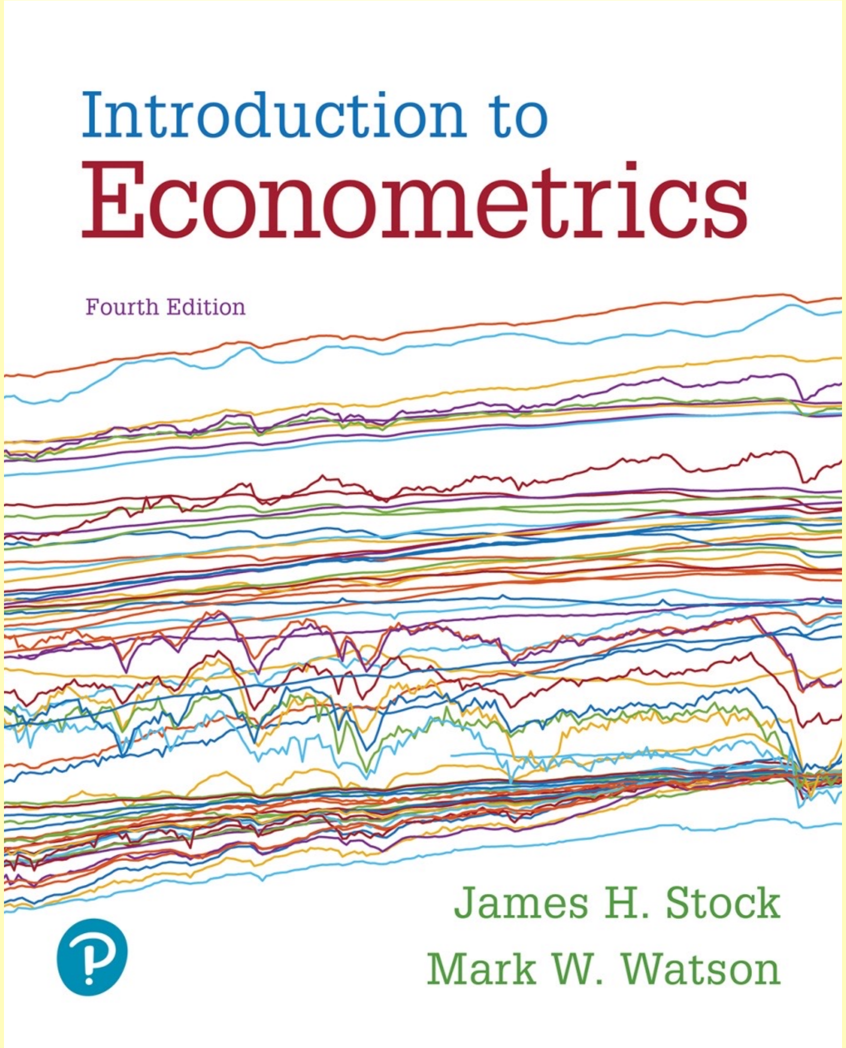
Introduction to Econometrics

Fourth Edition



Chapter 10

Regression with Panel Data

Outline

1. Panel Data: What and Why
2. Panel Data with Two Time Periods
3. Fixed Effects Regression
4. Regression with Time Fixed Effects
5. Standard Errors for Fixed Effects Regression
6. Application to Drunk Driving and Traffic Safety

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, (10.11)$$

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it}$$
$$+\gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it} (10.13)$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \delta_3 B3_t + \cdots + \delta_T BT_t + u_{it} \quad (10.18)$$

This model can equivalently be represented using n - 1 entity binary indicators and T - 1 time binary indicators, along with an intercept:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i +$$
$$\delta_2 B2_t + \delta_3 B3_t + \cdots + \delta_t BT_t + u_{it}, (10.20)$$

where β_0, β_1, γ_2, …, γ_n, δ_2,…, δ_t are unknown coefficients.

Panel Data: What and Why (SW Section 10.1)

A ***panel dataset*** contains observations on multiple entities (individuals, states, companies…), 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 (1 of 2)

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:

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

Notation for panel data (2 of 2)

Panel data with k regressors:

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

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, that is, all variables are observed for all entities (states) and all time periods (years)

Why are panel data useful?

With panel data we can control for factors that:

- Vary across entities 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

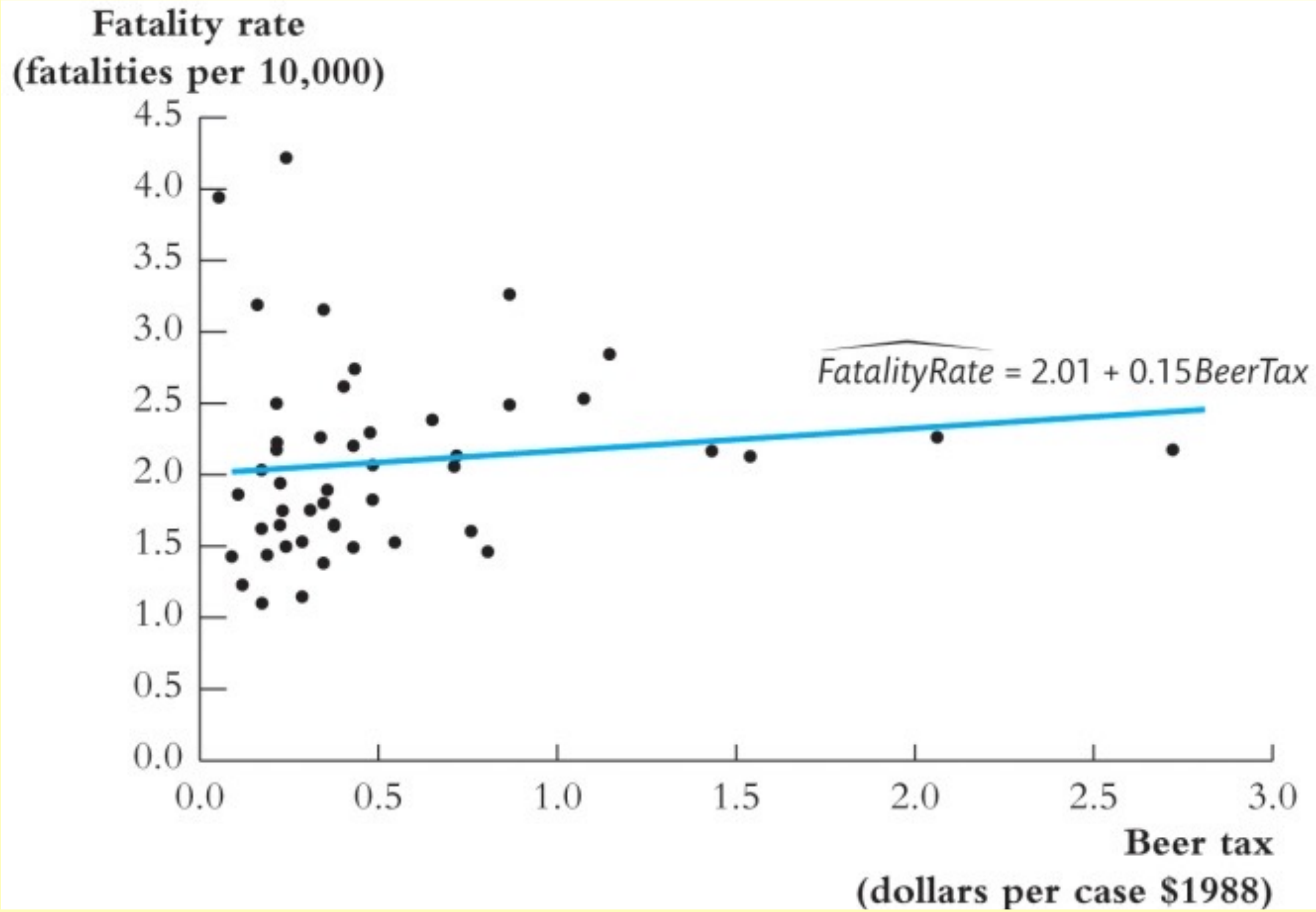
Observational unit: a year in a U.S. state

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

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?

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 (1 of 2)

Example #1: traffic density. Suppose:

I. High traffic density means more traffic deaths

II. (Western) states with lower traffic density have lower alcohol tax rate

- 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.

These omitted factors could cause omitted variable bias (2 of 2)

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.

- Then the two conditions for omitted variable bias are satisfied. Specifically, "high taxes" could pick up the effect of "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 (SW Section 10.2) (1 of 3)

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.

Panel Data with Two Time Periods (SW Section 10.2) (2 of 3)

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…

Panel Data with Two Time Periods (SW Section 10.2) (3 of 3)

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

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*
- This differences regression doesn't have an intercept – it was eliminated by the subtraction step

Example: Traffic deaths and beer taxes

1982 data:

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

1988 data:

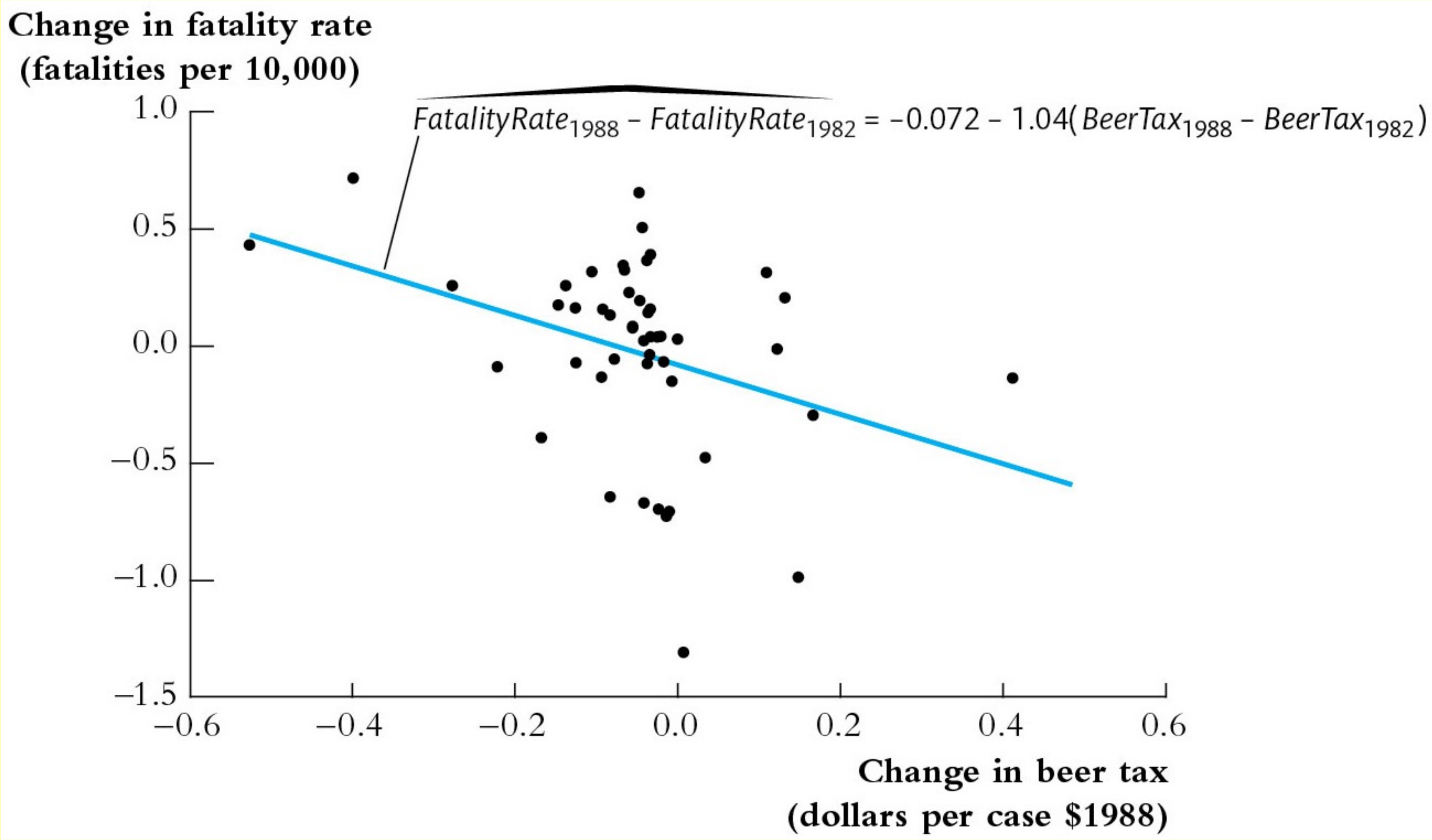
$$\widehat{FatalityRate} = \underset{(.11)}{1.86} + \underset{(.13)}{0.44}BeerTax \qquad (n = 48)$$

Difference regression (n = 48)

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

An intercept is included in this differences regression allows for the mean change in FR to be nonzero – more on this later…

Δ*FatalityRate* v. Δ*BeerTax*



Note that the intercept is nearly zero…

Fixed Effects Regression (SW Section 10.3) (1 of 3)

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, and Massachusetts.

Fixed Effects Regression (SW Section 10.3) (2 of 3)

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

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

$$\begin{aligned} Y_{CA,t} &= \beta_0 + \beta_1 X_{CA,t} + \beta_2 Z_{CA} + u_{CA,t} \\ &= (\beta_0 + \beta_2 Z_{CA}) + \beta_1 X_{CA,t} + u_{CA,t} \end{aligned}$$

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.

Fixed Effects Regression (SW Section 10.3) (3 of 3)

For TX:

$$
\begin{aligned}
Y_{TX,t} &= \beta_0 + \beta_1 X_{TX,t} + \beta_2 Z_{TX} + u_{TX,t} \\
&= (\beta_0 + \beta_2 Z_{TX}) + \beta_1 X_{TX,t} + u_{TX,t}
\end{aligned}
$$

or

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

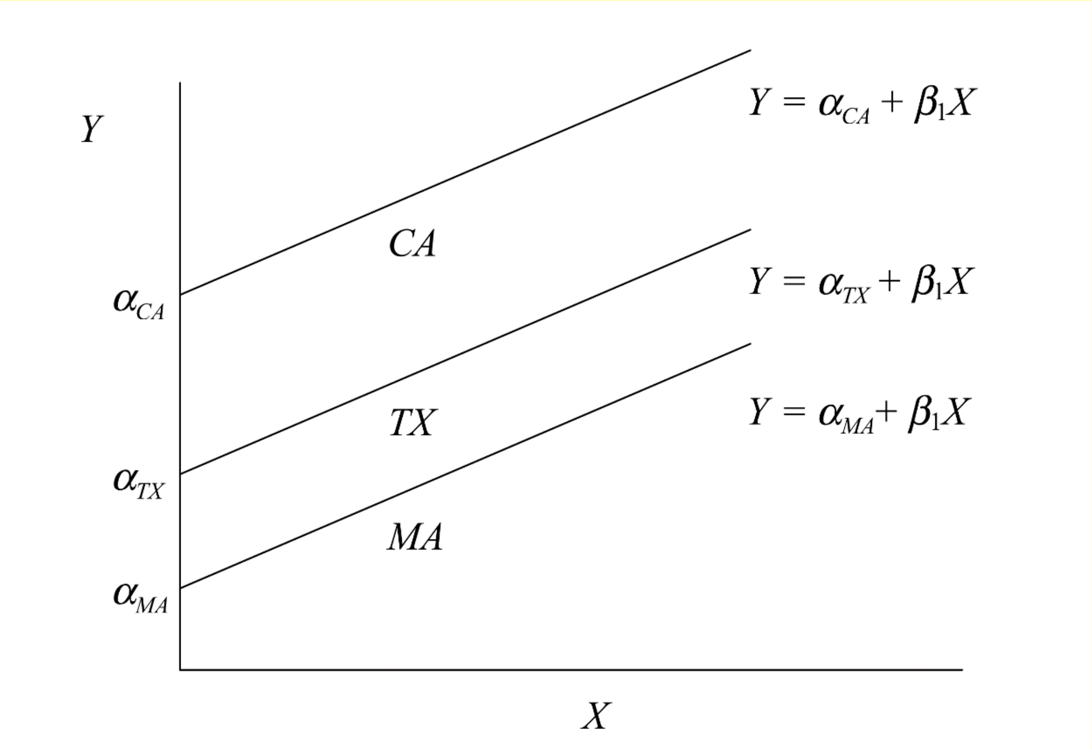
Collecting the lines for all three states:

$$
\begin{aligned}
Y_{CA,t} &= \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t} \\
Y_{TX,t} &= \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t} \\
Y_{MA,t} &= \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}
\end{aligned}
$$

or

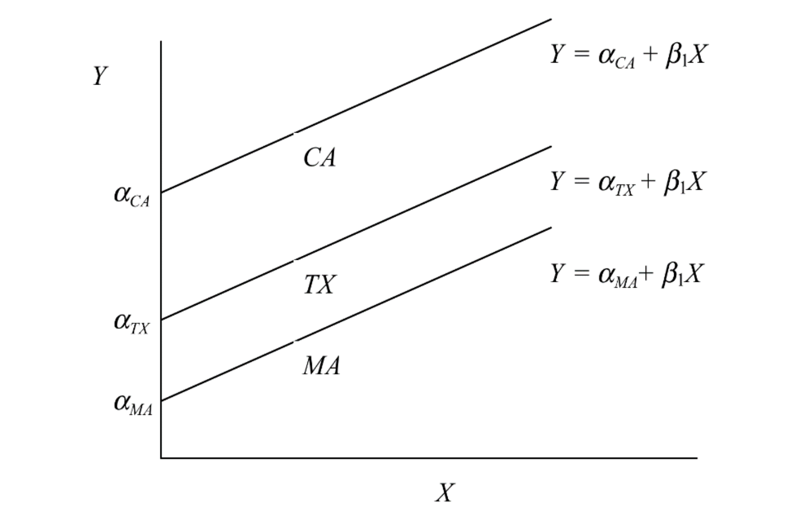
$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it},\ i = \text{CA, TX, MA},\ t = 1,\ldots,T$$

The regression lines for each state in a picture (1 of 2)



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

The regression lines for each state in a picture (2 of 2)



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

1. **"*n*-1 binary regressor" form**

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \ldots + \gamma_n Dn_i + u_{it}$$

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

2. **"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 (within Estimator)

Let's rewrite our unobserved effects model so that this is still firmly in our minds:

$$Y_{it} = \delta D_{it} + u_i + \varepsilon_{it}; t = 1,2, \dots, T$$

If we have data on multiple time periods, we can think of u_i as fixed effects to be estimated. OLS estimation with fixed effects yields

$$(\hat{\delta}, \hat{u}_1, \dots, \hat{u}_N) = argmin_{b,m_1,m_N} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - D_{it}b - m_i)^2$$

which amounts to including N individual dummies in regression of Y_{it} on D_{it}.

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

$$\sum_{i=1}^{N} \sum_{t=1}^{T} D'_{it}\left(Y_{it} - D_{it}\hat{\delta} - \hat{u}_i\right) = 0$$

$\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}'\left(Y_{it} - D_{it}\hat{\delta} - \hat{u}_i\right) = 0$ and $\sum_{t=1}^{l}\left(Y_{it} - D_{it}\hat{\delta} - \hat{u}_i\right)$, for i=1,..,N.

Therefore, for i=1,..,N, $\hat{u}_i = \frac{1}{T}\sum_{t=1}^{T}\left(Y_{it} - D_{it}\hat{\delta}\right) = \bar{Y}_i - \bar{D}_i\hat{\delta}$

where $\bar{D}_i = \frac{1}{T}\sum_{t=1}^{T} D_{it}$; $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$

Plug this result into the first FOC to obtain:

$$\hat{\delta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{it} - \bar{D}_i)'(D_{it} - \bar{D}_i)\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{it} - \bar{D}_i)'(Y_{it} - \bar{Y}_i)\right)$$

$$\hat{\delta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{D}_{it}'\ddot{D}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{D}_{it}'\ddot{Y}_{it}\right)$$

with time-demeaned variable $\ddot{D}_{it} = D_{it} - \bar{D}_i$, $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i$.

In case it isn't clear, though, running a regression with the time demeaned variables $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i$ and $\ddot{D}_{it} = D_{it} - \bar{D}_i$ is **numerically equivalent to** a regression of Y_{it} on D_{it} and unit-specific dummy variables.

Hence the reason this is sometimes called the "within" estimator, and sometimes called the "fixed effects" estimator. And when year fixed effects are included, the "two-way fixed effects" estimator. They are the same thing.

Even better, the regression with the time-demeaned variables is consistent for δ even when $C[D_{it}, u_i] = 0$ because time-demeaning eliminates the unobserved effects. Let's see this now:

$$Y_{it} = \delta D_{it} + u_i + \varepsilon_{it}$$

$$\bar{Y}_i = \delta \bar{D}_i + u_i + \bar{\varepsilon}_i$$

$$(Y_{it} - \bar{Y}_i) = \delta(D_{it} - \bar{D}_i) + (u_i - u_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$\ddot{Y}_{it} = \delta \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

Where'd the unobserved heterogeneity go?! It was deleted when we time-demeaned the data.

And as we said, including individual fixed effects does this time demeaning automatically so that you don't have to go to the actual trouble of doing it yourself manually.

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

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

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

- 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 (1 of 3)

The fixed effects regression model:

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

The entity 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 entity averages:

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

2. "Entity-demeaned" OLS regression (2 of 3)

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

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

- $\tilde{X}_{it}$ and $\tilde{Y}_{it}$ are "entity-demeaned" data
- 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.

2. "Entity-demeaned" OLS regression (3 of 3)

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

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

- First construct the entity-demeaned variables $\tilde{Y}_{it}$ and $\tilde{X}_{it}$
- Then estimate (2) by regressing $\tilde{Y}_{it}$ on $\tilde{X}_{it}$ using OLS
- This is like the "changes" approach, but instead Y_{it} is deviated from the state average instead of Y_{i1}.
- Standard errors need to be computed in a way that accounts for the panel nature of the data set (more later)
- This can be done in a single command in STATA

Before making the assumptions necessary for estimation, let's perform some useful algebra on (1). Whatever the properties of ν_i and ϵ_{it}, if (1) is true, it must also be true that

$$\overline{y}_i = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta} + \nu_i + \overline{\epsilon}_i \tag{2}$$

where $\overline{y}_i = \sum_t y_{it}/T_i$, $\overline{\mathbf{x}}_i = \sum_t \mathbf{x}_{it}/T_i$, and $\overline{\epsilon}_i = \sum_t \epsilon_{it}/T_i$. Subtracting (2) from (1), it must be equally true that

$$(y_{it} - \overline{y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + (\epsilon_{it} - \overline{\epsilon}_i) \tag{3}$$

These three equations provide the basis for estimating $\boldsymbol{\beta}$. In particular, `xtreg, fe` provides what is known as the fixed-effects estimator—also known as the within estimator—and amounts to using OLS to perform the estimation of (3). `xtreg, be` provides what is known as the between estimator and amounts to using OLS to perform the estimation of (2). `xtreg, re` provides the random-effects estimator and is a (matrix) weighted average of the estimates produced by the between and within estimators. In particular, the random-effects estimator turns out to be equivalent to estimation of

$$(y_{it} - \theta\overline{y}_i) = (1-\theta)\alpha + (\mathbf{x}_{it} - \theta\overline{\mathbf{x}}_i)\boldsymbol{\beta} + \{(1-\theta)\nu_i + (\epsilon_{it} - \theta\overline{\epsilon}_i)\} \tag{4}$$

where θ is a function of σ^2_ν and σ^2_ϵ. If $\sigma^2_\nu = 0$, meaning that ν_i is always 0, $\theta = 0$ and (1) can be estimated by OLS directly. Alternatively, if $\sigma^2_\epsilon = 0$, meaning that ϵ_{it} is 0, $\theta = 1$ and the within estimator returns all the information available (which will, in fact, be a regression with an R^2 of 1).

Example: Traffic deaths and beer taxes in STATA (1 of 3)

First let STATA know you are working with panel data by defining the entity variable (state) and time variable (year):

```
. xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 1982 to 1988
                delta:  1 unit
```

Example: Traffic deaths and beer taxes in STATA (2 of 3)

```
. xtreg vfrall beertax, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0407                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0934                                        max =         7
                                                F(1,47)            =      5.05
corr(u_i, Xb)  = -0.6885                        Prob > F           =    0.0294

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2918556    -2.25   0.029    -1.243011   -.0687358
       _cons |   2.377075   .1497966    15.87   0.000     2.075723    2.678427
------------------------------------------------------------------------------
```

- The panel data command xtreg with the option fe performs fixed effects regression. The reported intercept is arbitrary, and the estimated individual effects are not reported in the default output.
- The fe option means use fixed effects regression
- The vce(cluster state) option tells STATA to use clustered standard errors – more on this later

Example: Traffic deaths and beer taxes in STATA (3 of 3)

For n = 48, T = 7:

$$\widehat{FatalityRate} = -.66BeerTax + State\ fixed\ effects$$
$$(.29)$$

- 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

	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

By the way… how much do beer taxes vary?
(2 of 5)

	EXCISE TAX RATES ($ per gallon)	SALES TAXES APPLIED	OTHER TAXES
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

By the way… how much do beer taxes vary?

(3 of 5)

	EXCISE TAX RATES ($ per gallon)	SALES TAXES APPLIED	OTHER TAXES
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% alcohol – $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	

By the way… how much do beer taxes vary?

(4 of 5)

	EXCISE TAX RATES ($ per gallon)	SALES TAXES APPLIED	OTHER TAXES
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

By the way… how much do beer taxes vary? (5 of 5)

	EXCISE TAX RATES ($ per gallon)	SALES TAXES APPLIED	OTHER TAXES
Utah	0.41	Yes	over 3.2% alcohol – 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 (SW Section 10.4)

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 S_t denote the combined effect of variables which changes over time but not states ("safer cars").
- 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

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

This model can be recast as having an intercept that varies from one year to the next:

$$\begin{aligned} Y_{i,1982} &= \beta_0 + \beta_1 X_{i,1982} + \beta_3 S_{1982} + u_{i,1982} \\ &= (\beta_0 + \beta_3 S_{1982}) + \beta_1 X_{i,1982} + u_{i,1982} \\ &= \lambda_{1982} + \beta_1 X_{i,1982} + u_{i,1982,} \end{aligned}$$

where $\lambda_{1982} = \beta_0 + \beta_3 S_{1982}$ Similarly,

$$Y_{i,1983} = \lambda_{1983} + \beta_1 X_{i,1983} + u_{i,1983,}$$

where $\lambda_{1983} = \beta_0 + \beta_3 S_{1983,}$ etc.

Two formulations of regression with time fixed effects

1. "$T-1$ binary regressor" version:

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

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

2. "Time effects" version:

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

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,…,BT*
   - *B2* = 1 if *t* = year #2, = 0 otherwise
   - Regress *Y* on *X*, *B2,…,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

Estimation with both entity and time fixed effects

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

- When $T = 2$, computing the first difference and including an intercept is equivalent to (gives exactly the same regression as) including entity and time fixed effects.
- When $T > 2$, there are various equivalent ways to incorporate both entity and time fixed effects:
  - entity demeaning & $T - 1$ time indicators (this is done in the following STATA example)
  - time demeaning & $n - 1$ entity indicators
  - $T - 1$ time indicators & $n - 1$ entity indicators
  - entity & time demeaning

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, (10.11)$$

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it}$$
$$+\gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it} (10.13)$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \delta_3 B3_t + \cdots + \delta_T BT_t + u_{it} \quad (10.18)$$

This model can equivalently be represented using n - 1 entity binary indicators and T - 1 time binary indicators, along with an intercept:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + \delta_2 B2_t + \delta_3 B3_t + \cdots + \delta_t BT_t + u_{it}, (10.20)$$

where β_0, β_1, γ_2, …, γ_n, δ_2,…, δ_t are unknown coefficients.

```
. gen y83=(year==1983)                    First generate all the time binary variables
. gen y84=(year==1984)
. gen y85=(year==1985)
. gen y86=(year==1986)
. gen y87=(year==1987)
. gen y88=(year==1988)
. global yeardum "y83 y84 y85 y86 y87 y88"
. xtreg vfrall beertax $yeardum, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0803                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0876                                        max =         7
corr(u_i, Xb)  = -0.6781                        Prob > F           =    0.0009
                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3570783    -1.79   0.080    -1.358329    .0783691
         y83 |  -.0799029   .0350861    -2.28   0.027    -.1504869   -.0093188
         y84 |  -.0724206   .0438809    -1.65   0.106    -.1606975    .0158564
         y85 |  -.1239763   .0460559    -2.69   0.010    -.2166288   -.0313238
         y86 |  -.0378645   .0570604    -0.66   0.510    -.1526552    .0769262
         y87 |  -.0509021   .0636084    -0.80   0.428    -.1788656    .0770615
         y88 |  -.0518038   .0644023    -0.80   0.425    -.1813645    .0777568
       _cons |    2.42847   .2016885    12.04   0.000     2.022725    2.834215
-------------+----------------------------------------------------------------
```

Are the time effects jointly statistically significant?

```
.   test $yeardum

 ( 1)  y83 = 0
 ( 2)  y84 = 0
 ( 3)  y85 = 0
 ( 4)  y86 = 0
 ( 5)  y87 = 0
 ( 6)  y88 = 0

       F(  6,    47) =    4.22
            Prob > F =    0.0018
```

Yes

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

Under a panel data version of the least squares assumptions, the OLS fixed effects estimator of β_1 is normally distributed.

However, a new standard error formula needs to be introduced: the **"clustered" standard error** formula.

This new formula is needed because <u>observations for the same entity are not independent</u> (it's the same entity!), even though observations across entities are independent if entities are drawn by simple random sampling.

Here we consider the case of entity fixed effects. Time fixed effects can simply be included as additional binary regressors.

The Fixed Effects Regression Assumptions

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. $(X_{i1},\ldots, X_{iT}, u_{i1},\ldots, u_{iT}),\ i = 1,\ldots, n$, are i.i.d. draws from their joint distribution.
3. Large outliers are unlikely: (X_{it}, u_{it}) have finite fourth moments.
4. There is no perfect multicollinearity (multiple *X*'s)

Assumptions 3&4 are least squares assumptions 3&4

Assumptions 1&2 differ

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

- u_{it} has mean zero, given the entity fixed effect *and* the entire history of the X's for that entity
- 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 no 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.
  - Sometimes this "no feedback" assumption is plausible, sometimes it isn't. We'll return to it when we take up time series data.

Assumption #2: $(X_{i1},\ldots,X_{iT},u_{i1},\ldots,u_{iT})$, $i = 1,\ldots,n$, 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 are randomly sampled from their population by simple random sampling.
- 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 high beer tax this year is a good predictor of (highly correlated with) whether it will have a high beer tax next year. Similarly, the error term for an entity in one year is plausibly correlated with its value in the next year, that is, $\text{corr}(u_{it}, u_{it+1})$ is often plausibly nonzero.

Autocorrelation (serial correlation)

Suppose a variable Z is observed at different dates t, so observations are on Z_t, $t = 1,\ldots, T$. (Think of there being only one entity.) Then Z_t is said to be ***autocorrelated*** or ***serially correlated*** if $\text{corr}(Z_t, Z_{t+j}) \neq 0$ for some dates $j \neq 0$.

- "Autocorrelation" means correlation with itself.
- $\text{cov}(Z_t, Z_{t+j})$ is called the ***j^{th} autocovariance*** of Z_t.
- In the drunk driving example, u_{it} includes the omitted variable of annual weather conditions for state i. If snowy winters come in clusters (one follows another) then u_{it} will be autocorrelated (*why*?)
- In many panel data applications, u_{it} is plausibly autocorrelated.

Independence and autocorrelation in panel data in a picture

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

$\leftarrow$ Sampling is i.i.d. across entities $\rightarrow$

- If entities are sampled by simple random sampling, then $(u_{i1},\ldots, u_{iT})$ is independent of $(u_{j1},\ldots, u_{jT})$ for different entities $i \neq j$.
- But if the omitted factors comprising u_{it} are serially correlated, then u_{it} is serially correlated.

Under the LS assumptions for panel data

- The OLS fixed effect estimator $\hat{\beta}_1$ is unbiased, consistent, and asymptotically normally distributed
- However, the usual OLS standard errors (both homoskedasticity-only and heteroskedasticity-robust) will in general be wrong because they assume that u_{it} is serially uncorrelated.
  - In practice, the OLS standard errors often understate the true sampling uncertainty: if u_{it} is correlated over time, you don't have as much information (as much random variation) as you would if u_{it} were uncorrelated.
  - This problem is solved by using "**clustered**" standard errors.

Clustered Standard Errors

- Clustered standard errors estimate the variance of $\hat{\beta}_1$ when the variables (X's and u's) are i.i.d. across entities but are potentially autocorrelated within an entity.
- Clustered SEs are easiest to understand if we first consider the simpler problem of estimating the mean of Y using panel data…

Clustered SEs for the mean estimated using panel data (1 of 2)

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

The estimator of μ mean is $\bar{Y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it}$.

It is useful to write $\bar{Y}$ as the average across entities of the mean value for each entity:

$$\bar{Y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} Y_{it}\right) = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i,$$

where $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ is the sample mean for entity i.

Clustered SEs for the mean estimated using panel data (2 of 2)

Because observations are i.i.d. across entities, $(\bar{Y}_1, \ldots, \bar{Y}_n)$ are i.i.d. Thus, if n is large, the CLT applies and

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i \xrightarrow{d} N(0, \sigma_{\bar{Y}_i}^2/n), \text{ where } \sigma_{\bar{Y}_i}^2 = \text{var}(\bar{Y}_i).$$

- The SE of $\bar{Y}$ is the square root of an estimator of $\sigma_{\bar{Y}_i}^2/n$.
- The natural estimator of $\sigma_{\bar{Y}_i}^2$ is the sample variance of $\bar{Y}_i$, $s_{\bar{Y}_i}^2$. This delivers the clustered standard error formula for $\bar{Y}$ computed using panel data:

$$Clustered\ SE \text{ of } \bar{Y} = \sqrt{\frac{s_{\bar{Y}_i}^2}{n}}, \text{ where } s_{\bar{Y}_i}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \bar{Y})^2$$

What's special about clustered SEs?

- Not much, really – the previous derivation is the same as was used in Ch. 3 to derive the SE of the sample average, except that here the "data" are the i.i.d. entity averages $(\bar{Y}_1, \ldots, \bar{Y}_n)$ instead of a single i.i.d. observation for each entity.

- But in fact there is one key feature: in the cluster SE derivation we never assumed that observations are i.i.d. *within* an entity. Thus we have implicitly allowed for serial correlation within an entity.

- What happened to that serial correlation – where did it go? Answer: It determines $\sigma^2_{\bar{Y}_i}$, the variance of $\bar{Y}_i$…

Serial correlation in Y_{it} influences $\sigma^2_{\bar{Y}_i}$

$$\sigma^2_{\bar{Y}_i} = \text{var}(\bar{Y}_i)$$

$$= \text{var}\left(\frac{1}{T}\sum_{t=1}^{T} Y_{it}\right) = \frac{1}{T^2}\text{var}(Y_{i1} + Y_{i2} \ldots + Y_{iT})$$

$$= \frac{1}{T^2}\{\text{var}(Y_{i1}) + \text{var}(Y_{i2}) + \ldots + \text{var}(Y_{iT}) + 2\,\text{cov}(Y_{i1}, Y_{i2})$$

$$+ 2\,\text{cov}(Y_{i1}, Y_{i3}) + \ldots + 2\,\text{cov}(Y_{iT-1}, Y_{iT})\}$$

- If Y_{it} is serially uncorrelated, all the autocovariances = 0 and we have the usual (i.i.d.) formula.
- If these autocovariances are nonzero, the usual formula (which sets them to 0) will be wrong.
- If these autocovariances are positive, the usual formula understates the variance of $\bar{Y}_i$.

Serial correlation in Y_{it} enters $\sigma^2_{\bar{Y}_i}$ (2 of 3)

The "magic" of clustered SEs is that, by working at the level of the entities and their averages $\bar{Y}_i$, you never need to worry about estimating any of the underlying autocovariances – they are in effect estimated automatically by the cluster SE formula. Here's the math:

$Clustered\ SE$ of $\bar{Y} = \sqrt{s^2_{\bar{Y}_i}/n}$, where

$$s^2_{\bar{Y}_i} = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \bar{Y})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T}Y_{it} - \bar{Y}\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T}(Y_{it} - \bar{Y})\right)^2$$

Serial correlation in Y_{it} enters $\sigma^2_{\bar{Y}_i}$ (3 of 3)

Clustered SE of $\bar{Y} = \sqrt{s^2_{\bar{Y}_i}/n}$, where

$$s^2_{\bar{Y}_i} = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_i - \bar{Y})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T}Y_{it} - \bar{Y}\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{i=1}^{T}(Y_{it} - \bar{Y})\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\left(Y_{it} - \bar{Y}\right)\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left(Y_{is} - \bar{Y}\right)\right)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(Y_{is} - \bar{Y}\right)\left(Y_{it} - \bar{Y}\right)$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_{is} - \bar{Y}\right)\left(Y_{it} - \bar{Y}\right)\right]$$

- The final term in brackets, $\frac{1}{n-1}\sum_{i=1}^{n}(Y_{is} - \bar{Y})(Y_{it} - \bar{Y})$, estimates the autocovariance between Y_{is} and Y_{it}. Thus the clustered SE formula implicitly is estimating all the autocovariances, then using them to estimate $\sigma^2_{\bar{Y}_i}$!
- In contrast, the "usual" SE formula zeros out these autocovariances by omitting all the cross terms – which is only valid if those autocovariances are all zero.

Clustered SEs for the FE estimator in panel data regression

- The idea of clustered SEs in panel data is completely analogous to the case of the panel-data mean above – just a lot messier notation and formulas. See SW Appendix 10.2.
- Clustered SEs for panel data are the logical extension of HR SEs for cross-section. In cross-section regression, HR SEs are valid whether or not there is heteroskedasticity. In panel data regression, clustered SEs are valid whether or not there is heteroskedasticity and/or serial correlation.
- *By the way…* The term "clustered" comes from having a "cluster" of observations that are correlated within a cluster (within an entity), but not across clusters. The idea of clustering extends to other applications in which there is correlation within a cluster, but not across clusters.

Clustered SEs: Implementation in STATA

```
. xtreg vfrall beertax, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48
R-sq:  within  = 0.0407                         Obs per group: min =         7
       between = 0.1101                                        avg =       7.0
       overall = 0.0934                                        max =         7
                                                F(1,47)            =      5.05
corr(u_i, Xb)  = -0.6885                        Prob > F           =    0.0294

                                (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      vfrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736   .2918556    -2.25   0.029    -1.243011   -.0687358
       _cons |   2.377075   .1497966    15.87   0.000     2.075723    2.678427
------------------------------------------------------------------------------
```

- vce(cluster state) says to use clustered standard errors, where the clustering is at the state level (observations that have the same value of the variable "state" are allowed to be correlated, but are assumed to be uncorrelated if the value of "state" differs)

Mathematical Derivation of Two-Way Clustered Standard Errors in Panel Regression

We derive the asymptotic variance estimator for panel regression with two-way clustering (e.g., by firm *and* time), following Cameron, Gelbach & Miller (2011) and Thompson (2011).

1. Model Setup

Consider the panel regression:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \epsilon_{it},\ i=1,\ldots,N;\ t=1,\ldots,T,$$

where i indexes **groups** (e.g., firms), t indexes **time periods**,
Errors ϵ_{it} **may be correlated within groups and over time.**

2. Clustering Assumptions

Assume dependence in errors:

- **Within-group**: $\mathrm{Cov}(\epsilon_{it}, \epsilon_{is}) \neq 0$ for $t \neq s$ (same firm, different time).
- **Within-time**: $\mathrm{Cov}(\epsilon_{it}, \epsilon_{jt}) \neq 0$ for $i \neq j$ (same time, different firms).
- **Two-way clustering**: Accounts for both.

3. OLS Estimator and Residuals

The OLS estimator is:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \mathbf{x}_{it}^{\top} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} y_{it} \right).$$

$$\hat{\epsilon}_{it} = y_{it} - \mathbf{x}_{it}^{\top} \hat{\boldsymbol{\beta}}.$$

4. Variance Estimator

The **two-way clustered covariance matrix** is:

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \mathbf{Q}^{-1} \left(\mathbf{\Omega}_G + \mathbf{\Omega}_T - \mathbf{\Omega}_{GT} \right) \mathbf{Q}^{-1},$$

where:

1. $\mathbf{Q} = \frac{1}{NT} \sum_{i,t} \mathbf{x}_{it} \mathbf{x}_{it}^\top$ (Hessian),
2. $\mathbf{\Omega}_G = \frac{1}{NT} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \mathbf{x}_{it} \hat{\epsilon}_{it} \right) \left(\sum_{t=1}^{T} \mathbf{x}_{it} \hat{\epsilon}_{it} \right)^\top$ (group clustering),
3. $\mathbf{\Omega}_T = \frac{1}{NT} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \mathbf{x}_{it} \hat{\epsilon}_{it} \right) \left(\sum_{i=1}^{N} \mathbf{x}_{it} \hat{\epsilon}_{it} \right)^\top$ (time clustering),
4. $\mathbf{\Omega}_{GT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \mathbf{x}_{it}^\top \hat{\epsilon}_{it}^2$ (heteroskedasticity-only).

5. Intuition

- $\mathbf{\Omega}_G$: Captures **within-firm** correlation (over time).
- $\mathbf{\Omega}_T$: Captures **within-period** correlation (across firms).
- $\mathbf{\Omega}_{GT}$: Avoids double-counting the diagonal terms (when $i=j$ and $t=s$).

7. Key Properties

1. **Consistency**: Requires $N, T \to \infty$ (not necessarily balanced).
2. **Robustness**: Valid under arbitrary within-group *and* within-time dependence.
3. **Extensions**: For multi-way clustering ($M \geq 2$), use:

$$\widehat{\text{Var}}(\hat{\beta}) = \mathbf{Q}^{-1} \left(\sum_{m=1}^{M} \boldsymbol{\Omega}_m - (M-1)\boldsymbol{\Omega}_0 \right) \mathbf{Q}^{-1},$$

where $\boldsymbol{\Omega}_0$ is the heteroskedasticity-only term.

References

1. Cameron, Gelbach & Miller (2011), *"Robust Inference with Multiway Clustering"* (DOI).
2. Thompson (2011), *"Simple Formulas for Standard Errors that Cluster by Both Firm and Time"* (DOI).

This estimator is widely used in finance (firm-time clustering) and labor economics (worker-firm clustering).

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (1 of 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)

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (2 of 6)

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)

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (3 of 6)

MITT ROMNEY
GOVERNOR

KERRY HEALEY
LIEUTENANT GOVERNOR

FOR IMMEDIATE RELEASE:
October 28, 2005

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

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.

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (4 of 6)

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.

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (5 of 6)

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."

Application: Drunk Driving Laws and Traffic Deaths (SW Section 10.6) (6 of 6)

"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*
  - *Mandatory 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? (1 of 4)

- 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
    - 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
    - use time fixed effects

Why might panel data help? (2 of 4)

TABLE 10.1 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) [0.26, 0.46]	−0.66 (0.29) [−1.23, −0.09]	−0.64 (0.36) [−1.35, 0.07]	−0.45 (0.30) [−1.04, 0.14]	−0.69 (0.35) [−1.38, 0.00]	−0.46 (0.31) [−1.07, 0.15]	−0.93 (0.34) [−1.60, −0.26]
Drinking age 18		0.10		0.03 (0.07) [−0.11, 0.17]	−0.01 (0.08) [−0.17, 0.15]		0.04 (0.10) [−0.16, 0.24]
Drinking age 19				−0.02 (0.05) [−0.12, 0.08]	−0.08 (0.07) [−0.21, 0.06]		−0.07 (0.10) [−0.26, 0.13]
Drinking age 20				0.03 (0.05) [−0.07, 0.13]	−0.10 (0.06) [−0.21, 0.01]		−0.11 (0.13) [−0.36, 0.14]
Drinking age						0.00 (0.02) [−0.05, 0.04]	
Mandatory jail or community service?				0.04 (0.10) [−0.17, 0.25]	0.09 (0.11) [−0.14, 0.31]	0.04 (0.10) [−0.17, 0.25]	0.09 (0.16) [−0.24, 0.42]
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)

	Dependent variable:						
	fatal_rate						
	OLS	Linear Panel Regression					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
beertax	0.365***	−0.656**	−0.640*	−0.445	−0.690**	−0.456	−0.926***
	(0.053)	(0.289)	(0.350)	(0.291)	(0.345)	(0.301)	(0.337)
drinkagec[18,19)				0.028	−0.010		0.037
				(0.068)	(0.081)		(0.101)
drinkagec[19,20)				−0.018	−0.076		−0.065
				(0.049)	(0.066)		(0.097)
drinkagec[20,21)				0.032	−0.100*		−0.113
				(0.050)	(0.055)		(0.123)
drinkage						−0.002	
						(0.021)	
punishyes				0.038	0.085	0.039	0.089
				(0.101)	(0.109)	(0.101)	(0.161)
miles				0.00001	0.00002*	0.00001	0.0001***
				(0.00001)	(0.00001)	(0.00001)	(0.00005)
unemp				−0.063***		−0.063***	−0.091***
				(0.013)		(0.013)	(0.021)
log(income)				1.816***		1.786***	0.996
				(0.624)		(0.631)	(0.666)
Constant	1.853***						
	(0.047)						
Observations	336	336	336	335	335	335	95
R^2	0.093	0.041	0.036	0.360	0.066	0.357	0.659
Adjusted R^2	0.091	−0.120	−0.149	0.217	−0.134	0.219	0.157
Residual Std. Error	0.544 (df = 334)						
Note:							*p<0.1; **p<0.05; ***p<0.01

Why might panel data help? (3 of 4)

TABLE 10.1 (*Continued*)

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

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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

Why might panel data help? (4 of 4)

TABLE 10.1 (*Continued*)

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

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**							
Time effects = 0			4.22 (0.002)	10.12 (<0.001)	3.48 (0.006)	10.28 (<0.001)	37.49 (<0.001)
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 (<0.001)		31.96 (<0.001)	25.20 (<0.001)
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, 95% confidence intervals are given in square brackets under the coefficients, and *p*-values are given in parentheses under the *F*-statistics.

Empirical Analysis: Main Results (1 of 2)

- Sign of the beer tax coefficient changes when fixed state effects are included
- Time effects are statistically significant but including them doesn't have a big impact on the estimated coefficients
- Estimated effect of beer tax drops when other laws are included.
- 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 in the specifications which control for state economic conditions (unemployment rate, personal income)

Empirical Analysis: Main Results (2 of 2)

- In particular, the minimum legal drinking age has a small coefficient which is precisely estimated – reducing the MLDA doesn't seem to have much effect on overall driving fatalities.
- What are the threats to internal validity? How about:

1. Omitted variable bias
2. Wrong functional form
3. Errors-in-variables bias
4. Sample selection bias
5. Simultaneous causality bias

What do you think?

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 effect.
>
> Suppose funding and curricular issues are determined at the county level, and each county has several districts. If you are worried about OV bias resulting from unobserved county-level variables, you could include county effects (binary indicators, one for each county, omitting one county to avoid perfect multicollinearity).

Summary: Regression with Panel Data (SW Section 10.7) (1 of 2)

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

Summary: Regression with Panel Data (SW Section 10.7) (2 of 2)

- 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 entities
- Time lag effects can be important – we didn't model those in the beer tax application but they could matter
- You need to use clustered standard errors to guard against the often-plausible possibility u_{it} and u_{it} are autocorrelated