# Lecture Notes: Causal Inference Fall 2020

Claire Duquennois

7/28/2020

## Contents



# 1 Panel Data and Fixed Effects

Our ability to control for important omitted variables increases substantially when we can better control for unobserved confounders. There are a number of strategies econemetricians use to help control for unobservable confounds. These strategies require data that has a time or a cohort dimension in order to control for unobserved but fixed omitted variables, aka **fixed effects**.

Returning to our earlier example, suppose I am interested in the relationship between income and schooling. We saw that estimating the estimated coefficients from the following specification,

$$Income_i = \beta_0 + \beta_1 Schooling_i + \epsilon \tag{1}$$

could not be interpreted as causal estimates because of selection and omitted variable bias. We also saw that even if I add lots of controls for the type of demographic characteristics that are often included in datasets, there would still likely remain important omitted variables. Most datasets do not have measures of 'ability','enthusiasm', or 'grit' that could be important determinants of both a persons income and their schooling level. But what if I can control for all the unobservable characteristics of an individual, as long as they do not change over time?

## 1.1 Data Structures

Suppose I observe each individual multiple times and I track their schooling and their income. Instead of looking like this:

| Individual | Income | Schooling | Female |
|---|---|---|---|
| 1 | 22000 | 12 | 1 |
| 2 | 57000 | 16 | 1 |
| ... | ... | ... | ... |
| N | 15000 | 12 | 0 |

my data now looks like this:

| Individual | Income | Schooling | Female | Year |
|---|---|---|---|---|
| 1 | 22000 | 12 | 1 | 2001 |
| 1 | 23000 | 12 | 1 | 2002 |
| 2 | 57000 | 16 | 1 | 2001 |
| 2 | 63000 | 17 | 1 | 2002 |
| ... | ... | ... | ... | ... |
| N | 15000 | 12 | 0 | 2001 |
| N | 13000 | 12 | 0 | 2002 |

Unique observations must be identified by both the individual and time dimensons, thus the equation above requires the addition of time subscripts, to uniquely identify an observation.

$$Income_{it} = \beta_0 + \beta_1 Schooling_{it} + \epsilon. \tag{2}$$

We are now working with **panel data**, which consists of repeated observations of the same unit at multiple points in time. Panel data can be balanced, if you observe every unit the same number of times, or unbalanced, if some units are observed more often than others.[1]

## 1.2 Fixed Effects

Recall that we know we can control for the effect of being female on wages by adding an controling for the female indicator in the regression as follows. I was able to do this because I had multiple Female observation and multiple non-female observations.

$$Income_{it} = \beta_0 + \beta_1 Schooling_{it} + \beta_2 Female_i + \epsilon. \tag{3}$$

Now that I observe the same individual multiple times, I can do something similar by adding individual controls. I can create dummy variables set to 1 for observations about a particular individual, and 0 otherwise, and add these controls to my regression:

| Individual | Income | Schooling | Female | Year | Indiv1 | Indiv2 | ... | IndivN |
|---|---|---|---|---|---|---|---|---|
| 1 | 22000 | 12 | 1 | 2007 | 1 | 0 | 0 | 0 |
| 1 | 23000 | 12 | 1 | 2008 | 1 | 0 | 0 | 0 |
| 2 | 57000 | 16 | 1 | 2007 | 0 | 1 | 0 | 0 |
| 2 | 63000 | 17 | 1 | 2008 | 0 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| N | 15000 | 12 | 0 | 2007 | 0 | 0 | 0 | 1 |
| N | 13000 | 12 | 0 | 2008 | 0 | 0 | 0 | 1 |

$$Income_{it} = \beta_0 + \beta_1 Schooling_{it} + \beta_2 Female_i + \beta_{a1} Indiv1_i + \beta_{a2} Indiv2_i + ... + \beta_{aN-1} Indiv(N-1)_i + \epsilon. \tag{4}$$

[1] In the event that you are working with unbalanced data, attention should be given to why some units are oserved more frequently then others as there may be a selection process occuring.

Notes:

- These dummy variables only have an i subscript as the value of this dummy indicator only varies accross individuals, and not across time. (Question: What is the implied assumption if Female only has an i subscript?).
- Notice that only $(N-1)$ individual dummies are added. Why? Just as with any other variable, we must have an omitted cotegory to avoid the problem of multicollineariy.

What will these individual controls control for? $\beta_{a1}$ will control for the effect of being individual 1 on income that is not explained by that person's gender or schooling.Thus any **time invariant** characteristic that affects individual 1's income, such as ability, grit, enthusiasm... will be controlled for by adding this individual dummy variable.

These individual dummy variables are known as individual **fixed effects**. For notational convenience, rather than listing them all we instead add a greek letter with the correct subscript to our regression (say $\gamma_i$) as follows

$$Income_{it} = \beta_0 + \beta_1 Schooling_{it} + \beta_2 Female_i + \gamma_i + \epsilon. \tag{5}$$

Notice that given the data I can add an additional control that may be important. Suppose I am concerned that incomes in my data were severeley affected by the financial crisis in 2008. Just as above, I can control for any common effect a particular year had on all the individuals by adding year fixed effects $(\tau_t)$ as specified below,

$$Income_{it} = \beta_0 + \beta_1 Schooling_{it} + \beta_2 Female_i + \gamma_i + \tau_t + \epsilon. \tag{6}$$

## 1.3 A Simulation

Suppose you are a principle of a small school composed of four classrooms. You have just implemented a new option available to teachers for students to spend some small group reading time with a paraeducatior. You would like to know how this reading time is affecting reading scores. You have data for ten students in each class that tells you the class the student is in, whether they participated in small group reading and their reading score.

Below I construct a simulated dataset to show how the use of fixed effects can help us recover the true treatment effect.

I start by generating a vector of class identifiers and a random error term.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
class<-c(1,2,3,4)
scores<-as.data.frame(class)
scores<-rbind(scores,scores,scores,scores,scores,scores,scores,scores,scores,scores)
scores$error<-rnorm(40, mean=0, sd=5)
```

Next I simulate some selection into treatment. I generate a treatment vector where the probability of getting treated is higher for students in classrooms 3 and 4 then it is in classrooms 1 and 2.

```
scores$treat1<-rbinom(40,1,0.2)
scores$treat2<-rbinom(40,1,0.8)
scores$treat[scores$class%in%c(1,2)]<-scores$treat1[scores$class%in%c(1,2)]
scores$treat[scores$class%in%c(3,4)]<-scores$treat2[scores$class%in%c(3,4)]
```

I then generate a dummy variable for each classroom

```
scores<-scores%>%select(class,error,treat)
scores <- fastDummies::dummy_cols(scores, select_columns = "class")
```

Finally, I genrerate the simulated outcomes. The true treatment effect is set to 15. Notice that I am simulating a situation in which students in classrooms 1 and 2 have much higher reading scores then those in classrooms 3 and 4.

```
scores$score<-80+15*scores$treat+85*scores$class_2+ -30*scores$class_3+
  -35*scores$class_4+scores$error
```

I estimate the following three specifications:

$$
\begin{aligned}
Score_{ci} &= \beta_0 + \beta_1 Treat_{ci} + \epsilon \\
Score_{ci} &= \beta_0 + \beta_1 Treat_{ci} + \beta_2 Class2_c + \beta_3 Class3_c + \beta_4 Class4_c + \epsilon \\
Score_{ci} &= \beta_0 + \beta_1 Treat_{ci} + \kappa_c + \epsilon
\end{aligned}
$$

where $\kappa_i$ is a classroom fixed effect.

```
nofe<-felm(score~treat,scores)
dummies<-felm(score~treat+class_2+class_3+class_4, scores)
fe<-felm(score~treat|class,scores)

stargazer(nofe, dummies, fe, type='latex')
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Oct 05, 2020 - 4:45:24 PM

First, notice that the coefficient that does not control for the classroom the student is in is very biased. So much so that $\hat{\beta}_1$ is negative despite the fact that the true treatment effect, $\beta_1 = 15$. This is because the classes are an important omitted variable. we have that $cor(Score, Class3/4) < 0$ and $cor(Treat, Class3/4) > 0$ creating substantial downward bias.

We can correct for this in two (equivalent) ways: adding the dummy variables for the class to the regression, or adding a class fixed effect. Either approach returns an identical unbiased estimate such that $E[\hat{\beta}_1] = \beta_1$.

## 1.4 Fixed effects as demeaned data

To build intuition about how fixed effects work, it might be helpful to think about fixed effect as the **within estimator**, because it identifies $\beta$ using within-unit variation. In the example of our classroom reading scores, when estimatimating using the classroom fixed effects, we only using the variation that exists **within the classroom** to estimate the treatment effect. This is the equivalent of "correcting" our data by demeaning each observation using it's classroom mean, so that the corrected data represents deviations from the classroom mean. Our fixed effect estimation is

Table 4:

| | *Dependent variable:* | | |
|---|---|---|---|
| | score | | |
| | (1) | (2) | (3) |
| treat | −30.105** | 15.399*** | 15.399*** |
| | (14.090) | (2.021) | (2.021) |
| class_2 | | 85.779*** | |
| | | (2.129) | |
| class_3 | | −29.807*** | |
| | | (2.548) | |
| class_4 | | −34.473*** | |
| | | (2.548) | |
| Constant | 111.421*** | 79.881*** | |
| | (10.684) | (1.552) | |
| Observations | 40 | 40 | 40 |
| $R^2$ | 0.107 | 0.990 | 0.990 |
| Adjusted $R^2$ | 0.084 | 0.989 | 0.989 |
| Residual Std. Error | 44.051 (df = 38) | 4.739 (df = 35) | 4.739 (df = 35) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

$$y_{ci} = \beta_1 x_{ci} + \kappa_c + \epsilon_{ci}$$

For each class, the average accross the students is

$$\bar{y}_i = \beta_1 \bar{x}_i + \kappa_c + \bar{\epsilon}_i$$

Subtracting this from the fixed effect model gives

$$y_{ic} - \bar{y}_i = \beta_1 (x_{ic} - \bar{x}_i) + (\epsilon_{ic} - \bar{\epsilon}_i)$$

I do this in the with the following code. I first calculate the mean score, and the mean treatment, in each classroom. I then subtract these from each student's score and treatment indicator. Finally, I estimate my original model on the demeaned scores.

```
#calculateing the mean score in each classroom
cl_mean<-scores %>%
    group_by(class) %>%
    dplyr::summarize(Classmean = mean(score, na.rm=TRUE), treatmean=mean(treat, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#merging the means into full data
scores<-left_join(scores, cl_mean, by = "class")

#calculating the demeaned score
scores$demeansc<-scores$score-scores$Classmean
scores$demeantrt<-scores$treat-scores$treatmean

#running the basic regression on the demaned scores

regdemean<-felm(demeansc~demeantrt, scores)

stargazer(nofe, dummies, fe, regdemean, type='latex')
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Oct 05, 2020 - 4:45:25 PM

Notice that we get the same coefficient using the demeaned regression as with the fixed effects regression.[2]

## 1.5 Lets talk about variation

What would happen if none of the students in classes 1 and 2 went to the small reading group and all of the students in class 3 and 4 did?

Below I modify my simulated data to reflect this.

```
library(dplyr)
library(lfe)

class2<-c(1,2,3,4)
scores2<-as.data.frame(class2)
scores2<-rbind(scores2,scores2,scores2,scores2,scores2,scores2,scores2,scores2,scores2,scores2)
scores2$error<-rnorm(40, mean=0, sd=5)
```

[2]Note however that the standard errors on the demeaned regression are incorrect because the estimation does not take into account the fact that the cases are not independent of each other.

Table 5:

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | score | | | demeansc |
| | (1) | (2) | (3) | (4) |
| treat | −30.105** | 15.399*** | 15.399*** | |
| | (14.090) | (2.021) | (2.021) | |
| class_2 | | 85.779*** | | |
| | | (2.129) | | |
| class_3 | | −29.807*** | | |
| | | (2.548) | | |
| class_4 | | −34.473*** | | |
| | | (2.548) | | |
| demeantrt | | | | 15.399*** |
| | | | | (1.939) |
| Constant | 111.421*** | 79.881*** | | −0.000 |
| | (10.684) | (1.552) | | (0.719) |
| Observations | 40 | 40 | 40 | 40 |
| $R^2$ | 0.107 | 0.990 | 0.990 | 0.624 |
| Adjusted $R^2$ | 0.084 | 0.989 | 0.989 | 0.614 |
| Residual Std. Error | 44.051 (df = 38) | 4.739 (df = 35) | 4.739 (df = 35) | 4.548 (df = 38) |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

```
scores2$treat[scores2$class%in%c(1,2)]<-0
scores2$treat[scores2$class%in%c(3,4)]<-1

scores2<-scores2%>%select(class2,error,treat)
scores2 <- fastDummies::dummy_cols(scores2, select_columns = "class2")

scores2$score<-80+15*scores2$treat+85*scores2$class2_2+ -30*scores2$class2_3+ -35*scores2$class2_4+scor

nofe2<-felm(score~treat,scores2)
dummies2<-felm(score~treat+class2_2+class2_3+class2_4, scores2)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
#fe2<-felm(score~treat|class2,scores2)

stargazer(nofe2, dummies2,  type='latex')
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Oct 05, 2020 - 4:45:25 PM

Table 6:

| | *Dependent variable:* | |
|---|---|---|
| | score | |
| | (1) | (2) |
| treat | −58.264*** | −16.790*** |
| | (10.135) | (1.804) |
| class2_2 | | 87.568*** |
| | | (1.804) |
| class2_3 | | 4.619** |
| | | (1.804) |
| class2_4 | | |
| Constant | 121.159*** | 77.375*** |
| | (7.167) | (1.276) |
| Observations | 40 | 40 |
| $R^2$ | 0.465 | 0.992 |
| Adjusted $R^2$ | 0.451 | 0.991 |
| Residual Std. Error | 32.050 (df = 38) | 4.034 (df = 36) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

Clearly we have a problem. All of the treatment estimates are biased and the fixed effects regression refused to run entirely. This is because we simply do not have the variation to estimate the true effect. Because there is no variation in treatment within the classroom, it is not possible to estimate both the effect of the classroom and the effect of treatment since they confound each other.

## 1.6 Example: Crime and Unemployment

Suppose you are interested in thinking about the relationship between unemployment and crime. You have data on the crime and unemployment rates for 46 cities for 1982 and 1987. I start by using the data from the 1987 cross section and run the following simple regression of the crime rate on unemployment,

$$crimerate_i = \beta_0 + \beta_1 unemployment_i + \epsilon$$

```
#install.packages("wooldridge")
library(wooldridge)
```

```
## Warning: package 'wooldridge' was built under R version 3.6.3
```

```
library(lfe)
#note: this dataset comes from the wooldridge textbook. Conveniently there is an R package that
#includes all the wooldridge datasets.

crime<-data('crime2')
crime<-crime2

regcrime<-felm(crmrte~unem, crime[crime$year=="87",])
summary(regcrime)
```

```
##
## Call:
##    felm(formula = crmrte ~ unem, data = crime[crime$year == "87",      ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -57.55 -27.01 -10.56  18.01  79.75
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  128.378     20.757   6.185  1.8e-07 ***
## unem          -4.161      3.416  -1.218     0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.6 on 44 degrees of freedom
## Multiple R-squared(full model): 0.03262   Adjusted R-squared: 0.01063
## Multiple R-squared(proj model): 0.03262   Adjusted R-squared: 0.01063
## F-statistic(full model):1.483 on 1 and 44 DF, p-value: 0.2297
## F-statistic(proj model): 1.483 on 1 and 44 DF, p-value: 0.2297
```

Interpreting this coefficient on unemployment suggests that a higher unemployment level is associated with less crime. This seems backwards. The culprit? Probably omitted variables. The first solution that comes to mind is to control for more observable city characteristics that we can see in our data such as the area of the city, if the city is in the west, police officers per square mile, expenditure on law enforcement, per capita income... I estimate the following,

$$crmrte_i = \beta_0 + \beta_1 unemp_i + \beta_2 area_i + \beta_3 west_i + \beta_4 offarea_i + \beta_5 lawexp_i + \beta_6 pcinc_i + \epsilon$$

```
regcrime2<-felm(crmrte~unem+area+west+offarea+lawexpc+pcinc, crime[crime$year=="87",])
summary(regcrime2)
```

```
##
## Call:
##    felm(formula = crmrte ~ unem + area + west + offarea + lawexpc +      pcinc, data = crime[crime$y
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.847 -21.511  -6.829  18.940  75.114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 140.06017   51.16000   2.738  0.00927 **
## unem         -6.70024    3.71634  -1.803  0.07913 .
## area          0.05867    0.04757   1.233  0.22491
## west        -21.96336   12.27535  -1.789  0.08135 .
## offarea      -0.11442    0.66876  -0.171  0.86504
## lawexpc       0.02137    0.01859   1.149  0.25736
## pcinc        -0.00185    0.00352  -0.526  0.60215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.27 on 39 degrees of freedom
## Multiple R-squared(full model): 0.1587   Adjusted R-squared: 0.02932
## Multiple R-squared(proj model): 0.1587   Adjusted R-squared: 0.02932
## F-statistic(full model):1.227 on 6 and 39 DF, p-value: 0.3138
## F-statistic(proj model): 1.227 on 6 and 39 DF, p-value: 0.3138
```

We still get this puzzling result. We could continue to add more controls but there are so many variables about a city that could be correlated with both unemployment and crime that it seems unlikely we could observe them all. But what id we can **capture all unobserved, time invariant factors** about a city that might affect crime rates? If I use the data for 1987 and 1982, I can add city **fixed effects,**($\alpha_i$) to the regression I estimated above,

$$crmrte_{it} = \beta_0 + \beta_1 unemp_{it} + \beta_2 area_i + \beta_3 west_i + \beta_4 offarea_{it} + \beta_5 lawexp_{it} + \beta_6 pcinc_{it} + \alpha_i + \epsilon$$

```
#note: the data does not have a unique city identifier. I am assuming the area of the city is 1)time-in
crime <- transform(crime,city=as.numeric(factor(area)))
#I check that my assumptions were correct by seeing if I have 2 observations for 46 cities.
table(crime$city)
```

```
##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
##  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
##  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
```

```
regcrime3<-felm(crmrte~unem+area+west+offarea+lawexpc+pcinc|city, crime)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
summary(regcrime3)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
##
```

```
## Call:
##    felm(formula = crmrte ~ unem + area + west + offarea + lawexpc +      pcinc | city, data = crime)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -27.36  -6.85   0.00   6.85  27.36
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## unem     1.491297   0.795972   1.874   0.0680 .
## area           NA         NA      NA       NA
## west           NA         NA      NA       NA
## offarea  1.348882   1.805672   0.747   0.4592
## lawexpc -0.005076   0.013915  -0.365   0.7171
## pcinc    0.003821   0.001644   2.324   0.0251 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.66 on 42 degrees of freedom
## Multiple R-squared(full model): 0.8887   Adjusted R-squared: 0.7588
## Multiple R-squared(proj model):  0.18   Adjusted R-squared: -0.7767
## F-statistic(full model):6.843 on 49 and 42 DF, p-value: 1.888e-09
## F-statistic(proj model): 1.537 on 6 and 42 DF, p-value: 0.19
```

Now we get a coefficient on unemployment that makes a lot more sense. Notice that we were not able to estimate a coefficient for $area_i$ and $west_i$. This is because of the multicollinearity problem. These city characteristics are time-invariant. If we include city fixed effects, these factors will already be controlled for. When we control for the city fixed effect, there is no longer any remaining variation with which to estimate the impact of $area_i$ or $west_i$.

Suppose I am concerned about how national factors could be affecting all cities simultaneously. I can add a year fixed effect, ($\tau_t$) to control for these

$$crmrte_{it} = \beta_0 + \beta_1 unemp_{it} + \beta_2 area_i + \beta_3 west_i + \beta_4 offarea_{it} + \beta_5 lawexp_{it} + \beta_6 pcinc_{it} + \alpha_i + \tau_t + \epsilon$$

```
regcrime4<-felm(crmrte~unem+area+west+offarea+lawexpc+pcinc|city+year, crime)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
summary(regcrime4)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
##
## Call:
##    felm(formula = crmrte ~ unem + area + west + offarea + lawexpc +      pcinc | city + year, data =
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.641  -7.441   0.000   7.441  23.641
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
```

```
## unem     2.931904   1.133562   2.586   0.0133 *
## area           NA         NA      NA       NA
## west           NA         NA      NA       NA
## offarea  1.838022   1.785312   1.030   0.3093
## lawexpc -0.006982   0.013632  -0.512   0.6113
## pcinc   -0.005697   0.005683  -1.002   0.3220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.31 on 41 degrees of freedom
## Multiple R-squared(full model): 0.8964   Adjusted R-squared:  0.77
## Multiple R-squared(proj model): 0.1709   Adjusted R-squared: -0.8402
## F-statistic(full model):7.094 on 50 and 41 DF, p-value: 1.405e-09
## F-statistic(proj model): 1.408 on 6 and 41 DF, p-value: 0.2347
```

So can I enterpret the coefficient on this last regression as causal? We need to think carefully about what we have controlled for. The city fixed effects control for any time invariant factors that always affect the crime rates in a city in a similar way. This would be things like geography, street layouts, weather... The time fixed effects controls for any patterns that are common to all cities in a given year. This would be things like national policies such as the war on drugs, national interest rates... In addition to this we are also controling for some observable time varying variables: officers in an area, law enforcement expenditures per capita and income per capita. So are these estimates causal? What kind of ommitted variables should we still be concerned about? Any variable that changes within a city accross years and that is correlated with both unemployment and crime rates could still be biasing our results. This coule be things like school funding, criminalization of marijuana, housing costs... to name just a few.

Suppose I get ambition and want to control for all these factors as well. I decide I am going to generate a city-by-year fixed effect, ($\gamma_{it}$), to control for these time variant omitted variables. I estimate,

$$crmrte_{it} = \beta_0 + \beta_1 unemp_{it} + \beta_2 area_i + \beta_3 west_i + \beta_4 offarea_{it} + \beta_5 lawexp_{it} + \beta_6 pcinc_{it} + \gamma_{it} + \epsilon$$

```
crime$city_year<-paste(crime$city, crime$year, sep="_")

#Note: the following regression will not run!
#regcrime5<-felm(crmrte~unem+area+west+offarea+lawexpc+pcinc|city_year, crime)
#summary(regcrime5)
```

What happened?!? Because the variation in my outcome variable is at the city-by-year level, incluting a city-by-year fixed effect absorbe all of the identifying variation, making it impossible to estimate the effect of any of the other variables in my model. Thus I cannot include this type of fixed effect with this data set. If I had data on neighbourhood crime rates and neighbourhood unemployment I could.

---