these "bad" comparisons can produce:

- Biased estimates

- Wrong sign (negative when true effect is positive!)

Recent Literature and New Estimators

- This is a very active (and already large) literature at the moment

- Key papers identifying the problem:

- Goodman-Bacon (2021): Decomposition of TWFE into weighted 2x2 comparisons.

- de Chaisemartin & D'Haultfœuille (2020): Shows TWFE weights can be negative.

- Proposed solutions:

- Callaway & Sant'Anna (2021): Group-time ATTs, aggregated appropriately.

- Sun & Abraham (2021): Interaction-weighted estimator.

- Borusyak, Jaravel & Spiess (2024): Imputation approach.

Key Takeaways about TWFE and DiD:

- If two periods: TWFE is fine to estimate DiD models.

- If more than two periods: check if treatment is staggered. If it is, don't use naive TWFE.

  - Why? TWFE uses already-treated units as controls for later-treated units.

- If treatment effects are heterogeneous (vary by cohort or over time), this biases $\hat{\delta}$.

- If treatment effects are homogeneous, TWFE is still valid even with staggered timing.

- With staggered treatment and heterogeneous effects, use modern DiD estimators rather than naive TWFE.

Key References

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- de Chaisemartin, C. & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*.
- Callaway, B. & Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Sun, L. & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.
- Roth, J., Sant'Anna, P., Bilinski, A. & Poe, J. (2023). What's trending in difference-in-differences? *Journal of Econometrics*.

Estimating FE models in practice

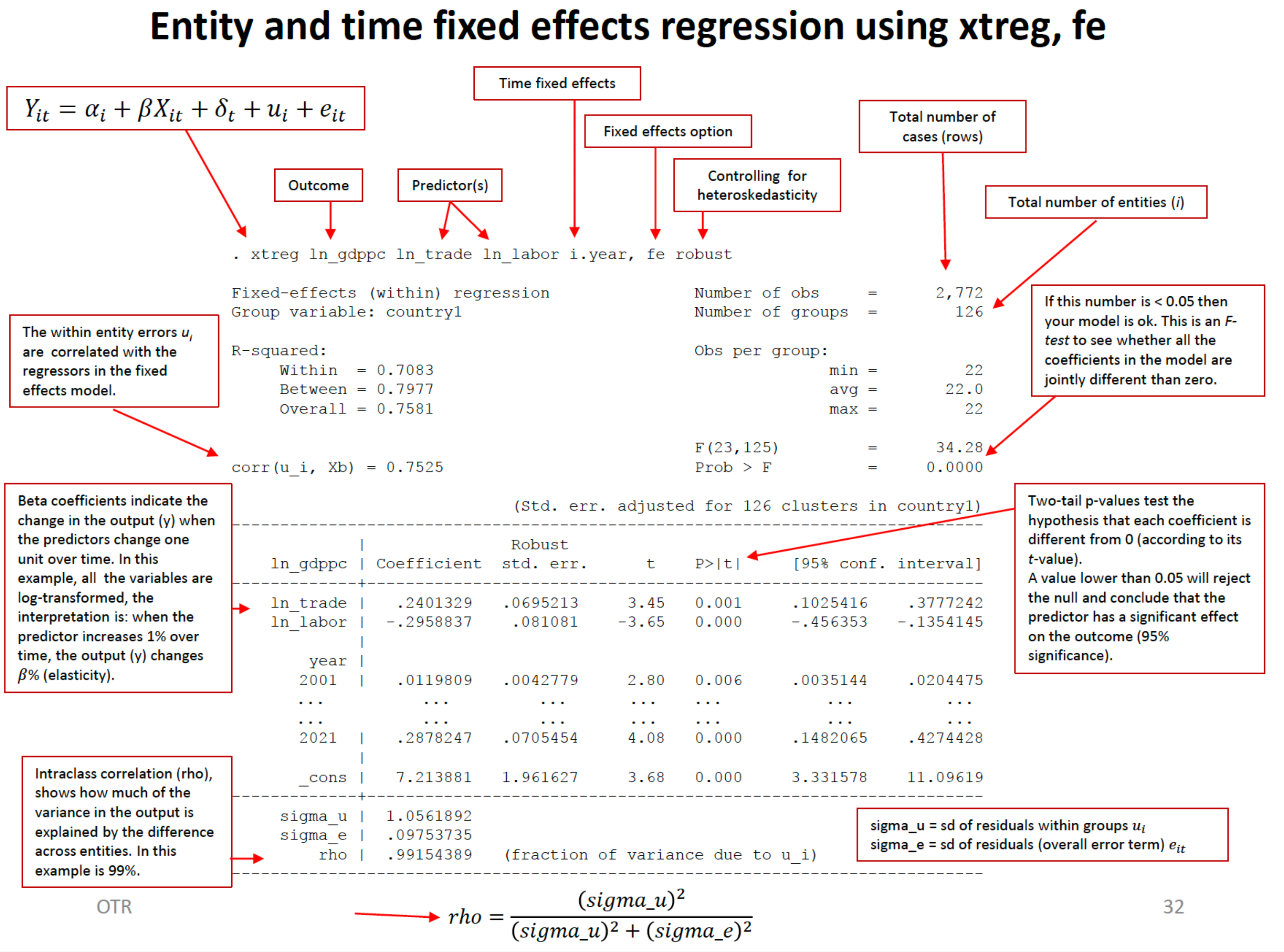
- You can estimate FE models using the software you prefer (STATA, R, Python . . .)

- Many economists use STATA, you will find a lot of examples, papers, replication packages written in STATA.

- See the website of the course for useful resources/examples.

Example: FE with stata

- Author: Oscar Torres-Reyna. Tip: use cluster s.e. (i.e., replace “robust” by vce(cluster country))



Entity and time fixed effects regression using xtreg, fe

$$Y_{it} = \alpha_i + \beta X_{it} + \delta_t + u_i + e_{it}$$

Outcome

Predictor(s)

Time fixed effects

Fixed effects option

Controlling for heteroskedasticity

Total number of cases (rows)

Total number of entities (*i*)

```
. xtreg ln_gdppc ln_trade ln_labor i.year, fe robust

Fixed-effects (within) regression               Number of obs     =      2,772
Group variable: country1                        Number of groups  =        126

R-squared:                                      Obs per group:
     Within  = 0.7083                                         min =         22
     Between = 0.7977                                         avg =       22.0
     Overall = 0.7581                                         max =         22

                                                F(23,125)         =      34.28
corr(u_i, Xb) = 0.7525                          Prob > F          =     0.0000

                           (Std. err. adjusted for 126 clusters in country1)
------------------------------------------------------------------------------
             |               Robust
    ln_gdppc | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    ln_trade |   .2401329   .0695213     3.45   0.001     .1025416    .3777242
    ln_labor |  -.2958837    .081081    -3.65   0.000     -.456353   -.1354145
             |
        year |
        2001 |   .0119809   .0042779     2.80   0.006     .0035144    .0204475
      ...        ...         ...        ...     ...         ...         ...
      ...        ...         ...        ...     ...         ...         ...
        2021 |   .2878247   .0705454     4.08   0.000     .1482065    .4274428
             |
       _cons |   7.213881   1.961627     3.68   0.000     3.331578    11.09619
-------------+----------------------------------------------------------------
     sigma_u |  1.0561892
     sigma_e |  .09753735
         rho |  .99154389   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The within entity errors u_i are correlated with the regressors in the fixed effects model.

If this number is < 0.05 then your model is ok. This is an *F-test* to see whether all the coefficients in the model are jointly different than zero.

Beta coefficients indicate the change in the output (y) when the predictors change one unit over time. In this example, all the variables are log-transformed, the interpretation is: when the predictor increases 1% over time, the output (y) changes β% (elasticity).

Two-tail p-values test the hypothesis that each coefficient is different from 0 (according to its *t*-value).
A value lower than 0.05 will reject the null and conclude that the predictor has a significant effect on the outcome (95% significance).

Intraclass correlation (rho), shows how much of the variance in the output is explained by the difference across entities. In this example is 99%.

sigma_u = sd of residuals within groups u_i
sigma_e = sd of residuals (overall error term) e_{it}

$$rho = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2}$$

4. Inference in Fixed Effects Models

Within Estimator

- Many text books devote considerable time to "efficiency" results (=whether this or that estimator has the smallest variance).

- Problem: these results are developed under very unrealistic assumptions! therefore they are not very useful.

- For instance, consider this assumption:

$$FE3 : Var(\varepsilon_i | X_i, c_i) = \sigma_\varepsilon^2 I_T$$

where I_T is the $T \times T$ identity matrix.

- Under FE3, the within estimator is efficient. But is FE3 a good/necessary assumption?

Within Estimator: Inference, II

- FE3 assumes two things:

  1) Homokedasticity and

  2) lack of serial correlation.

- Are these good assumptions?

- NO! they are very demanding

- Bottom line: never consider FE3 to be true in applications!

- or: don't worry about efficiency, worry about computing realistic standard errors.

Within Estimator: Inference, III

- In the following we relax the assumptions if FE3.

- Goal: compute standard errors of our β estimates that are robust to violations of FE3.

  - If 1) doesn't hold (but 2) does (heterokedasticity but no serial correlation): compute robust standard errors

  - If 1) and 2): compute clustered standard errors

Robust Standard Errors

- Robust standard errors= standard errors that take into account that there could be heteroskedasticity in the residual term.

- Always suspect heteroskedasticity in any regression you run (it's straightforward to compute s.e. that are robust to that).

- Under heteroskedasticity:

$$FE3' : \mathrm{Var}\big(\varepsilon_{it} \mid X_i, c_i\big) \;=\; \sigma^2_{\varepsilon,it} \;>\; 0, \quad \text{finite,}$$

and (no serial correlation)

$$\mathrm{Cov}\big(\varepsilon_{it},\, \varepsilon_{is} \mid X_i, c_i\big) \;=\; 0 \quad \forall\, s \neq t.$$

- Heteroskedasticity-Robust (Eicker–White) Variance:

$$\widehat{\mathrm{Var}}(\hat{\beta}) \;=\; (X_{\mathrm{within}}' X_{\mathrm{within}})^{-1} \left(\sum_{i,t} \hat{\varepsilon}_{it}^{\,2}\, X_{\mathrm{within},it}\, X_{\mathrm{within},it}' \right) (X_{\mathrm{within}}' X_{\mathrm{within}})^{-1},$$

where

$$X_{\mathrm{within},it} = X_{it} - \bar{X}_i, \qquad \hat{\varepsilon}_{it} = \tilde{\varepsilon}_{it}.$$

- **Interpretation:**
  - This adjusts for any form of heteroskedasticity in ε_{it}.
  - It does not account for correlation across t within each i (i.e., no clustering).
  - In software, this is often labeled `robust` or `HC` standard errors without clustering.

- Is this "enough" to get reasonable standard errors?

- In most instances, it's not

Clustered standard errors

- When using panel data you also have to suspect serial correlation. Why?
- We might assume that individuals are i.i.d. across themselves, but this assumption doesn't make sense within-individuals.
- Since an individual is correlated with herself over the T observations → serial correlation.
- We need to account for this in the standard errors.
- clustered standard errors: s.e. developed under the assumption that within-individuals there could be arbitrary correlation. This allows for serial correlation AND heteroskedasticity.

- In this case FE3 becomes:

$$FE'' : \mathrm{Var}\big(\varepsilon_i \mid X_i, c_i\big) \ = \ \Omega_{\varepsilon,i}(X_i),$$

which is positive definite (p.d.) and finite.

- FE" is good because s.e. derived under this assumption are also valid under FE and FE'!

- Under FE" you should compute clustered robust standard errors.

- This type of s.e. allow for heteroskedasticity AND within serial correlation.

- For more details on the computation of these s.e. see the notes in the website of the course.

5.2. Other estimation approaches for panel data models

Other estimation approaches: Roadmap

1. Pooled OLS

2. Between estimator

3. Random Effects

5.2. Other estimation approaches for panel data models

- All the methods that we'll see now DO NOT allow for correlation between the regressors and the fixed effects.

- As a result, they cannot help solving the OVB as FE can!

- They are only appropriate under stringent assumptions over c_i. Let's revise them quickly.

  1. Pooled OLS

  2. Between estimator

  3. Random Effects

Pooled OLS

- The model:

$$y_{it} = c + X_{it} + \varepsilon_{it} \qquad (1)$$

- c is assumed to be constant, therefore $corr(c, X_i) = 0$
- This method ignores the panel structure of the data
- As mentioned earlier, OLS can be employed.
- X_{it} can contain time invariant variables; c can be estimated consistently (as opposed to FE!)

$$\begin{pmatrix} \hat{c}_{\text{POLS}} \\ \hat{\beta}_{\text{POLS}} \end{pmatrix} = (W'W)^{-1}W'y,$$

where $W = [\iota_{NT}\ X]$ and ι_{NT} is an $NT \times 1$ vector of ones.

- But, big drawback: everything depends on $c_i = c$ being constant across i (very stringent assumption).

Between Estimator

Pooled OLS vs. Between Estimator

- **Pooled OLS** uses variation over both time and cross-sectional units to estimate β.
- **Between Estimator** uses just the cross-sectional variation.

■ How it works: consider the individual-Specific Effects Model:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}.$$

■ Average the data over time ($t = 1, \ldots, T$), it gives

$$\overline{y}_i = c_i + \overline{X}_i\beta + \overline{\varepsilon}_i,$$

which can be rewritten as the between model:

$$\overline{y}_i = c + \overline{X}_i\beta + \left(c_i - c + \overline{\varepsilon}_i\right), \quad i = 1, \ldots, N,$$

where $\overline{y}_i = \dfrac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{X}_i = \dfrac{1}{T}\sum_{t=1}^{T} X_{it}$, $\overline{\varepsilon}_i = \dfrac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$.

Between Estimator:

- OLS regression of $\overline{y}_i$ on an intercept and $\overline{X}_i$.
- Uses variation between individuals; analogous to cross-section regression (special case $T = 1$).
- Consistent if $\overline{X}_i$ is uncorrelated with $\left(c_i - c + \overline{\varepsilon}_i\right)$
- Inconsistent under fixed effects if c_i is correlated with X_{it} and hence $\overline{X}_i$.

Random Effects Models

- Consider the individual-specific effects model:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it},$$

- Key Random effects assumption: c_i and ε_{it} are uncorrelated.

- It would be possible to estimate this by pooled OLS (it is consistent)

- But notice that c_i is in the error term: heterokedasticity!

- Therefore, feasible GLS improves efficiency under the RE model.

Random Effects: Key Assumptions

Model Setup:

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it},$$

where c_i is unobserved and ε_{it} is idiosyncratic.

Assumption RE.1:

(a) Strict exogeneity:

$$E(\varepsilon_{it} \mid X_i, c_i) = 0 \quad \text{for all } t.$$

(b) Orthogonality between c_i and X_i:

$$E(c_i \mid X_i) = 0.$$

Why RE.1?

- Allows treating c_i as part of the error term.
- Ensures strict exogeneity needed for consistent GLS.

Random Effects: Estimation Procedure

Error Structure:

$$v_{it} = c_i + \varepsilon_{it}, \quad \text{with } W = E(v_i v_i') = \sigma_\varepsilon^2 I_T + \sigma_c^2 \mathbf{1}_T \mathbf{1}_T'$$

$$W = \begin{pmatrix} \sigma_c^2 + \sigma_\varepsilon^2 & \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_\varepsilon^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_\varepsilon^2 & \cdots & \sigma_c^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_\varepsilon^2 \end{pmatrix}_{T \times T} .$$

The matrix W has the <u>random effects structure</u>, depending on two parameters: σ_c^2 and σ_ε^2.

Assumptions for Efficiency:

- **RE.2:** Rank condition for consistent GLS: rank(X$'_i W^{-1} X_i) = K$

- **RE.3:** Constant conditional variances and homoskedasticity of c_i.

$$\text{(a) } E\Big[(\varepsilon_i \varepsilon_i') \mid X_i, c_i\Big] = \sigma^2_\varepsilon \, I_T.$$

$$\text{(b) } E\Big[c_i^2 \mid X_i\Big] = \sigma^2_c.$$

Estimation Steps:

1. Use pooled OLS to get an initial consistent estimate $\hat{\beta}_{\text{POLS}}$.
2. Compute residuals $\hat{v}_{it}$ and estimate σ_ε^2 and σ_c^2. [Check Wooldridge, page 734 for details]
3. Form the feasible GLS weight matrix
$$\widehat{W} = \hat{\sigma}_u^2 I_T + \hat{\sigma}_c^2 \mathbf{1}_T \mathbf{1}_T^\top .$$
4. Obtain the **Random Effects estimator**:
$$\hat{\beta}_{RE} = \Big(\sum_{i=1}^N X_i^\top \widehat{W}^{-1} X_i\Big)^{-1} \Big(\sum_{i=1}^N X_i^\top \widehat{W}^{-1} y_i\Big).$$

Properties:

- Two-step FGLS procedure.
- Consistent under RE.1 and rank conditions.
- Efficient under Assumptions RE.1–RE.3.

■ Variations:

- If RE.3 doesn't hold and there's heteroskasticity: use robust s.e. (sandwich variance-covariance matrix)
- Efficiency is lost if RE.3 fails
- You should always allow for deviations from RE.3 and compute standard errors accordingly, therefore efficiency is lost.

FE vs RE: Which to Use?

- **General advice:** Use Fixed Effects.

- FE is consistent whether or not Cov$(c_i, X_{it}) = 0$.

- RE requires the stronger assumption that c_i is uncorrelated with all regressors.

- FE is more robust — if in doubt, use FE.

- **When to consider RE:**

- You want to estimate the effect of a variable that *doesn't vary over time* (e.g., gender, race, country of birth).

- FE cannot identify effects of time-constant variables (they are absorbed by c_i).

- RE allows estimation of $z_i\gamma$, but only if you believe Cov$(c_i, z_i) = 0$.

RE or FE models?

- In theory, it's possible to test for FE vs RE (Hausman test)

- But in practice, the (standard) test is only valid under very stringent assumptions (homokedasticity, cannot include time dummies), so not very reliable either.

- Bottom line: FE models should be your default option!

RE or FE models?, II

Hausman Test

- Logic of the test:

- If the RE assumption is true (H_0), both the RE estimator and the FE estimators are consistent.

- if it's false, only the FE model is consistent (H_1).

- Therefore, under H_0, the difference between the RE and the FE estimators should be small. Under H_1, it should be large.

- The test rejects the null hypothesis if there are large deviations between the FE and the RE estimators.

Hausmann Test

- It's based on the (standardized) difference between the FE and the RE estimators.

$$H = (\hat{\beta}_{1,RE} - \hat{\beta}_{1,FE})' \left(\hat{V}[\hat{\beta}_{1,FE}] - \hat{V}[\hat{\beta}_{1,RE}]\right)^{-1} (\hat{\beta}_{1,RE} - \hat{\beta}_{1,FE}) \quad (2)$$

where V(.) denotes the variance of the relevant estimator.

- Under R3.1–RE.3 if H_0 is true: asymptotic distribution: χ^2.

- The test rejects H_0 (RE) if the value of the test is larger than the χ^2 critical value.

Hausman Test: A Caveat

- The standard Hausman test compares FE and RE estimators to test $H_0 : \text{Cov}(c_i, X_{it}) = 0$.
- **Problem:** The test assumes homoskedasticity and no clustering.
- Under heteroskedasticity or clustered errors, the standard test is invalid.
- There are robust alternatives (Wooldrige) to the standard tests
- **Practical advice:** If you suspect $\text{Cov}(c_i, X_{it}) \neq 0$, just use FE.

- **Robust alternative (Wooldridge):**
- Run the RE regression.
- Add within-transformed variables $(\tilde{X}_{it} = X_{it} - \bar{X}_i)$ as additional regressors.
- Test if coefficients on $\tilde{X}_{it}$ are jointly zero using cluster-robust standard errors.
- **Practical advice:** If you suspect Cov$(c_i, X_{it}) \neq 0$, just use FE.

Key Takeaways

- This handout introduces the basics of panel data models

- Advantage of panel data: allow to control for unobserved, time-invariant, heterogeneity across the units

- General tips:
  - Use FE models estimated within, FD or dummy variable approach
  - Other methods, such as RE, pooled OLS, between estimator, are not consistent in the general case!
  - Use s.e. that are valid under general assumptions: clustered s.e.

- Be careful with the interpretation of these models (they exploit within-unit variation exclusively!)

- The use of panel data models also has drawbacks
  - Measurement error problems can become more acute
  - Useful variation can be eliminated by the FE: estimates can be estimated very imprecisely, large s.e., etc.