Panel Data and Fixed Effects

04 Nov 2026 | Chapter 10

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Recall: the problem that IV addressed

The cigarette example used a tax to isolate price variation unrelated to omitted demand shocks.

Now imagine observing the same state repeatedly. Could comparing the state with itself remove a permanent driving culture? Would it remove a new enforcement campaign?

Explain the difference between a fixed characteristic and one that changes over time.

Beer taxes and traffic deaths

We observe 48 US states from 1982 through 1988: 336 state-year observations.

Y_{it}	Traffic deaths per 10,000 residents
X_{it}	Beer tax, in 1988 dollars per case
i and t	State and year

A permanent driving culture can differ across states. New enforcement can change within a state.

The first can be removed by state fixed effects. The second requires additional information or assumptions.

A cross-state comparison

```
. regress fatality beertax if year==1982, vce(robust)
Linear regression                               Number of obs     =         48
------------------------------------------------------------------------------
             |               Robust
    fatality | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |   .1484604   .1326054     1.12   0.269    -.1184604    .4153812
       _cons |   2.010381   .1495728    13.44   0.000     1.709307    2.311456
------------------------------------------------------------------------------
```

The positive slope describes a comparison *between states* in 1982. Permanent state differences can contribute to that relationship.

How could we compare changes within the same states instead?

A panel and its state-specific baseline

Constructed two-state panel: the same states are observed before and after.

State	Date	Tax X	Fatality rate Y
A	Before	.4	2.4
A	After	.6	2.1
B	Before	.3	1.8
B	After	.3	1.7

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}.$$

α_i collects the state's persistent baseline. u_{it} contains determinants that can change over time.

The before-after equation

For two dates, the same state's equations are

$$Y_{i2} = \alpha_i + \beta X_{i2} + u_{i2},$$
$$Y_{i1} = \alpha_i + \beta X_{i1} + u_{i1}.$$

Subtract the first date from the second:

$$\Delta Y_i = \beta \Delta X_i + \Delta u_i.$$

The fixed component cancels. A common time movement remains as an intercept when we compare changes across states.

Exercise: two changes and a comparison

Illustration: State A's tax rises from 0.4 to 0.6 and its fatality rate falls from 2.4 to 2.1. State B's tax stays at 0.3 and its rate falls from 1.8 to 1.7. Calculate each change and the difference between those changes.

Then calculate $(\Delta Y_A - \Delta Y_B)/(\Delta X_A - \Delta X_B)$. What common movement does this second comparison remove?

Debrief: the common change cancels

State	Tax change	Fatality-rate change
A	0.2	-0.3
B	0	-0.1

$$\frac{-0.3-(-0.1)}{0.2-0}=-1.$$

The comparison removes an additive movement shared by both states. A new campaign only in A can still confound it.

Stata: changes from 1982 to 1988

```
. regress dy dx, vce(robust)
Linear regression                               Number of obs     =         48
------------------------------------------------------------------------------
             |               Robust
          dy | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          dx |  -1.040973   .3550061    -2.93   0.005    -1.755563   -.3263823
       _cons |  -.0720371   .0653552    -1.10   0.276    -.2035903    .0595161
------------------------------------------------------------------------------
```

Here dy and dx are each state's 1988 value minus its 1982 value. The do-file shows the reshaping and change calculations.

The fitted slope is about -1.041 deaths per 10,000 residents per tax-dollar increase.

Fixed effects use all the observed years

For each state, average the model across its observed years:

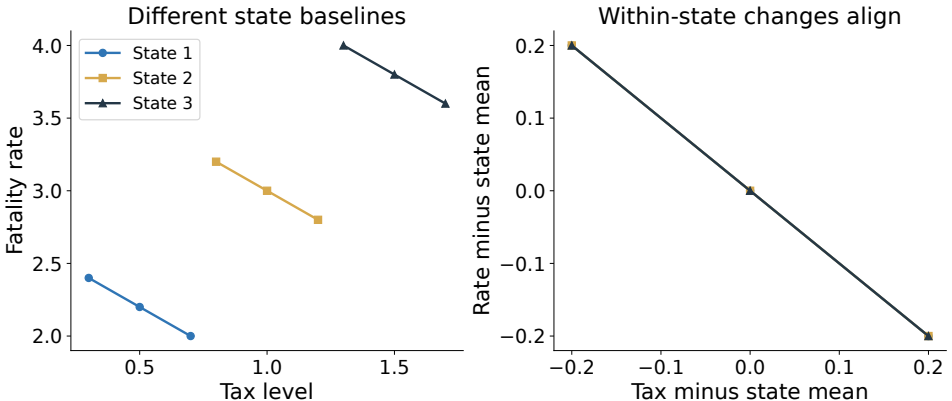
$$\overline{Y}_i = \alpha_i + \beta \overline{X}_i + \overline{u}_i.$$

Subtract that average from the observation for year t:

$$Y_{it} - \overline{Y}_i = \beta(X_{it} - \overline{X}_i) + (u_{it} - \overline{u}_i).$$

The state baseline disappears. The slope uses deviations from each state's own mean.

Levels and within-state deviations



Exercise: variation around a state mean

A state's taxes in three years are 0.3, 0.5 and 0.7. Its fatality rates are 2.4, 2.2 and 2.0. Calculate both means and the three demeaned values of each variable.

Could a characteristic that stays constant in all three years have a separate coefficient in this fixed-effects regression?

Debrief: demeaning and absorbed characteristics

The state means are $\overline{X}_i = 0.5$ and $\overline{Y}_i = 2.2$.

Year	$X_{it} - \overline{X}_i$	$Y_{it} - \overline{Y}_i$
1	-0.2	0.2
2	0	0
3	0.2	-0.2

Each observation is now measured relative to its state's own baseline.

A time-invariant state characteristic subtracts to zero, so its separate level contribution is absorbed by state fixed effects.

State and year fixed effects

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

- α_i allows each state its own persistent baseline.
- λ_t allows each year its own common shift.

The tax slope uses state-specific tax movements after accounting for common year patterns.

A nationwide safety improvement can be absorbed by year effects. An enforcement change in only one state can remain in u_{it}.

Exercise: what variation remains?

Consider a model with state and year fixed effects.

1. Can it separately estimate the effect of a state's coastline indicator, constant over time?
2. Can it separately estimate the effect of a national policy indicator that is identical for all states in each year?
3. A state's enforcement effort changes when its beer tax changes. Does either set of fixed effects automatically remove this problem?

Debrief: the identifying assumption

The coastline indicator is absorbed by state effects. The national policy indicator is absorbed by year effects.

The enforcement concern remains. A standard fixed-effects assumption is

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

with year effects included in the specification.

In words: after accounting for fixed state differences and common year shifts, remaining shocks must be unrelated to the state's tax history.

Fixed effects and clustered standard errors

Repeated observations of the same state may have correlated errors across years.

Fixed effects change the regression comparison by accounting for persistent state baselines.

Clustering by state allows within-state error dependence when estimating uncertainty. In this example there are 48 state clusters, rather than 336 independent observations.

Neither step automatically removes a changing omitted policy.

Stata: state fixed effects

```
. xtreg fatality beertax, fe vce(cluster state)
Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
                                  (Std. err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
    fatality | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2918556    -2.25   0.029    -1.243011   -.0687359
------------------------------------------------------------------------------
```

Selected raw rows. The estimated tax slope is about -0.656, using changes within states.

Stata: state and year fixed effects

```
. xtreg fatality beertax i.year, fe vce(cluster state)
Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
                                  (Std. err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
    fatality | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3570783    -1.79   0.080    -1.358329     .078369
------------------------------------------------------------------------------
```

Selected raw rows, omitting year coefficients. A ten-cent increase gives approximately

$$(-0.640)(0.10) = -0.064$$

deaths per 10,000 residents. The tax coefficient's p-value is 0.080.

A panel-regression explanation

In pairs, explain the two-way estimate in three sentences:

1. Name the comparison that identifies the fitted tax slope.
2. Report the fitted change for a ten-cent tax increase, with units.
3. Identify one remaining threat to a causal interpretation.

Next class begins with Quiz 4, followed by feedback and a panel application.