Lecture 10: Introduction to Panel Data

Introduction to Econometrics,Fall 2023

Zhaopeng Qu

Nanjing University Business School

May 17 2023

1 Panel Data: What and Why

2 Fixed Effects Model

3 Potential Threats of Internal Validity

4 Summary

Panel Data: What and Why

Introduction

What is Panel Data

- So far, we have only focused on data cross entities.Now it is the time to add time, which leads us to use **Panel Data**.
- **Panel data** refers to data with observations on *multiple entities, where each entity is observed at two or more points in time.*
- If the data set contains observations on the variables X and Y,then the data are denoted

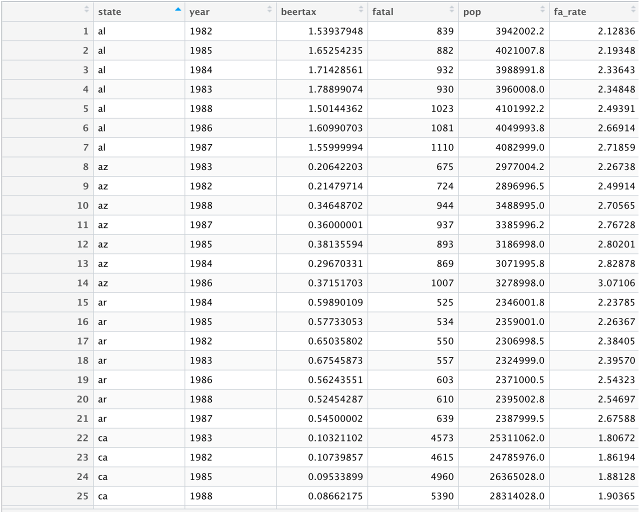
$$(X_{it}, Y_{it}),\ \ i = 1, ...n\ and\ t = 1, ..., T$$

  - the first subscript,i refers to the entity being observed
  - the second subscript,t refers to the date at which it is observed
- Extension: not necessarily involves time dimension
  - outcome of employee i in firm m (X_{im}, Y_{im}) $i = 1, ...n\ and\ m = 1, ..., M$

Introduction: Data Structure

- **Balanced** v.s **Unbalanced**
    - Balanced panel: *each unit of observation i is observed the same number of time periods, T. Thus, the total sample size is NT.*
    - Unbalanced panel: *each unit of observation i is observed an unequal number of time periods,* T_i, commonly some missing values for some entities at some periods.
- **Micro** v.s **Macro**
    - Micro: large N, and small T,more similar to cross-section data
    - Macro: small N, and large T,more similar to time series data
- In our class, we focus on **balanced** and **micro** panel data.

Example: Traffic Deaths and Alcohol Taxes



	state	year	beertax	fatal	pop	fa_rate
1	al	1982	1.53937948	839	3942002.2	2.12836
2	al	1985	1.65254235	882	4021007.8	2.19348
3	al	1984	1.71428561	932	3988991.8	2.33643
4	al	1983	1.78899074	930	3960008.0	2.34848
5	al	1988	1.50144362	1023	4101992.2	2.49391
6	al	1986	1.60990703	1081	4049993.8	2.66914
7	al	1987	1.55999994	1110	4082999.0	2.71859
8	az	1983	0.20642203	675	2977004.2	2.26738
9	az	1982	0.21479714	724	2896996.5	2.49914
10	az	1988	0.34648702	944	3488995.0	2.70565
11	az	1987	0.36000001	937	3385996.2	2.76728
12	az	1985	0.38135594	893	3186998.0	2.80201
13	az	1984	0.29670331	869	3071995.8	2.82878
14	az	1986	0.37151703	1007	3278998.0	3.07106
15	ar	1984	0.59890109	525	2346001.8	2.23785
16	ar	1985	0.57733053	534	2359001.0	2.26367
17	ar	1982	0.65035802	550	2306998.5	2.38405
18	ar	1983	0.67545873	557	2324999.0	2.39570
19	ar	1986	0.56243551	603	2371000.5	2.54323
20	ar	1988	0.52454287	610	2395002.8	2.54697
21	ar	1987	0.54500002	639	2387999.5	2.67588
22	ca	1983	0.10321102	4573	25311062.0	1.80672
23	ca	1982	0.10739857	4615	24785976.0	1.86194
24	ca	1985	0.09533899	4960	26365028.0	1.88128
25	ca	1988	0.08662175	5390	28314028.0	1.90365

Example: Traffic deaths and alcohol taxes

- Observational unit: *one* year in *one* U.S. state
  - Total 48 U.S. states, so N = the number of entities = 48
  - 7 years (1982,..., 1988),so T = the number of time periods = 7.
- Balanced panel, so total number of observations

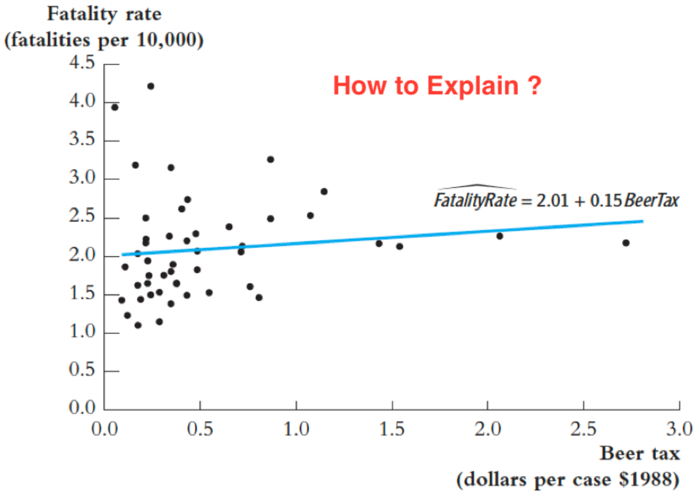
$$NT = 7 \times 48 = 336$$

- Variables:
  - Dependent Variable: **Traffic fatality rate** (# traffic deaths in that state in that year, per 10,000 state residents)
  - Independent Variable: **Tax on a case of beer**
  - Other Controls (legal driving age, drunk driving laws, etc.)
- A simple OLS regression model with $t = 1982, 1988$

$$FatalityRate_{it} = \beta_{0t} + \beta_{1t} BeerTax_{it} + u_{it}$$

U.S. traffic death data for 1982

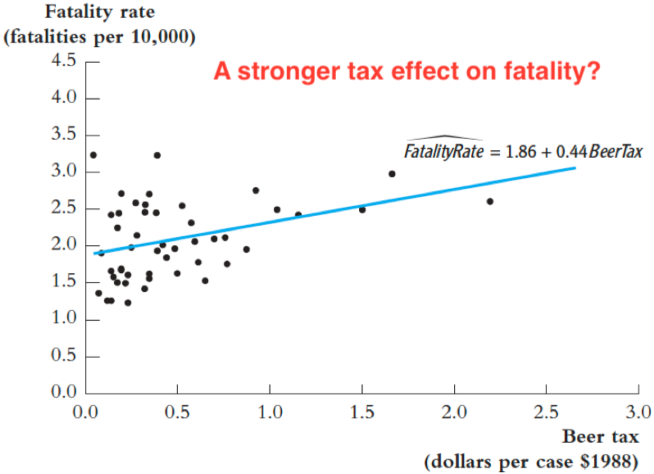
- Higher alcohol taxes, more traffic deaths



$\widehat{FatalityRate} = 2.01 + 0.15\,BeerTax$

U.S. traffic death data for 1988

- Still higher alcohol taxes, more traffic deaths



A stronger tax effect on fatality?

(b) 1988 data

Pooled Cross-Sectional Data(1982-1988)

- The positive relationship between alcohol taxes and traffic deaths might be due to using only two years data.Therefore,we run the following regression using full years data

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + u_{it}$$

- This is a simple OLS, only now sample size is $NT = 7 \times 48 = 336$
- If you we would like to control the time, in other words, we would like to strict our regression within every years and then make an average, then we should run

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \lambda T_t + u_{it}$$

Pooled Cross-Sectional Data(1982-1988)

- Still higher alcohol taxes, more traffic deaths(though some nonlinear pattern)

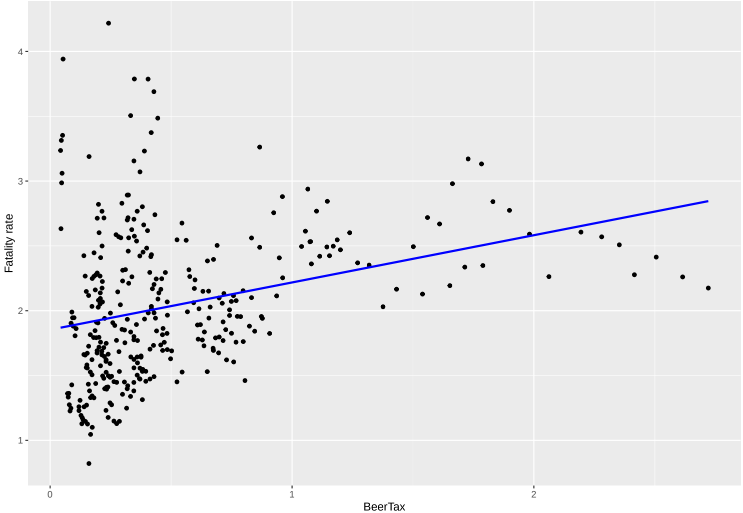


Table 1:

	Dependent Variable: Fatality Rate	
	Pooled OLS	Pooled OLS with Time
	(1)	(2)
beertax	0.365***	0.366***
	(0.053)	(0.053)
year_1983		−0.082
		(0.128)
year_1984		−0.072
		(0.121)
year_1985		−0.111
		(0.120)
year_1986		−0.016
		(0.121)
year_1987		−0.016
		(0.122)
year_1988		−0.001
		(0.119)
Constant	1.853***	1.895***
	(0.047)	(0.105)

Pooled Cross-Sectional Data(1982-1988)

- Could we are safety to make a conclusion:

  Higher beer tax cannot make less but more fatalities
- In other words : does the regression satisfy **OLS Assumption 1-4** to obtain an *unbiased* and *consistent* estimation for the conclusion?
- **Question**: are there some threatens to the internal validity of the estimate?

Pooled Cross-Sectional Data(1982-1988)

- **Assumption 1**, $E(u_i|X_i) = 0$ may not satisfied for some unobservables(**OVB**).
  - Some unobservable factors that determines the fatality rate may be correlated with *BeerTax*, such as **local cultural attitude** toward drinking and driving.
- **Assumption 2** random sampling is not satisfied for **serial correlation** of important variables.
  - Both *Beertax* and *Fatality rate* might be serial correlated between different periods.

Before-After Model

Simple Case: Panel Data with Two Time Periods

- Firstly let adjust our model with some unobservables

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

  where u_{it} is the error term and $i = 1, ...n \; and \; t = 1, ..., T$

- Z_i is the **unobservable factor** that determines the fatality rate in the i state but does not change over time.
- The omission of Z_i might cause omitted variable bias(**OVB**) but we don't have data on Z_i.
- 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.

Panel Data with Two Time Periods

- Consider the regressions for 1982 and 1988...

$$FatalityRate_{i1988} = \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988}$$
$$FatalityRate_{i1982} = \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982}$$

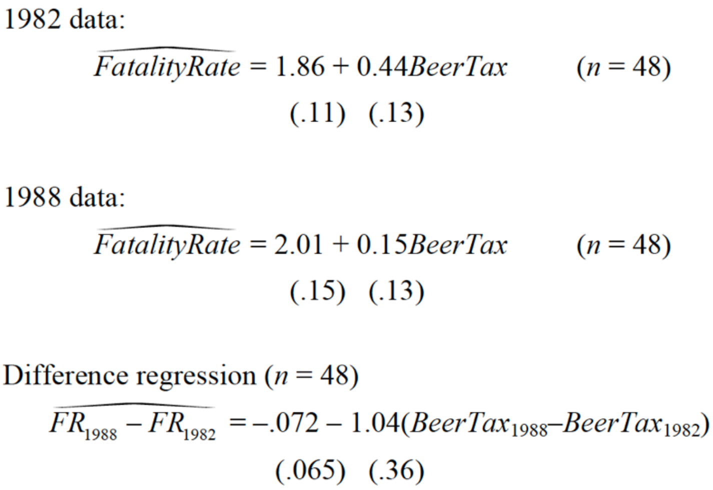
- Then make a difference

$$FatalityRate_{i1988} - FatalityRate_{i1982} =$$
$$\beta_1 (BeerTax_{i1988} - BeerTax_{i1982}) + (u_{i1988} - u_{i1982})$$

Panel Data with Two Time Periods

- Assumption: if $E(u_{it}|BeerTax_{it}, Z_{it}) = 0$,then $(u_{i1988} - u_{i1982})$ is uncorrelated with $(BeerTax_{i1988} - BeerTax_{i1982})$
- Then this “difference” equation can be estimated by OLS, even though Z_i isn’t observed.
- Intuition: because the omitted variable Z_i doesn’t change, it cannot be a determinant of the change in Y.

Case: Traffic deaths and beer taxes



1982 data:

$$\widehat{FatalityRate} = \underset{(.11)}{1.86} + \underset{(.13)}{0.44}BeerTax \qquad (n = 48)$$

1988 data:

$$\widehat{FatalityRate} = \underset{(.15)}{2.01} + \underset{(.13)}{0.15}BeerTax \qquad (n = 48)$$

Difference regression ($n = 48$)

$$\widehat{FR_{1988} - FR_{1982}} = \underset{(.065)}{-.072} - \underset{(.36)}{1.04}(BeerTax_{1988} - BeerTax_{1982})$$

Change in traffic deaths and change in beer taxes

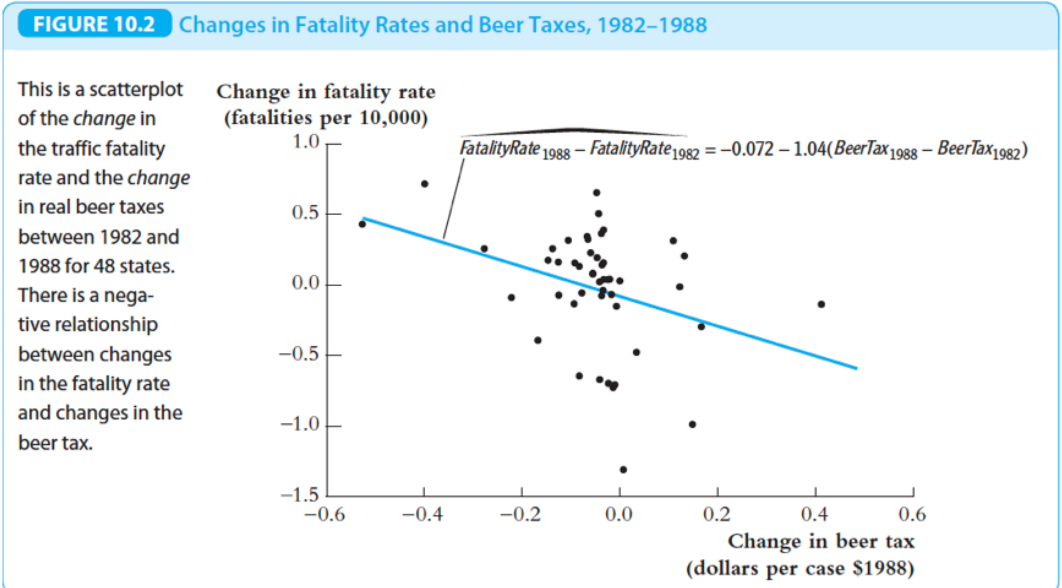


FIGURE 10.2 Changes in Fatality Rates and Beer Taxes, 1982–1988

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.

$$\widehat{FatalityRate_{1988} - FatalityRate_{1982}} = -0.072 - 1.04(BeerTax_{1988} - BeerTax_{1982})$$

Wrap up

- In contrast to the cross-sectional regression results, the estimated effect of a change in the real beer tax is **negative**, as predicted by economic theory.
- By examining changes in the fatality rate over time, the regression controls for some **unobservable but fixed factors** such as cultural attitudes toward drinking and driving.
- But there are many factors that influence traffic safety, and if they change over time and are correlated with the real beer tax, then their omission will still produce omitted variable bias(OVB).

Wrap up

- This “before and after” analysis works *when the data are observed in* ***two*** *different years*.
- Our data set, however, contains observations for **seven** different years,and it seems foolish to discard those potentially useful additional data.
- But the “before and after” method does not apply directly when $T > 2$. To analyze all the observations in our panel data set, we use a more general regression setting: **fixed effects**

Fixed Effects Model

Introduction

Introduction

- Fixed effects regression is a method for controlling for omitted variables in panel data when *the omitted variables vary across entities (states) but do not change over time.*
- Unlike the "before and after" comparisons,fixed effects regression can be used when there are **two or more time** observations for each entity.

Fixed Effects Regression Model

- The **dependent variable** (FatalityRate) and **independent variable** (BeerTax) denoted as Y_{it} and X_{it}, respectively. Then our model is

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it} \tag{11.1}$$

- Where Z_i is an **unobserved variable** that varies from one state to the next but **does not change over time**
    - eg. Z_i can still represent cultural attitudes toward drinking and driving.
- We want to estimate β_1, the effect on Y of X holding constant the unobserved state characteristics Z.

Fixed Effects Regression Model

- Because Z_i varies from one state to the next but is constant over time,then let $\alpha_i = \beta_0 + \beta_2 Z_i$,the Equation becomes

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{11.2}$$

- This is the **fixed effects regression model**, in which α_i are treated as *unknown intercepts* to be estimated, one for each state. The interpretation of α_i as a *state-specific intercept* in Equation (11.2).
- Because the intercept α_i can be thought of as the “effect” of being in entity i (in the current application, entities are states),the terms α_i,known as **entity fixed effects**.
- The variation in the entity fixed effects comes from omitted variables that, like Z_i in Equation (11.1), vary across entities but not over time.

Alternative : Fixed Effects by using binary variables

- How to estimate these parameters α_i.
- To develop the fixed effects regression model using binary variables, let $D1_i$ be a binary variable that equals 1 when i = 1 and equals 0 otherwise, let $D2_i$ equal 1 when i = 2 and equal 0 otherwise, and so on.
- Arbitrarily omit the binary variable $D1_i$ for the first group. Accordingly, the fixed effects regression model in Equation (7.2) can be written equivalently as

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \ldots + \gamma_n Dn_i + u_{it} \tag{7.3}$$

- Thus there are two equivalent ways to write the fixed effects regression model, Equations (7.2) and (7.3).
- In both formulations, the slope coefficient on X is the same from one state to the next.

Estimation and Inference

Estimation: Introduction

- In principle the binary variable specification of the fixed effects regression model can be estimated by OLS.
- But it is tedious to estimate so many fixed effects.If $n = 1000$, then you have to estimate $1000 - 1 = 999$ fixed effects.
- There are some special routines, which are equivalent to using OLS on the full binary variable regression, are *faster* because they employ some *mathematical simplifications* that arise in the algebra of fixed effects regression.

Estimation: The “entity-demeaned”

- Computes the OLS fixed effects estimator in two steps
- The **first** step:
  - take the average across times t of both sides of Equation (7.2);

$$\bar{Y}_i = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_t \tag{7.4}$$

  - demeaned: let Equation(7.2) minus (7.4)

$$Y_{it} - \bar{Y}_i = \beta_1 X_{it} - \bar{X}_i + (\alpha_i - \alpha_i) + u_{it} - \bar{u}_i$$

Estimation: The “entity-demeaned”

- Let

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$$

$$\tilde{X}_{it} = X_{it} - \bar{X}_i$$

$$\tilde{u}_{it} = u_{it} - \bar{u}_i$$

- Then the **second** step: accordingly,estimate

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \tag{7.5}$$

- Then the estimator is known as the **within estimator**. Because it matters not if a unit has consistently high or low values of Y and X. All that matters is how the variations around those mean values are correlated.
- In fact, this estimator is identical to the OLS estimator of β_1 without intercept obtained by estimation of the fixed effects model in Equation (7.3)

OLS estimator without intercept

- OLS estimator without intercept

$$Y_i = \beta_1 X_i + u_i$$

- The least squared term

$$\underset{b_1}{min} \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - b_1 X_i)^2$$

- F.O.C, thus differentiating with respect to β_1, we get

$$\sum_{i=1}^{n} 2\big(Y_i - b_1 X_i\big) X_i = 0$$

- At last,

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

Fixed effects estimator(I)

- The second step:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \tag{11.4}$$

- Then the fixed effects estimator can be obtained based on OLS estimator without intercept

$$\hat{\beta}_{demean} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{Y}_{it} \tilde{X}_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^{2}}$$

Fixed effect estimator(II)

- The fixed effects model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{7.2}$$

- Equivalence to

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + ... + \gamma_n Dn_i + u_{it} \tag{7.3}$$

- Then we can think of α_i as fixed effects or "nuisance parameters" to be estimated,thus yields

$$(\widehat{\beta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_n) = \underset{b,a_1,\ldots,a_n}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - bX_{it} - a_i)^2$$

  this amounts to including $n = n + 1 - 1$ dummies in regression of Y_{it} on X_{it}

Fixed effect estimator(II)

- The first-order conditions (FOC) for this minimization problem are:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\widehat{\beta}X_{it}-\widehat{\alpha}_i)X_{it}=0$$

- And

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\widehat{\beta}X_{it}-\widehat{\alpha}_i)=0$$

Fixed effect estimator(II)

- Therefore, for $i = 1, \ldots, N$,

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} (Y_{it} - \widehat{\beta} X_{it}) = \overline{Y}_i - \overline{X}_i \widehat{\beta},$$

where

$$\overline{X}_i \equiv \frac{1}{T} \sum_{t=1}^{T} X_{it}; \overline{Y}_i \equiv \frac{1}{T} \sum_{t=1}^{T} Y_{it}$$

Fixed effect estimator(II)

- Plug this result into the first FOC to obtain:

$$\begin{aligned}\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\widehat{\beta}X_{it}-\widehat{\alpha}_i)X_{it} &= \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-X_{it}\hat{\beta}-\overline{Y}_i+\overline{X}_i\hat{\beta})X_{it}\\ &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\overline{Y}_i)X_{it}\right)\\ &-\hat{\beta}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}-\overline{X}_i)X_{it}\right)=0\end{aligned}$$

Fixed effect estimator(II)

- Then we could obtain

$$
\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)}{\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\bar{Y})(X_{it}-\bar{X}_i)} \\
&= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{2}}
\end{aligned}
$$

  with time-demeaned variables $\tilde{X}_{it} \equiv X_{it} - \bar{X}$, $\tilde{Y}_{it} \equiv Y_{it} - \bar{Y}_i$

- which is same as we obtained in demeaned method.

Fixed effect estimator(III): first-differencing

- The fixed effects model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{11.2}$$

- Then implies

$$\begin{aligned} Y_{i1} &= \beta_1 X_{i1} + \alpha_i + u_{i1} \\ Y_{i2} &= \beta_1 X_{i2} + \alpha_i + u_{i2} \\ \vdots &= \vdots \\ Y_{iT} &= \beta_1 X_{iT} + \alpha_i + u_{iT} \end{aligned}$$

Fixed effect estimator(III): first-differencing

- Taking the differences between consecutive years

$$\begin{aligned} Y_{i2} - Y_{i1} &= \beta_1(X_{i2} - X_{i1}) + (u_{i2} - u_{i1}) \\ Y_{i3} - Y_{i2} &= \beta_1(X_{i3} - X_{i2}) + (u_{i3} - u_{i2}) \\ \vdots &= \vdots \\ Y_{iT} - Y_{i,T-1} &= \beta_1(X_{iT} - X_{i,T-1}) + (u_{iT} - u_{i,T-1}) \end{aligned}$$

Fixed effect estimator(III): first-differencing

- New notation,we use Δ represents the change from the preceding year,then

$$\begin{aligned}\Delta Y_{i2} &= \beta_1 \Delta X_{i2} + \Delta u_{i2} \\ \Delta Y_{i3} &= \beta_1 \Delta X_{i3} + \Delta u_{i3} \\ \vdots &= \vdots \\ \Delta Y_{iT} &= \beta_1 \Delta X_{iT} + \Delta u_{iT}\end{aligned}$$

- The first-difference fixed effect model is

$$\Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta u_{it} \; i = 1, ..., N, ; t = 2, ..., T \tag{11.5}$$

- Then first-difference estimator is

$$\hat{\beta}_{fd} = \frac{\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta Y_{it} \Delta X_{it}}{\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta X_{it}^2}$$

The Fixed Effects Regression Assumptions

- The simple fixed effect model

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, i = 1, ...n \ t = 1, ..., T$$

1. **Assumption 1**: u_{it} has conditional mean zero with X_{it}, or X_i at any time t and α_i

$$E(u_{it}|X_{i1}, X_{i2}, ..., X_{iT}, \alpha_i) = 0$$

2. **Assumption 2**: $(X_{i1}, X_{i2}, ..., X_{iT}, u_{i1}, u_{i2}, ..., u_{iT}), i = 1, 2, ..., n$ are *i.i.d.*
3. **Assumption 3**: Large outliers are unlikely.
4. **Assumption 4**: There is no perfect multicollinearity.

The Fixed Effects Regression Assumptions

- **Assumption 1**: u_{it} has conditional mean zero with X_{it}, or X_i at any time t and α_i,thus

$$E(u_{it}|X_{i1}, X_{i2}, ..., X_{iT}, \alpha_i) = 0$$

- u_{it} has mean zero, given the state fixed effect and the entire history of the Xs for that state.
- No feedback effect from u to future X
  - *Whether a state has a particularly high fatality rate this year does not subsequently affect whether it increases the beer tax.*

Fixed effect estimator(III): first-differencing

- When $T = 2$, FD and demean estimators and all test statistics are *identical*.
- When $T = 3$, FD and demean estimators are not the same, while both are consistent(T fixed as $N \to \infty$) if certain assumptions are satisfied.
- But if the strict exogenous assumption is not satisfied, then the demean estimator has more advantages over the FD estimator for having substantial less bias.

Statistical Properties of Fixed Effects Model

- Unbiasedness and Consistency

$$
\begin{aligned}
\widehat{\beta}_{fe-demean} &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{2}} \\
&= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}(\beta_1\tilde{X}_{it}+\tilde{u}_{it})}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{2}} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{2}}
\end{aligned}
$$

Statistical Properties

- Unbiasedness and Consistency

$$
\begin{aligned}
\hat{\beta}_{fd} &= \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta Y_{it}\Delta X_{it}}{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{it}^{2}} \\
&= \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{it}(\beta_1\Delta X_{it}+\Delta u_{it})}{\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta X_{it}^{2}} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{it}\Delta u_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta X_{it}^{2}}
\end{aligned}
$$

- It is very familiar: paralleling the derivation of OLS estimator, we could prove the estimator of fixed effects model is **unbiased** and **consistent**.

Statistical Properties

- Similarly, in panel data, if the fixed effects regression assumptions—holds, then the sampling distribution of the fixed effects OLS estimator is normal in large samples.
- Then the variance of that distribution can be estimated from the data, the square root of that estimator is the standard error,
- And the standard error can be used to construct t-statistics and confidence intervals.
- Statistical inference—testing hypotheses (including joint hypotheses using F-statistics) and constructing confidence intervals—proceeds in exactly the same way as in multiple regression with cross-sectional data.

Fixed Effects: goodness of fit

- Three measures of goodness of fit are commonly reported
  - *Within* R^2: demeaned Y_{it} and demeaned predicted $\hat{Y}_{it}$ using demeaned X_{it} and estimate coefficient $\hat{\beta}$
  - *Between* R^2: average Y_i and average predicted $\hat{\bar{Y}}_i$ using average $\bar{X}_i$ and estimate coefficient $\hat{\beta}$
  - *Overall* R^2: Y_{it} and predicted $\hat{Y}_{it}$

Fixed Effects: Extension to multiple X's.

- The multiple fixed effects regression model is

$$Y_{it} = \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \alpha_i + u_{it}$$

- Equivalently, the fixed effects regression can be expressed in terms of a common intercept

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} \\ & + \gamma_2 D2_i + \gamma_3 D3_i + \ldots + \gamma_n Dn_i + u_{it} \end{aligned}$$

Application to Traffic Deaths

- The OLS estimate of the fixed effects regression based on all 7 years of data (336 observations), is

$$\widehat{FatalityRate} = -\underset{(0.29)}{0.66}BeerTax + StateFixedEffects$$

- The estimated state fixed intercepts are not listed to save space and because they are not of primary interest.
- As predicted by economic theory,higher real beer taxes are associated with fewer traffic deaths, which is the opposite of what we found in the initial cross-sectional regressions.

Application to Traffic Deaths

- Recall: The result in Before-After Model is

$$\widehat{FR_{1988} - FR_{1982}} = -.072 - 1.04\left(\text{BeerTax}_{1988} - \text{BeerTax}_{1982}\right)$$

$$(.065)(.36)$$

- The magnitudes of estimate coefficients are not identical,because they use different data.
- And because of the additional observations, the standard error now is also smaller than before-after model.

Extension: Both Entity and Time Fixed Effects

Regression with Time Fixed Effects

- Just as fixed effects for each entity can control for variables that are constant over time but differ across entities, so can **time fixed effects** control for variables that *are constant across entities but evolve over time*.
    - Like *safety improvements in new cars* as an **omitted variable** that changes over time but has the same value for all states.
- Now our regression model with **time fixed effects**

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_3 S_t + u_{it}$$

- where S_t is *unobserved* and where the single t subscript emphasizes that safety changes over time but is constant across states. Because $\beta_3 S_3$ represents variables that determine Y_{it}, if S_t is correlated with X_{it}, then omitting S_t from the regression leads to omitted variable bias.

Time Effects Only

- Although S_t is unobserved, its influence can be eliminated because it varies over time but not across states, just as it is possible to eliminate the effect of Z_i, which varies across states but not over time.
- Similarly,the presence of S_t leads to a regression model in which each time period has its own intercept,thus

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}$$

- This model has a different intercept, λ_t, for each time period, which are known as **time fixed effects**.The variation in the time fixed effects comes from omitted variables that vary over time but not across entities.

Time Effects Only

- Just as the **entity fixed effects** regression model can be represented using $n - 1$ binary indicators, the time fixed effects regression model be represented using $T - 1$ binary indicators too:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \delta_2 B2_t + \ldots + \delta_T BT_t + \alpha_i + u_{it} \tag{11.18}$$

  - where $\delta_2, \delta_3, \ldots, \delta_T$ are unknown coefficients
  - where $B2_t = 1$ if $t = 2$ and $B2_t = 0$ otherwise and so forth.
- Nothing new, just a another form of Fixed Effects model with another explanation.

Table 2:

	Dependent Variable: Fatality Rate	
	Pooled OLS	Pooled OLS with Time
	(1)	(2)
beertax	0.365***	0.366***
	(0.053)	(0.053)
year_1983		−0.082
		(0.128)
year_1984		−0.072
		(0.121)
year_1985		−0.111
		(0.120)
year_1986		−0.016
		(0.121)
year_1987		−0.016
		(0.122)
year_1988		−0.001
		(0.119)
Constant	1.853***	1.895***
	(0.047)	(0.105)

Both Entity and Time Fixed Effects

- If some omitted variables are constant over time but vary across states (such as cultural norms) while others are constant across states but vary over time (such as national safety standards)
- Then, combined entity and time fixed effects regression model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

  - where α_i is the **entity fixed effect** and λ_t is the **time fixed effect**.

- This model can equivalently be represented as follows

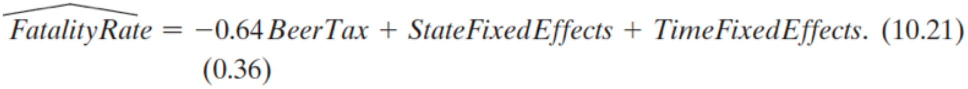
$$\begin{aligned} Y_{it} =& \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \ldots + \gamma_n Dn_i \\ &+ \delta_2 B2_t + \delta_3 B3_t + \ldots + \delta_T BT_i + u_{it} \end{aligned}$$

Both Entity and Time Fixed Effects: Estimation

- The time fixed effects model and the entity and time fixed effects model are both **variants** of *the multiple regression model*.
- Thus their coefficients can be estimated by OLS by including the additional time and entity binary variables.
- Alternatively,first deviating Y and the X's from their entity and time-period means and then by estimating the multiple regression equation of deviated Y on the deviated X's.

Application to traffic deaths

- This specification includes the beer tax, 47 state binary variables (state fixed effects), 6 single-year binary variables (time fixed effects), and an intercept, so this regression actually has $1 + 47 + 6 + 1 = 55$ right-hand variables!



$$\widehat{FatalityRate} = \underset{(0.36)}{-0.64}\,BeerTax + StateFixedEffects + TimeFixedEffects. \quad (10.21)$$

- When time effects are included,this coefficient is less precisely estimated, it is still significant only at the 10%, but not the 5%.
- This estimated relationship between the real beer tax and traffic fatalities is immune to omitted variable bias from variables that are constant either over time or across states.

Potential Threats of Internal Validity

Measurement error in FE

Recall: Classical measurement error of X

- The true model is

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

with $E[u_i|X_i] = 0$

- Due to the **classical measurement error**,we only have X_i^* thus

$$X_i^* = X_i + w_i$$

with $E[w_i|X_i] = 0$

- Then we have to estimate the model is

$$Y_i = \beta_0 + \beta_1 X_i^* + e_i$$

where $e_i = -\beta_1 w_i + u_i$

Recall: Classical measurement error of X

- Similar to OVB bias in simple OLS model, we had derived that

$$plim\left(\hat{\beta}_1\right) = \beta_1 \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2}$$

- Then we have

$$plim\left(\hat{\beta}_1\right) = \beta_1 \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2} \leq \beta_1$$

- The classical measurement error β_1 is biased towards 0, which is also called **attenuation bias**

Measurement error in FE

- Suppose we will estimate a fixed effect model

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

- Unfortunately, our measurement of X is not accurate, suppose it satisfies the classical measurement error, thus

$$X_{it}^* = X_{it} + w_{it}$$

  with $E[w_{it}|X_{it}] = 0$

- Then we estimate

$$Y_{it} = \beta_1 X_{it}^* + \alpha_i + e_{it}$$

  with $e_{it} = -\beta_1 w_{it} + u_{it}$

Measurement error in FE

- First difference estimator for fixed effect

$$\Delta Y_{it} = \beta_1 \Delta X^*_{it} + \Delta e_{it}$$

with $\Delta e_{it} = -\beta_1 \Delta w_{it} + \Delta u_{it}$

- Following the formula of ME in Simple OLS regression,we have

$$plim(\hat{\beta}_1) = \beta_1 \frac{\sigma^2_{\Delta X}}{\sigma^2_{\Delta X} + \sigma^2_{\Delta w}}$$

- Assume that time series X_t is stationary, which means that the expectation and variance are both constant.

$$\begin{aligned} \sigma^2_{\Delta X} &= Var(X_{it}) - 2Cov(X_{it}, X_{i,t-1}) + Var(X_{i,t-1}) \\ &= 2\sigma^2_X - 2\rho\sigma^2_X \\ &= 2\sigma^2_X(1-\rho) \end{aligned}$$

Measurement error in FE

- Similarly, define r to be the autocorrelation coefficient in w_{it},then the **attenuation bias** in fixed effect model is

$$plim(\hat{\beta}) = \beta \frac{\sigma_X^2(1-\rho)}{\sigma_X^2(1-\rho) + \sigma_w^2(1-r)}$$

- If both X_{it} and w_{it} are uncorrelated over time(t), then $\rho = 0$ and $r = 0$, the bias equals to the one in simple OLS case.
- If measurement error is uncorrelated over time, but X_{it} are correlated over time, thus $\rho \neq 0$ and $r = 0$.Then we have

$$plim(\hat{\beta}) = \beta \frac{\sigma_X^2(1-\rho)}{\sigma_X^2(1-\rho) + \sigma_w^2} < \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2}$$

- It means that *attenuation bias in fixed-effect model* will be **larger** than the bias in OLS. In other words, measurement error will be magnified in a FE model.

Autocorrelation

Autocorrelation

- An important difference for a key assumption
  - **Cross-Section**: Assumption 2 holds: i.i.d sample.
  - **Times series**: $Y_1, \ldots, Y_T$ are dependent(not independent) and may not be identically distributed.
  - **Panel data**: independent across entities but no such restricition **within** an entity.
- Observations of a time series are typically correlated. This type of correlation is called **autocorrelation** or **serial correlation**.

Autocorrelation

- The covariance between Y_t and its j^{th} lag, Y_{t-j}, is called the j^{th} *autocovariance* of the series Y_t.

$$j^{th}\text{autocovariance} = Cov(Y_t, Y_{t-j})$$

- The j^{th} *autocorrelation coefficient*, also called the *serial correlation coefficient*, measures the correlation between Y_t and Y_{t-j}.

$$j^{th}\text{autocorrelation} = \rho_j = \rho_{Y_t, Y_{t-j}} = \frac{Cov(Y_t, Y_{t-j})}{\sqrt{Var(Y_t)Var(Y_{t-j})}}$$

Autocorrelated in Panel Data

- In the traffic fatality example, X_{it}, the beer tax in state i in year t,is autocorrelated:
  - Most of the time, the legislature does not change the beer tax, so if it is high one year relative to its mean value for state i,it will tend to be high the next year,too.

Autocorrelated in Panel Data

- Similarly,u_{it} would be also autocorrelated. It consists of time-varying factors that are determinants of Y_{it} but are not included as regressors, and some of these omitted factors might be autocorrelated. It can formally be expressed as

$$Cov(u_{it}, u_{is}|X_{it}, X_{is}, \alpha_i) \neq 0 \, for \, t \neq s$$

  - eg. a downturn in the local economy and a road improvement project.

Autocorrelated in Panel Data

- If the regression errors are autocorrelated, then the usual heteroskedasticity-robust standard error formula for cross-section regression is not valid.
- The result: an analogy of **heteroskedasticity**.
- OLS panel data estimators of β are unbiased and consistent but the standard errors will be wrong
  - usually the OLS standard errors understate the true uncertainty
- This problem can be solved by using **“heteroskedasticity and autocorrelation-consistent(HAC) standard errors”**

Standard Errors for Fixed Effects Regression

- The standard errors used are one type of HAC standard errors, **clustered standard errors**.
- The term **clustered** arises because these standard errors allow the regression errors to have an arbitrary correlation within a cluster, or grouping, but assume that the regression errors are uncorrelated across clusters.
- In the context of panel data,each cluster consists of an entity.Thus **clustered standard errors** allow for heteroskedasticity and for arbitrary autocorrelation *within an entity*, but treat the errors as *uncorrelated across entities*.
- Like **heteroskedasticity-robust standard errors** in regression with cross-sectional data, **clustered standard errors** are valid whether or not there is heteroskedasticity,autocorrelation,or both.

Application: Drunk Driving Laws and Traffic Deaths

Application: Drunk Driving Laws and Traffic Deaths

- Two ways to cracks down on Drunk Driving
  1. toughening driving laws
  2. raising taxes
- Both driving laws and economic conditions could be omitted variables,it is better to put them into the regression as covariates.
- Besides, In two way fixed effect model, controlling both unobservable variables simultaneously that
  - do not change over time
  - do not vary across states

Application: Drunk Driving Laws and Traffic Deaths

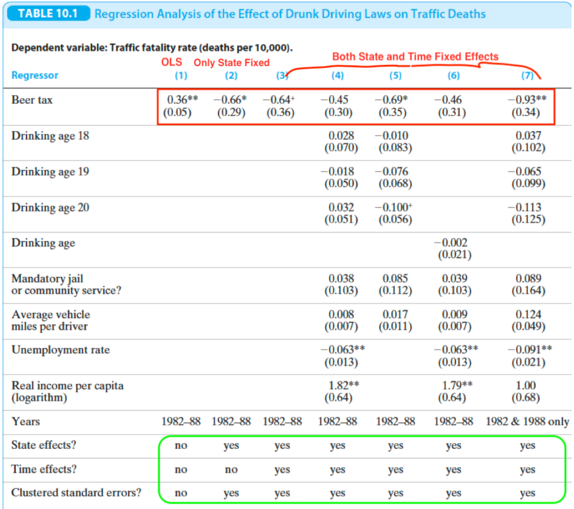


TABLE 10.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

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

OLS

Only State Fixed

Both State and Time Fixed Effects

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

OVB and 2SLS in FE

Introduction

- Recall our basic FE model is

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

  - where α_i is the **entity fixed effect**, which control individual unobservables persistent in time.

- But what if *some individual unobservables correlated with X are also changing in time*? thus we will still suffer an OVB bias even we use FE model. More specifically,

$$E(u_{it}|X_{it}, \alpha_i) \neq 0$$

- Then our FE estimator will be **biased** again due to the OVB problem.

Instrumental Variable in FE

- Recall the Within Estimator

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{1it} + \tilde{u}_{it}$$

- Then the Assumption

$$E(\tilde{u}_{it} \mid \tilde{X}_{1it}) = 0$$

- Two Assumptions of IV
  - **Relevance**

$$Cov(\tilde{Z}_{1it}, \tilde{X}_{1it}) \neq 0$$

  - **Exogeneity**

$$Cov(\tilde{Z}_{it}, \tilde{u}_{it}) = 0$$

First-Stage and Reduced Form

- First stage

$$X_{it} = \gamma_1 Z_{it} + \alpha_i + \varepsilon_{it}$$

- Second stage

$$Y_{it} = \beta_1 \hat{X}_{it} + \alpha_i + u_{it}$$

- Reduced form

$$Y_{it} = \delta_1 Z_{it} + \alpha_i + \epsilon_{it}$$

2SLS estimator in FE

- Recall 2sls estimator

$$\hat{\beta}_{demean} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{Y}_{it} \tilde{X}_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}^{2}}$$

- Then 2sls estimator in FE is

$$\hat{\beta}_{demean} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{Y}_{it} \tilde{Z}_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it} \tilde{Z}_{it}}$$

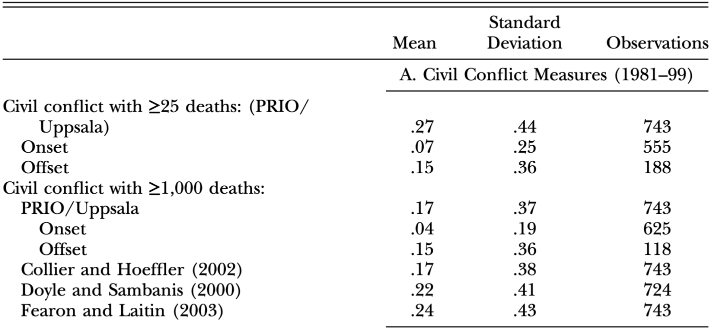
Application: Miguel, Satyanath and Sewrgenti(2004)

Introduction

- *"Economic Shocks and Civil Conflict: An Instrumental Variables Approach",Journal of Political Economy, 2004, vol(112),no.4*
- **Topic**: *Economic Shocks and Civil Conflict*
  - Civil wars have resulted in 3 times as many deaths as wars between states since WW II (Fearon and Laitin 2003).
  - Sub-Saharan Africa: 29 of 43 countries suffered from civil conflict during the 1980s and 1990s.
  - Previous research highlights the **association between economic conditions and civil conflict** rather than a causal relationship.
- **Empirical Strategy**: Use *exogenous variation in rainfall as an IV* for income growth to estimate the impact of economic growth on civil conflict.

Data and Measurement

- Armed Conflict Data: PRIO/Uppsala database
  - small conflicts($>=$ 25) deaths per year
  - conflicts with the standard 1,000-death per year
  - coded as a dummy variable(Yes=1,No=0)

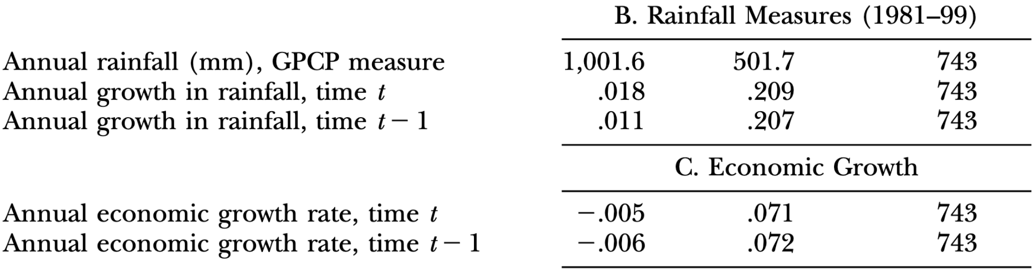


	Mean	Standard Deviation	Observations
	A. Civil Conflict Measures (1981–99)		
Civil conflict with ≥25 deaths: (PRIO/ Uppsala)	.27	.44	743
Onset	.07	.25	555
Offset	.15	.36	188
Civil conflict with ≥1,000 deaths:			
PRIO/Uppsala	.17	.37	743
Onset	.04	.19	625
Offset	.15	.36	118
Collier and Hoeffler (2002)	.17	.38	743
Doyle and Sambanis (2000)	.22	.41	724
Fearon and Laitin (2003)	.24	.43	743

Data and Measurement

- Rainfall Data: Global Precipitation Climatology Project (GPCP) database of monthly rainfall estimates.
- The principal measure of a rainfall shock is the **proportional change** in rainfall from the previous year,

$$\Delta R_{it} = \frac{R_{it} - R_{i,t-1}}{R_{i,t-1}}$$



	B. Rainfall Measures (1981–99)		
Annual rainfall (mm), GPCP measure	1,001.6	501.7	743
Annual growth in rainfall, time t	.018	.209	743
Annual growth in rainfall, time $t-1$	.011	.207	743
	C. Economic Growth		
Annual economic growth rate, time t	−.005	.071	743
Annual economic growth rate, time $t-1$	−.006	.072	743

Empirical Strategy: OLS-FE

- Simple OLS-FE Model

$$\text{conflict}_{it} = X'_{it}\beta + \gamma_0 \text{ growth }_{it} + \alpha_i + year_t + \epsilon_{it}$$

- Extended OLS-FE Model: one-period-lagged effect and a country-specific time trends,$\delta_i \times \text{year}_t$

$$\text{conflict }_{it} = X'_{it}\beta + \gamma_0 \text{ growth }_{it} + \gamma_1 \text{ growth }_{i,t-1} + \alpha_i + \delta_i \times \text{ year }_t + \epsilon_{it}$$

- Potential bias?

Empirical Strategy: IV-FE

- **First-stage**: Take rainfall shocks as IVs

$$\text{growth}_{it} = X_{it}'\beta + c_0\Delta R_{it} + c_1\Delta R_{i,t-1} + \alpha_i + \delta_i \times \text{ year }_t + \epsilon_{it}$$

Empirical Strategy: IV-FE

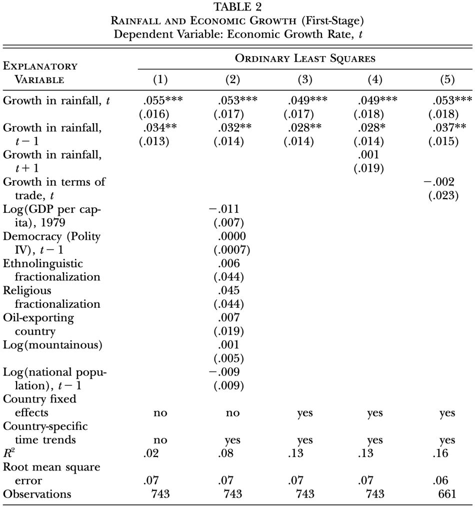


TABLE 2

Rainfall and Economic Growth (First-Stage)

Dependent Variable: Economic Growth Rate, t

Explanatory Variable	Ordinary Least Squares				
	(1)	(2)	(3)	(4)	(5)
Growth in rainfall, t	.055***	.053***	.049***	.049***	.053***
	(.016)	(.017)	(.017)	(.018)	(.018)
Growth in rainfall, $t-1$	.034**	.032**	.028**	.028*	.037**
	(.013)	(.014)	(.014)	(.014)	(.015)
Growth in rainfall, $t+1$				.001	
				(.019)	
Growth in terms of trade, t					−.002
					(.023)
Log(GDP per capita), 1979		−.011			
		(.007)			
Democracy (Polity IV), $t-1$		.0000			
		(.0007)			
Ethnolinguistic fractionalization		.006			
		(.044)			
Religious fractionalization		.045			
		(.044)			
Oil-exporting country		.007			
		(.019)			
Log(mountainous)		.001			
		(.005)			
Log(national population), $t-1$		−.009			
		(.009)			
Country fixed effects	no	no	yes	yes	yes
Country-specific time trends	no	yes	yes	yes	yes
R^2	.02	.08	.13	.13	.16
Root mean square error	.07	.07	.07	.07	.06
Observations	743	743	743	743	661

Empirical Strategy: IV-FE

- Reduced-Form:

$$\text{conflict}_{it} = X_{it}'\beta + b_0 \Delta R_{it} + b_1 \Delta R_{i,t-1} + \alpha_i + \delta_i \times \text{ year }_t + u_{it}$$

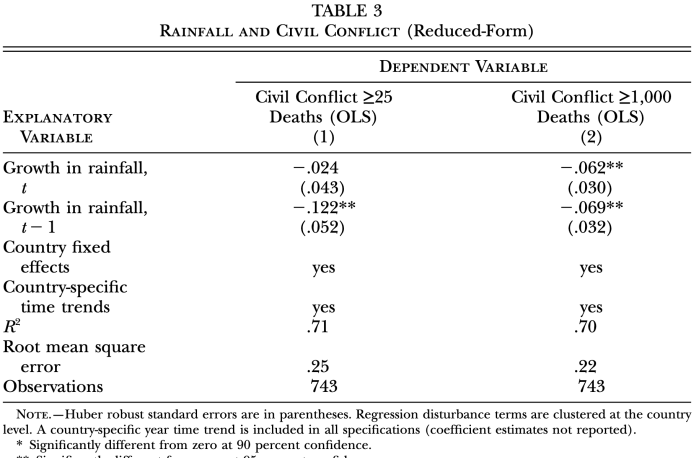


TABLE 3
Rainfall and Civil Conflict (Reduced-Form)

Explanatory Variable	Dependent Variable: Civil Conflict ≥25 Deaths (OLS) (1)	Dependent Variable: Civil Conflict ≥1,000 Deaths (OLS) (2)
Growth in rainfall, t	−.024	−.062**
	(.043)	(.030)
Growth in rainfall, $t-1$	−.122**	−.069**
	(.052)	(.032)
Country fixed effects	yes	yes
Country-specific time trends	yes	yes
R^2	.71	.70
Root mean square error	.25	.22
Observations	743	743

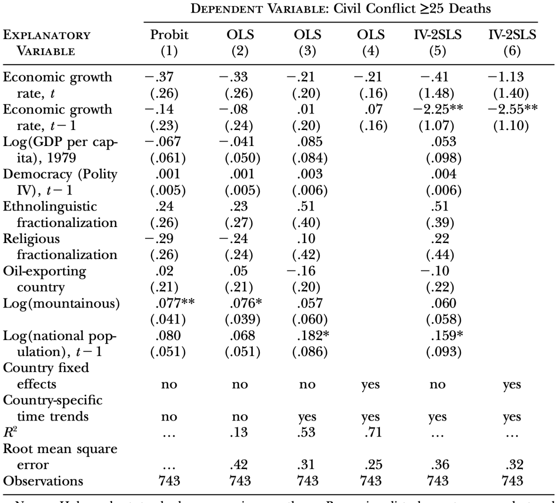
Note.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. A country-specific year time trend is included in all specifications (coefficient estimates not reported).

* Significantly different from zero at 90 percent confidence.

** Significantly different from zero at 95 percent confidence.

Empirical Strategy: IV-FE

- **Second-stage**: $\text{conflict}_{it} = X'_{it}\beta + \gamma_0 \widehat{\text{growth}}_{it} + \gamma_1 \widehat{\text{growth}}_{i,t-1} + \alpha_i + \delta_i \times \text{year}_t + \epsilon_{it}$



Explanatory Variable	Dependent Variable: Civil Conflict ≥25 Deaths					
	Probit (1)	OLS (2)	OLS (3)	OLS (4)	IV-2SLS (5)	IV-2SLS (6)
Economic growth rate, t	−.37 (.26)	−.33 (.26)	−.21 (.20)	−.21 (.16)	−.41 (1.48)	−1.13 (1.40)
Economic growth rate, $t-1$	−.14 (.23)	−.08 (.24)	.01 (.20)	.07 (.16)	−2.25** (1.07)	−2.55** (1.10)
Log(GDP per capita), 1979	−.067 (.061)	−.041 (.050)	.085 (.084)		.053 (.098)	
Democracy (Polity IV), $t-1$	.001 (.005)	.001 (.005)	.003 (.006)		.004 (.006)	
Ethnolinguistic fractionalization	.24 (.26)	.23 (.27)	.51 (.40)		.51 (.39)	
Religious fractionalization	−.29 (.26)	−.24 (.24)	.10 (.42)		.22 (.44)	
Oil-exporting country	.02 (.21)	.05 (.21)	−.16 (.20)		−.10 (.22)	
Log(mountainous)	.077** (.041)	.076* (.039)	.057 (.060)		.060 (.058)	
Log(national population), $t-1$	.080 (.051)	.068 (.051)	.182* (.086)		.159* (.093)	
Country fixed effects	no	no	no	yes	no	yes
Country-specific time trends	no	no	yes	yes	yes	yes
R^2	...	.13	.53	.71	...	...
Root mean square error	...	.42	.31	.25	.36	.32
Observations	743	743	743	743	743	743

Main Results

- OLS: Contemporaneous and lagged economic growth rates are negatively, though not statistically significantly, correlated with conflicts.
- IV-2SLS with country controls: -2.25 (s.e 1.07) on lagged growth, which is significant at 5$\%$ significant level.
- IV-2SLS with FE: -2.55 (s.e 1.10) on lagged growth, which is significant at 5$\%$ significant level.
- Economic significance: The size of the estimated impact is huge.
  - 1$\%$ point decline in GDP increases the likelihood of civil conflict by over 2$\%$ points.
  - 5$\%$ point decline in GDP increases the likelihood of civil conflict by over 12$\%$ points,which amounts to an increase of almost one-half(average is 27).

Summary

Wrap up

- We’ve showed that how panel data can be used to control for **unobserved omitted variables** that differ across entities but are **constant** over time.
- The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics.
- Double fixed Effects model, thus both entity and time fixed effects can be included in the regression to control for variables that vary across entities but are constant over time and for variables that vary over time but are constant across entities.

Wrap up

- Despite these virtues, one shortcoming of fixed effect model is that it will **exaggerate the attenuation bias** as when X is measured with some errors.
- Second,fixed effect model eliminate the OVB bias with demean or differences. But in the mean time, it also **diminishes the variations of Xs** significantly, which will make the estimate less precise.
  - If the treatment variable of the interest is also constant, then it will gone when you use fixed effect model.
- Last but not least, entity and time fixed effects regression *cannot* control for *omitted variables* that *vary both across entities and over time*.There remains a need for new methods that can eliminate the influence of unobserved omitted variables.