ECON 7310 Elements of Econometrics
Week 8: Model for Panel Data

David Du[1]

[1] University of Queensland

Draft

Outline

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 CA school districts in 1999 and 2000, for 840 observations total.
  - Data on 50 U.S. states, each observed in 3 years, for 150 observations total.
  - Data on 1000 individuals, in 4 different months, for 4000 observations total.
- A double subscript distinguishes entities and time periods
  - If we have 1 regressor, the data are:

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

  - More generally, if we have k regressor, the data are:

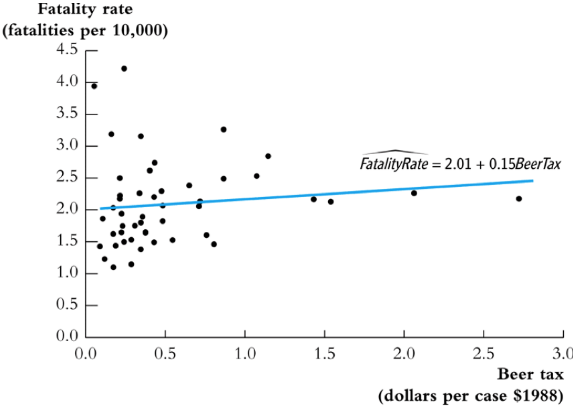
$$(X_{i1t}, \ldots, X_{ikt}, Y_{it}), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T$$

- Some jargon...
  - Another term for panel data is **longitudinal data**
  - **balanced panel:** all variables are observed for all entities and all time periods

Why are panel data useful?

- With panel data we can control for (unobserved) factors that:
  1. may cause omitted variable bias
  2. vary across entities i but do not vary over time t (or, the other way around)
- Example of a panel data set: Traffic deaths and alcohol taxes
  - 48 U.S. states, so $n =$ # of entities $= 48$
  - 7 years (1982,..., 1988), so $T =$ # of time periods $= 7$
  - Balanced panel, so total # observations $= 7 \times 48 = 336$
  - For each state i and each year t, we observe Traffic fatality rate (# traffic deaths in state i in year t, per 10,000 state residents), Tax on a case of beer, and other variables (legal driving age, drunk driving laws, etc.)

U.S. traffic death data for 1982:



(a) 1982 data

- Higher alcohol taxes, more traffic deaths?
- There can be omitted factors that cause omitted variable bias

Omitted factors

- Example 1: “traffic density.”
  - high traffic density means more traffic accidents, and more traffic deaths.
  - Also, (Western) states with lower traffic density have lower alcohol taxes
- Example 2: Cultural attitudes towards drinking and driving
  - arguably are a determinant of traffic deaths; and
  - potentially are correlated with the beer tax.
- Both cases satisfy the two conditions for omitted variable bias
- We can eliminate the omitted factors using the structure of the panel data if the factors do not change over time (at least within the sample period)

Panel Data with Two Time Periods SW Section 10.2

- Consider the panel data model,

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

  where Z_i is a factor that does not change over time (density), at least during the years on which we have data.
- Suppose $E[u_{it}|BeerTax_{it}, Z_i] = 0$ but Z_i is not observed.
- Then, its omission could result in omitted variable bias. However, the effect of Z_i can be eliminated using $T = 2$ years (or more).
- 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.

- We have two regression equations, one for 1988 and the other for 1982

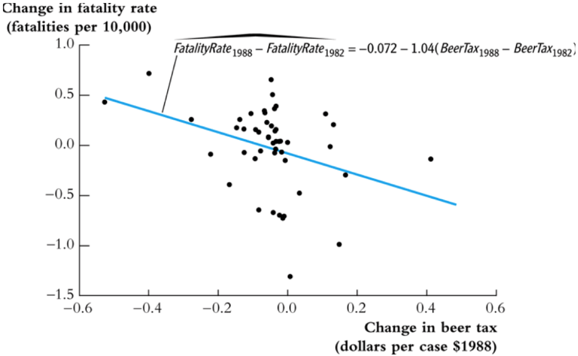
$$
\begin{aligned}
FatalityRate_{i,1988} &= \beta_0 + \beta_1 BeerTax_{i,1988} + \beta_2 Z_i + u_{i,1988} \\
FatalityRate_{i,1982} &= \beta_0 + \beta_1 BeerTax_{i,1982} + \beta_2 Z_i + u_{i,1982}
\end{aligned}
$$

- We take the difference to eliminate the effect from Z_i;

$$
\begin{aligned}
(FatalityRate_{i,1988} - FatalityRate_{i,1982}) =& \beta_1 (BeerTax_{i,1988} - BeerTax_{i,1982}) \\
&+ (u_{i,1988} - u_{i,1982})
\end{aligned}
$$

- The new error term, $(u_{i,1988} - u_{i,1982})$, is not correlated with either $BeerTax_{i,1988}$ or $BeerTax_{i,1982}$.
- Hence, an OLS regression of (the <u>change</u> in *FatalityRate*) on (the <u>change</u> in *BeerTax*) would result in a consistent and unbiased estimator for β_1.

$\Delta FatalityRate$ vs. $\Delta BeerTax$



- Note that the intercept is included in this regression and its estimate is nearly zero
  - adding an irrelevant variable → estimation less efficient (larger SE)
  - we might actually need an intercept; more on this later.

Fixed Effects Regression SW Section 10.3

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

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

- We can rewrite this in two equivalent ways:
  - "$n - 1$ binary regressor" regression model
  - "Fixed Effects" regression model
- We first rewrite this in "fixed effects" form. Suppose we have $n = 3$ states: California (CA), Texas (TX), and Massachusetts (MA).
- For $i = CA$, we rewrite the model above as follow;

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

- So, α_{CA} 'picks up' Z_{CA}, unobserved factors like 'traffic density' and 'driving/drinking culture' in CA, which may cause omitted variable bias.

- We can do the same for TX and MA. Then, we have

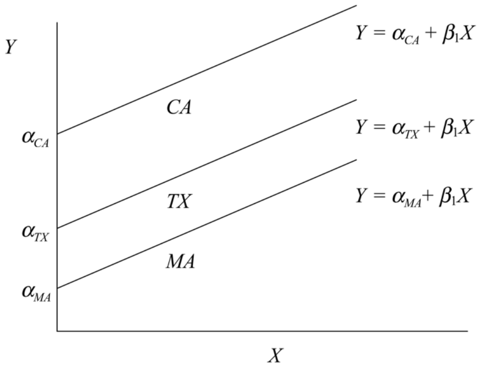
$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$
$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}$$
$$Y_{MA,t} = \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}$$

  Or,

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

- So, we have three regressions with a common slope β_1 and state-specific intercepts α_i for $i = CA, TX, MA$.
- Here, α_i is called the fixed effect (or <u>state</u> fixed effect in <u>this</u> example)

The regression lines for each state in a picture



- Recall that we can re-write the fixed effect form using binary regressors;

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

  where DTX_i is the dummy for TX and DCA_i is for CA.
- **Question:** Why *DMA* not included?

Fixed Effects Regression Estimation

- We can easily generalize this to n observations:
  Fixed effects form or, equivalently, regression with $n-1$ dummies.
- Now, we have three estimation strategies;
  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 studied 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 is not too big

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

$$Y_{it} = \alpha_1 + \beta_1 X_{it} + \gamma_2 D2_i + \cdots + \gamma_n Dn_i + u_{it}$$

where $D2_i = 1$ if i is 2 (e.g., State #2), otherwise it is zero, etc.

- First create the binary variables $D2_i, \ldots, Dn_i$. (how about $D1$?)
- Then estimate the coefficients by OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroskedasticity-robust standard errors)
- This is impractical when n is large (for example if $n = 1000$ workers)

2. "Entity-demeaned" OLS regression

- The Fixed Effect regression model:

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

  where the FE, α_i, absorbs unobserved factors that may result in omitted variable bias in estimation of β_1.
- In order to delete out α_i, we take the sample average over t for each i;

$$\overline{Y}_i = \beta_1 \overline{X}_i + \alpha_i + \overline{u}_i$$

  where $\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and similarly for $\overline{X}_i$ and $\overline{u}_i$, i.e., they are sample averages over time for each entity $i = 1, \ldots, n$.
- Then, take the mean deviation for each entity $i = 1, \ldots, n$;

$$(Y_{it} - \overline{Y}_i) = \beta_1 (X_{it} - \overline{X}_i) + (u_{it} - \overline{u}_i)$$

- Finally, estimate β_1 via OLS without an intercept.
  Then, this FE estimator is unbiased and consistent.

Entity-demeaned OLS regression, ctd.

The entity demeaned regression model can be written as

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

where $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i$ and $\widetilde{X}_{it} = X_{it} - \overline{X}_i$

- First construct the entity-demeaned variables $\widetilde{Y}_{it}$ and $\widetilde{X}_{it}$
- Then estimate β_1 by regressing $\widetilde{Y}_{it}$ on $\widetilde{X}_{it}$ using OLS
- 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

Example: Traffic deaths and beer taxes in STATA

- 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

```
. 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

Regression with Time Fixed Effects SW Section 10.4

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

- Suppose safety improvements (air bags, etc) in new cars are introduced nationally at some t's in the sample period.
- These serve to reduce traffic fatalities in all states and also 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

- If there was no entity FE, the model would be given as

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

- That is, the **time fixed effects regression model** is

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

  where $\lambda_1, \dots, \lambda_T$ are known as time fixed effects.
- This model can be equivalently written with $T - 1$ time dummies

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

  where $B2_t = 1$ if t is 2, otherwise it is zero, etc.
- Estimation and inference is parallel to the entity FE case above.
  1. "$T - 1$" binary regressor" OLS regressions
  2. "time-demeaned" OLS regression

Estimation with both entity and time fixed effects

- We may have both entity FEs and time FEs. Then, the **entity and time fixed effects regression model** is

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

- When $T = 2$, computing the first difference and including an intercept is equivalent to including entity and time fixed effects.
- When $T > 2$, there are a number of alternative algorithms to estimate this model;
  - entity demeaning & $T - 1$ time indicators (see the STATA example below)
  - time demeaning & $n - 1$ entity indicators
  - $T - 1$ time indicators & $n - 1$ entity indicators
  - entity & time demeaning

Example: Traffic deaths and beer taxes in STATA

```
. 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
-------------+----------------------------------------------------------------
```

Example: Traffic deaths and beer taxes in STATA

```
.   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
```

FE Regression Assumptions and SEs for FE Regression SW Section 10.5

- In panel data, errors can be correlated over time within an entity.
- This does not introduce bias into the FE estimator, but it affects the variance of the estimator (just like heteroskedasticity).
- Hence, we have to adjust the way to compute SEs of the FE estimators.
- Here, we study FE regression assumptions under which FE estimator is consistent and asymptotically normally distributed (as $n \to \infty$).
- Then, we describe clustered standard errors, which have been used in the examples in this chapter.

Model and Assumptions SW Section 10.5

- Consider the regression model with entity fixed effects,

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

  where

  1. $E[u_{it}|X_{i1}, \ldots, X_{iT}, \alpha_i] = 0$,
     - This assumption implies there is no omitted variable bias.
     - u_{it} is not correlated with any of $(X_{i1}, \ldots, X_{iT})$, i.e., the whole history
  2. $(X_{i1}, \ldots, X_{iT}, u_{i1}, \ldots, u_{iT})$, $i = 1, \ldots, n$ are i.i.d draws,
     - This is i.i.d. across entities, but correlation is allowed within an entity over t.
     - If X_{it} is correlated with X_{is} for $t \neq s$, X_{it} is autocorrelated or serially correlated.
     - Example: beer tax of CA in 1982 will be correlated with beer tax of CA in 1983.
     - Also, a major road improvement would reduce traffic accidents for many years.
  3. Large outliers are unlikely: (X_{it}, u_{it}) have nonzero finite fourth moments,
  4. There is no perfect multicollinearity.

- Under these assumptions, the FE estimator is unbiased, and it is consistent and asymptotically normally distributed.

HAC standard errors SW Section 10.5

- As before u_{it} are heteroskedastic over i (and t). In addition to this, u_{it} are now likely to be autocorrelated omitted variableer t for each i.
- For valid statistical inference, we should use SEs that are robust to both heteroskedasticity and autocorrelation (HAC): HAC standard errors.
- The SEs we use here are one type of HAC SEs, called **clustered SEs**, which allows arbitrary serial correlation within each 'cluster' by i.
- Like heteroskedasticity-robust SEs in regression with cross-sectional data, clustered SEs are valid whether or not there is heteroskedasticity or autocorrelation or both for large n.

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)

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 1 am and 3 am have been drinking
- A drunk driver is 13 times as likely to cause a fatal crash as a non-drinking driver

Application: Drunk Driving Laws and Traffic Deaths SW Section 10.6

Public policy issues:

- Drunk driving causes massive externalities (sober drivers are killed, society bears medical costs, 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 drunk driving panel data set
$n = 48$ U.S. states, $T = 7$ years $(1982, \ldots, 1988)$ (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 driving whilst intoxicated (DWI) violation:
  - Mandatory Jail
  - Mandatory Community Service
  - otherwise, sentence will just be a monetary fine
- Vehicle miles per driver (US Department of Transportation)
- State economic data (real per capita income, etc.)

Why might panel data help?

- Potential omitted variable 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 omitted variable 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

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* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes

***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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)
$\overline{R}^2$	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, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the $^{+}$10%, *5%, or **1% significance level.

Empirical Analysis: Main Results

- Sign of the beer tax coefficient changes when state FEs 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)
- In particular, the minimum legal drinking age (MLDA) has a small coefficient which is not precisely[1] estimated – reducing the MLDA doesn't seem to have much effect on overall driving fatalities.

[1] The textbook says it is 'precisely' estimated, which is a typo.

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 each group has its own intercept.
- Example: Class size effect on Test Score.
  - Suppose funding and curricular issues are determined at the county level, and each county has several districts.
  - If you are worried about omitted variable bias resulting from unobserved county-level variables, you could include county effects.
  - That is, include binary indicators, one for each county, omitting one county to avoid perfect multicollinearity