Regression with Panel Data

Ming-Ching Luoh

2022.3.21.

- Multiple regression is a powerful tool for controlling the effect of variables on which we have data.
- If the data are not available for some of the variables, however, they can not be included in the regression and the OLS estimators of the regression coefficients could have omitted variable bias
- This chapter describes a method for controlling some types of omitted variables without actually observing them.

- This method requires a specific type of data, called panel data, in which each observational unit, or entity, is observed at two or more periods.
- By studying *changes* in the dependent variable over time, it is possible to eliminate the effect of omitted variables that differ across entities but are constant over time.

Panel Data

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

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.

Notations for panel data

- A double subscript distinguishes **entities** (states) and **time periods** (years)
- i = entity (state), n = number of entities, so $i = 1, \cdots, n$.
- t = time period (year), T = number of time periods so $t = 1, \cdots, T$
- Data: Suppose we have 1 regressor. The data are

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

Panel data with k regressors:

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

n = number of entities (states)
T = number of time periods (years)
Some terminologies.

- Another term for panel data is **longitudinal data**.
- **balanced panel**: no missing observations.
- **unbalanced panel**: some entities (states) are not observed for some 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.

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: Traffic Deaths and Alcohol Taxes

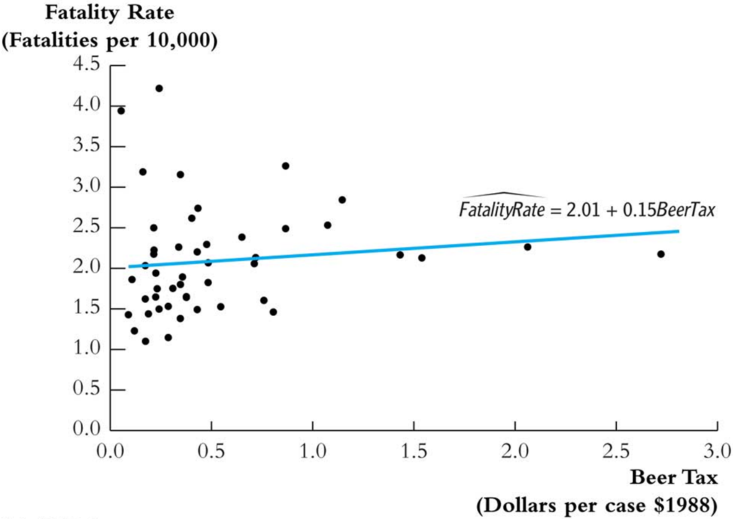
Observational unit: a year in a U.S. state

- 48 U.S. states, so n = no. of entities = 48.
- 7 years (1982,$\cdots$, 1988), so T = # of time periods = 7.
- balanced panel, total # observations = $7 \times 48 = 336$.

Variables:

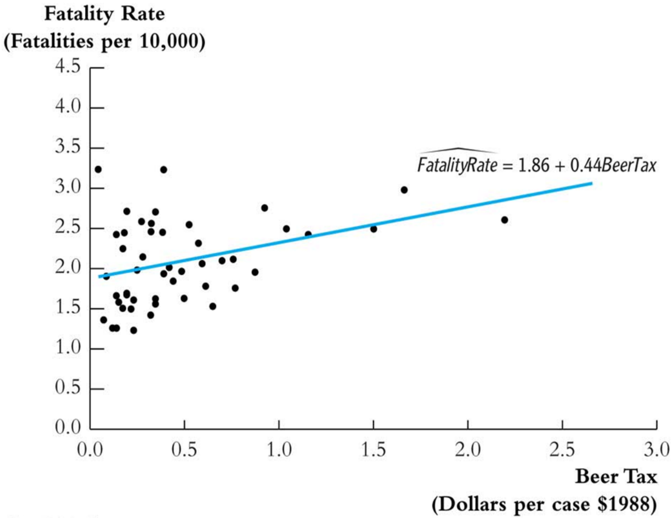
- Traffic fatality rate (# traffic deaths in that state in that year, per 10,000 state residents).
- Tax on a case of beer.
- Other (legal driving age, drunk driving laws, etc.).

Traffic death data for 1982



Higher alcohol taxes, more traffic deaths?

Traffic death data for 1988



Higher alcohol taxes, more traffic deaths?

Why might there be *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) are correlated with the beer tax, so beer taxes could be picking up cultural differences.

- Then the two conditions for omitted variable bias are satisfied. Specifically, "high taxes" could reflect "cultural attitudes towards drinking."
- 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: "Before and After" Comparisons

Consider the panel data model,

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

Z_i is a factor that does not change over time, 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 when $T = 2$.

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.

- Consider fatality rates in 1988 and 1982:

$$\begin{aligned} FatalRate_{i1988} &= \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988} \\ FatalRate_{i1982} &= \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982} \end{aligned}$$

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

  Subtracting 1988 - 1982 (that is, calculating the change), eliminates the effect of Z_i.

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

so

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

- 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 is not 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:

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

1988 data:

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

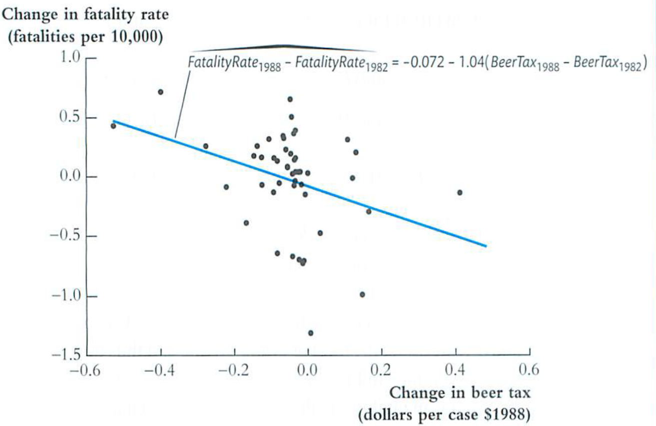
Difference regression (n = 48)

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

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

FIGURE 10.2 Changes in Fatality Rates and Beer Taxes from 1982 to 1988

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in the real beer tax between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.



- In contrast to the cross-sectional regression results, the estimated effect of a change in the beer tax is negative, as predicted by economic theory.
- According to this estimated coefficient, an increase in the beer tax by \$1 per case reduces the traffic fatality rate by 1.04 deaths per 10,000 people.
- This estimated effect is very large: The average fatality rate is approximately 2 in these data.
- Traffic fatalities can be cut in half merely by increasing the real tax on beer by \$1 per case.

Fixed Effects Regression

- Fixed effects regression is a method for omitted variables in panel data when the omitted variables vary across entities (states) but do not change over time.
- Unlike the “before and after" comparisons for two-period data, fixed effects regression can be used when there are two or more observations for each entity.

- The fixed effects regression model has n different intercepts, one for each entity.
- These intercepts can be represented by a set of binary (or indicator) variables.
- These binary variables absorb the influences of all omitted variables that differ from one entity to the next but are constant over time.

$$\begin{aligned} Y_{it} &= \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, \\ & \quad i = 1, \cdots, n, t = 1, \cdots, T \end{aligned}$$

We can rewrite this in two useful ways:

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

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

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

Population regression for California (i = CA):

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

- $\alpha_{CA} = \beta_0 + \beta_2 Z_{CA}$ 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:

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

where $\alpha_{TX} = \beta_0 + \beta_2 Z_{TX}$.
Collecting the lines for all three states:

$$
\begin{aligned}
Y_{it} &= \color{red}{\alpha_i} + \beta_1 X_{it} + u_{it}, \\
& \quad i = CA, TX, MA, t = 1, \cdots, T
\end{aligned}
$$

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, $\quad = 0$ otherwise.
- $DTX_i = 1$ if state is TX, $\quad = 0$ otherwise.
- Leave out DMA_i (why?)

Summary:
Two ways to write the fixed effects model

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

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 D_{2i} + \cdots + \beta_n D_{ni} + u_i$$

where $D_{2i} = 1$ if $i = 2$ (state #2), 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 Regression: Estimation

Three estimation methods:

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

- These three methods produce identical estimates of the regression coefficients, and identical standard errors.
- We already did the "changes" specification— but this only works for $T = 2$.
- 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

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 D_{2i} + \cdots + \beta_n D_{ni} + u_i$$

where $D_{2i} = 1$ if $i = 2$ (state #2), etc.

- First create the binary variables $D_{2i}, \cdots, D_{ni}$.
- Then estimate it 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:

$$
\begin{aligned}
& Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} \\
= \;& \beta_1\left(X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}\right) + \left(u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}\right) \\
\tilde{Y}_{it} = \;& \beta_1 \tilde{X}_{it} + \tilde{u}_{it}
\end{aligned}
$$

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.

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

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

- Construct the demeaned variables $\tilde{Y}_{it}$ and $\tilde{X}_{it}$.
- Estimate by regressing $\tilde{Y}_{it}$ on $\tilde{X}_{it}$ using OLS.
- This is like the “changes", 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.

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
```

```
. 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

For n = 48, T = 7:

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

- 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 $(T = 2, n = 48)$:

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

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:

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

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

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

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

Two formulations for time fixed effects

1. "Binary regressor" formulation:

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

where $B2_t = 1$ if $t = 2$ (year #2), etc.

2. "Time effects" formulation:

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

Time fixed effects: estimation methods

1. "T-1 binary regressors" OLS regression

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

- Create binary variables $B2, \cdots, BT$.
- $B2 = 1$ if $t =$ year #2, $= 0$ otherwise.
- Regress Y on X, $B2, \cdots, BT$ 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.

Both Entity and Time Fixed Effects

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

1. “Binary regressor" formulation:

$$\begin{aligned} Y_{it} &= \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \cdots + \gamma_n Dn_i \\ &\quad + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it} \end{aligned}$$

2. “State and time effects" formulation:

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

entity and time effects: estimation methods

1. "n-1 and T-1 binary regressors" OLS regression

- Create binary variables $D2, \cdots, Dn$.
- Create binary variables $B2, \cdots, BT$.
- Regress Y on $X, D2, \cdots, Dn, B2, \cdots, BT$ using OLS.
- What about D_1 and B_1?

2. "State- and year-demeaned" OLS regression

- Deviate Y_{it}, X_{it} from year and state averages.
- Estimate by OLS using "year- and state-demeaned" data.

These two methods can be combined too.

STATA example: Traffic deaths

```
. 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

For a single X:

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

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

Assumptions 3&4 are identical; 1, 2, differ; 5 is new.

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

- 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).
- 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.

Assumption #2: $(X_{i1}, \cdots, X_{iT}, Y_{i1}, \cdots, Y_{iT})$, $i = 1, \cdots, n$, 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.
- 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:
$corr(u_{it}, u_{is}|X_{it}, X_{is}, \alpha_i) = 0$ for $t \neq s$.

- 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.

What if assumption #5 fails: $corr(u_{it}, u_{is}|X_{it}, X_{is}, \alpha_i) \neq 0$

- 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 much random variation as you would were u_{it} uncorrelated.
- This problem is solved by using "heteroskedasticity and autocorrelation-consistent (HAC) standard errors".

Standard errors: (Appendix 10.2)

"Clustered" standard errors for $\bar{Y}$

Recall the derivation of the variance of $\bar{Y}$ for Y_i i.i.d:

$$\begin{aligned}
\mathrm{Var}(\bar{Y}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(Y_1 + Y_2 + \cdots + Y_n\right) \\
&= \frac{1}{n^2}\left(\mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_n)\right) \\
&\quad + \frac{1}{n^2}\left(2\mathrm{Cov}(Y_1, Y_2) + \cdots + 2\mathrm{Cov}(Y_{n-1}, Y_n)\right) \\
&= \frac{1}{n^2}\left(\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + \cdots + \mathrm{Var}(Y_n)\right) \\
&= \frac{\sigma_Y^2}{n}
\end{aligned}$$

What about panel data when $\{Y_{it}\}$ are possibly correlated within an entity over time, but are independent across entities?

$$\begin{aligned}
\bar{Y} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it} \\
\mathrm{Var}(\bar{Y}) &= \mathrm{Var}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it}\right)
\end{aligned}$$

Consider the special case $T = 2$:

$$\begin{aligned}
\mathrm{Var}(\bar{Y}) &= \mathrm{Var}\left(\frac{1}{nT}\left[(Y_{11}+Y_{12})+(Y_{21}+Y_{22})+\cdots+(Y_{n1}+Y_{n2})\right]\right) \\
&= \frac{1}{(nT)^2}\mathrm{Var}\left[(Y_{11}+Y_{12})+(Y_{21}+Y_{22})+\cdots+(Y_{n1}+Y_{n2})\right] \\
&= \frac{1}{(nT)^2}\left[\mathrm{Var}(Y_{11}+Y_{12})+\mathrm{Var}(Y_{21}+Y_{22})+\cdots+\mathrm{Var}(Y_{n1}+Y_{n2})\right] \\
&= \frac{\mathrm{Var}(Y_{i1}+Y_{i2})}{nT^2}
\end{aligned}$$

because Y_{it} is i.i.d. **across entities**.

The formula for the general case (general T) is,

$$\mathrm{Var}(\bar{Y}) = \frac{\mathrm{Var}\left(\sum_{t=1}^{T} Y_{it}\right)}{nT^2} \qquad (*)$$

If Y_{it} is i.i.d. overt time, then all the covariance terms over time drop out and we have the usual expression,

$$\mathrm{Var}(\bar{Y}) = \frac{\mathrm{Var}(Y_{it})}{nT}$$

But if there is correlation over time within entities, then the correct variance formula is $(*)$. This means that we need a new formula for the standard error of $\bar{Y}$.

Standard error of $\bar{Y}$ in panel data if there is correlation over time within entities, but independent across entities:

$$SE(\bar{Y}) = \sqrt{\frac{\hat{\mathrm{Var}}\left(\sum_{t=1}^{T} Y_{it}\right)}{nT^2}} \quad (**)$$

where $\widehat{\mathrm{Var}}\left(\sum_{t=1}^{T} Y_{it}\right)$ is the sample variance of $\sum_{t=1}^{T} Y_{it}$ (computed over $i = 1, \cdots n$).

The formula $(**)$ is the **"clustered" standard error** formula for $\bar{Y}$ in panel data— where the clustering is by entity.

Clustered SEs for the OLS fixed effects estimator

First get the large-n sampling distribution of the fixed effects estimator:

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

OLS fixed effects estimator:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2} \\
\hat{\beta}_1 - \beta_1 &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2} \\
&= \frac{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2}
\end{aligned}$$

Sampling distribution of the fixed effects estimator, ctd.

Fact:

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

so

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

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

By the CLT,

$$\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i = \sqrt{n}\frac{\sum_{i=1}^{n}\eta_i}{n} = \sqrt{n}\bar{\eta} \xrightarrow{d} N(0, \sigma_\eta^2)$$

where σ_η^2 is the variance of η_i. Therefore,

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

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

Next, obtain standard error of $\hat{\beta}_1$.

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

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

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

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

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

- We 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: $\hat{\sigma}_\eta^2$ when u_{it}, u_{is} are correlated, so assumptions 5 fails.

$$
\begin{aligned}
\sigma_\eta^2 &= \text{Var}\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\tilde{v}_{it}\right) \\
&= \text{Var}\left(\frac{\tilde{v}_{i1}+\tilde{v}_{i2}+\cdots+\tilde{v}_{iT}}{\sqrt{T}}\right) \\
&\neq \text{Var}(\tilde{v}_{it})
\end{aligned}
$$

- Recall $\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X,Y)$.
- 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.

Variance:

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

Variance estimator:

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

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:

- Clustered SEs are robust to both heteroskedasticity and serial correlation of the error term. Clustered SEs are valid whether T is large or small.
- If the errors are serially correlated, the usual hetero-robust SEs are wrong.
- So, if the serial correlation is concern, we should use clustered standard errors.
- Serial correlation is almost always a concern.

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)

Drunk Driving Laws and Traffic Deaths

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).

Public policy issues

- Drunk driving causes massive externalities (sober drivers are killed, 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 drunk driving panel data set

n = 48 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 violation:
  - Mandatory Jail
  - Mandatory Community Service
  - otherwise, sentence will just be a monetary fine
- Vehicle miles per driver
- 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

  $\Rightarrow$ 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

  $\Rightarrow$ 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) [0.26, 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)
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

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

- 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 in the specifications which control for state economic conditions (unemployment rate, personal income).

- In particular, the minimum legal drinking age has a small coefficient— reducing the MLDA doesn't seem to have much effect on overall driving fatalities.

Extensions of the "n-1 binary regressor" approach

- 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.
- 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.

Summary

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 estimation 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.

- You need to use clustered standard errors to guard against the often-plausible possibility u_{it} is autocorrelated.