Chapter 10 - Regression with panel data

Ercio Munoz

November 1, 2018

In this chapter we use panel data, so we have observations of a group (of size n $\geq$ 2) of entities over time (T $\geq$ 2 periods of time). Specifically, we are going to work with a panel of 48 states over 7 years.

```r
library(foreign)
library(plm) # panel data linear model
```

```
## Loading required package: Formula
```

```r
a =
"http://fmwww.bc.edu/ec-p/data/stockwatson/fatality.dta"
pd = read.dta(a)
attach(pd)
# Some descriptive statistics of the variables we will use
summary(pd[,c("state","year","beertax",
              "mrall")])
```

```
##      state          year         beertax            mrall
##  AL     :  7   Min.   :1982   Min.   :0.04331   Min.   :8.212e-05
##  AZ     :  7   1st Qu.:1983   1st Qu.:0.20885   1st Qu.:1.624e-04
##  AR     :  7   Median :1985   Median :0.35259   Median :1.956e-04
##  CA     :  7   Mean   :1985   Mean   :0.51326   Mean   :2.040e-04
##  CO     :  7   3rd Qu.:1987   3rd Qu.:0.65157   3rd Qu.:2.418e-04
##  CT     :  7   Max.   :1988   Max.   :2.72076   Max.   :4.218e-04
##  (Other):294
```

```r
# Looking the first 9 observations
head(pd[,c("state","year","beertax",
              "mrall")],10)
```

```
##    state year   beertax       mrall
## 1     AL 1982 1.5393795 0.000212836
## 2     AL 1983 1.7889907 0.000234848
## 3     AL 1984 1.7142856 0.000233643
## 4     AL 1985 1.6525424 0.000219348
## 5     AL 1986 1.6099070 0.000266914
## 6     AL 1987 1.5599999 0.000271859
## 7     AL 1988 1.5014436 0.000249391
## 8     AZ 1982 0.2147971 0.000249914
## 9     AZ 1983 0.2064220 0.000226738
## 10    AZ 1984 0.2967033 0.000282878
```

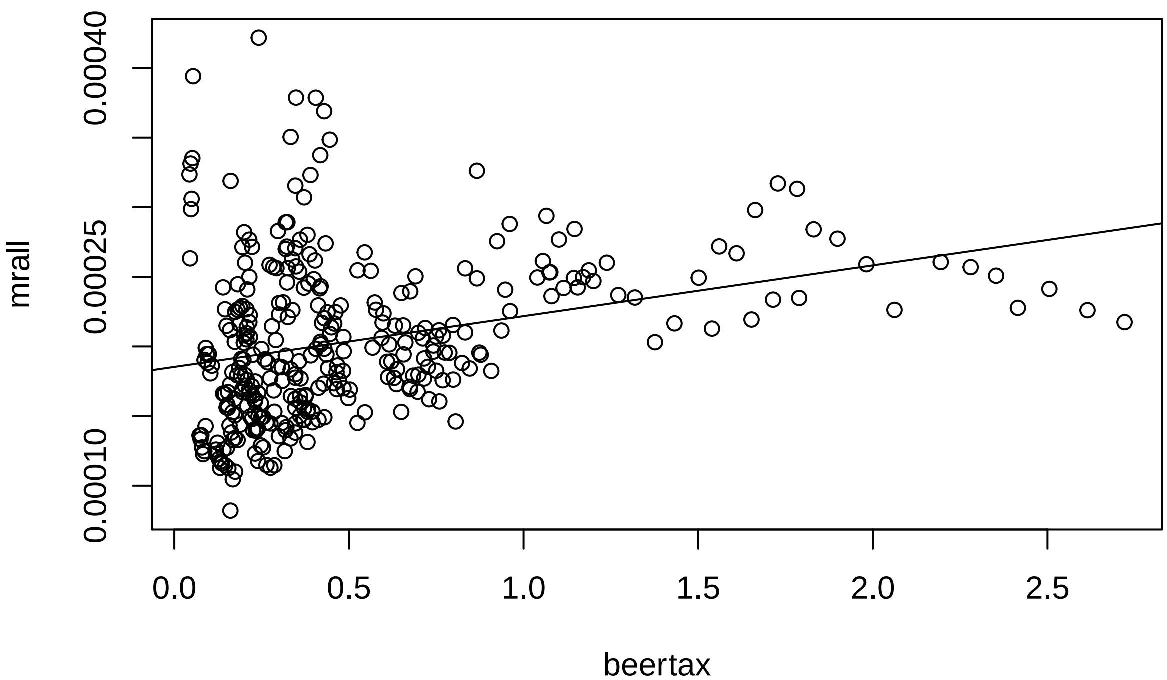
First, we estimate a ols regression between fatality rate and beer tax ignoring the panel data structure of our data set. This is usually called "pooled ols".

```r
# Model using ols
ols = lm(mrall ~ beertax)
summary(ols)
```

```
##
## Call:
## lm(formula = mrall ~ beertax)
```

```
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.091e-04 -3.777e-05 -9.436e-06  2.855e-05  2.276e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.853e-04  4.357e-06  42.539  < 2e-16 ***
## beertax     3.646e-05  6.217e-06   5.865 1.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.437e-05 on 334 degrees of freedom
## Multiple R-squared:  0.09336,    Adjusted R-squared:  0.09065
## F-statistic: 34.39 on 1 and 334 DF,  p-value: 1.082e-08
```

```r
plot(beertax,mrall)
abline(ols)
```



We find a positive and statistically significant coefficient. Now let's do a regression that includes state fixed effects using n-1 binary regressors (this estimator is also called least square dummy variable [LSDV] estimator):

```r
# Fixed effects using n-1 binary regressors
fe1 = lm(mrall ~ beertax + factor(state) - 1, data=pd)
summary(fe1)
```

```
##
## Call:
## lm(formula = mrall ~ beertax + factor(state) - 1, data = pd)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.870e-05 -8.284e-06 -1.270e-07  7.955e-06  8.978e-05
```

```
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## beertax          -6.559e-05  1.879e-05  -3.491 0.000556 ***
## factor(state)AL  3.478e-04  3.134e-05  11.098  < 2e-16 ***
## factor(state)AZ  2.910e-04  9.254e-06  31.445  < 2e-16 ***
## factor(state)AR  2.823e-04  1.321e-05  21.364  < 2e-16 ***
## factor(state)CA  1.968e-04  7.401e-06  26.594  < 2e-16 ***
## factor(state)CO  1.993e-04  8.037e-06  24.802  < 2e-16 ***
## factor(state)CT  1.615e-04  8.391e-06  19.251  < 2e-16 ***
## factor(state)DE  2.170e-04  7.746e-06  28.016  < 2e-16 ***
## factor(state)FL  3.210e-04  2.215e-05  14.489  < 2e-16 ***
## factor(state)GA  4.002e-04  4.640e-05   8.625 4.43e-16 ***
## factor(state)ID  2.809e-04  9.877e-06  28.437  < 2e-16 ***
## factor(state)IL  1.516e-04  7.848e-06  19.318  < 2e-16 ***
## factor(state)IN  2.016e-04  8.867e-06  22.736  < 2e-16 ***
## factor(state)IA  1.934e-04  1.022e-05  18.918  < 2e-16 ***
## factor(state)KS  2.254e-04  1.086e-05  20.753  < 2e-16 ***
## factor(state)KY  2.260e-04  8.046e-06  28.089  < 2e-16 ***
## factor(state)LA  2.631e-04  1.627e-05  16.171  < 2e-16 ***
## factor(state)ME  2.370e-04  1.601e-05  14.805  < 2e-16 ***
## factor(state)MD  1.771e-04  8.246e-06  21.480  < 2e-16 ***
## factor(state)MA  1.368e-04  8.648e-06  15.818  < 2e-16 ***
## factor(state)MI  1.993e-04  1.166e-05  17.089  < 2e-16 ***
## factor(state)MN  1.580e-04  9.363e-06  16.880  < 2e-16 ***
## factor(state)MS  3.449e-04  2.094e-05  16.472  < 2e-16 ***
## factor(state)MO  2.181e-04  9.252e-06  23.576  < 2e-16 ***
## factor(state)MT  3.117e-04  9.441e-06  33.017  < 2e-16 ***
## factor(state)NE  1.955e-04  1.055e-05  18.534  < 2e-16 ***
## factor(state)NV  2.877e-04  8.106e-06  35.492  < 2e-16 ***
## factor(state)NH  2.223e-04  1.411e-05  15.751  < 2e-16 ***
## factor(state)NJ  1.372e-04  7.333e-06  18.709  < 2e-16 ***
## factor(state)NM  3.904e-04  1.015e-05  38.449  < 2e-16 ***
## factor(state)NY  1.291e-04  7.563e-06  17.070  < 2e-16 ***
## factor(state)NC  3.187e-04  2.517e-05  12.661  < 2e-16 ***
## factor(state)ND  1.854e-04  1.019e-05  18.191  < 2e-16 ***
## factor(state)OH  1.803e-04  1.019e-05  17.691  < 2e-16 ***
## factor(state)OK  2.933e-04  1.843e-05  15.913  < 2e-16 ***
## factor(state)OR  2.310e-04  8.117e-06  28.453  < 2e-16 ***
## factor(state)PA  1.710e-04  8.648e-06  19.776  < 2e-16 ***
## factor(state)RI  1.213e-04  7.753e-06  15.640  < 2e-16 ***
## factor(state)SC  4.035e-04  3.548e-05  11.372  < 2e-16 ***
## factor(state)SD  2.474e-04  1.412e-05  17.519  < 2e-16 ***
## factor(state)TN  2.602e-04  9.162e-06  28.398  < 2e-16 ***
## factor(state)TX  2.560e-04  1.085e-05  23.589  < 2e-16 ***
## factor(state)UT  2.314e-04  1.545e-05  14.972  < 2e-16 ***
## factor(state)VT  2.512e-04  1.397e-05  17.975  < 2e-16 ***
## factor(state)VA  2.187e-04  1.466e-05  14.917  < 2e-16 ***
## factor(state)WA  1.818e-04  8.233e-06  22.084  < 2e-16 ***
## factor(state)WV  2.581e-04  1.077e-05  23.971  < 2e-16 ***
## factor(state)WI  1.718e-04  7.746e-06  22.185  < 2e-16 ***
## factor(state)WY  3.249e-04  7.233e-06  44.922  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.899e-05 on 287 degrees of freedom
## Multiple R-squared:  0.9931, Adjusted R-squared:  0.992
## F-statistic: 847.8 on 49 and 287 DF,  p-value: < 2.2e-16
```

We can also estimate this model using the within transformation. In other words, we transform the data applying entity de-meaning and then run OLS on the transformed model:

```
# Fixed effects regression (index refers to the subscripts it)
fe2 = plm(mrall~beertax,data=pd,
  index=c("state", "year"),model="within")
summary(fe2)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = mrall ~ beertax, data = pd, model = "within", index = c("state",
##     "year"))
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -5.87e-05 -8.28e-06 -1.27e-07  7.95e-06  8.98e-05
##
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)
## beertax -6.5587e-05  1.8785e-05 -3.4915 0.000556 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.0785e-07
## Residual Sum of Squares: 1.0345e-07
## R-Squared:      0.040745
## Adj. R-Squared: -0.11969
## F-statistic: 12.1904 on 1 and 287 DF, p-value: 0.00055597
```

Note that we now get a negative and statistically significant coefficient for beer taxes.

Suppose we are not sure whether the model should have fixed effects. In this case we can test whether the fixed effects are statistically significant with a F test:

```
# Testing for fixed effects, null: OLS better than fixed effects model
pFtest(fe2, ols)
```

```
##
##  F test for individual effects
##
## data:  mrall ~ beertax
## F = 52.179, df1 = 47, df2 = 287, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

We reject the null hypothesis that all the fixed effects are equal to zero (p-value less than 0.01).

Similarly we can estimate a regression with entity and time fixed effects (adding T-1 binary regressors and applying entity de-meaning):

```r
library(foreign); library(plm)
a =
"http://fmwww.bc.edu/ec-p/data/stockwatson/fatality.dta"
pd = read.dta(a)
# by generating the dummy varibles by year:
fe3 = plm(mrall ~ beertax + factor(year) - 1,data=pd,
  index=c("state", "year"),model="within")
summary(fe3)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = mrall ~ beertax + factor(year) - 1, data = pd,
##     model = "within", index = c("state", "year"))
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -5.96e-05 -8.10e-06  1.43e-07  8.23e-06  8.39e-05
##
## Coefficients: (1 dropped because of singularities)
##                     Estimate  Std. Error t-value Pr(>|t|)
## beertax          -6.3998e-05  1.9738e-05 -3.2424 0.001328 **
## factor(year)1982  5.1804e-06  3.9623e-06  1.3074 0.192145
## factor(year)1983 -2.8099e-06  3.9731e-06 -0.7072 0.480005
## factor(year)1984 -2.0617e-06  3.9590e-06 -0.5208 0.602945
## factor(year)1985 -7.2173e-06  3.9045e-06 -1.8485 0.065587 .
## factor(year)1986  1.3939e-06  3.8772e-06  0.3595 0.719474
## factor(year)1987  9.0176e-08  3.8470e-06  0.0234 0.981316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.0785e-07
## Residual Sum of Squares: 9.9193e-08
## R-Squared:      0.080251
## Adj. R-Squared: -0.096497
## F-statistic: 3.50261 on 7 and 281 DF, p-value: 0.0012789
```

We can test the time fixed effects jointly with a F test comparing the model with only entity fixed effects versus the model with both time and entity fixed effects:

```r
# Testing for time effects, null: fixed effects better than fixed effects plus time effects model
pFtest(fe3,fe2)
```

```
##
##  F test for individual effects
##
## data:  mrall ~ beertax + factor(year) - 1
## F = 2.0117, df1 = 6, df2 = 281, p-value = 0.0642
## alternative hypothesis: significant effects
```

We fail to reject that the T-1 time fixed effects are all equal to zero at the standard 5% of confidence (p-value greater than 0.05).

Note that the previous models use OLS standard errors, we can compute standard errors robust to hereroskedasticity and autocorrelation using:

```r
fe4 = plm(mrall~beertax,data=pd,
  index=c("state", "year"),model="within")
summary(fe4)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = mrall ~ beertax, data = pd, model = "within", index = c("state",
##     "year"))
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -5.87e-05 -8.28e-06 -1.27e-07  7.95e-06  8.98e-05
##
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)
## beertax -6.5587e-05  1.8785e-05 -3.4915 0.000556 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.0785e-07
## Residual Sum of Squares: 1.0345e-07
## R-Squared:      0.040745
## Adj. R-Squared: -0.11969
## F-statistic: 12.1904 on 1 and 287 DF, p-value: 0.00055597
```

```r
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
coeftest(fe4, vcovHC(fe4,type="HC0",cluster="group"))
```

```
##
## t test of coefficients:
##
##            Estimate  Std. Error t value Pr(>|t|)
## beertax -6.5587e-05  2.8837e-05 -2.2744  0.02368 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```