ECON2228.05 Econometric Methods

Module 6
Panel Data Models: Fixed Effects and Random Effects

Alberto Cappello

Department of Economics, Boston College

Spring 2025

Panel Data: What and Why

- 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:

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

- The first subscript, i, refers to the entity being observed, and the second subscript, t, refers to the time at which it is observed.
- Panel data allows the study of both between-entity and within-entity variation.

Introduction

- Panel data refers to data with observations on multiple entities, where each entity is observed at two or more points in time.
- Balanced vs 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 vs 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.

Fatality Rate and State Beer-Tax from 1982 to 1988

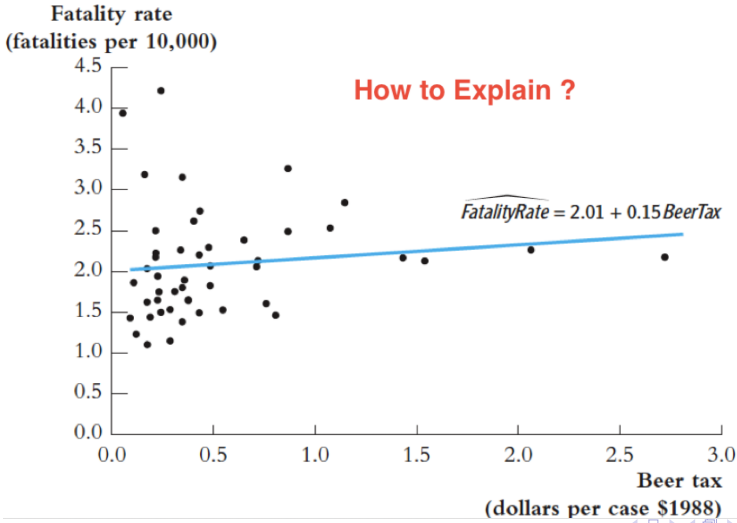
	state	year	beertax	fatal	pop	fa_rate
1	al	1982	1.53937948	839	3942002.2	2.12836
2	al	1985	1.65254235	882	4021007.8	2.19348
3	al	1984	1.71428561	932	3988991.8	2.33643
4	al	1983	1.78899074	930	3960008.0	2.34848
5	al	1988	1.50144362	1023	4101992.2	2.49391
6	al	1986	1.60990703	1081	4049993.8	2.66914
7	al	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

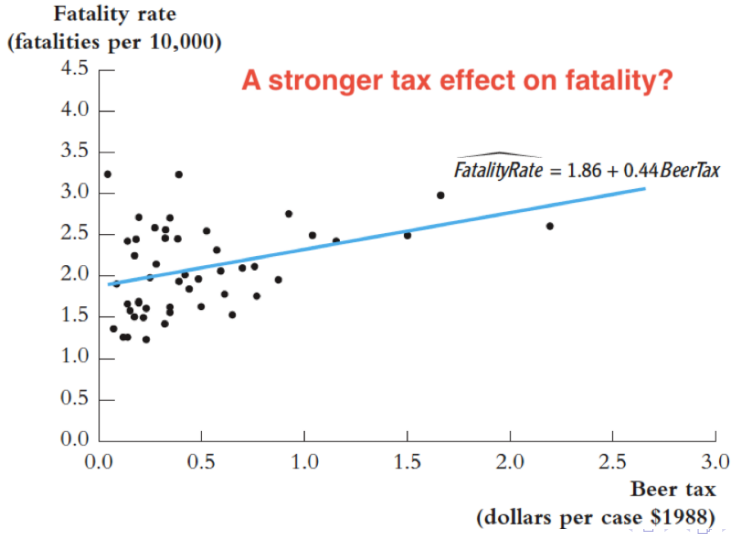
- Observational unit: one year in one U.S. state
- Total 48 U.S. states, so N = the number of entities = 48
- 7 years (1982,..., 1988), so T = the number of time periods = 7
- Balanced panel, so total number of observations $NT = 7 \times 48 = 336$
- Variables:
  - Dependent Variable: Traffic fatality 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 $t = 1982, 1988$:

$$\text{FatalityRate}_{it} = \beta_0 t + \beta_1 t\, \text{BeerTax}_{it} + u_{it}$$

Traffic death data for 1982



Traffic death data for 1988



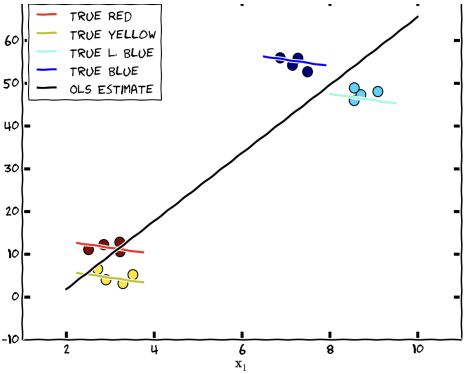
Fixed Effects: Unobserved Time Invariant Factors

- OLS Assumption $E(u_{it}|Z_i) = 0$ may not be satisfied for some unobservables omitted variable Z_i.
- Unobservable factors Z_i that determine the fatality rate may be correlated with BeerTax, such as local (state-specific) attitudes toward drinking and driving.
- Firstly, adjust our model with some unobservables Z_i:

$$\text{FatalityRate}_{it} = \beta_0 + \beta_1 \text{BeerTax}_{it} + \beta_2 Z_i + u_{it}$$

where Z_i is some unobservable, state-specific, time-invariant factor called **Fixed Effect**.

- The omission of Z_i might cause omitted variable bias (OVB), but we don't have data on Z_i.



Fixed Effects Illustration

Panel Data with Two Time Periods

- Consider the regressions for 1982 and 1988:

$$\text{FatalityRate}_{i1988} = \beta_0 + \beta_1 \text{BeerTax}_{i1988} + \beta_2 Z_i + u_{i1988}$$

$$\text{FatalityRate}_{i1982} = \beta_0 + \beta_1 \text{BeerTax}_{i1982} + \beta_2 Z_i + u_{i1982}$$

Panel Data with Two Time Periods

- Consider the regressions for 1982 and 1988:

$$\text{FatalityRate}_{i1988} = \beta_0 + \beta_1 \text{BeerTax}_{i1988} + \beta_2 Z_i + u_{i1988}$$

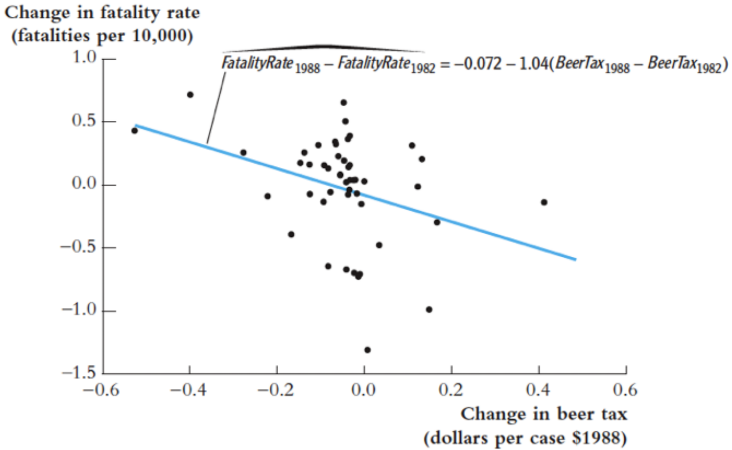
$$\text{FatalityRate}_{i1982} = \beta_0 + \beta_1 \text{BeerTax}_{i1982} + \beta_2 Z_i + u_{i1982}$$

- 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.
- Then take the difference:

$$\underbrace{\text{FatalityRate}_{i1988} - \text{FatalityRate}_{i1982}}_{\Delta FR_i} = \beta_1 \underbrace{(\text{BeerTax}_{i1988} - \text{BeerTax}_{i1982})}_{\Delta BT_i} + \underbrace{(u_{i1988} - u_{i1982})}_{\Delta u_i}$$

- Notice that the unobserved time invariant factor Z_i cancel out (and β_0 as well).
- Assumption: If $E(u_{it}|\text{BeerTax}_{it}, Z_i) = 0$, then Δu_i is uncorrelated with ΔBT_i.
- Then this "difference" equation can be estimated by OLS, even though Z_i isn't observed.
- Intuition: Because Z_i does not change over time, it cannot be a determinant of the change in ΔFR.

After-Before Regression



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).
- 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 Regression Model

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

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

- Where Z_i is an unobserved variable that varies from one state to the next but does not change over time.
- For example, Z_i could 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 fixed effects regression model

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

- α_i are unknown individual specific fixed effects (FE) to be estimated.
- The interpretation of α_i as a state-specific intercept in the example.
- The variation in the entity fixed effects comes from omitted variables that, like Z_i, vary across entities but not over time.
- An alternative way to write the fixed effects model is by using binary (dummy) variables for each entity:

$$Y_{it} = \beta_0 + \beta_1 \text{BeerTax}_{it} + \beta_2 X_{it} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}$$

where D_i is a binary indicator variable for each entity i, taking the value of 1 if the observation belongs to entity i, and 0 otherwise.

Estimation: The "entity-demeaned" estimator

- In principle the binary variable specification of the fixed effects regression model can be estimated by OLS. However, it is tedious to estimate so many fixed effects.
- The first step: Take the average across times t of both sides of the fixed effect regression:

$$\overline{Y_i} = \beta_1 \overline{X_i} + \alpha_i + \overline{u_i}$$

- Demeaned Equation: subtract the above equation from the the fixed effect regression

$$Y_{it} - \overline{Y_i} = \beta_1 (X_{it} - \overline{X_i}) + (\alpha_i - \alpha_i) + u_{it} - \overline{u_i}$$

- Define:

$$\tilde{Y}_{it} = Y_{it} - \overline{Y_i}, \quad \tilde{X}_{it} = X_{it} - \overline{X_i}, \quad \tilde{u}_{it} = u_{it} - \overline{u_i}$$

- Estimate the Demeaned Equation:

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

- The estimator $\hat{\beta}_1$ is known as the demeaned estimator or within estimator. It does not matter 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.

Fixed effects estimator-Within estimator

- The second step of estimating within estimator is (12.5), thus:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \quad (12.5)$$

- Therefore, the fixed effects estimator (demeaned) can be obtained based on the formula of an OLS estimator without intercept, thus:

$$\hat{\beta}_{\text{demean}} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{Y}_{it}\tilde{X}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it}^2} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} (Y_{it} - \overline{Y_i})(X_{it} - \overline{X_i})}{\sum_{i=1}^{N}\sum_{t=1}^{T} (X_{it} - \overline{X_i})^2}$$

- This estimator is identical to the OLS estimator of β_1 without intercept obtained by estimation of the fixed effects model with the individual dummy variables.

Fixed effect estimator: First-differencing

- Recall our fixed effects model is

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

- Then implies:

$$Y_{i1} = \beta_1 X_{i1} + \alpha_i + u_{i1}$$
$$Y_{i2} = \beta_1 X_{i2} + \alpha_i + u_{i2}$$
$$\dots$$
$$Y_{iT} = \beta_1 X_{iT} + \alpha_i + u_{iT}$$

- Taking the differences between consecutive years:

$$Y_{i2} - Y_{i1} = \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$$
$$Y_{i3} - Y_{i2} = \beta_1 (X_{i3} - X_{i2}) + (u_{i3} - u_{i2})$$
$$\dots$$
$$Y_{iT} - Y_{i,T-1} = \beta_1 (X_{iT} - X_{i,T-1}) + (u_{iT} - u_{i,T-1})$$

Fixed effect estimator(III): first-differencing

- New notation, we use Δ to represent the change from the preceding year, then:

$$\Delta Y_{i2} = \beta_1 \Delta X_{i2} + \Delta u_{i2}$$

$$\Delta Y_{i3} = \beta_1 \Delta X_{i3} + \Delta u_{i3}$$

$$\dots$$

$$\Delta Y_{iT} = \beta_1 \Delta X_{iT} + \Delta u_{iT}$$

- The first-difference fixed effect model is:

$$\Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta u_{it} \quad \text{for} \quad i = 1, \dots, N, \quad t = 2, \dots, T \quad (12.6)$$

- Then, the 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:

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

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

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

  ie the error term has mean zero, given the state fixed effect and the entire history of the X and the unobserved individual FE α_i.
- Assumption 2: $(X_{i1}, X_{i2}, \ldots, X_{iT}, u_{i1}, u_{i2}, \ldots, u_{iT})$, for $i = 1, 2, \ldots, n$, are i.i.d.
- Assumption 3: There is no perfect multicollinearity.

Statistical Properties

- Under these assumptions, we can prove the estimator of fixed effects (demeaned and first-differencing) model is unbiased and consistent.
- If the fixed effects regression assumptions hold, then the sampling distribution of the fixed effects OLS estimator is normal in large samples.
- The variance of the estimator can be estimated from the data, and its square root gives the standard error.
- The standard error is 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

- The overall R^2 measures the proportion of the total variation in the dependent variable (i.e., across both time and entities) explained by the independent variables. It combines the *within* and *between* variations into a single measure.

$$R^2_{\text{overall}} = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\beta}X_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y})^2}$$

- The between R^2 measures the proportion of the variation in the dependent variable that is explained by the independent variables across different entities, but not over time for each entity.

$$R^2_{\text{between}} = 1 - \frac{\sum_{i=1}^{N}(\bar{y}_i - \hat{\beta}\bar{X}_i)^2}{\sum_{i=1}^{N}(\bar{y}_i - \bar{y})^2}$$

- The within R^2 uses demeaned data and therefore focuses on how well the model explains the variation over time relative to the mean of each individual.

$$R^2_{\text{within}} = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i - \hat{\beta}(X_{it} - \bar{X}_i))^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2}$$

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:

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

where $Time_t$ is unobserved and represents variables that change over time but are constant across states. If $Time_t$ is correlated with X_{it}, then omitting $Time_t$ from the regression leads to omitted variable bias.

Time Effects

- Similarly, the presence of time fixed effects 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.
- Just as the entity fixed effects regression model, the time fixed effects regression model can be represented using $T - 1$ binary indicators $Time_t$:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 Time_2 + \cdots + \delta_T Time_T + \alpha_i + u_{it}$$

where $\delta_2, \delta_3, \ldots, \delta_T$ are unknown coefficients.

Both Entity and Time Fixed Effects

- Example: 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, the 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 using indicator variables:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \cdots + \gamma_n D_{ni} + \delta_2 Time_2 + \delta_3 Time_3 + \cdots + \delta_T Time_T + u_{it}$$

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!
- The regression equation is:

$$\text{FatalityRate}_{it} = \beta_0 + \beta_1 \text{BeerTax}_{it} + \sum_{i=1}^{47} \alpha_i D_i + \sum_{t=1}^{6} \lambda_t \mathit{Time}_t + u_{it}$$

  where α_i are state fixed effects and λ_t are time fixed effects.
- 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.

Autocorrelation

- Does it make sense to assume zero autocorrelation with panel data? In other words, can we assume that

$$Cov(u_{it}, u_{is}) = 0 \quad \text{for} \quad t \neq s$$

  No, panel data observations of are typically correlated. over time. This type of correlation is called autocorrelation or serial correlation.

- The covariance between Y_t and its jth lag, Y_{t-j}, is called the jth autocovariance of the series Y_t:

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

- The jth autocorrelation coefficient, also called the serial correlation coefficient, measures the correlation between Y_t and Y_{t-j}:

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

Autocorrelated in Panel Data

- In the traffic fatality example, Y_{it}, the fatality rate in state i in year t, is autocorrelated: If it is high one year relative to its mean value for state i, it will tend to be high the next year too.
- Then, u_{it} would also be autocorrelated, as it consists of time-varying factors that are determinants of Y_{it} but are not included as regressors. Some of these omitted factors might be autocorrelated.

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

- Example: A downturn in the local economy and a road improvement project.

Standard Errors for Fixed Effects Regression

- 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."
- The standard errors (SE) used for FE regression are one type of HAC SE, called clustered SE.
- 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 over time, but treat the errors as uncorrelated across entities.

Application: Drunk Driving Deaths and Beer-Tax

- Two ways to crack 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 a two-way fixed effect model, we control both unobservable variables simultaneously that:
  - Do not change over time.
  - Do not vary across states.

Application: Drunk Driving Deaths and Beer-Tax

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

Regressor	OLS (1)	Only State Fixed (2)	(3)	Both State and Time Fixed Effects (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

Case Study: Instrumental Variable in FE model

- "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.

Instrumental Variable in FE

- 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 controls for individual unobservables persistent in time.
- Recall the Within Estimator:
$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{1it} + \tilde{u}_{it}$$
- The Assumption:
$$E(\tilde{u}_{it}|\tilde{X}_{1it}) = 0$$
- fails if X_{it} is endogenous, and our FE estimator will be biased.
- We need an instrumental variable Z_{it} that satisfies:
  - Relevance:
$$\text{Cov}(\tilde{Z}_{1it}, \tilde{X}_{1it}) \neq 0$$
  - Exogeneity:
$$\text{Cov}(\tilde{Z}_{it}, \tilde{u}_{it}) = 0$$

Empirical Strategy: OLS-FE

- Simple OLS-FE Model:

$$\text{conflict}_{it} = X_{it}\beta + \gamma_0 \text{growth}_{it} + \alpha_i + \text{year}_t + \epsilon_{it}$$

- But what if some individual unobservables correlated with X are also changing in time? Thus, we will still suffer from an OVB bias even if we use the FE model.
- Extended OLS-FE Model: one-period-lagged effect and a country-specific time trend:

$$\text{conflict}_{it} = X_{it}\beta + \gamma_0 \text{growth}_{it} + \gamma_1 \text{growth}_{i,t-1} + \alpha_i + \delta_i \times \text{year}_t + \epsilon_{it}$$

Empirical Strategy: IV-FE

- 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:

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

- First-stage: Take rainfall shocks as IVs:

$$\text{growth}_{it} = X_{it}\beta + c_0 \Delta R_{it} + c_1 \Delta R_{i,t-1} + \alpha_i + \delta_i \times \text{year}_t + \epsilon_{it}$$

- Second-stage: use predicted growth $\hat{\text{growth}}$ to predict conflict

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

First Stage

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 (Second-Stage)

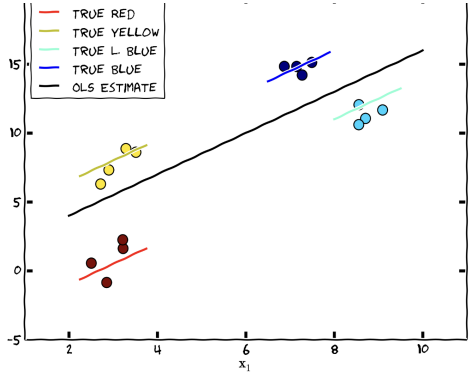
	Dependent Variable: Civil Conflict ≥25 Deaths					
Explanatory Variable	Probit (1)	OLS (2)	OLS (3)	OLS (4)	IV-2SLS (5)	IV-2SLS (6)
Economic growth rate, t	−.37	−.33	−.21	−.21	−.41	−1.13
	(.26)	(.26)	(.20)	(.16)	(1.48)	(1.40)
Economic growth rate, $t-1$	−.14	−.08	.01	.07	−2.25**	−2.55**
	(.23)	(.24)	(.20)	(.16)	(1.07)	(1.10)
Log(GDP per capita), 1979	−.067	−.041	.085		.053	
	(.061)	(.050)	(.084)		(.098)	
Democracy (Polity IV), $t-1$	.001	.001	.003		.004	
	(.005)	(.005)	(.006)		(.006)	
Ethnolinguistic fractionalization	.24	.23	.51		.51	
	(.26)	(.27)	(.40)		(.39)	
Religious fractionalization	−.29	−.24	.10		.22	
	(.26)	(.24)	(.42)		(.44)	
Oil-exporting country	.02	.05	−.16		−.10	
	(.21)	(.21)	(.20)		(.22)	
Log(mountainous)	.077**	.076*	.057		.060	
	(.041)	(.039)	(.060)		(.058)	
Log(national population), $t-1$	.080	.068	.182*		.159*	
	(.051)	(.051)	(.086)		(.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	…	…

Main Results

- OLS: Contemporaneous and lagged economic growth rates are negatively, though not statistically significantly, correlated with conflicts.
- IV 2SLS with country controls: $-2.5(1.10)$ on lagged growth, which is significant at 5% 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).

Random Effects (RE)

- We studied the fixed-effects (FE) model for panel-data estimation.
- We now consider the random-effects (RE) model as an alternative
- It has some advantages over fixed effects:
  - More degrees of freedom
  - Allows regressors that do not vary across time
- It also has a big disadvantage: It is often inconsistent
  - Hausman test can test for this



Random Effects Illustration

Setup of random-effects (RE) model

- As before,

$$y_{it} = \beta_0 + a_i + \beta_1 x_{1it} + \beta_2 x_{2it} + u_{it}$$

- As in FE model, intercept terms a_i differs across cross-sectional entities, and are time invariant for a given entity.
- In random-effects model, intercepts a_i are assumed to be uncorrelated with the other regressors:

$$\mathbb{E}[a_i | x_{1it}, x_{2it}] = 0$$

- Therefore, a_i can be considered as part of the error term
- Hence, define the error term ν_{it} as the sum of two component: (i) one time invariant component a_i, and (ii) one time varying component u_{it}

$$\nu_{it} = a_i + u_{it} \quad \text{with} \quad \text{Var}(a_i) = \sigma_a^2 \quad , \text{Var}(u_{it}) = \sigma_u^2 \quad \text{and cov}(a_i, a_j) = 0 \forall \ i \neq j$$

- and write RE model as

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \nu_{it}$$

Regression in RE model

- Notice that

$$\text{cov}(\nu_{it}, \nu_{jt}) = 0 \quad \forall\ i \neq j \quad \text{and} \quad \forall t$$

but

$$\text{cov}(\nu_{it}, \nu_{is}) \neq 0 \qquad \forall i \quad \text{and} \quad \forall\ s \neq t$$

which implies that the OLS Assumption of zero autocorrelation does not hold because for each entity the error terms are correlated over time due to the common time invariant term a_i.

- Error-components model: v has pattern of autocorrelation between observations on same i
- The autocorrelation is given by

$$\text{corr}(v_{it}, v_{is}) = \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}$$

- Estimate ρ with correlation of residuals within units, then do feasible GLS

Assumptions of RE model

$$\text{var}(v_{it}) = \sigma_u^2 + \sigma_a^2, \quad \text{cov}(v_{it}, v_{is}) = \sigma_a^2, \ t \neq s$$
$$\text{cov}(v_{it}, v_{js}) = 0, \ i \neq j$$
$$\text{cov}(u_{it}, x_{kit}) = 0, \quad k = 1, 2, \ldots, K$$
$$\text{cov}(a_i, x_{kit}) = 0, \quad k = 1, 2, \ldots, K$$

- Last assumption is most likely to be problematic
- Is random effect really uncorrelated with regressors?
- Random effect is unobserved variables that have values specific to i
- We would often expect to be correlated with x values specific to i
- RE estimators are biased and inconsistent if last assumption fails

Feasible GLS estimator for RE model

- Use OLS residuals to estimate σ_a^2 and σ_u^2
- Calculate:

$$\theta = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_a^2}}$$

- Quasi-de-mean model to get v^* error term that satisfies all OLS assumptions (including zero autocorrelation)

$$\begin{aligned} y_{it}^* &\equiv y_{it} - \theta \bar{y}_i \\ x_{0it}^* &= 1 - \theta \\ x_{kit}^* &\equiv x_{kit} - \theta \bar{x}_{ki} \quad k = 1, 2, \ldots, K \\ v_{it}^* &\equiv v_{it} - \theta \bar{v}_i \end{aligned}$$

- Random effects = fixed effects if $\theta = 1$ (as $T \to \infty$)

Problem with random effects

Assumptions of RE model

$$\text{var}(v_{it}) = \sigma_u^2 + \sigma_a^2, \quad \text{cov}(v_{it}, v_{is}) = \sigma_a^2, \ t \neq s$$
$$\text{cov}(v_{it}, v_{js}) = 0, \ i \neq j$$
$$\text{cov}(u_{it}, x_{kit}) = 0, \quad k = 1, 2, \ldots, K$$
$$\text{cov}(a_i, x_{kit}) = 0, \quad k = 1, 2, \ldots, K$$

- Last assumption is most likely to be problematic. Random effect is unobserved variables that have values specific to i. We would often expect to be correlated with x values specific to i
- RE estimators are biased and inconsistent if last assumption fails
- Hausman test compares results of RE and FE estimator
  - Null hypothesis is that RE is similar to FE and valid
  - Reject if results of FE and RE differ significantly
- Common practice: (i) FE is better since it relies on weaker assumptions, (ii) RE are harder to justify and cannot be used if Hausman test rejects its validity.