Section 3
Difference-in-Difference Estimator

Difference-in-Difference Estimator

- Intuitive identification of effect of a program/policy:

1) Compare participants ("treated") and non-participants ("untreated") (in cross-section): simple "differences estimator" (OLS)

   → Problem of unobserved differences between treated and untreated that are correlated with outcomes

Example: Effect of job training program on earnings

Those who participate in job training program are more motivated to work anyways, so would earn more than non-participants even without training program → overestimate effect of program

Difference-in-Difference Estimator

- Intuitive identification of effect of a program/policy:

2) Compare outcome of individuals who participate before and after "treatment" (in panel data set):

   → Problem of time-trends (e.g. business cycles)

Example: Effect of job training on employment

If there is a recession in the time after the job training, then underestimate the effect of the job training.

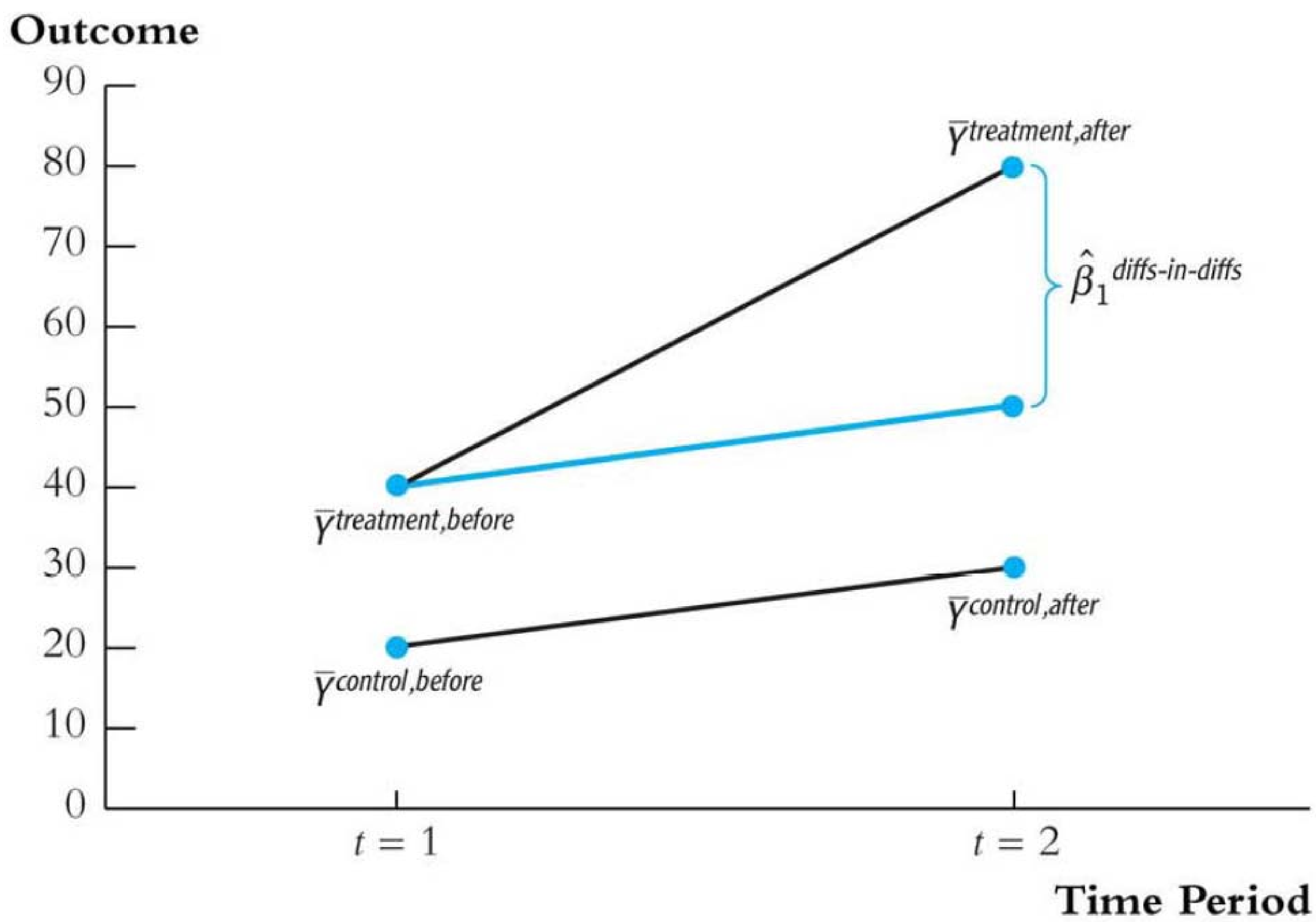
→ Solution: Differences-in-Differences Estimator (DID)

(differences out time-constant (level) differences between treatment and control and time-trends)

Condition: Panel Data, where entities (e.g. individuals or states) are observed at two or more points in time.

Difference-in-Difference Estimator

1) Graphic illustration of the Differences-in-Differences Estimator (DID)



$$\hat{\beta}_1^{diffs-in-diffs} = (\overline{Y}^{treat,after} - \overline{Y}^{treat,before}) - (\overline{Y}^{control,after} - \overline{Y}^{control,before})$$

Difference-in-Difference Estimator

<u>Main assumption of DID:</u>

Counterfactual LEVELS for treated and non-treated can be different,

But their TIME VARIATION is similar:

$\boxed{E(\ Y0(t1) - Y0(t0) \mid D=1) = E(\ Y0(t1) - Y0(t0) \mid D=0)}$ [E(Y0(t1)|D=1) is counterf]

→ "In the absence of treatment, change in treated outcome would have been as change in non-treated outcome,

i.e. changes in the economy or life-cycle etc (unrelated to treatment) affect the two groups in a similar way."

<u>Implies relaxing</u> assumption: $E(\ Y0 \mid X,D=0) = E(\ Y0 \mid X,D=1)$ and

$E(\ Y1 \mid X,D=0) = E(\ Y1 \mid X,D=1)$

→ Selectivity bias is allowed even conditional on X, BUT only through an individual fixed effect (i.e. time constant).

- Main assumption of DID:

  $$\boxed{E(Y0(t1) - Y0(t0) \mid D=1)= E(Y0(t1) - Y0(t0) \mid D=0)}$$

- Show how this assumption is used to generate a “control group” that can be substituted in for the missing counterfactual

  $$\begin{aligned} TTE &= E(Y1(t1) - Y0(t1) \mid D=1) \\ &= E(Y1(t1) - Y0(t0) + Y0(t0) - Y0(t1) \mid D=1) \\ &= E(Y1(t1) - Y0(t0) \mid D=1) - E(Y0(t1) - Y0(t0) \mid D=1) \quad \text{[2. term unobs]} \\ &= E(Y1(t1) - Y0(t0) \mid D=1) - E(Y0(t1) - Y0(t0) \mid D=0) \quad \text{[2. term obs]} \end{aligned}$$

Difference-in-Difference Estimator

2) Formally

→ Show that this can be written in a regression framework with individual fixed effects and time fixed effects

Y(it) = a(t)+b*D(it)+m(i)+u(it), where a(t)=time FE, m(i)=indiv FE

Then

Y(i1) -Y(i0) = [a(1)-a(0)]+ b*[D(i1)-D(i0)] +[u(i1)-u(i0)]

and the fixed-effect estimator reduces to

b= E [Y(i1)-Y(i0)|D=1] – E [Y(i1)-Y(i0)|D=0]

→ Sample version of this is the simple DID estimator

→ Assumption in this framework: E [u(i1)-u(i0)|D=1] = E [u(i1)-u(i0)|D=0]

Brief Intro to Panel Data

- Panel data with k regressors
  {X1(it),…, Xk(it), Y(it)},
  i=1,…, n (number of entities), t=1,.., T (number of time periods)

- Another term for panel data is longitudinal data
- Balanced panel: no missing observations
- Unbalanced panel: some entities (states or indiv) are not observed for some time periods

Why are Panel Data useful?
Entity Fixed Effects

With panel data we can control for factors that:

- Vary across entities (indiv or states), but do not vary over time
- Could cause omitted variable bias if they are omitted
- Are unobserved or unmeasured – and therefore cannot be included in the regression

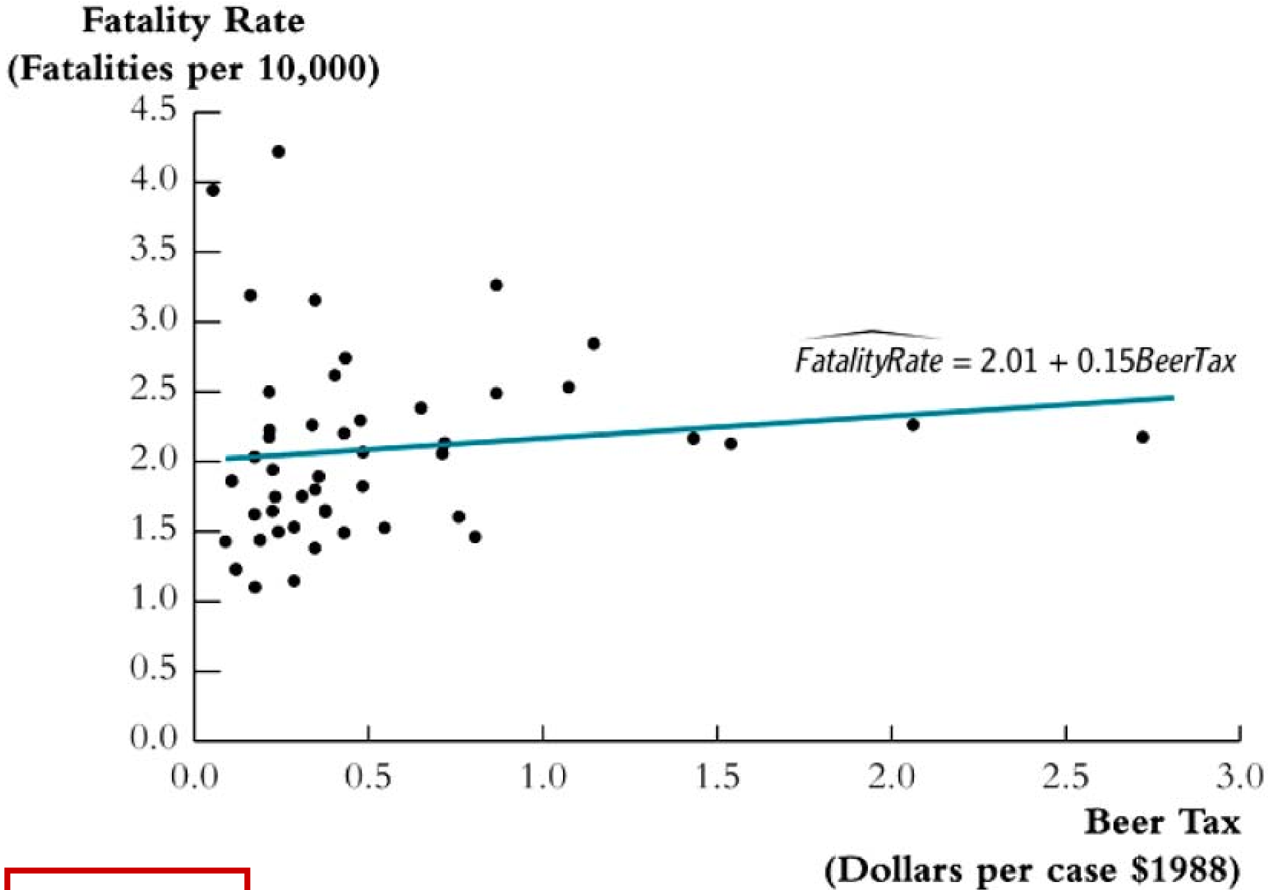
Key Idea:

If an omitted variable does not change over time, then any changes in Y over time cannot be caused by the omitted variable

Example: Can alcohol taxes be a means to reduce traffic deaths?

FIGURE 8.1 **The Traffic Fatality Rate and the Tax on Beer**

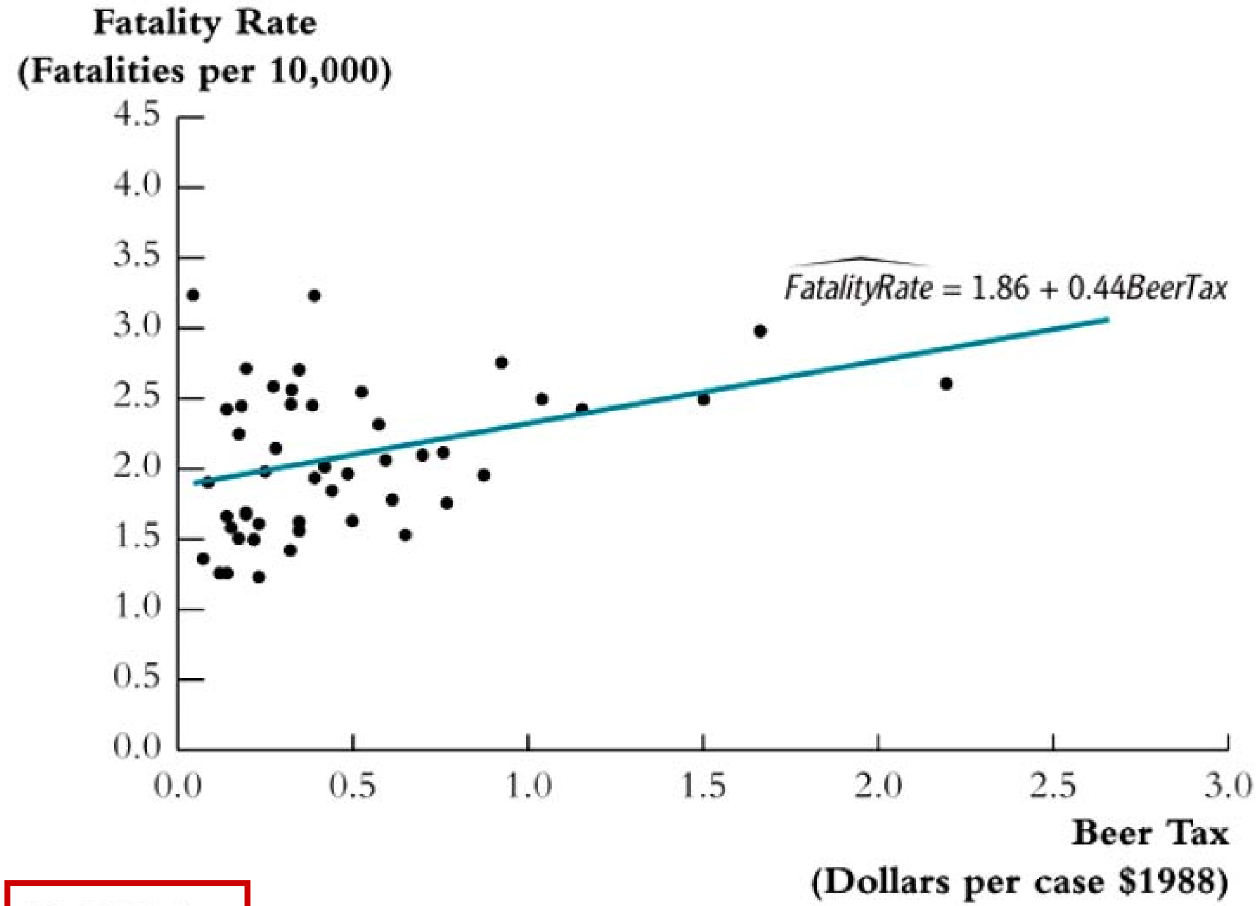
Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(a) 1982 data

FIGURE 8.1 **The Traffic Fatality Rate and the Tax on Beer**

Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(b) 1988 data

Omitted variable bias

- Why might there be more traffic deaths in states that have higher alcohol taxes?
- Other factors that determine traffic fatality rate:
  - Density of cars on the road
  - “Culture” around drinking and driving, etc
- These omitted factors could cause omitted variable bias.
- Example: traffic density, suppose
  1. High traffic density means more traffic deaths
  2. (Western) states with lower traffic density have lower alcohol taxes

  → Two conditions for omitted variable bias are satisfied. Specifically, “high taxes” could reflect “high traffic density”, so the OLS coefficient would be biased upwards.

→ Panel data allows us to eliminate omitted variable bias when the omitted variables are constant over time within a given state.

- Consider the panel data model

  FatalityRate(it)=a+b*BeerTax(it)+c*Z(i)+u(it),

Where Z(i) is a factor that does not change over time (eg traffic 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 T=2 years.

- 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.

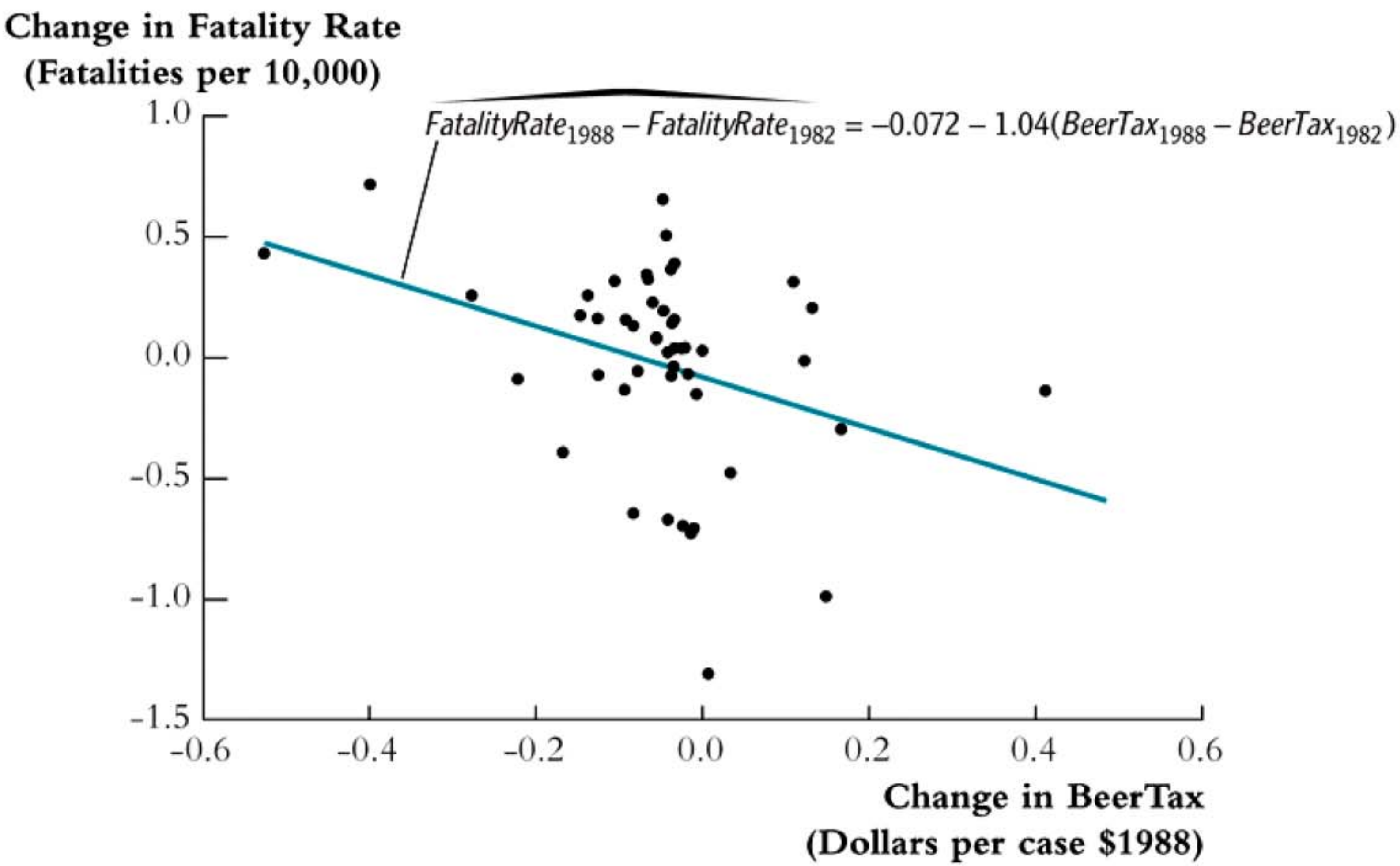
  [FatalityRate(i1988)=a+b*BeerTax(i1988)+c*Z(i)+u(i1988)]

\- [FatalityRate(i1982)=a+b*BeerTax(i1982)+c*Z(i)+u(i1982)]

- This difference eqn can be estimated by OLS, even though Z(i) is not observed.

FIGURE 8.2 **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.



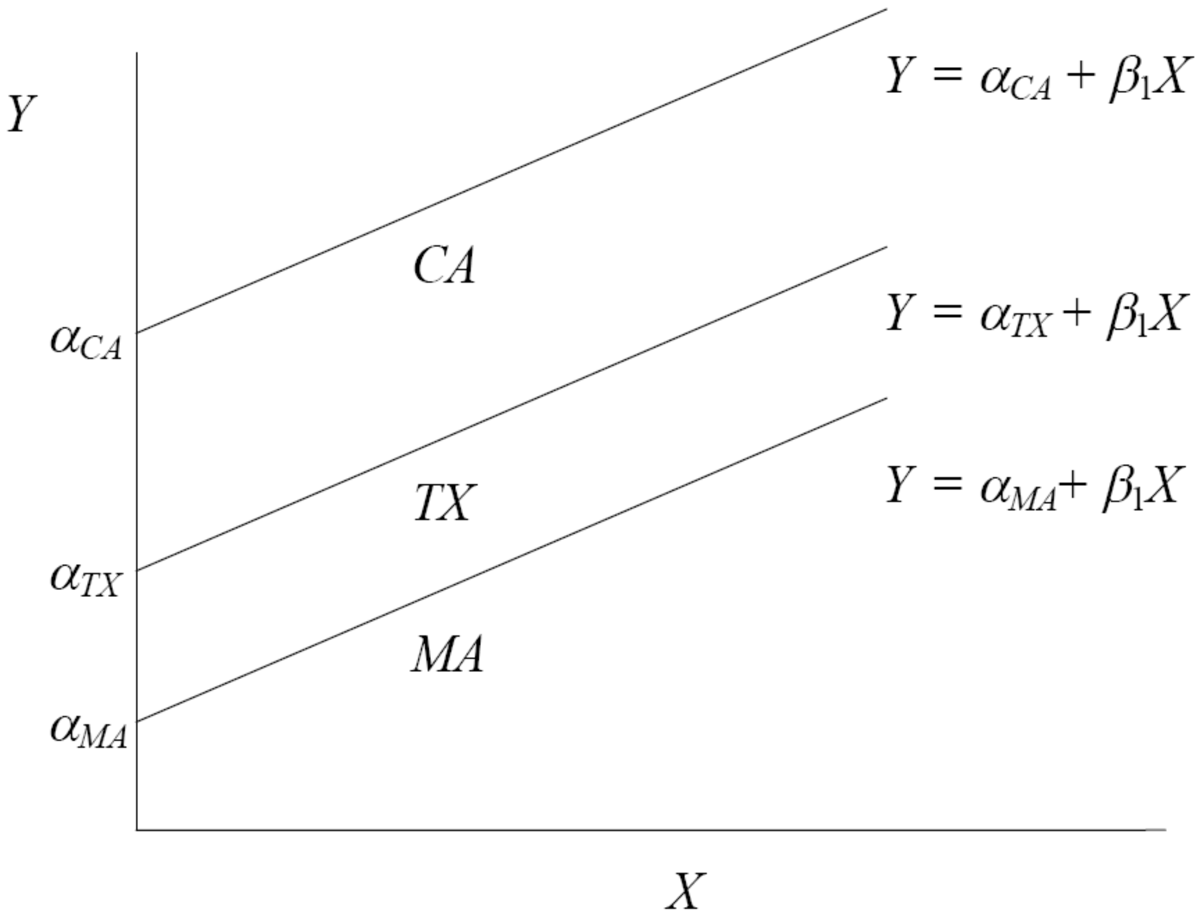
- What if you have more than 2 time periods (T>2)?
- For i=1,…,n and t=1,…, T

  Y(it) = a+b*X(it)+c*Z(i)+u(it)

  we can rewrite this in two useful ways:

1. “Fixed Effects” regression model

   Y(it) = a(i)+b*X(it)+u(it) → intercept a(i) is unique for each state, slope b is the same in all states


Y
α_{CA}
α_{TX}
α_{MA}
CA
TX
MA
$Y = \alpha_{CA} + \beta_1 X$
$Y = \alpha_{TX} + \beta_1 X$
$Y = \alpha_{MA} + \beta_1 X$
X

- What if you have more than 2 time periods ($T>2$)?
- For $i=1,\ldots,n$ and $t=1,\ldots, T$

  $Y(it) = a+b*X(it)+c*Z(i)+u(it)$

  we can rewrite this in two useful ways:

1. "Fixed Effects" regression model

   $Y(it) = a(i)+b*X(it)+u(it)$ → intercept a(i) is unique for each state, slope b is the same in all states

2. "n-1 binary regressors" regression model

   $Y(it) = a+b*X(it)+c2*D2(i)+c3*D3(i)+\ldots+cn*Dn(i)+u(it)$, where $D2(i)=1\{i=2\}$, i.e. D2(i) is 1 if the ith observation is from state 2

Three estimation methods:

1. “n-1” binary regressors” OLS regression
2. “Entity-demeaned” OLS regression
3. “Changes” specification (only works for T=2)

→ These three methods produce identical estimates of the regression coefficients and identical standard errors.

Ad 1. Y(it) = a+b*X(it)+c2*D2(i)+c3*D3(i)+…+cn*Dn(i)+u(it)

- First create the binary variables, D2(i),…, Dn(i)
- Then estimate above equation by OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroscedasticity-robust standard errors!!)
- Impractical when n is very large, although STATA automates dummy variable creation

Ad 2. $(Y_{it} - \overline{Y}_i) = a + b(X_{it} - \overline{X}_i) + u_{it}$

- First construct the demeaned variables
- Then estimate above equation by OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroscedasticity-robust standard errors!!)
- This is like the “changes” approach, but Y(it) is deviated from the state average instead of Y(i1)

Estimation can be done easily in STATA:

- “areg” automatically demeans the data (useful when n large)
- The reported intercept is the average of the n-1 dummy variables (no clear interpretation)

```
areg fatality beertax, absorb(state) r

Regression with robust standard errors                 Number of obs =     336
                                                       F(  1,   287) =   10.41
                                                       Prob > F      =  0.0014
                                                       R-squared     =  0.9050
                                                       Adj R-squared =  0.8891
                                                       Root MSE      =  .18986

------------------------------------------------------------------------------
             |               Robust
    fatality |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2032797    -3.23   0.001    -1.055982   -.2557655
       _cons |   2.377075   .1051515    22.61   0.000     2.170109    2.584041
-------------+----------------------------------------------------------------
       state |   absorbed                                      (48 categories)
```

Why are Panel Data useful?
Time Fixed Effects

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

- Safer cars (air bags, etc); changes in national laws
- These produce intercepts that change over time
- Let these changes ("safer cars") be denoted by the variable, S(t), which changes over time but not states
- The resulting population regression model is

  Y(it)=a+b*X(it)+c*S(t)+u(it)

→ The intercept varies from one year to the next, m(1982)=a+c*S(1982)

Why are Panel Data useful?
Time Fixed Effects

Again two formulations for time fixed effects:

1. “Binary regressor” formulation: “T-1 binary regressors” OLS regression
2. “Time effects” formulation: “Year demeaned” OLS regression (deviate Y(it) and X(it) from year averages), then estimate by OLS

Time and entity fixed effects or “back to DID”

$Y(it) = a(t)+\boxed{b*T(it)}+m(i)+u(it),$

where T(it)=1 if in ***treatment*** group and ***after*** treatment, 0 otherwise

or

$Y(it) = a+ \boxed{b*D(it)*Z(it)} +c*Z(it)+d*D(it)+u(it),$

where D(it)=1 if in ***treatment*** group, 0 otherwise

Z(it)=1 if in ***“after”*** period, 0 in “before” period

D(it)*Z(it)=1 if in ***treatment*** group in ***“after”*** pd (interaction effect)

→ b is the Diff-in-Diff estimator

Estimation

Various equivalent ways to allow for both entity and time fixed effects:

- Differences and intercept (T=2 only)
- Entity (or time) demeaning and T-1 time (or N-1 entity) indicators
- T-1 time indicators and n-1 entity indicators
- Entity and time demeaning

Estimation

- Under assumptions that are basically extensions of the least squares assumptions, the OLS fixed effects estimator of b is normally distributed.
- BUT there are some subtleties associated with computing standard errors that do not come up with cross-sectional data
- Outline:

1. Fixed effects regression assumptions
2. Standard errors for fixed effects regression

Fixed-Effects Regression Assumptions

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

where $X_{it} = (X_{1it}, ..., X_{Kit})'$ is a $k \times 1$ dimensional vector and α_i are entity specific unobservables potentially correlated with X_{it}

1. $\mathbb{E}(u_{it}|X_{i1}, \ldots, X_{iT}, \alpha_i) = 0$
2. We observe an i.i.d. sample $\{W_i\}_{i=1}^n$ where

$$W_i = \{Y_{i1}, .., Y_{iT}, X_{i1}', ..., X_{iT}'\}$$

3. Let $\tilde{X}_{it} = \left(X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}\right)$ ("deviations from time average"). The matrix

$$\mathbb{E}\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}'\right)$$

has rank k

Fixed-Effects Regression Assumptions

Assumption #1: $\mathbb{E}(u_{it}|X_{i1},\ldots,X_{iT},\alpha_i) = 0$

- u_{it} has mean zero, given the state fixed effect and the entire history of the X's for that state.
- This is an extension of the previous multiple regression Assumption #1
- This means there are no omitted lagged effects (any lagged effects of X must enter explicitly)
- Also, there is not feedback from u to future X:
  - Whether a state has a particularly high fatality rate this year doesn't subsequently affect whether it increases the beer tax.

Fixed-Effects Regression Assumptions

Assumption #2: We observe an i.i.d. sample $\{W_i\}_{i=1}^{n}$ where

$$W_i = \{Y_{i1}, .., Y_{iT}, X'_{i1}, ..., X'_{iT}\}$$

- This is an extension of Assumption #2 for multiple regression with cross-section data
- This is satisfied if entities (states, individuals) are randomly sampled from their population by simple random sampling, then data for those entities are collected over time.
- This does not require observations to be i.i.d. over time for the same entity – that would be unrealistic (whether a state has a beer tax this year is strongly related to whether it will have a high tax next year). In fact, want to allow $Corr(u_{it}, u_{is}) \neq 0$ but will discuss case where correlation is zero as well.

Assumption #2

$$\begin{array}{cccccc} & i=1 & i=2 & i=3 & \cdots & i=n \\ t=1 & u_{11} & u_{21} & u_{31} & \cdots & u_{n1} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ t=T & u_{1T} & u_{2T} & u_{3T} & \cdots & u_{nT} \end{array}$$

- Sampling is i.i.d. across entities (so r.vs belonging to two different cols are independent)
- However, within a col, the error terms are not independent (could be correlated). Will need to take this into account when constructing standard errors, but not for consistency (think of this as analogous to the heteroscedasticity problem)

Fixed-Effects Regression Assumptions

Assumption #3: $\mathbb{E}\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}'\right)$ has rank k

- This is the no multicollinearity assumption for the fixed effect case.
- Note that this applies to the demeaned regressors. If X_{it} contains an element that does not vary over time for any entity i, then the corresponding element in $\tilde{X}_{it}$ would be identically zero and the rank condition would fail.
- This assumption therefore shows explicitly why time constant variables are not allowed in fixed effects analysis (unless they are interacted with time-varying variables)

Consistency, Normality and Standard Errors of Fixed Effects Estimator

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

$$\frac{1}{T}\sum_{t=1}^{T} Y_{it} = \frac{1}{T}\sum_{t=1}^{T} X_{it}'\beta + \alpha_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

$$Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \left(X_{it}' - \frac{1}{T}\sum_{t=1}^{T} X_{it}'\right)\beta + \left(u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}\right)$$

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

FE estimator is OLS applied to the above equation

Consistency of Fixed Effects Estimator

Substituting in the equation for $\tilde{Y}_{it}$ into the formula for $\hat{\beta}_{FE}$ we obtain

$$\begin{aligned}\hat{\beta}_{FE} &= \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\left(\tilde{X}_{it}'\beta+\tilde{u}_{it}\right)\right)\\ &= \beta + A_n\end{aligned}$$

where

$$A_n = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)$$

Consistency of Fixed Effects Estimator

Substituting in the equation for $\tilde{Y}_{it}$ into the formula for $\hat{\beta}_{FE}$ we obtain

$$\begin{aligned}\hat{\beta}_{FE} &= \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\left(\tilde{X}_{it}'\beta+\tilde{u}_{it}\right)\right)\\ &= \beta + A_n\end{aligned}$$

where

$$A_n = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)$$

Consistency of Fixed Effects Estimator

Since $\mathbb{E}\left(u_{it} | X_{i1}, \ldots X_{iT}, \alpha_i\right) = 0$

$$\mathbb{E}\left(\tilde{u}_{it} \tilde{X}_{it}\right) = 0$$

for $t = 1 \ldots T$ so that by a LLN for i.i.d. random vectors

$$\boxed{\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it} \tilde{u}_{it}} \rightarrow \mathbb{E}\left(\sum_{t=1}^{T} \tilde{X}_{it} \tilde{u}_{it}\right) = \sum_{t=1}^{T} \mathbb{E}\left(\tilde{X}_{it} \tilde{u}_{it}\right) = 0$$

and likewise

$$\boxed{\left(\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it} \tilde{X}_{it}'\right)} \rightarrow \mathbb{E}\left(\sum_{t=1}^{T} \tilde{X}_{it} \tilde{X}_{it}'\right)$$

which is well-behaved by the full rank assumption

Consistency of Fixed Effects Estimator

Putting the previous results together, we conclude that $A_n \longrightarrow 0$ so that $\boxed{\hat{\beta}_{FE} \longrightarrow \beta}$

Normality of Fixed Effects Estimator

We now write

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) = \sqrt{n}A_n$$

we will show

$$\sqrt{n}A_n \Rightarrow \mathcal{N}\left(0, V\right)$$

for some (soon to be specified) variance matrix V
Then,

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) \Rightarrow \mathcal{N}\left(0, V\right)$$

and as long as we have a good guess $\hat{V}$ for V we will do inference as if $\hat{\beta} - \beta$ has the $\mathcal{N}\left(0, \hat{V}/n\right)$ distribution

Normality of Fixed Effects Estimator

$$\sqrt{n}A_n = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)$$

Two parts:

- By a CLT for the sample average of i.i.d. random vectors

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} \Rightarrow \mathcal{N}\left(\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right), Var\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)\right)$$

Recall that $\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right) = 0$ so the limiting distribution is

$$\mathcal{N}\left(0, \mathbb{E}\left(\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)'\right)\right)$$

Normality of Fixed Effects Estimator

- From earlier analysis of consistency, we know

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1} \to \left(\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\right)^{-1}$$

$$\sqrt{n}A_n \Rightarrow \mathcal{N}\left(0, V\right)$$

where

$$V = \left(\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\right)^{-1}\mathbb{E}\left(\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)'\right)\left(\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{i}\right.\right.$$

Variance-Covariance Matrix

Define the matrices

$$\mathbf{X}_i = \begin{bmatrix} \tilde{X}'_{i1}\big|_{1\times k} \\ \cdot \\ \cdot \\ X'_{iT} \end{bmatrix} \quad \tilde{\mathbf{u}}_i = \begin{bmatrix} \tilde{u}_{i1} \\ \cdot \\ \cdot \\ \tilde{u}_{iT} \end{bmatrix}$$

where $\mathbf{X}_i$ has dimension $T \times k$ and $\tilde{\mathbf{u}}_i$ has dimension $T \times 1$.

$$\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it}\right) = \mathbf{X}'_i\tilde{\mathbf{u}}_i$$

$$\mathbb{E}\left(\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it}\right)\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it}\right)'\right) = \mathbb{E}\left(\mathbf{X}'_i\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}'_i\mathbf{X}_i\right)$$

Use this expression to study serial correlation in the error term

Variance-Covariance Matrix

- In general, want to allow for error terms to be correlated over time for an entity and this makes the formula for asymptotic variance complication.
- This is messy – but you get something for it – you can have correlation of the error for an entity from one time period to the next. This would arise if the omitted variables that make up u_{it} are correlated over time as we often have reason to believe.
- Other names for this asymptotic variance: Heteroscedasticity- and autocorrelation-consistent asymptotic variance(autocorrelation is correlation with other time periods – uit and uis correlated). Clustered standard errors, because there is a grouping, or "cluster," within which the error term is possibly correlated, but outside of which (across groups) it is not.

Variance-Covariance Matrix

Extensions:

o The clusters can be other groupings, not necessarily time

o For example, you can allow for correlation of uit between individuals within a given group, as long as there is independence across groups – for example i runs over individuals, the clusters can be families (correlation of uit for i within same family, not between families).

Variance-Covariance Matrix - special case: no correlation across time within entities

If the error terms are not correlated with each other, conditional on $\mathbf{X}_i$,

$$\mathbb{E}\left(\tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i' | \mathbf{X}_i\right) = \sigma^2 \mathbf{I}_{T \times T}$$

where $\mathbf{I}_{T \times T}$ is the $T \times T$ identity matrix. Then, the form of V simplifies greatly

$$\mathbb{E}\left(\mathbf{X}_i' \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i' \mathbf{X}_i\right) = \sigma^2 \mathbb{E}\left(\mathbf{X}_i' \mathbf{X}_i\right)$$

so that V simplifies to

$$V = \sigma^2 \left(\mathbb{E}\left(\mathbf{X}_i' \mathbf{X}_i\right)\right)^{-1}$$

Analogous to Conditional Homoscedasticity case

Variance-Covariance Matrix

- Last critical bit is an estimator for the variance V. As usual, apply analogy principle and replace population expectations with sample averages, so

$$\mathbb{E}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right) \cong \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right) = V_{1n}$$

- However, estimating $\mathbb{E}\left(\mathbf{X}_i'\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i'\mathbf{X}_i\right)$ by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i'\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i'\mathbf{X}_i\right)$$

is infeasible since $\mathbf{u}_i$ is not observed.

Variance-Covariance Matrix

Replace $\tilde{\mathbf{u}}_i$ by $\hat{\mathbf{u}}_i = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{FE}$ and estimate $\mathbb{E}\left(\mathbf{X}_i' \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i' \mathbf{X}_i\right)$ by

$$V_{2n} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{X}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{X}_i\right)$$

and under an appropriate fourth moment assumption V_{2n} will be consistent for $\mathbb{E}\left(\mathbf{X}_i' \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i' \mathbf{X}_i\right)$. Putting these together, estimate V by

$$\hat{V} = V_{1n}^{-1} V_{2n} V_{1n}^{-1}$$

Case 1: Allowing for Serial Correlation

→ Heteroscedasticity and autocorrelation-consistent asymptotic variance (HAC)

```
areg fatality beertax, absorb(state) r cluster(state)

Regression with robust standard errors                 Number of obs =      336
                                                       F(  1,    47) =     4.34
                                                       Prob > F      =   0.0427
                                                       R-squared     =   0.9050
                                                       Adj R-squared =   0.8891
                                                       Root MSE      =   .18986

                              (standard errors adjusted for clustering on state)
------------------------------------------------------------------------------
             |               Robust
    fatality |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .3148476    -2.08   0.043    -1.289265    -.022482
       _cons |   2.377075   .1615974    14.71   0.000     2.051983    2.702167
-------------+----------------------------------------------------------------
       state |   absorbed                                      (48 categories)
```

Case 2: No Serial Correlation

```
areg fatality beertax, absorb(state) r

Regression with robust standard errors                 Number of obs =      336
                                                       F(  1,    287) =   10.41
                                                       Prob > F      =  0.0014
                                                       R-squared     =  0.9050
                                                       Adj R-squared =  0.8891
                                                       Root MSE      =  .18986

------------------------------------------------------------------------------
             |               Robust
    fatality |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2032797    -3.23   0.001    -1.055982   -.2557655
       _cons |   2.377075   .1051515    22.61   0.000     2.170109    2.584041
-------------+----------------------------------------------------------------
       state |   absorbed                                      (48 categories)
```

- Point Estimates are Identical (Why?)
- Note that standard errors come down by a lot

Case 3: No Serial Correlation and Conditional Homoscedasticity

```
. areg fatality beertax, absorb(state)

                                                    Number of obs =      336
                                                    F(  1,    287) =   12.19
                                                    Prob > F      =   0.0006
                                                    R-squared     =   0.9050
                                                    Adj R-squared =   0.8891
                                                    Root MSE      =   .18986

------------------------------------------------------------------------------
    fatality |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736     .18785    -3.49   0.001    -1.025612   -.2861352
       _cons |   2.377075   .0969699    24.51   0.000     2.186212    2.567937
-------------+----------------------------------------------------------------
       state |          F(47, 287) =     52.179   0.000           (48 categories)
```

- Point Estimates are Identical
- Standard Errors come down even further

$\hat{V} / n$

Allowing Serial Correlation and Heteroscedasticity

```
symmetric e(V)[2,2]
             beertax       _cons
beertax      .099129
  _cons   -.05087855   .02611372
```

No Serial Correlation and Heteroscedasticity

```
symmetric e(V)[2,2]
             beertax       _cons
beertax   .04132264
  _cons  -.02127132   .01105685
```

No Serial Correlation and Homoscedasticity

```
             beertax       _cons
beertax   .03528762
  _cons  -.01811158   .00940316
```

Conclusions based on different assumptions

	(1)	(2)	(3)	(4)
BeerTax	-.656** (.203)	-.656^{+} (.315)	-.640* (.255)	-.640^{++} (.386)
State effects?	Yes	Yes	Yes	Yes
Time effects?	No	No	Yes	Yes
F testing time effects = 0	–	2.47 (.024)	–	3.61 (.005)
Clustered SEs?	No	Yes	No	Yes

Significant at the **1% *5% $^{+}$10% level

This is a hard call – what would you conclude?

Additions to DID

- Note: It is also possible to use repeated cross-sections instead of panel data under certain conditions, e.g. group composition has to be stable (see, e.g., Meyer (1995) or Abadie (2005))
- Application of DID: read Card and Krueger (1995)
- Caveats and extensions:
  - Endogenous treatment (Besley/Case (2000)) → example: DID assumptions exclude the possibility that a state increases the alcohol tax because of high rate of traffic fatalities in the past
  - Parallel trends: Trends can be different in the two groups if the distribution of X is different (Abadie (2005) "Semi-parametric DID")

    → approach mixes "matching" and "diff-in-diff" (discuss later)
  - Inference (Bertrand et al (2004)): when residual autocorrelation over time is not accounted for, the variance may be underestimated

    → additional potential solutions in addition to approach presented before (heteroscedasticity and autocorrelation-consistent asymptotic variance)