Applied Statistics and Econometrics

Giuseppe Ragusa
Luiss University

Lecture 13: Panel Data - Part I

Outline

- Panel Data: What and Why
- Panel Data with Two Time Periods
- Fixed Effects Regression
- Regression with Time Fixed Effects
- Standard Errors for Fixed Effects Regression
- Application to Drunk Driving and Traffic Safety

Panel Data: What and Why

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

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

- $i=$ entity (state), n = number of entities, so $i = 1, \dots, n$
- $t=$ time period (year), T = number of time periods, so $t = 1, \dots, T$

Data:

- Suppose we have 1 regressor. The data are:

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

- With k regressors

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

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 it*

Example of a panel data set: Traffic deaths and alcohol taxes

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

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

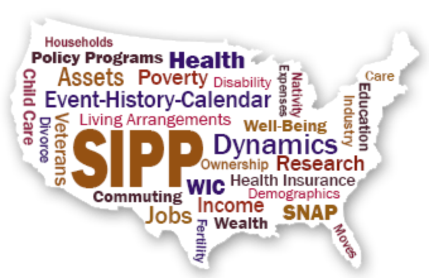
Variables:

- 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.)

Panel Study of Income Dynamics

The PSID began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the US.

Information on these individuals and their descendants has been collected continuously, including data covering employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics.



SIPP is source of information for income and welfare program participation.

SIPP collects information for assistance received either directly as money or indirectly as in-kind benefits.

SIPP collects data topics: economic well-being, family dynamics, education, assets, health insurance, childcare, and food security.

"SEC. 413. RESEARCH, EVALUATIONS, AND NATIONAL STUDIES.

"(a) RESEARCH.—The Secretary shall conduct research on the benefits, effects, and costs of operating different State programs funded under this part, including time limits relating to eligibility for assistance. The research shall include studies on the effects of different programs and the operation of such programs on welfare dependency, illegitimacy, teen pregnancy, employment rates, child well-being, and any other area the Secretary deems appropriate. The Secretary shall also conduct research on the costs and benefits of State activities under section 409.

"(b) DEVELOPMENT AND EVALUATION OF INNOVATIVE APPROACHES TO REDUCING WELFARE DEPENDENCY AND INCREASING CHILD WELL-BEING.—

"(1) IN GENERAL.—The Secretary may assist States in developing, and shall evaluate, innovative approaches for reducing welfare dependency and increasing the well-being of minor children living at home with respect to recipients of assistance under programs funded under this part. The Secretary may provide funds for training and technical assistance to carry out the approaches developed pursuant to this paragraph.

"(2) EVALUATIONS.—In performing the evaluations under paragraph (1), the Secretary shall, to the maximum extent feasible, use random assignment as an evaluation methodology.

"(c) DISSEMINATION OF INFORMATION.—The Secretary shall develop innovative methods of disseminating information on any research, evaluations, and studies conducted under this section, including the facilitation of the sharing of information and best practices among States and localities through the use of computers and other technologies.

"(d) ANNUAL RANKING OF STATES AND REVIEW OF MOST AND LEAST SUCCESSFUL WORK PROGRAMS.—

"(1) ANNUAL RANKING OF STATES.—The Secretary shall rank annually the States to which grants are paid under section 403 in the order of their success in placing recipients of assistance under the State program funded under this part into long-term private sector jobs, reducing the overall welfare caseload, and, when a practicable method for calculating this information becomes available, diverting individuals from formally applying to the State program and receiving assistance.

Art. 32.

Pubblicità e valutazione dell'impatto delle misure

1. Al fine di promuovere una maggiore consapevolezza pubblica, in particolare presso i giovani delle scuole superiori, degli istituti tecnici superiori e delle università, sulle opportunità imprenditoriali legate all'innovazione e alle materie oggetto della presente sezione, la Presidenza del Consiglio dei Ministri, su proposta del Ministero dell'istruzione, dell'università e della ricerca e del Ministero dello sviluppo economico, promuove, entro 60 giorni dalla data di conversione in legge del presente decreto, un concorso per sviluppare una campagna di sensibilizzazione a livello nazionale. Agli adempimenti previsti dal presente comma si provvede nell'ambito delle risorse umane, strumentali e finanziarie disponibili a legislazione vigente.

2. Al fine di monitorare lo stato di attuazione delle misure di cui alla presente sezione volte a favorire la nascita e lo sviluppo di start-up innovative e di valutarne l'impatto sulla crescita, l'occupazione e l'innovazione, è istituito presso il Ministero dello sviluppo economico un sistema permanente di monitoraggio e valutazione, che si avvale anche dei dati forniti dall'Istituto nazionale di statistica (ISTAT) e da altri soggetti del Sistema statistico nazionale (Sistan).

3. Il sistema di cui al comma 2 assicura, con cadenza almeno annuale, rapporti sullo stato di attuazione delle singole misure, sulle conseguenze in termini microeconomici e macroeconomici, nonché sul grado di effettivo conseguimento delle finalità di cui all'articolo 25, comma 1. Dagli esiti del monitoraggio e della valutazione di cui al presente articolo sono desunti elementi per eventuali correzioni delle misure introdotte dal presente decreto-legge.

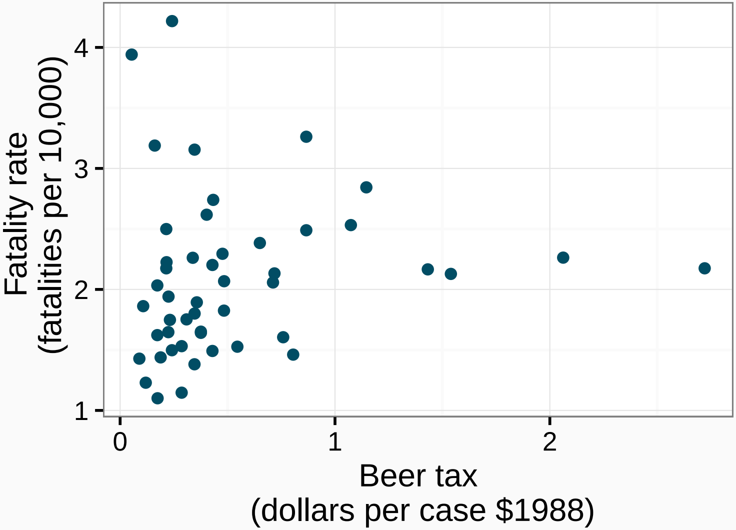
4. Allo scopo di assicurare il monitoraggio e la valutazione indipendenti dello stato di attuazione delle misure di cui alla presente sezione, l'ISTAT organizza delle banche dati informatizzate e pubbliche, rendendole disponibili gratuitamente.

5. Sono stanziate risorse pari a 150 mila euro per ciascuno degli anni 2013, 2014 e 2015, destinate all'ISTAT, per provvedere alla raccolta e all'aggiornamento regolare dei dati necessari per compiere una valutazione dell'impatto, in particolare sulla crescita, sull'occupazione, e sull'innovazione delle misure previste nella presente sezione, coerentemente con quanto indicato nel presente articolo.

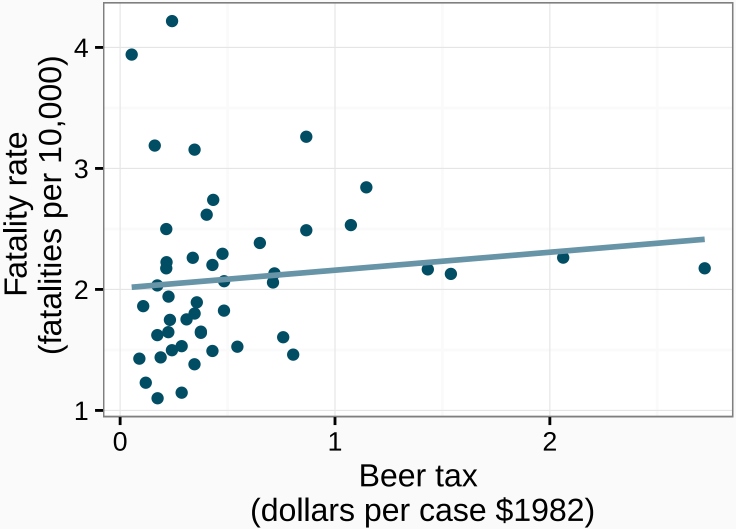
6. L'ISTAT provvede ad assicurare la piena disponibilità dei dati di cui al presente articolo, assicurandone la massima trasparenza e accessibilità, e quindi la possibilità di elaborazione e ripubblicazione gratuita e libera da parte di soggetti terzi.

7. Avvalendosi anche del sistema permanente di monitoraggio e valutazione previsto al comma 2, il Ministro dello sviluppo economico presenta entro il primo marzo di ogni anno una relazione sullo stato di attuazione delle disposizioni contenute nella presente sezione, indicando in particolare l'impatto sulla crescita e l'occupazione e formulando una valutazione comparata dei benefici per il sistema economico nazionale in relazione agli oneri derivanti dalle stesse disposizioni, anche ai fini di eventuali modifiche normative. La prima relazione successiva all'entrata in vigore del presente decreto è presentata entro il 1° marzo 2014.

U.S. traffic death data for 1982:



U.S. traffic death data for 1982:



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

Other factors that determine traffic fatality rate:

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

These omitted factors could cause omitted variable bias.

Example #1

Traffic Density:

1. High traffic density means more traffic deaths
2. (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:

1. arguably are a determinant of traffic deaths; and
2. 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

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, ctd.

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

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

Panel Data with Two Time Periods, ctd.

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

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:

$$\widehat{FatalityRate} = \underset{(0.12)}{1.86} + \underset{(0.14)}{0.44}\, BeerTax$$

1988 Data:

$$\widehat{FatalityRate} = \underset{(0.15)}{2.01} + \underset{(0.14)}{0.15}\, BeerTax$$

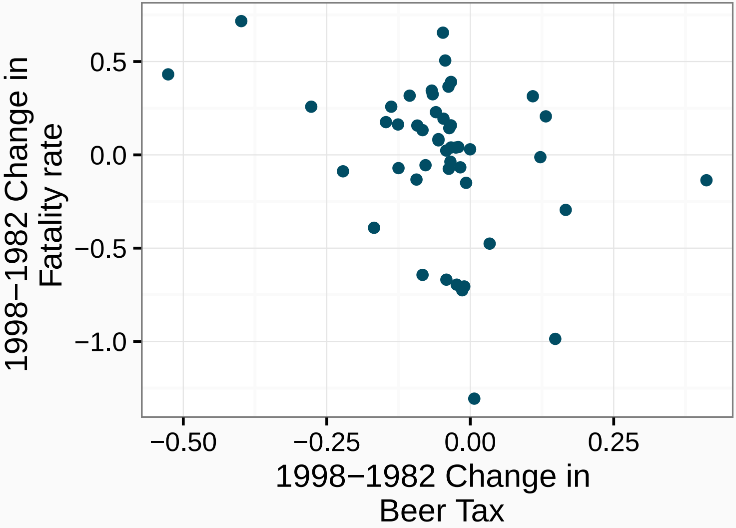
Example: Traffic deaths and beer taxes, ctd.

$$\Delta FatalityRate_i = FatalityRate_{i1988} - FatalityRate_{i1982}$$
$$\Delta BeerTax_i = BeerTax_{i1988} - BeerTax_{i1982}$$

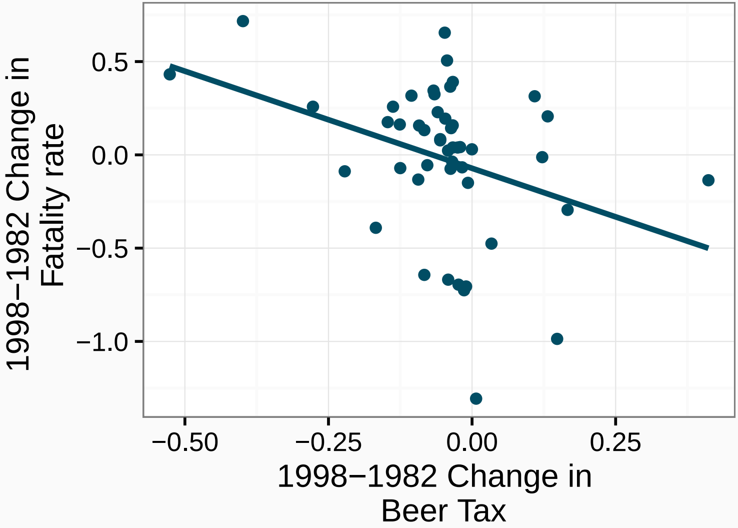
```r
Dfatalities <- with(fatalities, data.frame(
  Dfatalityrate = fatalityrate[year==1988]-fatalityrate[year==1982],
  Dbeertax      = beertax[year==1988]-beertax[year==1982]
))
summary(Dfatalities)
```

```
##  Dfatalityrate       Dbeertax
##  Min.   :-1.31   Min.   :-0.53
##  1st Qu.:-0.13   1st Qu.:-0.09
##  Median : 0.04   Median :-0.04
##  Mean   :-0.02   Mean   :-0.05
##  3rd Qu.: 0.24   3rd Qu.:-0.02
##  Max.   : 0.72   Max.   : 0.41
```

Example: Traffic deaths and beer taxes, ctd.



Example: Traffic deaths and beer taxes, ctd.



Example: Traffic deaths and beer taxes, ctd.

```
lm1 <- lm(Dfatalityrate ~ Dbeertax - 1, data = Dfatalities)
summary_rob(lm1)
```

```
##
## Coefficients:
##          Estimate Std. Error z value Pr(>|z|)
## Dbeertax   -0.869      0.269   -3.23   0.0012
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.4 on 47 degrees of freedom
## Multiple R-squared:  0.0942, Adjusted R-squared:  0.0749
## F-statistic:   NA on NA and Inf DF,  p-value: NA
```

- We can include an intercept which allows for the mean change in fatalityrate to be nonzero – more on this later. . .

Example: Traffic deaths and beer taxes, ctd.

$$\mathit{Dfatalityrate} = -\underset{(0.29)}{0.87}\,\mathit{Dbeertax}$$

- What is the interpretation of the coefficients?
- How do we perform test of hypothesis?
- How do we construct confidence intervals?

Fixed Effects Regression

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

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

We can rewrite this in two useful 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, Texas, and Massachusetts.

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

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

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

Population regression for **Texas** (that is, $i = T$):

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

Or

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

- $\alpha_{TX} = (\beta_0 + \beta_2 Z_{TX})$ doesn't change over time
- α_{TX} is the intercept for TX, and β_1 is the slope

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

Population regression for **Massachusetts** (that is, $i = MA$):

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

Or

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

- $\alpha_{MA} = (\beta_0 + \beta_2 Z_{MA})$ doesn't change over time
- α_{MA} is the intercept for MA, and β_1 is the slope

Collecting the lines for all three states:

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

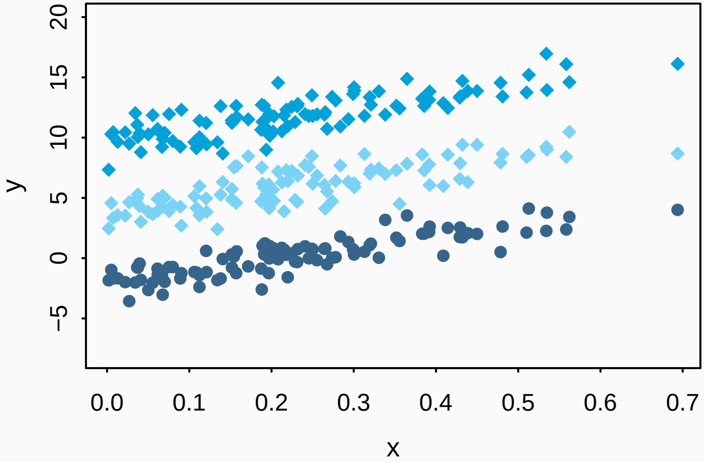
or

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

Interpretation:

It is a model with common slope (β_1) and state-specific intercept

$$
(\alpha_{CA}, \alpha_{TX}, \alpha_{MA})
$$



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

In binary regressor form:

We can write the fixed effect model in a binary regression form:

$$Y_{it} = \beta_0 + DCA_i + DTX_i + \beta_1 X_{it} + u_{it}, \quad i = \{CA, TX, MA\},\ t = 1, \ldots, T$$

where

$$DCA_i = \begin{cases} 1 & \text{if } i = CA \\ 0 & \text{otherwise,} \end{cases} \quad DTX_i = \begin{cases} 1 & \text{if } i = TX \\ 0 & \text{otherwise} \end{cases}$$

- Leave out DMA_i. Why?

Summary: Two ways to write the fixed effects model

1. "n-1 binary regressor" form

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it}$$

   where

$$D_{2i} = \begin{cases} 1 & \text{for } i = 2 \text{ (state \#2)} \\ 0 & \text{otherwise} \end{cases}$$

2. "Fixed effects" form:

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

   - α_i is called a "fixed effect" – it is the constant (fixed) effect of being in i
   - if i ranges over states, α_i is often calles "state fixed effects"
   - if i ranges over individuals, α_i is often called "individual fixed effects"

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

"n-1 binary regressors" OLS regression

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it}$$

where

$$D_{2i} = \begin{cases} 1 & \text{for } i = 2 \text{ (state \#2)} \\ 0 & \text{otherwise} \end{cases}$$

- First create the binary variables $D_{2i}, \ldots, D_{ni}$
- Then estimate (1) by OLS
- Inference (hypothesis tests, confidence intervals) is as usual (using heteroskedasticity-robust standard errors)
- This becomes impractical when n is very large (for example if $n = 1000$ workers)
- In `R` this is even easier if the identifier of i is a `factor`

Example: Traffic deaths and beer taxes, ctd.

```
## The state variable in fatalities is a factor
class(fatalities$state)
```

```
## [1] "factor"
```

```
summary(fatalities$state)
```

```
## al az ar ca co ct de fl ga id il in ia ks ky la me md ma mi mn
##  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## ms mo mt ne nv nh nj nm ny nc nd oh ok or pa ri sc sd tn tx ut
##  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## vt va wa wv wi wy
##  7  7  7  7  7  7
```

Example: Traffic deaths and beer taxes, ctd.

If the variable indexing the i is a factor, estimation with the $n-1$ method is really easy

```
lm_dummy <- lm(fatalityrate ~ beertax + state, data = fatalities)
summary_rob(lm_dummy)


##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.4776     0.3508    9.91  < 2e-16
## beertax        -0.6559     0.2033   -3.23  0.00125
## az             -0.5677     0.2952   -1.92  0.05446
## ar             -0.6550     0.2403   -2.73  0.00641
## ca             -1.5095     0.3325   -4.54  5.6e-06
## co             -1.4843     0.3240   -4.58  4.6e-06
## ct             -1.8623     0.3069   -6.07  1.3e-09
## de             -1.3076     0.3329   -3.93  8.6e-05
## fl             -0.2681     0.1445   -1.86  0.06356
## ga              0.5246     0.1786    2.94  0.00331
## id             -0.6690     0.2821   -2.37  0.01773
## il             -1.9616     0.3189   -6.15  7.7e-10
```

Example: Traffic deaths and beer taxes, ctd.

To be sure that $n-1$ is equivalent to take differences we do this

```
lm_diff <- lm(fatalityrate ~ beertax + state, data = fatalities, subset = (year ==
    1988 | year == 1982))
summary_rob(lm_diff)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.63225    0.44175    8.22  < 2e-16
## beertax     -0.86892    0.26870   -3.23  0.00122
## az          -0.78600    0.40684   -1.93  0.05336
## ar          -0.65630    0.30302   -2.17  0.03032
## ca          -1.66516    0.41793   -3.98  6.8e-05
## co          -1.62363    0.53119   -3.06  0.00224
## ct          -1.86841    0.39549   -4.72  2.3e-06
## de          -1.26766    0.44274   -2.86  0.00419
## fl          -0.20039    0.21189   -0.95  0.34430
## ga           0.89384    0.30303    2.95  0.00318
## id          -0.72620    0.35874   -2.02  0.04294
```

"Entity-demeaned" OLS regression

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

The entity averages satisfy:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} Y_{it} &= \frac{1}{T}\sum_{t=1}^{T}(\alpha_i + \beta_1 X_{it} + u_{it}) \\
&= \alpha_i + \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{it} + \frac{1}{T}\sum_{t=1}^{T} u_{it}
\end{aligned}$$

Subtracting

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

- α_i is gone....

Entity-demeaned OLS regression, ctd.

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

where

$$\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}, \;\; \tilde{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}, \;\; \tilde{u}_{it} = u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

- First construct the entity-demeaned variables and
- Then estimate (2) by regressing on 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)

A R package for Panel Data

- The package plm has many functions to estimate panel data model
- The main function is called plm
- The following code estimates a panel data regression of y on x with fixed effect using the within transformation

```
m.fe <- plm(y   x, data = data, method = "within", effect = "individual")
```

- The same model can be estimated using the first difference approach

```
m.fd <- plm(y   x, data = data, method = "fd", effect = "individual")
```

- The summary_rob can be applied to these models

Entity-demeaned OLS regression, ctd.

```r
summary_rob(plm(fatalityrate ~ beertax + 1, data = fatalities, model = "within"))
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = fatalityrate ~ beertax + 1, data = fatalities,
##     model = "within")
##
## Balanced Panel: n=48, T=7, N=336
##
## Coefficients :
##         Estimate Std. Error t-value Pr(>|t|)
## beertax   -0.656      0.188   -3.49  0.00057 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Adj. R-Squared :  -0.12
## F-statistic: 12.1511 on 1 and 47 DF, p-value: 0.00107
```

Summary

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

Statistical inference: like multiple regression.

- Advantages
  1. You can control for unobserved variables that: o vary across states but not over time,
  2. More observations give you more information
  3. Estimation involves relatively straightforward extensions of multiple regression
- Yet to be cover:
  1. What about time effects, i.e. unobservables that vary over time, but are constant across individuals?
  2. Standard errors might be too low (errors might be correlated over time)