Econ 422 – Lecture Notes

(These notes are slightly modified versions of lecture notes provided by Stock and Watson, 2007. They are for instructional purposes only and are not to be distributed outside of the classroom.)

Regression with Panel Data

A ***panel dataset*** contains observations on multiple entities (individuals), where each entity is observed at two or more points in time.

Hypothetical examples:

- Data on 420 California school districts in 1999 *and again* in 2000, for 840 observations total.
- Data on 50 U.S. states, each state is observed in 3 years, for a total of 150 observations.
- Data on 1000 individuals, in four different months, for 4000 observations total.

Notation for panel data

A double subscript distinguishes entities (states) and time periods (years)

i = entity (state), n = number of entities,

so $i = 1,\ldots,n$

t = time period (year), T = number of time periods

so $t = 1,\ldots,T$

Data: Suppose we have 1 regressor. The data are:

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

Panel data notation, ctd.

Panel data with k regressors:

$$(X_{1it}, X_{2it},\ldots,X_{kit}, Y_{it}),\ i = 1,\ldots,n,\ t = 1,\ldots,T$$

n = number of entities (states)

T = number of time periods (years)

Some jargon…

- Another term for panel data is ***longitudinal data***
- ***balanced panel***: no missing observations (all variables are observed for all entites [states] and all time periods [years])

Why are panel data useful?

With panel data we can control for factors that:

- Vary across entities (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 using multiple regression

Here's the 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 of a panel data set: Traffic deaths and alcohol taxes

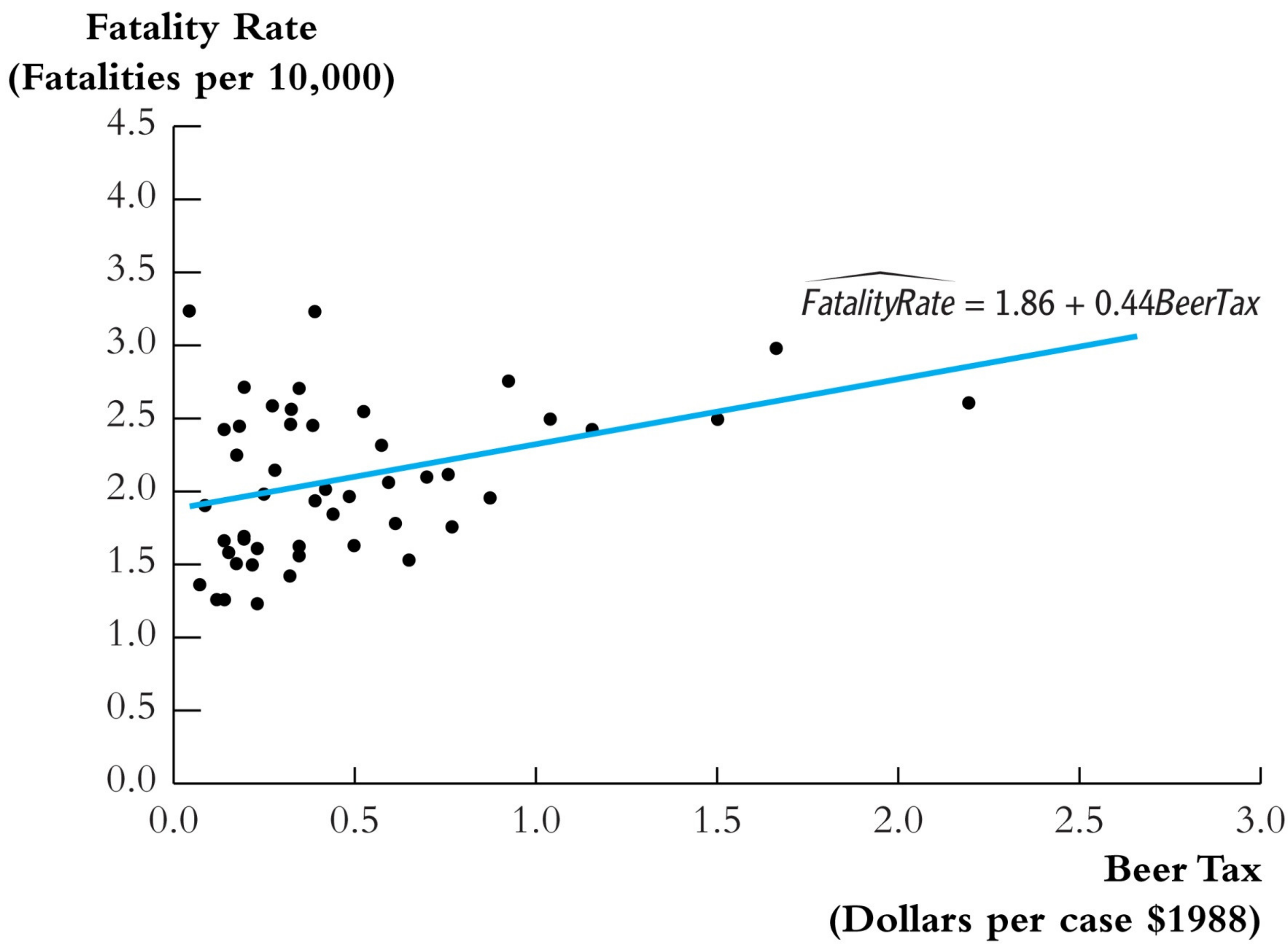
Observational unit: a year in a U.S. state

- 48 U.S. states, so n = # of entities = 48
- 7 years (1982,…, 1988), so T = # of time periods = 7
- Balanced panel, so total # observations = 7×48 = 336

Variables:

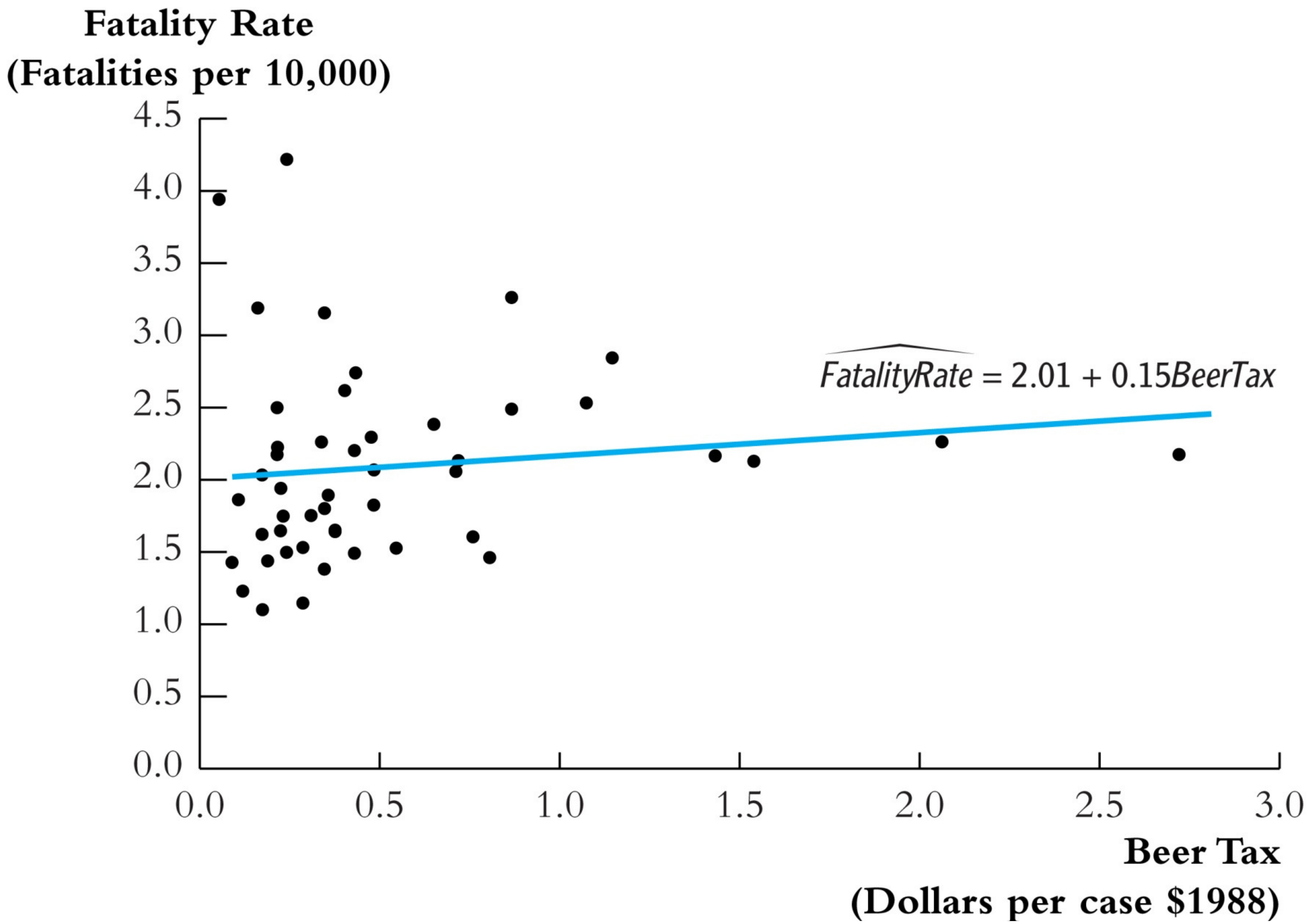
- Traffic fatality rate (# traffic deaths in that state in that year, per 10,000 state residents)
- Tax on a case of beer
- Other (legal driving age, drunk driving laws, etc.)

U.S. traffic death data for 1982:



Higher alcohol taxes, more traffic deaths?

U.S. traffic death data for 1988



Higher alcohol taxes, more traffic deaths?

Why might there be higher *more* traffic deaths in states that have higher alcohol taxes?

Other factors that determine traffic fatality rate:

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

These omitted factors could cause omitted variable bias.

Example #1: traffic density. Suppose:

(i) High traffic density means more traffic deaths
(ii) (Western) states with lower traffic density have lower alcohol taxes

- Then the two conditions for omitted variable bias are satisfied. Specifically, “high taxes” could reflect “high traffic density” (so the OLS coefficient would be biased positively – high taxes, more deaths)
- Panel data lets us eliminate omitted variable bias when the omitted variables are constant over time within a given state.

Example #2: cultural attitudes towards drinking and driving:

(i) arguably are a determinant of traffic deaths; and
(ii) potentially are correlated with the beer tax, so beer taxes could be picking up cultural differences (omitted variable bias).

- Then the two conditions for omitted variable bias are satisfied. Specifically, "high taxes" could reflect "cultural attitudes towards drinking" (so the OLS coefficient would be biased)
- Panel data lets us eliminate omitted variable bias when the omitted variables are constant over time within a given state.

Panel Data with Two Time Periods

Consider the panel data model,

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

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

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

The key idea:

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

The math: consider fatality rates in 1988 and 1982:

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

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

Suppose $E(u_{it}|BeerTax_{it}, Z_i) = 0$.

Subtracting 1988 – 1982 (that is, calculating the change), eliminates the effect of Z_i…

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

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

so

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

- The new error term, $(u_{i1988} - u_{i1982})$, is uncorrelated with either $BeerTax_{i1988}$ or $BeerTax_{i1982}$.
- This "difference" equation can be estimated by OLS, even though Z_i isn't observed.
- The omitted variable Z_i doesn't change, so it cannot be a determinant of the *change* in Y

Example: Traffic deaths and beer taxes

1982 data:

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

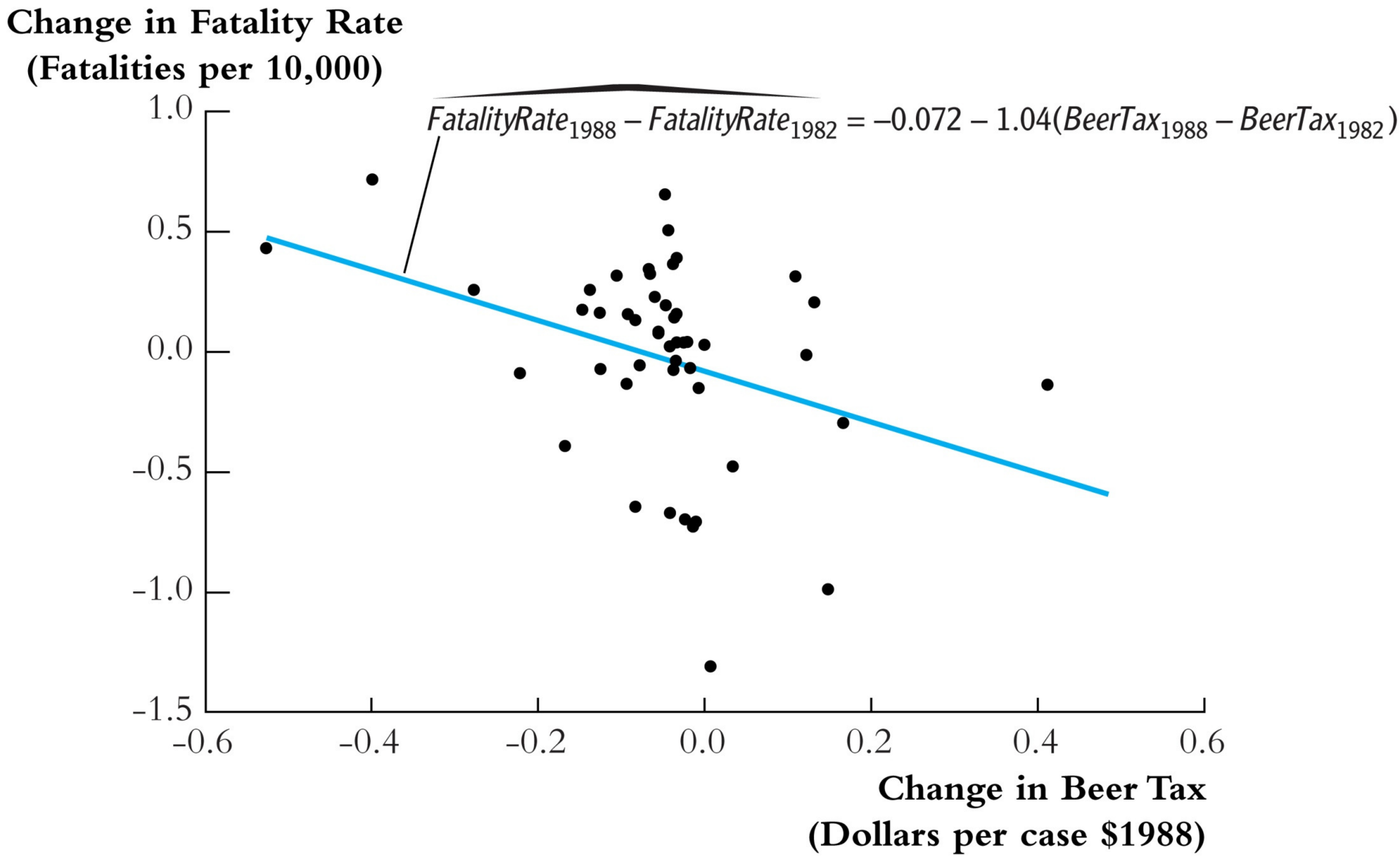
1988 data:

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

Difference regression (n = 48)

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

$\Delta FatalityRate$ v. $\Delta BeerTax$:



Fixed Effects Regression

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

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it},\ i = 1,\ldots,n,\ T = 1,\ldots,T$$

We can rewrite this in two useful ways:

1. "*n*-1 binary regressor" regression model
2. "Fixed Effects" regression model

We first rewrite this in "fixed effects" form. Suppose we have $n = 3$ states: California, Texas, Massachusetts.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_i,\ i = 1,\ldots,n,\ T = 1,\ldots,T$$

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

$$Y_{CA,t} = \beta_0 + \beta_1 X_{CA,t} + \beta_2 Z_{CA} + u_{CA,t}$$

$$= (\beta_0 + \beta_2 Z_{CA}) + \beta_1 X_{CA,t} + u_{CA,t}$$

or

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

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

For TX:

$$Y_{TX,t} = \beta_0 + \beta_1 X_{TX,t} + \beta_2 Z_{TX} + u_{TX,t}$$

$$= (\beta_0 + \beta_2 Z_{TX}) + \beta_1 X_{TX,t} + u_{TX,t}$$

or

$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}, \text{ where } \alpha_{TX} = \beta_0 + \beta_2 Z_{TX}$$

Collecting the lines for all three states:

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

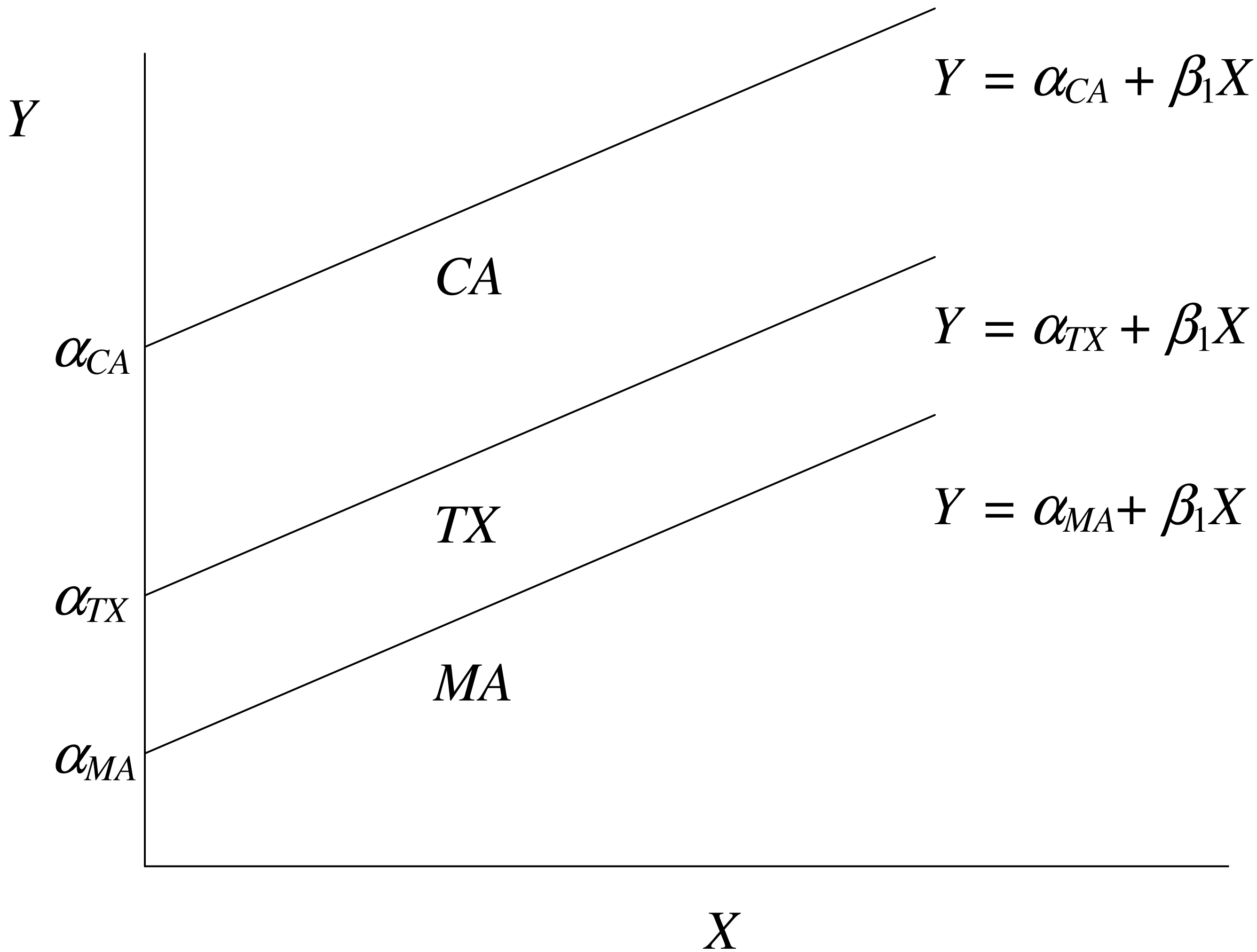
$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}$$

$$Y_{MA,t} = \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}$$

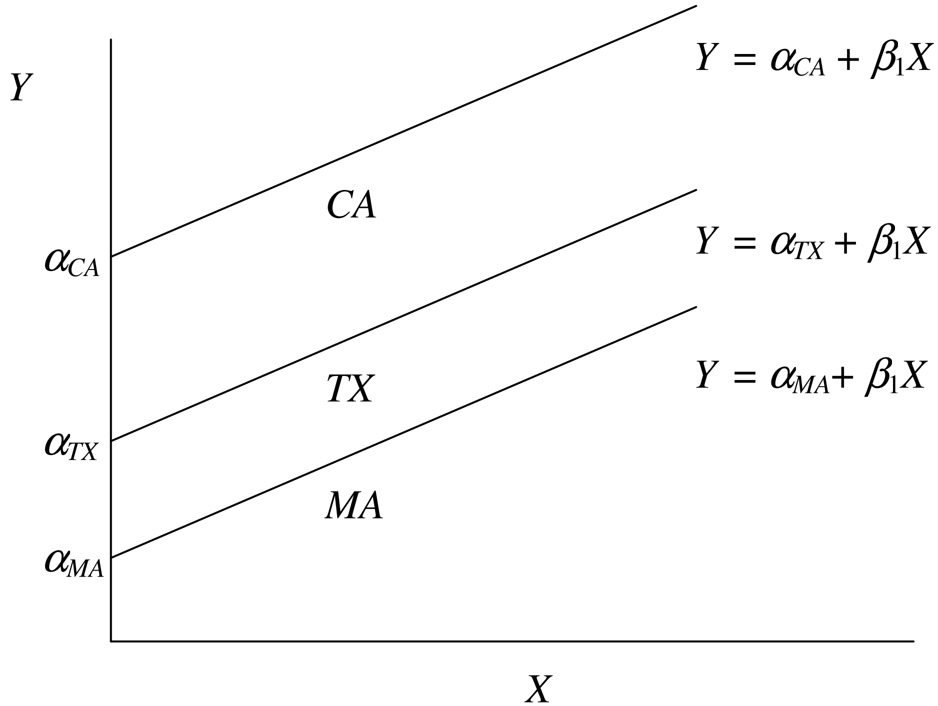
or

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it},\ i = \text{CA, TX, MA},\ T = 1,\ldots,T$$

The regression lines for each state in a picture



Recall that shifts in the intercept can be represented using binary regressors…



In binary regressor form:

$$Y_{it} = \beta_0 + \gamma_{CA}DCA_i + \gamma_{TX}DTX_i + \beta_1 X_{it} + u_{it}$$

- DCA_i = 1 if state is *CA*, = 0 otherwise
- DTX_t = 1 if state is *TX*, = 0 otherwise
- leave out DMA_i (*why*?)

Summary: Two ways to write the fixed effects model

"*n*-1 binary regressor" form

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

where $D2_i = \begin{cases} 1 \text{ for } i\text{=2 (state \#2)} \\ 0 \text{ otherwise} \end{cases}$, etc.

"Fixed effects" form:

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

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

Fixed Effects Regression: Estimation

Three estimation methods:

1. "n-1 binary regressors" OLS regression
2. "Entity-demeaned" OLS regression
3. "Changes" specification, without an intercept (only works for $T = 2$)

- These three methods produce identical estimates of the regression coefficients, and identical standard errors.
- We already did the "changes" specification (1988 minus 1982) – but this only works for $T = 2$ years
- Methods #1 and #2 work for general T
- Method #1 is only practical when n isn't too big

1. "*n*-1 binary regressors" OLS regression

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it} \qquad (1)$$

where $D2_i = \begin{cases} 1 \text{ for } i\text{=2 (state \#2)} \\ 0 \text{ otherwise} \end{cases}$ etc.

- First create the binary variables $D2_i, \ldots, Dn_i$
- Then estimate (1) by OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroskedasticity-robust standard errors)
- This is impractical when *n* is very large (for example if *n* = 1000 workers)

2. "Entity-demeaned" OLS regression

The fixed effects regression model:

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

The state averages satisfy:

$$\frac{1}{T}\sum_{t=1}^{T} Y_{it} = \alpha_i + \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{it} + \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

Deviation from state averages:

$$Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_1 \left(X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it} \right) + \left(u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \right)$$

Entity-demeaned OLS regression, ctd.

$$Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_1 \left(X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it} \right) + \left(u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \right)$$

or

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

where $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\tilde{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$

- For i=1 and t = 1982, $\tilde{Y}_{it}$ is the difference between the fatality rate in Alabama in 1982, and its average value in Alabama averaged over all 7 years.

Entity-demeaned OLS regression, ctd.

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \qquad (2)$$

where $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.

- First construct the demeaned variables $\tilde{Y}_{it}$ and $\tilde{X}_{it}$
- Then estimate (2) by regressing $\tilde{Y}_{it}$ on $\tilde{X}_{it}$ using OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroskedasticity-robust standard errors)
- This is like the "changes" approach, but instead Y_{it} is deviated from the state average instead of Y_{i1}.
- This can be done in a single command in STATA

Example: Traffic deaths and beer taxes in STATA

```
. areg vfrall 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
      vfrall |      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)
```

- "areg" automatically de-means the data
- this is especially useful when *n* is large
- the reported intercept is arbitrary

***Example*, ctd.** For $n = 48$, $T = 7$:

$$\widehat{FatalityRate} = -.66BeerTax + State\ fixed\ effects$$
$$(.20)$$

- Should you report the intercept?
- How many binary regressors would you include to estimate this using the "binary regressor" method?
- Compare slope, standard error to the estimate for the 1988 v. 1982 "changes" specification ($T = 2$, $n = 48$) (*note that this includes an intercept – return to this below*):

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

By the way… how much do beer taxes vary?

Beer Taxes in 2005

Source: Federation of Tax Administrators

http://www.taxadmin.org/fta/rate/beer.html

	EXCISE TAX RATES ($ per gallon)	SALES TAXES APPLIED	OTHER TAXES
Alabama	$0.53	Yes	$0.52/gallon local tax
Alaska	1.07	n.a.	$0.35/gallon small breweries
Arizona	0.16	Yes	
Arkansas	0.23	Yes	under 3.2% - $0.16/gallon; $0.008/gallon and 3% off- 10% on-premise tax
California	0.20	Yes	
Colorado	0.08	Yes	
Connecticut	0.19	Yes	
Delaware	0.16	n.a.	
Florida	0.48	Yes	2.67¢/12 ounces on-premise retail tax

Georgia	0.48	Yes	$0.53/gallon local tax
Hawaii	0.93	Yes	$0.54/gallon draft beer
Idaho	0.15	Yes	over 4% - $0.45/gallon
Illinois	0.185	Yes	$0.16/gallon in Chicago and $0.06/gallon in Cook County
Indiana	0.115	Yes	
Iowa	0.19	Yes	
Kansas	0.18	--	over 3.2% - {8% off- and 10% on-premise}, under 3.2% - 4.25% sales tax.
Kentucky	0.08	Yes*	9% wholesale tax
Louisiana	0.32	Yes	$0.048/gallon local tax
Maine	0.35	Yes	additional 5% on-premise tax
Maryland	0.09	Yes	$0.2333/gallon in Garrett County
Massachusetts	0.11	Yes*	0.57% on private club sales
Michigan	0.20	Yes	
Minnesota	0.15	--	under 3.2% - $0.077/gallon. 9% sales tax
Mississippi	0.43	Yes	
Missouri	0.06	Yes	
Montana	0.14	n.a.	
Nebraska	0.31	Yes	
Nevada	0.16	Yes	
New Hampshire	0.30	n.a.	
New Jersey	0.12	Yes	
New Mexico	0.41	Yes	

New York	0.11	Yes	$0.12/gallon in New York City
North Carolina	0.53	Yes	$0.48/gallon bulk beer
North Dakota	0.16	--	7% state sales tax, bulk beer $0.08/gal.
Ohio	0.18	Yes	
Oklahoma	0.40	Yes	under 3.2% - $0.36/gallon; 13.5% on-premise
Oregon	0.08	n.a.	
Pennsylvania	0.08	Yes	
Rhode Island	0.10	Yes	$0.04/case wholesale tax
South Carolina	0.77	Yes	
South Dakota	0.28	Yes	
Tennessee	0.14	Yes	17% wholesale tax
Texas	0.19	Yes	over 4% - $0.198/gallon, 14% on-premise and $0.05/drink on airline sales
Utah	0.41	Yes	over 3.2% - sold through state store
Vermont	0.265	no	6% to 8% alcohol - $0.55; 10% on-premise sales tax
Virginia	0.26	Yes	
Washington	0.261	Yes	
West Virginia	0.18	Yes	
Wisconsin	0.06	Yes	
Wyoming	0.02	Yes	
Dist. of Columbia	0.09	Yes	8% off- and 10% on-premise sales tax
U.S. Median	$0.188		

Regression with 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:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \beta_3 S_t + u_{it}$$

Time fixed effects only

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

In effect, the intercept varies from one year to the next:

$$Y_{i,1982} = \beta_0 + \beta_1 X_{i,1982} + \beta_3 S_{1982} + u_{i,1982}$$
$$= (\beta_0 + \beta_3 S_{1982}) + \beta_1 X_{i,1982} + u_{i,1982}$$

or

$$Y_{i,1982} = \mu_{1982} + \beta_1 X_{i,1982} + u_{i,1982}, \quad \mu_{1982} = \beta_0 + \beta_3 S_{1982}$$

Similarly,

$$Y_{i,1983} = \mu_{1983} + \beta_1 X_{i,1983} + u_{i,1983}, \quad \mu_{1983} = \beta_0 + \beta_3 S_{1983}$$

etc.

Two formulations for time fixed effects

1. "*T*-1 binary regressor" formulation:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \ldots\ \delta_T BT_t + u_{it}$$

where $B2_t = \begin{cases} 1 \text{ when } t\text{=2 (year \#2)} \\ 0 \text{ otherwise} \end{cases}$, etc.

2. "Time effects" formulation:

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

Time fixed effects: estimation methods

1. "T-1 binary regressor" OLS regression

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_{it} + \ldots \delta_T BT_{it} + u_{it}$$

- Create binary variables $B2,\ldots,BT$
- $B2 = 1$ if t = year #2, = 0 otherwise
- Regress Y on X, $B2,\ldots,BT$ using OLS
- Where's $B1$?

2. "Year-demeaned" OLS regression
   - Deviate Y_{it}, X_{it} from *year* (not state) averages
   - Estimate by OLS using "year-demeaned" data

These two methods can be combined…

```
. gen y83=(year==1983);
. gen y84=(year==1984);
. gen y85=(year==1985);
. gen y86=(year==1986);
. gen y87=(year==1987);
. gen y88=(year==1988);
. areg vfrall beertax y83 y84 y85 y86 y87 y88, absorb(state) r;

Regression with robust standard errors                 Number of obs =     336
                                                       F(  7,    281) =    3.70
                                                       Prob > F      =  0.0008
                                                       R-squared     =  0.9089
                                                       Adj R-squared =  0.8914
                                                       Root MSE      =  .18788
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .2547149    -2.51   0.013    -1.141371   -.1385884
         y83 |  -.0799029   .0502708    -1.59   0.113    -.1788579    .0190522
         y84 |  -.0724206   .0452466    -1.60   0.111     -.161486    .0166448
         y85 |  -.1239763   .0460017    -2.70   0.007     -.214528   -.0334246
         y86 |  -.0378645   .0486527    -0.78   0.437    -.1336344    .0579055
         y87 |  -.0509021   .0516113    -0.99   0.325    -.1524958    .0506917
         y88 |  -.0518038     .05387    -0.96   0.337    -.1578438    .0542361
       _cons |    2.42847   .1468565    16.54   0.000     2.139392    2.717549
-------------+----------------------------------------------------------------
       state |   absorbed                                      (48 categories)
```

(there are other ways to do this in STATA)

Combined entity and time fixed effects

- When $T = 2$, computing the first difference and including an intercept is equivalent (gives exactly the same regression) as the previous STATA command.
- So there are various equivalent ways to allow for both entity and time fixed effects:
  - differences & intercept ($T = 2$ only) – this is what we did initially
  - entity demeaning & $T - 1$ time indicators
  - time demeaning & $n - 1$ entity indicators
  - $T - 1$ time indicators & $n - 1$ entity indicators
  - entity & time demeaning

The Fixed Effects Regression Assumptions and Standard Errors for Fixed Effects Regression (SW Section 10.5 and App. 10.2)

Under assumptions that are basically extensions of the least squares assumptions, the OLS fixed effects estimator of β_1 is normally distributed. However, there are some subtleties associated with computing standard errors that don't come up with cross-sectional data.

Outline

A. The fixed effects regression assumptions

B. Standard errors for fixed effects regression – in two cases, one of which is new.

A. Extension of LS Assumptions to Panel Data

Consider a single X:

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

1. $E(u_{it}|X_{i1},\ldots,X_{iT},\alpha_i) = 0$.
2. $(X_{i1},\ldots,X_{iT}, Y_{i1},\ldots,Y_{iT})$, i =1,…,n, are i.i.d. draws from their joint distribution.
3. (X_{it}, Y_{it}) have finite fourth moments.
4. There is no perfect multicollinearity (multiple X's)
5. $\text{corr}(u_{it},u_{is}|X_{it},X_{is},\alpha_i) = 0$ for $t \neq s$.

Assumptions 3&4 are least squares LS assumptions 3&4

Assumptions 1&2 differ

Assumption 5 is new

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

- 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.
  - We'll return to this when we take up time series data.

Assumption #2: $(X_{i1},\ldots,X_{iT}, Y_{i1},\ldots,Y_{iT})$, $i = 1,\ldots,n$, are i.i.d. draws from their joint distribution.

- 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).

Assumption #5: $\text{corr}(u_{it}, u_{is} | X_{it}, X_{is}, \alpha_i) = 0$ for $t \neq s$

- We haven't seen this before.
- This says that (given X), the error terms are uncorrelated over time within a state.
- For example, $u_{CA,1982}$ and $u_{CA,1983}$ are uncorrelated
- Is this plausible? What enters the error term?
  - Especially snowy winter
  - Opening major new divided highway
  - Fluctuations in traffic density from local economic conditions
- Assumption #5 requires these omitted factors entering u_{it} to be uncorrelated over time, within a state.

Assumption #5 in a picture:

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

← Sampling is i.i.d. across entities →

(by Assumption #2)

Assumption #5: u's are uncorrelated over time, same entity

- Is this plausible?
- The u's consist of omitted factors – are they uncorrelated over time?

What if Assumption #5 fails: so corr(u_{it},u_{is}|X_{it},X_{is},α_i) $\neq 0$?

- A useful analogy is heteroskedasticity.
- OLS panel data estimators of β_1 are unbiased, consistent
- The OLS standard errors will be wrong – usually the OLS standard errors understate the true uncertainty
- Intuition: if u_{it} is correlated over time, you don't have as much information (as much random variation) as you would were u_{it} uncorrelated.
- This problem is solved by using "heteroskedasticity and autocorrelation-consistent standard errors"

B. Standard Errors

B.1 First get the large-*n* approximation to the sampling distribution of the FE estimator

Fixed effects regression model: $\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$

OLS fixed effects estimator: $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}$$

so: $$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}$$

Sampling distribution of fixed effects estimator, ctd.

Fact:

$$\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} = \sum_{t=1}^{T}\tilde{X}_{it}u_{it} - \left[\sum_{t=1}^{T}\left(X_{it}-\bar{X}_{i}\right)\right]\bar{u}_{i} = \sum_{t=1}^{T}\tilde{X}_{it}u_{it}$$

so

$$\sqrt{nT}\,(\hat{\beta}_1 - \beta_1) = \frac{\sqrt{\frac{1}{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{v}_{it}}{\hat{Q}_{\tilde{X}}^{2}} = \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i}{\hat{Q}_{\tilde{X}}^{2}}$$

where $\eta_{\mathrm{i}} = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}$, $\tilde{v}_{it} = \tilde{X}_{it}u_{it}$, and $\hat{Q}_{\tilde{X}}^{2} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{2}$.

By the CLT,

$$\sqrt{nT}\,(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_{\eta}^{2}/Q_{\tilde{X}}^{4})$$

where $\xrightarrow{d}$ means converges in distribution and $\hat{Q}_{\tilde{X}}^{2} \xrightarrow{p} Q_{\tilde{X}}^{2}$.

Sampling distribution of fixed effects estimator, ctd.

$$\sqrt{nT}\,(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_\eta^2/Q_{\tilde{X}}^4), \text{ where } \sigma_\eta^2 = \text{var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right)$$

B.2 Obtain Standard Error:

- Standard error of $\hat{\beta}_1$: $SE(\hat{\beta}_1) = \sqrt{\frac{1}{nT}\frac{\hat{\sigma}_\eta^2}{\hat{Q}_{\tilde{X}}^4}}$
- Only part we don't have: what is $\hat{\sigma}_\eta^2$?
  - Case I: u_{it}, u_{is} uncorrelated
  - Case II: u_{it}, u_{is} correlated

Case I: $\hat{\sigma}_B^2$ when u_{it}, u_{is} are uncorrelated

$$\sigma_\eta^2 = \text{var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right) = \text{var}\left(\frac{\tilde{v}_{i1}+\tilde{v}_{i2}+\ldots+\tilde{v}_{iT}}{\sqrt{T}}\right)$$

- Recall var(X + Y) = var(X) + var(Y) + 2cov(X,Y).
- When u_{it} and u_{is} are uncorrelated, cov$(\tilde{v}_{it},\tilde{v}_{is})$ = 0 unless t = s, so all the covariance terms are zero and

$$\sigma_\eta^2 = \frac{1}{T}\times T\,\text{var}\left(\tilde{v}_{it}\right) = \text{var}(\tilde{v}_{it})$$

- You can use the usual (hetero-robust) SE formula for standard errors if T isn't too small. This works because the usual hetero-robust formula is for uncorrelated errors – which is the case here.

Case II: u_{it} and u_{is} are correlated – so Assumption 5 fails

$$\sigma_\eta^2 = \operatorname{var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right)$$

$$= \operatorname{var}\left(\frac{\tilde{v}_{i1} + \tilde{v}_{i2} + ... + \tilde{v}_{iT}}{\sqrt{T}}\right)$$

$$\neq \operatorname{var}(\tilde{v}_{it})$$

- Recall $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)$
- If u_{it} and u_{is} are correlated, we have some nonzero covariances!! So in general we don't get any further simplifications.
- However, we can still compute standard errors – but using a different method: "clustered" standard errors.

Case II: Clustered Standard Errors

Variance:

$$\sigma_\eta^2 = \mathrm{var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right)$$

Variance estimator:

$$\hat{\sigma}^2_{\eta,clustered} = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\hat{\tilde{v}}_{it}\right)^2, \text{ where } \hat{\tilde{v}}_{it} = \tilde{X}_{it}\hat{u}_{it}.$$

Clustered standard error:

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{nT}\frac{\hat{\sigma}^2_{\eta,clustered}}{\hat{Q}^4_{\tilde{X}}}}$$

Comments on clustered standard errors:

- The clustered *SE* formula is NOT the usual (hetero-robust) SE formula!

- OK 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.

Comments on clustered standard errors, ctd.

- The Case II standard error formula goes under various names:
  - ***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.
  - ***Heteroskedasticity- and autocorrelation-consistent standard errors*** (autocorrelation is correlation with other time periods – u_{it} and u_{is} correlated)

Comments on clustered standard errors, ctd.

- Extensions:
  - The clusters can be other groupings, not necessarily time
  - For example, you can allow for correlation of u_{it} 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 u_{it} for i within same family, not between families).

Implementation in STATA

Case I: treat u_{it} and u_{is} as uncorrelated

```
. areg vfrall beertax, absorb(state) robust;

Linear regression, absorbing indicators                    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
      vfrall |      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)
```

Case II: treat u_{it} and u_{is} as possibly correlated

```
. areg vfrall beertax, absorb(state) robust cluster(state);

Linear regression, absorbing indicators                 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

                                  (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      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)
```

Coefficients are identical (*why*?)

Pretty big difference in the standard errors!

Try adding year effects:

```
. areg vfrall beertax y83 y84 y85 y86 y87 y88, absorb(state) r;

Regression with robust standard errors                Number of obs =      336
                                                      F(  7,   281) =     3.70
                                                      Prob > F      =   0.0008
                                                      R-squared     =   0.9089
                                                      Adj R-squared =   0.8914
                                                      Root MSE      =   .18788
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .2547149    -2.51   0.013    -1.141371   -.1385884
         y83 |  -.0799029   .0502708    -1.59   0.113    -.1788579    .0190522
         y84 |  -.0724206   .0452466    -1.60   0.111     -.161486    .0166448
         y85 |  -.1239763   .0460017    -2.70   0.007     -.214528   -.0334246
         y86 |  -.0378645   .0486527    -0.78   0.437    -.1336344    .0579055
         y87 |  -.0509021   .0516113    -0.99   0.325    -.1524958    .0506917
         y88 |  -.0518038     .05387    -0.96   0.337    -.1578438    .0542361
       _cons |    2.42847   .1468565    16.54   0.000     2.139392    2.717549
-------------+----------------------------------------------------------------
       state |   absorbed                                     (48 categories)

.   test $yeardum;

       F(  6,   281) =    2.47
            Prob > F =    0.0243
```

```
.  areg vfrall beertax $yeardum, absorb(state) r cluster(state);

Linear regression, absorbing indicators                Number of obs =     336
                                                       F(  7,    47) =    3.74
                                                       Prob > F      =  0.0027
                                                       R-squared     =  0.9089
                                                       Adj R-squared =  0.8914
                                                       Root MSE      =  .18788

                                  (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3857867    -1.66   0.104    -1.416083    .1361229
         y83 |  -.0799029   .0379069    -2.11   0.040    -.1561617   -.003644
         y84 |  -.0724206   .0474088    -1.53   0.133    -.1677948    .0229537
         y85 |  -.1239763   .0497587    -2.49   0.016    -.2240779   -.0238747
         y86 |  -.0378645   .0616479    -0.61   0.542    -.1618841    .0861552
         y87 |  -.0509021   .0687224    -0.74   0.463    -.1891536    .0873495
         y88 |  -.0518038   .0695801    -0.74   0.460    -.1917809    .0881732
       _cons |    2.42847   .2179038    11.14   0.000     1.990104    2.866836
-------------+----------------------------------------------------------------
       state |   absorbed                                     (48 categories)

   .   test $yeardum;

        F(  6,    47) =    3.61
             Prob > F =    0.0050
```

Fixed Effects Regression Results

Dependent variable: *Fatality rate*

	(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

$^{++}$ Significant at the 10% level using normal but not Student *t* critical values

This is a hard call – what would you conclude?

Summary: *SE*s for Panel Data in a picture:

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

← *i.i.d. sampling across entities* →

- Intuition #1: This is similar to heteroskedasticity – you make an assumption about the error, derive *SEs* under that assumption, if the assumption is wrong, so are the *SEs*
- Intuition #2: If u_{21} and u_{22} are correlated, there is less information in the sample than if they are not – and *SEs* need to account for this (usual *SEs* are typically too small)
- Hetero-robust (or homosk-only) *SE*s don't allow for this correlation, but clustered *SE*s do.

Application: Drunk Driving Laws and Traffic Deaths
(SW Section 10.6)

Some facts

- Approx. 40,000 traffic fatalities annually in the U.S.
- 1/3 of traffic fatalities involve a drinking driver
- 25% of drivers on the road between 1am and 3am have been drinking (estimate)
- A drunk driver is 13 times as likely to cause a fatal crash as a non-drinking driver (estimate)

Drunk driving laws and traffic deaths, ctd.

Public policy issues

- Drunk driving causes massive externalities (sober drivers are killed, society bears medical costs, etc. etc.) – there is ample justification for governmental intervention
- Are there any effective ways to reduce drunk driving? If so, what?
- What are effects of specific laws:
  - mandatory punishment
  - minimum legal drinking age
  - economic interventions (alcohol taxes)

The Commonwealth of Massachusetts
Executive Department
State House • Boston, MA 02133
(617) 725-4000

MITT ROMNEY
GOVERNOR

KERRY HEALEY
LIEUTENANT GOVERNOR

FOR IMMEDIATE RELEASE:
October 28, 2005

CONTACT:
Julie Teer
Laura Nicoll
(617) 725-4025

ROMNEY CELEBRATES THE PASSAGE OF MELANIE'S BILL

Legislation puts Massachusetts in line with federal standards for drunk driving

Governor Mitt Romney today signed into law the toughest drunk driving legislation in the Commonwealth's history.

Named in honor of 13-year-old Melanie Powell, the new law will stiffen penalties for drunk driving offenses in Massachusetts and close loopholes in the legal system that allow repeat drunk drivers to get back behind the wheel.

“Today we honor those who have lost their lives in senseless drunk driving tragedies and act to save the lives we could otherwise lose next year,” said Romney. “We have Melanie’s Law today because the citizens of the Commonwealth cared enough to make it happen.”

The new measure gives prosecutors the power to introduce certified court documents to prove that a repeat offender has been previously convicted of drunk driving. In addition, the mandatory minimum jail sentence for any individual found guilty of manslaughter by motor vehicle will be increased from 2 ½ to five years.

Repeat offenders will be required to install an interlock device on any vehicle they own or operate. These devices measure the driver’s Blood Alcohol Content (BAC) and prevent the car from starting if the driver is intoxicated. Any individual who tampers with the interlock device could face a jail sentence.

For the first time, Massachusetts will be in compliance with federal standards for drunk driving laws.

Romney was joined by Tod and Nancy Powell, the parents of Melanie Powell, and her grandfather, Ron Bersani to celebrate the passage of the new drunk driving measure.

“Today we should give thanks to all of those who have worked so hard to make this day possible," said Bersani. "Governor Romney and the Legislative leadership have advanced the fight against repeat drunk driving to heights that seemed unattainable just six months ago.

Under the law, stiff penalties will be established for individuals who drive while drunk with a child under the age of 14 in the vehicle and those who drive with a BAC of .20 or higher, more than twice the legal limit.

Romney thanked the Legislature for enacting a tough bill that cracks down on repeat drunk driving offenders in Massachusetts.

"Public safety is one of our top priorities and Melanie's Law will go a long way towards making our citizens and roadways safer," said Speaker Salvatore F. DiMasi. "I commend the my colleagues in the Legislature and the Governor for taking comprehensive and quick action on this very important issue."

"Today we are sending a powerful message that Massachusetts is serious about keeping repeat drunken drivers off the road," said House Minority Leader Bradley H. Jones Jr. "I am proud of the Governor, Lieutenant Governor, and my legislative colleagues for joining together to pass tough laws to make our roadways safer."

"I am pleased and proud that the Legislature did the right thing in the end and supported a Bill worthy of Melanie's name and the sacrifices made by the Powell family and all victims of drunk drivers," said Senator Robert L. Hedlund. "Melanie's Law will save lives and it would not have been accomplished if not for the tireless efforts and advocacy of the families."

Representative Frank Hynes added, "I'd like to commend Ron, Tod, and Nancy for their tireless work in support of Melanie's bill. As a family, they were able to turn the horrific tragedy in their lives into a greater measure of safety for all families on Massachusetts roadways."

###

The drunk driving panel data set

n = 48 U.S. states, T = 7 years (1982,…,1988) (balanced)

Variables

- Traffic fatality rate (deaths per 10,000 residents)
- Tax on a case of beer (*Beertax*)
- Minimum legal drinking age
- Minimum sentencing laws for first DWI violation:
  - *Mandatory Jail*
  - *Manditory Community Service*
  - otherwise, sentence will just be a monetary fine
- Vehicle miles per driver (US DOT)
- State economic data (real per capita income, etc.)

Why might panel data help?

- Potential OV bias from variables that vary across states but are constant over time:
  - culture of drinking and driving
  - quality of roads
  - vintage of autos on the road

    $\Rightarrow$ use state fixed effects
- Potential OV bias from variables that vary over time but are constant across states:
  - improvements in auto safety over time
  - changing national attitudes towards drunk driving

    $\Rightarrow$ use time fixed effects

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

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

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66** (0.20)	−0.64* (0.25)	−0.45* (0.22)	−0.70** (0.25)	−0.46* (0.22)	−0.45 (0.32)
Drinking age 18				0.028 (0.066)	−0.011 (0.064)		0.028 (0.076)
Drinking age 19				−0.019 (0.040)	−0.078 (0.049)		−0.019 (0.054)
Drinking age 20				0.031 (0.046)	−0.102* (0.046)		0.031 (0.055)
Drinking age						−0.002 (0.017)	
Mandatory jail?				0.013 (0.032)	−0.026 (0.065)		0.013 (0.018)
Mandatory community service?				0.033 (0.115)	0.147 (0.137)		0.033 (0.144)
Mandatory jail or community service?						0.039 (0.084)	
Average vehicle miles per driver				0.008 (0.008)	0.017 (0.010)	0.009 (0.008)	0.008 (0.007)
Unemployment rate				−0.063** (0.012)		−0.063** (0.012)	−0.063** (0.014)
Real income per capita (logarithm)				1.81** (0.47)		1.79** (0.45)	1.81* (0.69)
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	no	no	no	no	no	yes

F-statistics and p-values testing exclusion of groups of variables:

Time effects = 0			2.47 (0.024)	11.44 (< 0.001)	2.28 (0.037)	11.62 (< 0.001)	8.64 (< 0.001)
Drinking age coefficients = 0				0.48 (0.696)	2.09 (0.102)		0.30 (0.825)
Jail, community service coefficients = 0				0.17 (0.845)	0.59 (0.557)		0.28 (0.758)
Unemployment rate, income per capita = 0				38.29 (< 0.001)		40.15 (< 0.001)	25.88 (< 0.001)
$\overline{R}^2$	0.090	0.889	0.891	0.926	0.893	0.926	0.926

These regressions were estimated using panel data for 48 U.S. states from 1982 to 1988 (336 observations total), described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the *5% level or **1% significance level.

Empirical Analysis: Main Results

- Sign of beer tax coefficient changes when fixed state effects are included
- Fixed time effects are statistically significant but do not have big impact on the estimated coefficients
- Estimated effect of beer tax drops when other laws are included as regressor
- The only policy variable that seems to have an impact is the tax on beer – not minimum drinking age, not mandatory sentencing, etc. – *however the beer tax is not significant even at the 10% level using clustered SEs*.
- The other economic variables have plausibly large coefficients: more income, more driving, more deaths

Digression: extensions of the "*n*-1 binary regressor" idea

The idea of using many binary indicators to eliminate omitted variable bias can be extended to non-panel data – the key is that the omitted variable is constant for a group of observations, so that in effect it means that each group has its own intercept.

Example: Class size problem.

Suppose funding and curricular issues are determined at the county level, and each county has several districts. Resulting omitted variable bias could be addressed by including binary indicators, one for each county (omit one to avoid perfect multicollinearity).

Summary: Regression with Panel Data (SW Section 10.7)

Advantages and limitations of fixed effects regression

Advantages

- You can control for unobserved variables that:
  - vary across states but not over time, and/or
  - vary over time but not across states
- More observations give you more information
- Estimation involves relatively straightforward extensions of multiple regression

- Fixed effects regression can be done three ways:
  1. "Changes" method when $T = 2$
  2. "n-1 binary regressors" method when n is small
  3. "Entity-demeaned" regression
- Similar methods apply to regression with time fixed effects and to both time and state fixed effects
- Statistical inference: like multiple regression.

Limitations/challenges

- Need variation in X over time within states
- Time lag effects can be important
- You should use heteroskedasticity- and autocorrelation-consistent (clustered) standard errors if you think u_{it} could be correlated over time