ECON5128 MRes Econometrics 2

Weeks 1–2: Panel data

Santiago Montoya-Blandón

January 14 & 21, 2025

Outline

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Introduction

- A *panel dataset* contains observations on multiple units (individuals, states, companies, etc.) where *each unit is observed at two or more points in time*
- Hypothetical examples:
  - Data on 420 English school districts in 1999 and again in 2000, for 840 observations total.
  - Data on 50 U.S. states, each state is observed across 3 years, for a total of 150 observations.
  - Data on income for 1000 Scottish individuals, in four different months, for 4000 observations total.
- Another term for panel data is *longitudinal data*

Notation for panel data

- A double subscript distinguishes units and time periods:
  - i = unit (e.g., state), n = number of total units $\Rightarrow i = 1, \ldots, n$
  - t = time period (e.g., year), T = number of time periods $\Rightarrow t = 1, \ldots, T$
- The data with 1 regressor x and an outcome y can be represented as

$$(x_{it}, y_{it}), \quad i = 1, \ldots, n, \; t = 1, \ldots, T$$

- Panel data with k regressors:

$$(x_{1it}, x_{2it}, \ldots, x_{kit}, y_{it}), \quad i = 1, \ldots, n, \; t = 1, \ldots, T$$

- A panel is called *balanced* if there are no missing observations: all variables are observed for all units and time periods

Panel Data: Balanced Example

Unit ID	Time	Wage (£ per hour)	Sex	Experience
1	Jul 2018	12.0	F	5
1	Jul 2020	14.5	F	7
1	Jul 2022	17.0	F	9
2	Jul 2018	11.2	M	0
2	Jul 2020	11.2	M	2
2	Jul 2022	12.5	M	4
⋮	⋮	⋮	⋮	⋮
n	Jul 2018	25.3	F	12
n	Jul 2020	28.0	F	14
n	Jul 2022	29.7	F	16

Panel Data: Unbalanced Example

Unit ID	Time	Wage (£ per hour)	Sex	Experience
1	Jul 2018	12.0	F	5
1	Jul 2020	14.5	F	7
1	Jul 2022	17.0	F	9
2	Jul 2018	—	M	0
2	Jul 2020	11.2	M	2
2	Jul 2022	12.5	M	4
⋮	⋮	⋮	⋮	⋮
n	Jul 2018	—	F	—
n	Jul 2020	28.0	F	—
n	Jul 2022	29.7	F	—

Why are panel data useful?

With panel data we can control for factors that:

(i) Vary across units but do not vary over time

(ii) Could cause omitted variable bias if they are omitted

(iii) Time-constant unobservables that cannot be included in the regression

Key ideas

1. If an omitted variable does not change over time, then any changes in outcome (y) over time cannot be caused by the omitted variable
2. Panel data lets us eliminate omitted variable bias if the omitted variables are constant over time within a given unit

Example panel data set: U.S. traffic deaths and alcohol taxes

Observational unit: a year in a U.S. state

- 48 U.S. states, so $n = \#$ of units $= 48$
- 7 years $(1982, \ldots, 1988)$, so $T = \#$ of time periods $= 7$
- Balanced panel, so total $\#$ observations $= 48 \times 7 = 336$

Variables:

- Outcome y: Traffic fatality rate measured as $\#$ traffic deaths per 10,000 state residents in state i in year t
- Main covariate of interest x_1: Tax on beer (in 1988 USD$ per case)
- Other covariates $x_2, \ldots, x_k$: legal driving age, drunk driving laws, etc.

Example: U.S. traffic deaths and alcohol taxes in **1982**

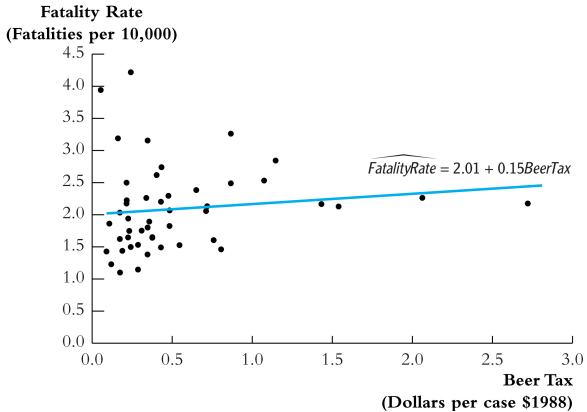

Figure: *1982 data: Higher alcohol taxes, more traffic deaths?*

Example: U.S. traffic deaths and alcohol taxes in **1988**

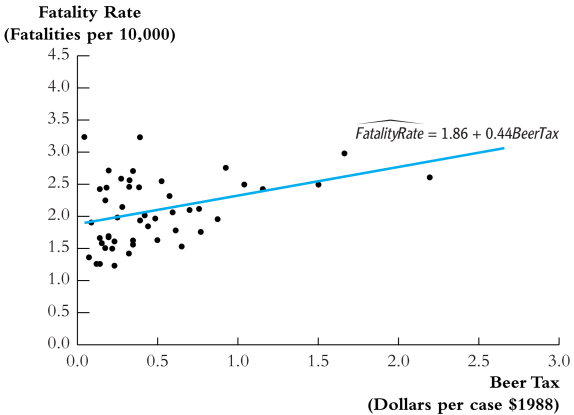

Figure: *1988 data: Higher alcohol taxes, more traffic deaths?*

Example: U.S. traffic deaths and alcohol taxes

Question for you

Why do we observe higher traffic deaths in states that have higher alcohol taxes?

Example: U.S. traffic deaths and alcohol taxes

Question for you

Why do we observe higher traffic deaths in states that have higher alcohol taxes?

Other factors that determine traffic fatality rate:

- Quality (age) of automobiles
- Quality of roads
- "Culture" around drinking and driving
- Density of cars on the road

These omitted factors could cause omitted variable bias.

Econometric issue: Omitted variable bias

- Suppose the true model is

$$y_i = \beta_x x_i + \beta_z z_i + \varepsilon_i$$

- but we estimate

$$y_i = b_x x_i + u_i$$

- How does omitting z_i from the model affect the least-squares (LS) estimate of b_x? Is $\widehat{b}_x$ a reliable estimate of the true β_x?
- Model in matrix notation: $y = X\beta_x + Z\beta_z + \varepsilon$

$$\begin{aligned}\widehat{b}_x &= (X^\top X)^{-1} X^\top y = (X^\top X)^{-1} X^\top (y = X\beta_x + Z\beta_z + \varepsilon) \\ &= \beta_x + \beta_z (X^\top X)^{-1} X^\top Z + (X^\top X)^{-1} X^\top \varepsilon\end{aligned}$$

Econometric issue: Omitted variable bias

$$\widehat{b}_x = \beta_x + \beta_z \left(x'x\right)^{-1} x'z + \left(x'x\right)^{-1} x'\varepsilon$$

Thus, $\widehat{b}_x$ is an unbiased estimate of β_x (i.e. $\mathbb{E}[\widehat{b}_x] = \beta_x$) if

1. $\mathbb{E}[x_i \varepsilon_i] = 0$; and either
   - 1.1 $\beta_z = 0$; or
   - 1.2 $\mathbb{E}[x_i z_i] = 0$

In words,

1. Observed regressor x_i and error ε_i are uncorrelated (usual assumption in linear regression model, satisfied when no further omitted variables)
   - 1.1 Exclusion restriction: omitted factor z_i not directly relevant for outcome y_i
   - 1.2 Observed (x_i) and omitted (z_i) regressors are uncorrelated

Example #1: Traffic density

Suppose:

(i) High traffic density (z) means more traffic deaths (y)

(ii) (Western) states with lower traffic density have lower alcohol taxes (x)

Omitted variable bias exists in this scenario as traffic density is a confounder:

- From (i), we have $\beta_z > 0$
- From (ii), we have $\text{Corr}(x_i, z_i) > 0$
- Specifically, large taxes could reflect high traffic density
- OLS coefficient would be *biased* positively
- If sufficiently biased, we could have $\beta_x > 0 \Rightarrow$ larger taxes, more deaths

Example #2: Cultural attitudes towards drinking and driving

(i) Culture towards drinking and driving (z) is arguably a determinant of traffic deaths (y)

(ii) These attitudes are potentially correlated with the beer tax (x)

- States where drinking and driving are not frowned upon might experience higher deaths ($\beta_z \neq 0$)
- Attitudes towards drinking and driving are shaped by state values, and laxer states are likely to also have laxer tax systems ($\text{Corr}(x_i, z_i) \neq 0$)
- Omitted variable bias expected to occur again: taxes could pick up the effect of cultural attitudes towards drinking
- OLS coefficient of a regression of traffic deaths on taxes would be biased

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Example: U.S. traffic deaths and alcohol taxes

Consider the panel data model

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + \varepsilon_{it}$$

Z_i is a factor that does not change over time (density), at least during the years on which we have data.

- Suppose Z_i is not observed, so its omission could result in omitted variable bias.
- The effect of Z_i can be eliminated using just $T = 2$ years.

Example: U.S. traffic deaths and alcohol taxes

The key idea:

Any change in the fatality rate from 1982 to 1988 cannot be caused by Z_i, because Z_i (by assumption) does not change between 1982 and 1988.

The math: Consider fatality rates in 1988 and 1982,

$$FatalityRate_{i,1982} = \beta_0 + \beta_1 BeerTax_{i,1982} + \beta_2 Z_i + \varepsilon_{i,1982}$$
$$FatalityRate_{i,1988} = \beta_0 + \beta_1 BeerTax_{i,1988} + \beta_2 Z_i + \varepsilon_{i,1988}$$

Subtracting 1988-1982, eliminates the effect of Z_i:

$$FatalityRate_{i,1988} - FatalityRate_{i,1982} = \beta_1(BeerTax_{i,1988} - BeerTax_{i,1982}) + (\varepsilon_{i,1988} - \varepsilon_{i,1982})$$

Example: U.S. traffic deaths and alcohol taxes

$$\underbrace{FatalityRate_{i,1988} - FatalityRate_{i,1982}}_{\Delta FatalityRate_i} =$$

$$\beta_1 \underbrace{(BeerTax_{i,1988} - BeerTax_{i,1982})}_{\Delta BeerTax_i} + \underbrace{(\varepsilon_{i,1988} - \varepsilon_{i,1982})}_{\Delta \varepsilon_i}$$

- Previous equation can be estimated by OLS even without data on Z_i
- The new error term, $\Delta\varepsilon_i := \varepsilon_{i1988} - \varepsilon_{i1982}$, is assumed uncorrelated with either $BeerTax_{i,1988}$ or $BeerTax_{i,1982}$
- This differences regression *does not have an intercept* — it was eliminated by the subtraction step
- ALL information that is constant across time is also discarded and its effects cannot be estimated (e.g., effect of culture on traffic deaths)

Example: Traffic deaths and beer taxes

1982 data:

$$\underset{}{FatalityRate} = \underset{(0.15)}{2.01} + \underset{(0.13)}{0.15}\, BeerTax \qquad (n = 48)$$

1988 data:

$$\underset{}{FatalityRate} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44}\, BeerTax \qquad (n = 48)$$

Difference regression:

$$FR_{1988} - FR_{1982} = \underset{(0.065)}{-0.72} + \underset{(0.36)}{-1.04}\,(BT_{1988} - BT_{1982}) \quad (n = 48)$$

An intercept is included in this differences regression that allows for the mean change in FR to be nonzero - more on this later...

$\Delta FatalityRate$ vs $\Delta BeerTax$

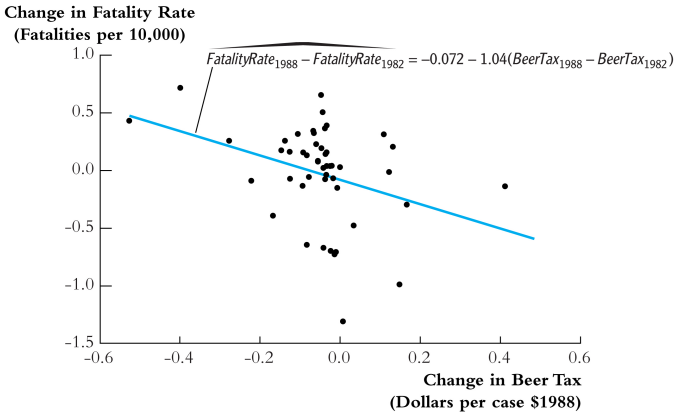

Figure: *Note that the intercept is nearly zero...*

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Panel Data with Multiple Time Periods

What if you have more than 2 time periods ($T > 2$)?

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + \varepsilon_{it}, \quad i = 1, \ldots, n; t = 1, \ldots, T$$

We can rewrite this in two useful ways:

1. Regression with fixed effects
2. Regression with $n - 1$ binary regressors

Fixed Effects Transformation

Population regression for California (that is, $i = CA$):

$$\begin{aligned} y_{CA,t} &= \beta_0 + \beta_1 x_{CA,t} + \beta_2 z_{CA} + \varepsilon_{CA,t} \\ &= (\beta_0 + \beta_2 z_{CA}) + \beta_1 x_{CA,t} + \varepsilon_{CA,t} \\ &= \alpha_{CA} + \beta_1 x_{CA,t} + \varepsilon_{CA,t} \end{aligned}$$

- $\alpha_{CA} := \beta_0 + \beta_2 z_{CA}$ does not change over time
- α_{CA} is the intercept for CA, and β_1 is the slope
- The intercept is unique to CA, but the slope is the same in all the states: parallel lines

Regression lines for each state in a picture

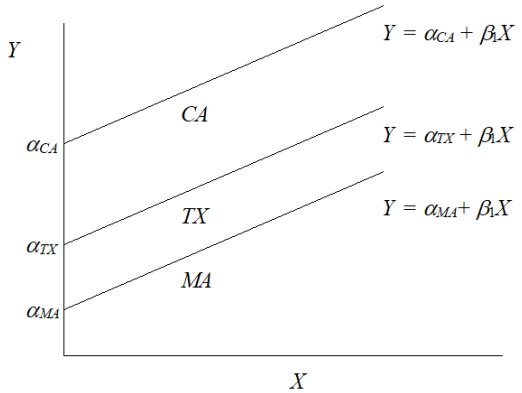

Figure: *Recall that shifts in the intercept can be represented using binary regressors...*

Dummy Variable Representation

Equivalently, we can write the model in binary regressor form:

$$y_{it} = \beta_0 + \gamma_{CA} DCA_i + \gamma_{TX} DTX_i + \beta_1 x_{it} + \varepsilon_{it}$$

- $DCA_i = 1$ if state is CA, it is $= 0$ otherwise
- $DTX_i = 1$ if state is TX, it is $= 0$ otherwise
- leave out DMA_i (why?)
- Estimate by OLS using robust standard errors and tests
- Dummy variable approach is impractical when n is very large (ok for $n = 50$ US states, but what if $n = 1000$?)

Summary: Two ways to write the fixed effects model

1. Dummy variable form:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma_1 D_{1,i} + \cdots + \gamma_{n-1} D_{n-1,i} + \varepsilon_{it},$$

where $D_{j,i} = \begin{cases} 1, & \text{if } j = i, \quad i,j \in \{1,\ldots,n\} \\ 0, & \text{otherwise} \end{cases}$

2. Fixed effects form:

$$y_{it} = \beta_1 x_{it} + \alpha_i + \varepsilon_{it}$$

α_i is called a *unit fixed effect* or *unit effect* – it is the constant (fixed) effect of being in unit (state) i

3. Connection between representations: $\alpha_i = \begin{cases} \beta_0 + \gamma_i, & \text{if } i \in \{1,\ldots,n-1\} \\ \beta_0, & \text{if } i = n \end{cases}$

Regression with Time Fixed Effects

An omitted variable might vary over time but not across states:

- Example: Safer cars (air bags, etc.); changes in national laws
- These produce intercepts that change over time
- Let s_t denote the combined effect of variables which changes over time but not states ("safer cars")
- The resulting population regression model is:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_3 s_t + \varepsilon_{it}$$

Time fixed effects only

Population regression at a specific year, say $t = 1982$:

$$\begin{aligned} y_{i,1982} &= \beta_0 + \beta_1 x_{i,1982} + \beta_3 s_{1982} + \varepsilon_{i,1982} \\ &= (\beta_0 + \beta_3 s_{1982}) + \beta_1 x_{i,1982} + \varepsilon_{i,1982} \\ &= \lambda_{1982} + \beta_1 x_{i,1982} + \varepsilon_{i,1982} \end{aligned}$$

- $\lambda_{1982} := \beta_0 + \beta_3 s_{1982}$ does not change over units
- λ_{1982} is the intercept in 1982, and β_1 is the slope
- The intercept is specific to the year 1982, but the slope is the same in all the states: parallel lines interpretation is preserved

Summary: Two ways to write time fixed effects

1. Dummy variable formulation:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \delta_2 B_{2,t} + \cdots + \delta_T B_{T,t} + \varepsilon_{it},$$

where $B_{s,t} = \begin{cases} 1, & \text{if } s = t, \quad t, s \in \{1, \ldots, T\} \\ 0, & \text{otherwise} \end{cases}$

2. Time effects formulation:

$$y_{it} = \beta_1 x_{it} + \lambda_t + \varepsilon_{it}$$

λ_t is called a *time fixed effect* or *time effect* – it is the constant (fixed) effect of all aggregate shocks occurring at time t

3. Connection between representations: $\lambda_t = \begin{cases} \beta_0, & \text{if } t = 1 \\ \beta_0 + \delta_t, & \text{if } t \in \{2, \ldots, T\} \end{cases}$

TWFE: Two-way fixed effects model

- Using both unit and time effects brings us to the *two-way fixed effects* formulation
- Workhorse panel data specification in econometrics (adding k covariates)

$$y_{it} = \alpha_i + \lambda_t + \beta_1 x_{1,it} + \cdots + \beta_k x_{k,it} + \varepsilon_{it}$$

- So far, no assumptions on statistical behaviour of α_i or λ_t, nor their relationship to controls $x_1, \ldots, x_k$
- Similarly, no assumptions on number of units (n) nor time periods (T)
- Estimates of slope coefficients and their properties heavily depend on assumptions and estimators used

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Sources of Variation

- Main challenge in estimating TWFE specification are the fixed effects
- Requires us to consider *sources of variation* in panel data:

$$\underbrace{\bar{y}_i := \frac{1}{T}\sum_{t=1}^{T} y_{it}}_{\text{Unit level}}, \quad \underbrace{\bar{y}_t := \frac{1}{n}\sum_{i=1}^{n} y_{it}}_{\text{Time level}}, \quad \text{and} \quad \underbrace{\bar{y} := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}}_{\text{Total mean}}$$

- Following variance decomposition (ANOVA) holds in panel data:

$$\underbrace{\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}-\bar{y})^2}_{\text{Total}} = \underbrace{\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2}_{\text{Within}} + \underbrace{\sum_{i=1}^{n} T(\bar{y}_i-\bar{y})^2}_{\text{Between}}$$

- Understanding standard panel data estimators requires a small digression to pooled Ordinary Least Squares (OLS)

Pooled Ordinary Least Squares (OLS)

- Consider simplest model with no time-constant confounders (Z_i) and no aggregate confounders (s_t)
- Equivalent to standard model with no unit or time effects:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \varepsilon_{it}$$

- Pooled OLS estimation: use OLS ignoring panel structure!

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x})^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x})(y_{it} - \bar{y}) \right)$$

- Intuition: treat all $n \times T$ observations as independent in standard OLS
- Easy to compute and always available general estimator: $\hat{\beta} = (X'X)^{-1}X'Y$

Estimation with Unit Effects

- If no unit effects, we could use pooled OLS to recover slope coefficients
- How to estimate slope coefficients if unit effects are present?

$$y_{it} = \alpha_i + \beta_1 x_{it} + \varepsilon_{it}$$

- Commonly known as *Fixed Effects transformation*: uses unit-demeaned data
- Note the unit-level averages $\bar{y}_i$ under the model satisfy:

$$\frac{1}{T}\sum_{t=1}^{T} y_{it} = \frac{1}{T}\sum_{t=1}^{T}(\alpha_i + \beta_1 x_{it} + \varepsilon_{it}) = \alpha_i + \beta_1 \frac{1}{T}\sum_{t=1}^{T} x_{it} + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$$

- Taking deviations from unit-level means removes unit effects!

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \beta_1 \left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \right) + \left(\varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it} \right)$$

Fixed Effects Transformation

- For any variable v_{it}, fixed effects or unit-demeaned transformation defined as

$$\widetilde{v}_{it} := v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it} = v_{it} - \bar{v}_i$$

- Deviations from unit-level means can be written as:

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \beta_1 \left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \right) + \left(\varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it} \right)$$

$$\widetilde{y}_{it} = \beta_1 \widetilde{x}_{it} + \widetilde{\varepsilon}_{it}$$

Fixed Effects Transformation

- For any variable v_{it}, fixed effects or unit-demeaned transformation defined as

$$\widetilde{v}_{it} := v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it} = v_{it} - \bar{v}_i$$

- Deviations from unit-level means can be written as:

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \beta_1 \left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \right) + \left(\varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it} \right)$$

$$\widetilde{y}_{it} = \beta_1 \widetilde{x}_{it} + \widetilde{\varepsilon}_{it}$$

- Example for $i = 1$ (Alabama) and $t = 1982$: $\widetilde{y}_{it}$ is the difference between the fatality rate in Alabama in 1982 and the average fatality rate in Alabama averaged over all 7 years

Fixed Effects Estimator

$$\widetilde{y}_{it} = \beta_1 \widetilde{x}_{it} + \widetilde{\varepsilon}_{it}$$

- As unit effects are no longer present, equation above can be estimated using pooled OLS on unit-demeaned variables $\widetilde{x}_{it}$ and $\widetilde{y}_{it}$
- Results in the *Fixed Effects Estimator*:

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \widetilde{x}_{it}^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \widetilde{x}_{it} \widetilde{y}_{it} \right)$$

- Also known as *Within Estimator* as it uses within-unit variation (recall ANOVA decomposition)
- As $\widetilde{\varepsilon}_{it}$ is a complicated object, require stringent assumptions
- Standard errors need to be computed in a way that accounts for the panel nature of the data set (more later)

Example: Traffic deaths and beer taxes in STATA

- Previous procedure can be easily accomplished in a single command in most software
- Example in STATA: let STATA know you are working with panel data by defining the unit variable (state) and time variable (year):

```
. xtset state year;
       panel variable:  state (strongly balanced)
        time variable:  year, 1982 to 1988
                delta:  1 unit
```

```
. xtreg vfrall beertax, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0407                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0934                                        max =         7
                                                F(1,47)            =      5.05
corr(u_i, Xb)  = -0.6885                        Prob > F           =    0.0294

                                    (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2918556    -2.25   0.029    -1.243011   -.0687358
       _cons |   2.377075   .1497966    15.87   0.000     2.075723    2.678427
------------------------------------------------------------------------------
```

- The panel data command xtreg with the option fe performs fixed effects regression. The reported intercept is arbitrary, and the estimated individual effects are not reported in the default output.
- The fe option means use fixed effects regression
- The vce(cluster state) option tells STATA to use clustered standard errors – more on this later

Dummy Variable Estimation

- Recall dummy variable representation of unit effects:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma_1 D_{1,i} + \cdots + \gamma_{n-1} D_{n-1,i} + \varepsilon_{it}$$

- Dummy Variable estimator: use pooled OLS directly on this equation
- If n (cross-sectional dimension) is small, this is feasible
- Computational advances are able to implement fast versions for medium-sized n
- Following exercise ties together both approaches:

Exercise for you

Prove that the slope coefficient $\hat{\beta}_1$ obtained by using Dummy Variables or the Fixed Effects transformation is identical. Hint: Frisch-Waugh-Lovell theorem.

(First) Difference Estimation

- Take the unit effect representation at two time periods $s, t \in \{1, \ldots, T\}$:

$$y_{is} = \alpha_i + \beta_1 x_{is} + \varepsilon_{is} \quad \text{and} \quad y_{it} = \alpha_i + \beta_1 x_{it} + \varepsilon_{it}$$

- Difference between any two time periods removes unit effects:

$$y_{it} - y_{is} = \beta_1 (x_{it} - x_{is}) + (\varepsilon_{it} - \varepsilon_{is})$$

- Popular First Difference (FD) transformation uses t and $t-1$:

$$y_{it} - y_{i,t-1} = \beta_1 (x_{it} - x_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad \Rightarrow \quad \Delta y_{it} = \beta_1 \Delta x_{it} + \Delta \varepsilon_{it}$$

- First Difference estimator: use pooled OLS on difference equation
- One time series observation is lost by differencing
- *Exercise for you*: prove FE estimator is equivalent to FD estimator if $T = 2$

Estimation with Time Effects

- How to estimate slope coefficients if time effects are present?

$$y_{it} = \lambda_t + \beta_1 x_{it} + \varepsilon_{it}$$

- Similar to before, we can now use *time demeaned data* for estimation
- Time-level averages $\bar{y}_t$ under the model satisfy:

$$\frac{1}{n}\sum_{i=1}^{n} y_{it} = \frac{1}{n}\sum_{i=1}^{n}(\lambda_t + \beta_1 x_{it} + \varepsilon_{it}) = \lambda_t + \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_{it} + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{it}$$

- Taking deviations from time-level means removes time effects

$$y_{it} - \frac{1}{n}\sum_{i=1}^{n} y_{it} = \beta_1 \left(x_{it} - \frac{1}{n}\sum_{i=1}^{n} x_{it} \right) + \left(\varepsilon_{it} - \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{it} \right)$$

Estimation with Time Effects

$$y_{it} - \bar{y}_t = \beta_1(x_{it} - \bar{x}_t) + (\varepsilon_{it} - \bar{\varepsilon}_t)$$

- Construct deviations from time-level averages $y_{it} - \bar{y}_t$ and $x_{it} - \bar{x}_t$
- Using pooled OLS on the equation above provides estimates of the slope coefficients in the presence of time effects
- Dummy variables per time unit are also feasible if T is not too large

Estimating TWFE Specification

- We can now easily estimate the specification with both unit and time effects:

$$y_{it} = \alpha_i + \lambda_t + \beta_1 x_{it} + \varepsilon_{it}$$

- Four possible combinations depending on the relative sizes of cross-sectional dimension (n) and time dimension (T):
  1. Unit-demeaning vs $n - 1$ dummy variables
  2. Time-demeaning vs $T - 1$ dummy variables
- All approaches yield identical slope coefficient estimates
- First Differences slope coefficients do not equal those of unit-demeaning if $T > 2$
- After defining the estimators, we can then discuss inference

```
. gen y83=(year==1983);      First generate all the time binary variables
. gen y84=(year==1984);
. gen y85=(year==1985);
. gen y86=(year==1986);
. gen y87=(year==1987);
. gen y88=(year==1988);
. global yeardum "y83 y84 y85 y86 y87 y88";
. xtreg vfrall beertax $yeardum, fe vce(cluster state);

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0803                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0876                                        max =         7
corr(u_i, Xb)  = -0.6781                        Prob > F           =    0.0009
                                    (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3570783    -1.79   0.080    -1.358329    .0783691
          y83 |  -.0799029   .0350861    -2.28   0.027    -.1504869   -.0093188
          y84 |  -.0724206   .0438809    -1.65   0.106    -.1606975    .0158564
          y85 |  -.1239763   .0460559    -2.69   0.010    -.2166288   -.0313238
          y86 |  -.0378645   .0570604    -0.66   0.510    -.1526552    .0769262
          y87 |  -.0509021   .0636084    -0.80   0.428    -.1788656    .0770615
          y88 |  -.0518038   .0644023    -0.80   0.425    -.1813645    .0777568
        _cons |    2.42847   .2016885    12.04   0.000     2.022725    2.834215
------------------------------------------------------------------------------
```

Are the time effects jointly statistically significant?

```
. test $yeardum;

( 1)  y83 = 0
( 2)  y84 = 0
( 3)  y85 = 0
( 4)  y86 = 0
( 5)  y87 = 0
( 6)  y88 = 0

       F(  6,    47) =    4.22
            Prob > F =    0.0018
```

Answer is yes!

Lecture outline

Panel Data Assumptions

- Under a panel data version of the least squares assumptions, the OLS Fixed Effects estimator of β_1 is asymptotically normally distributed
- We require assumptions on the cross-sectional (n) and time (T) dimensions
- Textbook treatment of the model assumes n to be large and T to be small
- Asymptotic results will assume units are independent in the cross section, with n growing while T remains fixed (i.e., $n/T \to \infty$)
- Small time dimension T means time effects can be simply included as additional binary regressors
- Under standard assumptions, a new standard error formula needs to be introduced for inference: "clustered" standard errors

Assumptions for Panel Data

Consider a single x for the moment:

$$y_{it} = \beta_1 x_{it} + \alpha_i + \varepsilon_{it}$$

Least Squares assumptions for this setting:

1. $\{(y_{i,1}, x_{i,1}), \ldots, (y_{i,T}, x_{i,T})\}$ are independent and identically distributed (i.i.d.) draws from their joint distribution
2. $\mathbb{E}[\varepsilon_{it}|x_{is}, \alpha_i] = 0$, for all $s = 1, \ldots, t, \ldots, T$
3. (y_{it}, x_{it}) have finite fourth moments
4. There is no perfect multicollinearity (in the case of multiple x variables)

Assumption #1: Random sampling

- Extension of similar assumption for multiple regression OLS with cross-section data
- Units (individuals, countries, firms, etc.) are assumed to be sampled randomly from the population
- We then capture the time history of each randomly sampled unit to form our panel data
- This assumption does **not** require observations to be i.i.d. over time — that would be unrealistic as observations within each unit are naturally correlated
- Example: high beer taxes yesterday are good predictors (correlated with) high beer taxes today due to persistence: $\text{Corr}(x_{it}, x_{i,t-1}) \neq 0$
- Similarly, the error term (unobservables) affecting a state's traffic deaths in one year are likely correlated across time: $\text{Corr}(\varepsilon_{it}, \varepsilon_{i,t-1}) \neq 0$

Assumption #2: Exogeneity

$$\mathbb{E}[\varepsilon_{it}|x_{is}, \alpha_i] = 0, \quad \text{for all } s, t \in \{1, \ldots, T\}$$

- In words: unobservables ε_{it} cannot be predicted (are not correlated with) the entire history of covariates x conditional on the unit fixed effects
- Intuitively: after controlling for all unit-specific characteristics α_i, there are no remaining confounders (including *past or future* values of covariates!)
- This version is formally known as *strict exogeneity*
- Another extension of the equivalent multiple regression assumption
- All relevant dynamics in x ($x_{i,t-1}, \ldots, x_{i,t-p}$) need to be included explicitly
- Rules out lags of y in x as they natural confounders: $\text{Corr}(y_{i,t-1}, \varepsilon_{it}) \neq 0$
- Also rules out feedback from ε to future x:
  - Example: a year of high traffic fatalities (high ε) should not affect whether beer taxes (x) are increased
  - Plausability of exogeneity assumption is specific to each application

Under LS Assumptions for Panel Data

- Under these assumptions, standard proof of consistency follows through
- Unit-demeaning removes the unit fixed effects and having linearly independent regressors guarantees identification
- Strict exogeneity and random sampling are sufficient to guarantee consistency of the FE $\hat{\beta}_1$
- Further moment assumptions yield asymptotic normality of FE
- However, usual OLS standard errors (even heteroskedasticity-robust ones) will in general be invalid because they assume that ε_{it} is serially uncorrelated
  - OLS standard errors understate the true sampling uncertainty: if ε_{it} is correlated over time, you have less information (less random variation) compared to what would happen if ε_{it} were uncorrelated
  - This problem is solved by using "clustered" standard errors

Clustered Standard Errors in Panel Data

- As a simplified scenario, consider obtaining inference for the mean of y_{it}
- That is, begin by considering a panel data model with only a constant:

$$y_{it} = \mu + \varepsilon_{it}$$

- The estimator of μ is the sample average: $\overline{y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}$
- Write $\overline{y}$ as the average across units of the unit-level average

$$\overline{y} = \frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T} y_{it}\right) = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i, \tag{1}$$

Clustered Standard Errors in Panel Data

- Due to random sampling assumption, $(\overline{y}_1, \ldots, \overline{y}_n)$ are i.i.d with $\mathbb{E}[\overline{y}_i] = \mu$
- Define the unit-level variance as $\sigma^2_{\overline{y}_i} \coloneqq \text{Var}(\overline{y}_i)$
- For large n, an appropriate Central Limit Theorem (CLT) implies

$$\sqrt{n}(\overline{y} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\overline{y}_i}\right) \;\Rightarrow\; \text{SE}(\overline{y}) = \frac{\sigma^2_{\overline{y}_i}}{n}$$

- Pooled OLS estimate of SE uses: $\widetilde{\sigma}^2_{\overline{y}_i} = \frac{1}{nT-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \overline{y})^2$
- Clustered SE estimate uses: $\hat{\sigma}^2_{\overline{y}_i} = \frac{1}{n-1}\sum_{i=1}^{}(\overline{y}_i - \overline{y})^2$
- Final clustered standard error estimate: $\widehat{\text{SE}}(\overline{y}) = \sqrt{\hat{\sigma}^2_{\overline{y}_i}/n}$

Clustered Standard Errors in Panel Data

- Difference between pooled and clustered SEs is the use of unit-level averages
- Recall observations are not assumed i.i.d. across time for the same unit
- This simple correction adds a key feature: we have implicitly allowed for serial correlation within an unit
- The clustered SE estimator is naturally robust to autocorrelation!
- How? Formula for $\hat{\sigma}^2_{\overline{y}_i}$ estimates autocorrelations at all lags

Serial Correlation in Panel Data

- Autocovariance of order j: $\text{Cov}(y_{it}, y_{i,t-j})$
- Asymptotic variance of $\overline{y}$ includes serial correlation:

$$\sigma^2_{\overline{y}_i} = \text{Var}(\overline{y}_i) = \text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} y_{it}\right) = \frac{1}{T^2}\text{Var}(y_{i1} + y_{i2} + \cdots + y_{iT})$$

$$= \frac{1}{T^2}\left[\text{Var}(y_{i1}) + \cdots + \text{Var}(y_{iT})\right] + 2\,\text{Cov}(y_{i1}, y_{i2})$$

$$+ \cdots + 2\,\text{Cov}(y_{i1}, y_{iT}) + \cdots + 2\,\text{Cov}(y_{i,T-1}, y_{iT})$$

- If y_{it} is serially uncorrelated, all autocovariances will equal 0 and pooled estimate is still valid
- If these autocovariances are non-zero, pooled estimate is incorrect
- Example: if $\text{Cov}(y_{it}, y_{i,t-j}) > 0$ for all j (all autocovariances positive, common in practice), pooled estimate *underestimates* true variance

Details: Clustered Standard Errors in Panel Data

- Let us now expand formula for $\hat{\sigma}^2_{\overline{y}_i}$:

$$
\begin{aligned}
\hat{\sigma}^2_{\overline{y}_i} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{y}_i - \overline{y}\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \overline{y}\right)\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \overline{y}\right)\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left(y_{is} - \overline{y}\right)\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(y_{it} - \overline{y}\right)\left(y_{is} - \overline{y}\right)
\end{aligned}
$$

Details: Clustered Standard Errors in Panel Data

- Finally:

$$\hat{\sigma}^2_{\overline{y}_i} = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\frac{1}{n-1} \sum_{i=1}^{n} (y_{it} - \overline{y})(y_{is} - \overline{y}) \right]$$

- For any $s, t \in \{1, \ldots, T\}$, term in brackets is the sample autocovariance of order $|t - s|$
- This is the standard estimate of $\text{Cov}(y_{is}, y_{it})$
- All autocovariances are already naturally included in variance formula!
- Pooled OLS ignores the cross-products across time, missing any pattern of serial correlation

Clustered SEs for FE estimator in Panel Data Regression

- Idea of clustered SEs in a regression is analogous to the previous example of estimating the mean only
- Extension simply requires additional notation and formulas (found in most standard textbooks)
- Clustered SEs for panel data are the logical extension of heteroskedasticity robust (HR) SEs for cross-section:
  - In cross-section regression, HR-SEs are valid whether or not there is heteroskedasticity
  - In panel data regression, clustered SEs are valid whether or not there is heteroskedasticity and/or serial correlation
  - Term "clustered" comes from allowing correlation within a "cluster" of observations (e.g., *within* units), but *not across* clusters

Clustered SEs: Implementation in STATA

```
. xtreg vfrall beertax, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0407                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0934                                        max =         7
                                                F(1,47)            =      5.05
corr(u_i, Xb)  = -0.6885                        Prob > F           =    0.0294

                                    (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2918556    -2.25   0.029    -1.243011   -.0687358
       _cons |   2.377075   .1497966    15.87   0.000     2.075723    2.678427
------------------------------------------------------------------------------
```

- `vce(cluster state)` says to use clustered standard errors, where the clustering is at the state level (observations that have the same value of the variable "state" are allowed to be correlated, but are assumed to be uncorrelated if the value of "state" differs)

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Facts and Policy Questions

- Approx. 40,000 traffic fatalities annually in the U.S.
- 1/3 of traffic fatalities involve a drinking driver
- 25% of drivers on the road between 1am and 3am have been drinking (estimate)
- A drunk driver is 13 times as likely to cause a fatal crash as a non-drinking driver (estimate)
- Drunk driving causes massive externalities (sober drivers are killed, society bears medical costs, etc.) — room for policy intervention
- Are there any effective ways to reduce drunk driving? If so, what?
- What are effects of existing specific laws?
  - Mandatory punishment
  - Minimum legal drinking age
  - Economic interventions (alcohol taxes)

Panel Dataset

- Data on $n = 48$ U.S. states for $T = 7$ years (1982, ..., 1988)
- Main outcome: Traffic fatality rate (deaths per 10,000 residents)
- Main covariate of interest: Tax on a case of beer (BeerTax)
- Minimum legal drinking age
- Minimum sentencing laws for first DWI/DUI violation:
  - Mandatory Jail
  - Mandatory Community Service
  - Otherwise, sentence will just be a monetary fine
- Vehicle miles per driver (US DOT)
- State economic data (real per capita income and unemployment rate)

Why might panel data help?

- Potential omitted variable bias from variables that vary across states but are constant over time:
  - Culture of drinking and driving
  - Quality of roads
  - Age of cars on the road
- All these factors imply we should likely include *state fixed effects*
- Potential omitted variable bias from variables that vary over time but are constant across states:
  - Improvements in auto safety over time
  - Changing national attitudes towards drunk driving
- Dummy variables for years 1983, ..., 1988 to control for these time effects

TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent variable: traffic fatality rate (deaths per 10,000).

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes

***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**

Time effects = 0			4.22 (0.002)	10.12 (< 0.001)	3.48 (0.006)	10.28 (< 0.001)	37.49 (< 0.001)
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 (< 0.001)		31.96 (< 0.001)	25.20 (< 0.001)
$\overline{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the *10%, **5%, or ***1% significance level.

Empirical Analysis: Main Results

- Sign of the beer tax coefficient changes when state fixed effects are included
- Time effects are statistically significant but including them does not have a big impact on estimated coefficients
- Estimated effect of beer tax drops when other laws are included
- Only policy variable that seems to have an impact in traffic deaths is the tax on beer — not minimum drinking age, nor mandatory sentencing, etc.
- However, beer tax is not significant at 10% level using clustered SEs for some specifications that control for state economic conditions (unemployment rate and personal income)

Empirical Analysis: Validity of Assumptions

Question for you

What are the threats to internal validity of this model?

Empirical Analysis: Validity of Assumptions

Question for you

What are the threats to internal validity of this model?

1. Omitted variable bias from additional confounders (must be time-varying!)
2. Wrong functional form
3. Errors-in-variables bias (measurement error)
4. Sample selection bias
5. Simultaneous (reverse) causality bias

Lecture outline

1. Panel Data: What and Why?
2. Panel Data with two time periods
3. Panel Data with multiple time periods
4. Estimators for Linear Panel Data Models
5. Fixed Effects Regression Assumptions and Standard Errors
6. Application: Drunk Driving Laws and Traffic Deaths
7. Summary: Panel Data Regression

Advantages of Fixed Effects Regression

- You can control for unobserved variables that:
  - Vary across states but not over time, and/or
  - Vary over time but not across states
- More observations per unit give you more information
- Estimation involves straightforward and computationally simple extensions of multiple regression
- Fixed effects regression can be carried out in two equivalent ways:
  1. $n-1$ binary regressors
  2. "unit-demeaned" regression
- (First) Difference approach is an alternative with different statistical properties
- Similar methods apply to regression with time fixed effects and both effects
- General statistical inference: introduce clustered SEs

Limitations of Fixed Effects Regression

- Need variation in controls across time within units
- Example: policy evaluation requires a policy to change (vary) within analysis window, otherwise impossible to estimate its effect
- All time-constant information is discarded along with the unit effects and their impacts cannot be estimated
- Example: cannot estimate effect of quality of roads on traffic deaths
- Stringent strict exogeneity assumption required for inference rules out dynamic panel models
- Unit-demeaning does not lead to valid panel estimators for non-linear models
- Subject to all usual caveats of standard OLS (endogeneity, sample selection, causality, etc.)