Regression with Panel Data

ECON 3500
Spring 2026

Learning objectives

- Understand differences between types of data
- Estimate first-difference regressions with time series and panel models
- Estimate and interpret regressions with fixed effects
- Understand the difference between fixed effects and pooled OLS estimates

Table of Contents

Types of data

First differencing

Fixed effects and pooled OLS

Examples

Activity

Types of data

- Cross-sectional
  - Multiple units, one time period
- Time series
  - One unit, multiple time periods
- Panel / longitudinal
  - Multiple units, multiple time periods
  - Same units over time!
- Repeated cross-section
  - Multiple units, multiple time periods
  - Different units over time!

Examples! :)

- First: `ssc install bcuse`
- Cross-section: `bcuse wage1`
  - Survey of 526 workers, with wage, educ, exper, others
- Time series: `bcuse intdef`
  - Annual data (1948-2003), with 3-month T-bill rate, inflation, and deficit
- Panel: `bcuse wagepan`
  - 545 individuals observed each year from 1980 to 1987, with wage, hours, educ, others
- Repeated cross-section: `bcuse kielmc`
  - House sales before and after an incinerator was built, with square feet, age, rooms, distance from incinerator

Panels and repeated cross-sections

- Panel data
  - Same entities over time
  - Independent cross-sectionally: Bob in 1990 is independent of Jane in 1990
  - Dependent temporally: Bob in 1990 is related to Bob in 1991
- Repeated cross-section
  - Different entities over time. Optimally a random draw from the population at each point
  - Independent over time and cross-sectionally: Bob in 1990 is independent of Jane in 1990, and Bob/Jane in 1990 is independent of Charles in 1991
- Which one do we want? Depends on your question:
  - We intend to estimate the effect of a new drug on individuals' all-cause mortality
  - We intend to estimate the effect of a zoning reform on the income distribution of a neighborhood

Table of Contents

Types of data

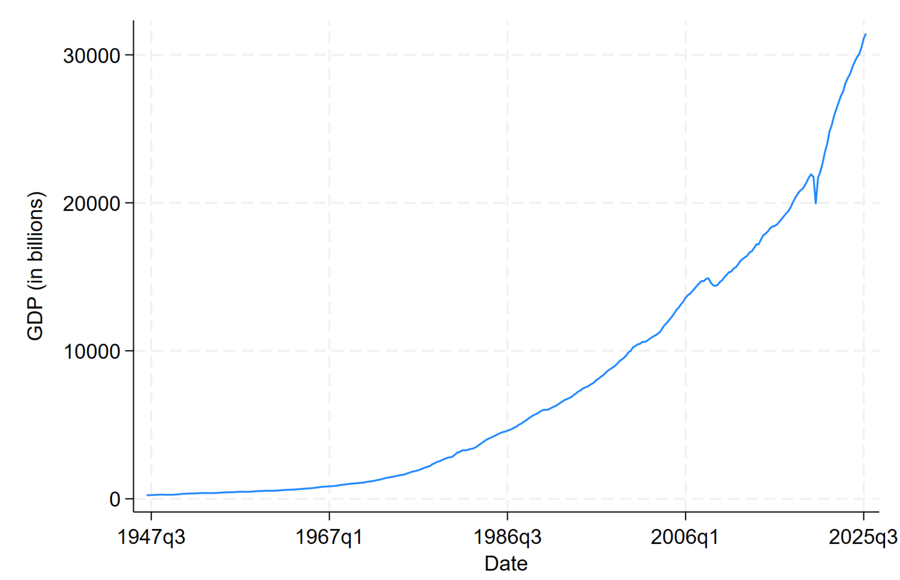
First differencing

Fixed effects and pooled OLS

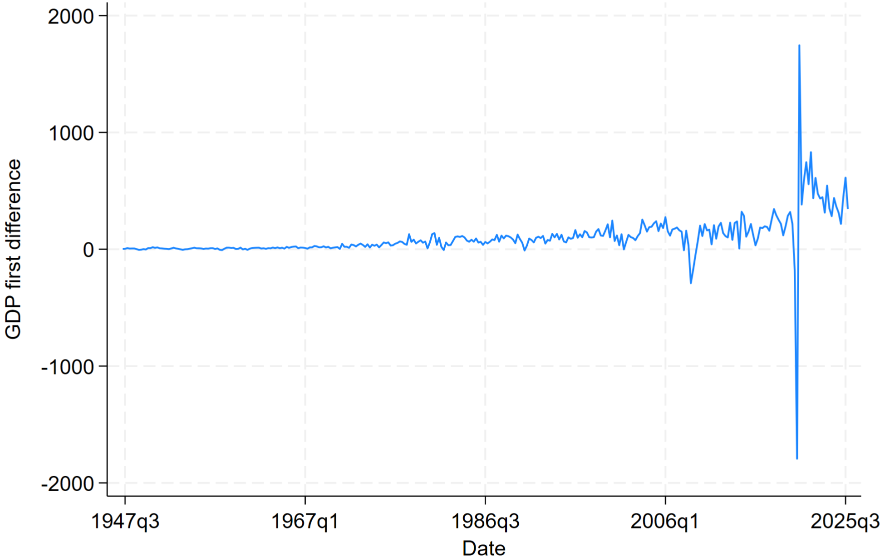
Examples

Activity

First differencing



First differencing



First differencing

- Instead of looking at the *level* of a variable, we are looking at the *change* in that variable
- $\Delta x_t = x_t - x_{t-1}$
- How does the *change* in unemployment rate affect the *change* in crime levels?
$\Delta \widehat{crmrte}_t = 15.4 + 2.22 \Delta unemp_t$
- How does this differ from the not-differenced $\widehat{crmrte}_t = \beta_0 + \beta_1 unemp_t$?
- They ask different questions:
    - Levels: In periods when unmemployment is 1pp higher, crime tends to be β_1 higher/lower
    - Differences: When unemployment increases by 1pp from one period to the next, crime tends to change by 2.22
- A note: if there is little variation over time, our first-differenced estimator will be less precise

Table of Contents

Types of data

First differencing

Fixed effects and pooled OLS

Examples

Activity

Motivation for fixed effects

- Data on annual GDP and pollution levels of 27 European countries from 1990 to 2010
- $pollution_{it} = \beta_0 + \beta_1 GDP_{it} + u_{it}$
- This asks: do countries with higher GDP have higher pollution?
- (this is called pooled OLS)
- But what if I want to know what the effect of a change in GDP is on pollution *within* a country?
- Pooled OLS does not handle that there are 27 different countries
- Now what? Fixed effects!!!

Fixed effects

- Data on annual GDP and pollution levels of 20 European countries from 1990 to 2010
- Unit fixed effects: $pollution_{it} = \alpha_i + \beta_1 GDP_{it} + u_{it}$
- Giving each country their own intercept
- Equivalent to centering each country at their mean: with FEs, we are looking at deviations from their mean
- Now, β_1 asks: on average, when a country's GDP deviates from its mean by 1, how much does pollution tend to deviate from its mean?

Fixed effects estimation

- Pooled OLS: $pollution_{it} = \gamma_0 + \gamma_1 GDP_{it} + u_{it}$
- $\gamma_1 = \frac{\text{cov}(pollution_{it}, GDP_{it})}{\text{var}(GDP_{it})} = \frac{\mathbb{E}[(pollution_{it} - \overline{pollution}_{it}, GDP_{it} - \overline{GDP}_{it})]}{\mathbb{E}[(GDP_{it} - \overline{GDP}_{it})^2]}$

- Unit fixed effects: $pollution_{it} = \alpha_i + \beta_1 GDP_{it} + u_{it}$
- $(pollution_{it} - \overline{pollution_i}) = \beta_1 (GDP_{it} - \overline{GDP}_i) + (u_{it} - \overline{u}_i)$
- $\beta_1 = \frac{\mathbb{E}[(pollution_{it} - \overline{pollution}_i)(GDP_{it} - \overline{GDP}_i)]}{\mathbb{E}[(GDP_{it} - \overline{GDP}_i)^2]}$
- Because $\alpha_i - \bar{\alpha}_i = 0$, the FE drops out

- What's the difference?
- Pooled OLS looks at variations around overall mean
- Fixed effects look at variations around each unit's mean
- animation

Fixed effects example

Pooled OLS: Is sunny weather associated with more steps? (between-estimator, does not control for individual mean)

	Mon	Tue	Wed	Thu	Fri	Mean
Weather	Sunny	Rainy	Sunny	Cloudy	Rainy	
Bob	8200	5100	7800	6400	5500	6600
Jane	12400	9800	12100	10500	9200	10800

Fixed effects: Is sunny weather associated with a person walking more steps than that person's average? (within-estimator, controls for individual mean)

	Mon	Tue	Wed	Thu	Fri	Mean
Weather	Sunny	Rainy	Sunny	Cloudy	Rainy	
Bob	$+1600$	-1500	$+1200$	-200	-1100	0
Jane	$+1600$	-1000	$+1300$	-300	-1600	0

Pooled OLS and fixed effects

- Pooled OLS is the "between" estimator: looks at variations between units
- Fixed effects is the "within" estimator: looks at variations within units

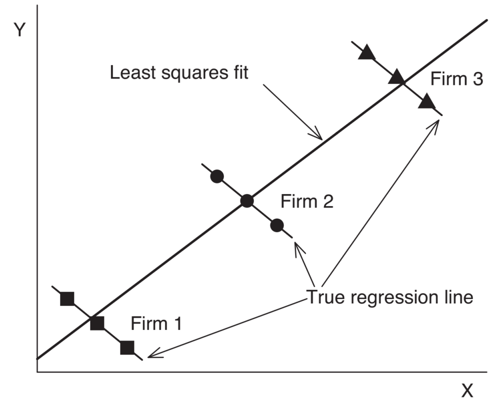


FIGURE 17.1 Scatter plot and pooled regression line

Some notes

- Fixed effects control for factors which vary between units but not within units
- Only covers stuff that does not vary over time: absorbed by the fixed effect
- Fixed effects and pooled OLS can have opposite signs!! (Simpson's paradox)
- You can use multiple sets of fixed effects (e.g., unit and time)
- Fixed effects cannot solve reverse causality/simultaneity

Table of Contents

Types of data

First differencing

Fixed effects and pooled OLS

Examples

Activity

How to read subscripts

- Cross-sectional: $y_i = \beta_0 + \beta_1 x_i + u_i$
- Time series: $y_t = \beta_0 + \beta_1 x_t + u_t$
- Panel: $y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}$
- Panel with unit FEs: $y_{it} = \alpha_i + \beta_1 x_{it} + u_{it}$
- With the person next to you,
- write a possible regression equation for your project, with subscripts!

Does income affect democracy?

TABLE 3—FIXED EFFECTS RESULTS USING POLITY MEASURE OF DEMOCRACY

	Base sample, 1960–2000								
	Five-year data					Annual data	Ten-year data		Twenty-year data
	Pooled OLS (1)	Fixed effects OLS (2)	Anderson-Hsiao IV (3)	Arellano-Bond GMM (4)	Fixed effects OLS (5)	Fixed effects OLS (6)	Fixed effects OLS (7)	Arellano-Bond GMM (8)	Fixed effects OLS (9)
	Dependent variable is democracy								
Democracy_{t-1}	0.749	0.449	0.582	0.590		[0.00]	0.060	0.309	–0.516
	(0.034)	(0.063)	(0.127)	(0.106)			(0.091)	(0.134)	(0.165)
Log GDP per capita $_{t-1}$	0.053	–0.006	–0.413	–0.351	–0.011	[0.53]	0.007	–0.368	–0.1260
	(0.010)	(0.039)	(0.127)	(0.127)	(0.055)		(0.070)	(0.190)	(0.164)
Hansen *J* test			[0.03]	[0.03]				[0.01]	
AR(2) test			[0.39]	[0.39]				[0.38]	
Implied cumulative effect of income	0.211	–0.011	–0.856	–0.856			0.007	–0.533	–0.083
	[0.00]	[0.89]	[0.00]	[0.00]			[0.92]	[0.04]	[0.45]
Observations	854	854	747	747	880	3701	419	302	168
Countries	136	136	114	114	136	134	114	107	100
R-squared	0.77	0.82			0.77	0.96	0.77		0.87

Do unions increase wages?

Table III. Wage regressions with union effects

Variable	[1] OLS	[2] OLS	[3] FE	[4] FE	[5] OLS	[6] OLS	[7] OLS	[8] OLS
Constant	0·224	0·388*			0·273	0·360*	0·282	0·363*
	(0·128)	(0·158)			(0·156)	(0·158)	(0·156)	(0·159)
Union	0·146*	0·177*	0·079*	0·080*	0·392*	0·389*	0·285*	0·311*
	(0·026)	(0·026)	(0·018)	(0·018)	(0·087)	(0·084)	(0·088)	(0·085)
School	0·090*	0·073*			0·083*	0·070*	0·082*	0·070*
	(0·008)	(0·010)			(0·010)	(0·010)	(0·010)	(0·010)
Exper	0·076*	0·057*	0·112*	0·111*	0·051*	0·049*	0·053*	0·050*
	(0·011)	(0·018)	(0·008)	(0·008)	(0·018)	(0·018)	(0·018)	(0·018)
Exper2	−0·0022*	−0·0018	−0·0041*	−0·0041*	−0·0016	−0·0015	−0·0017	−0·0016
	(0·0008)	(0·0009)	(0·0005)	(0·0006)	(0·0009)	(0·0009)	(0·0009)	(0·0009)
Hisp	−0·059	−0·047			−0·079	−0·061	−0·063	−0·049
	(0·042)	(0·042)			(0·042)	(0·042)	(0·042)	(0·042)
Black	−0·155*	−0·126*			−0·189*	−0·154*	−0·171*	−0·141*
	(0·044)	(0·044)			(0·046)	(0·045)	(0·046)	(0·046)
Rural	−0·131*	−0·114*	0·050	0·048	−0·131*	−0·113*	−0·131*	−0·116*
	(0·031)	(0·032)	(0·032)	(0·027)	(0·032)	(0·032)	(0·032)	(0·032)
Mar	0·110*	0·102*	0·040*	0·038*	0·102*	0·094*	0·107*	0·097*
	(0·024)	(0·024)	(0·017)	(0·017)	(0·024)	(0·024)	(0·024)	(0·024)
Health	−0·058	−0·032	−0·017	−0·010	−0·036	−0·011	−0·037	−0·011
	(0·062)	(0·062)	(0·044)	(0·044)	(0·062)	(0·042)	(0·062)	(0·042)
C_i	=	=			−0·050*	−0·030	−0·051*	−0·033
					(0·024)	(0·023)	(0·025)	(0·023)
C_{it}	=	=	=	=	−0·109*	−0·113*	−0·072*	−0·084*
					(0·031)	(0·031)	(0·032)	(0·032)
C_i^2	=	=			=	=	0·034*	0·027*
							(0·013)	(0·012)
C_{it}^2	=	=	=	=	=	=	−0·0005	−0·0008
							(0·0058)	(0·0061)

Did more equitable land ownership increase birth rates in 17-19th century Japan?

$$Y_{h,v,t} = \beta_0 + \beta_1 f(Landownership_{h,v,t}) + \beta_2 X_{h,v,t} + \theta_{v,t} + \varepsilon_{h,v,t}. \quad (9)$$

Here, Y denotes the demographic variable of interest, v denotes village, h denotes household, t denotes time, and X is a set of control variables. I include a village-year fixed effect to control for factors such as changes in local food prices or disease that can affect fertility or mortality in certain years. I do not include individual or household fixed effects in my main regression tables because most of the variation is between households.[5] This is because landownership was a slow moving variable as I show in a later regression. Nonetheless, I show the results with individual/household fixed effects in the Supplemental Appendix G.

	Number of births per 1,000 HHs	
	OLS	IV
	(1)	(2)
Landownership	6.643	11.269
	(2.132)	(2.999)
Landownership^2	−0.074	−0.397
	(0.108)	(0.190)
Village-year FE	Yes	Yes

Table of Contents

Types of data

First differencing

Fixed effects and pooled OLS

Examples

Activity