Panel Data and DID

Jing Bu

12/27/2018

目录

1 Panel Data

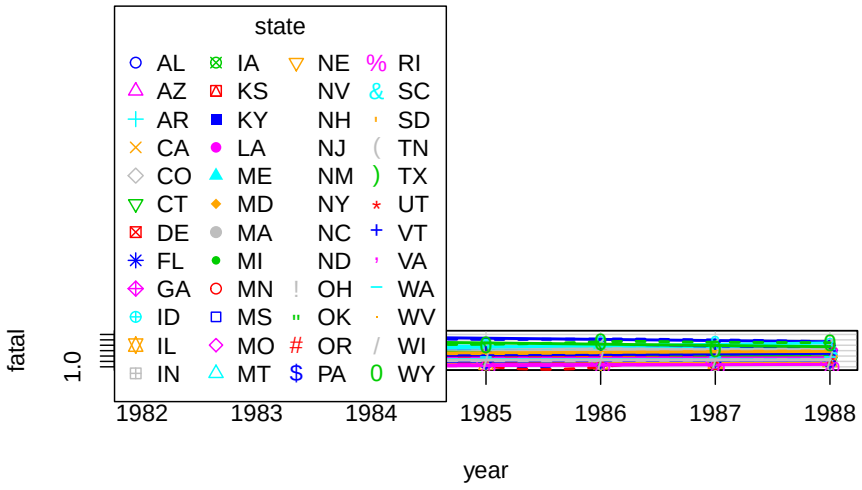
- Panel data (also known as longitudinal or cross-sectional time-series data) is a dataset in which the behavior of entities are observed across time. These entities could be states, companies, individuals, countries, etc.

1.1 Exploring panel data

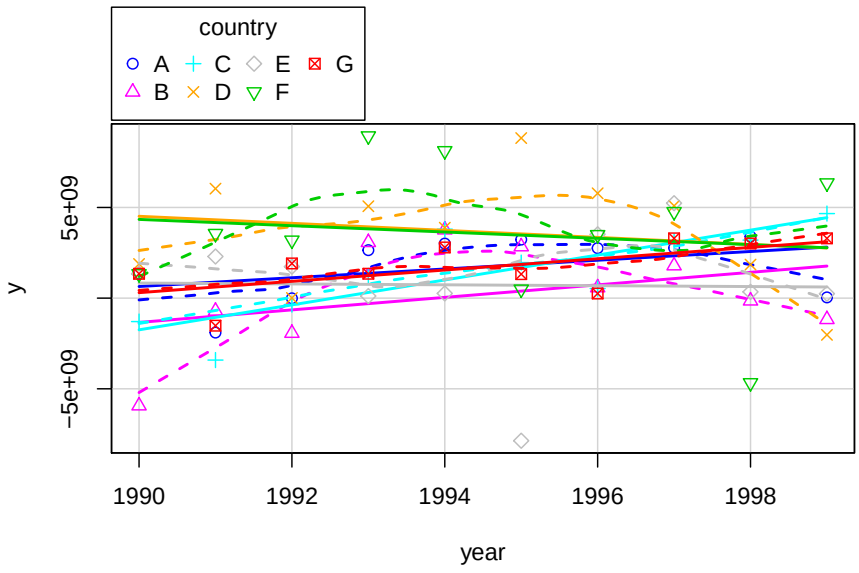
```
library("foreign")
Panel <-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assista
#install.packages("car")
library("car")
```

```
## Loading required package: carData
```

```
scatterplot(fatal~year|state,data=Panel)
```




```
library("foreign")
library("car")
Panel1 <-read.dta("http://dss.princeton.edu/training/Panel101.dta")
scatterplot(y~year|country, data=Panel1)
```



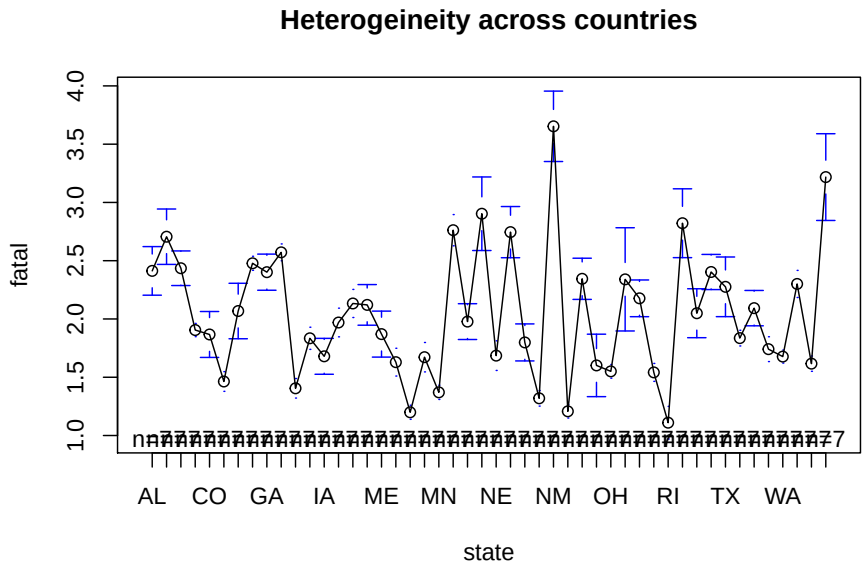
1.2 Fixed Effects Model

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

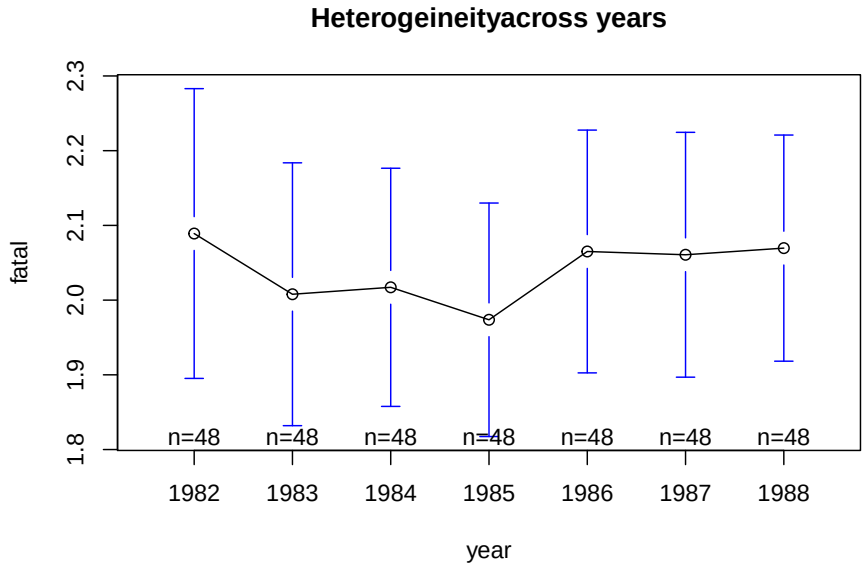
```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess
```

```
plotmeans(fatal ~ state, main="Heterogeineity across countries", data=Panel)
```




```
library("gplots")
plotmeans(fatal ~ year, main="Heterogeineityacross years", data=Panel)
```



1.2.1 OLS

```
ols<-lm(fatal ~ beertax+spircons+unrate+perinck, data=Panel)
summary(ols)
```

```
##
## Call:
## lm(formula = fatal ~ beertax + spircons + unrate + perinck, data = Panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22581 -0.35100 -0.05238  0.27829  1.94364
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.11867    0.29700  13.868  < 2e-16 ***
## beertax      0.09720    0.06155   1.579 0.115256
## spircons     0.16235    0.04325   3.754 0.000206 ***
## unrate      -0.02910    0.01272  -2.289 0.022731 *
## perinck     -0.15843    0.01699  -9.327  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4793 on 331 degrees of freedom
## Multiple R-squared:  0.3019, Adjusted R-squared:  0.2934
## F-statistic: 35.78 on 4 and 331 DF,  p-value: < 2.2e-16
```

- Regular OLS regression does not consider heterogeneity across groups or time

1.2.2 Least squares dummy variable model

```
fixed.dum <-lm(fatal ~ beertax+spircons+unrate+perinck +factor(state), data=Panel)
summary(fixed.dum)
```

```
##
## Call:
## lm(formula = fatal ~ beertax + spircons + unrate + perinck +
##     factor(state), data = Panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44379 -0.07923  0.00079  0.06761  0.56862
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.269226   0.450406   2.818 0.005172 **
## beertax          -0.484073   0.162511  -2.979 0.003145 **
## spircons          0.816965   0.079212  10.314  < 2e-16 ***
## unrate           -0.029050   0.009027  -3.218 0.001441 **
## perinck           0.104710   0.020599   5.083 6.74e-07 ***
## factor(state)AZ -1.146089   0.237083  -4.834 2.19e-06 ***
## factor(state)AR -0.342280   0.188359  -1.817 0.070246 .
## factor(state)CA -2.478906   0.287670  -8.617 4.86e-16 ***
## factor(state)CO -2.359016   0.264732  -8.911  < 2e-16 ***
## factor(state)CT -3.517465   0.301913 -11.651  < 2e-16 ***
## factor(state)DE -2.621819   0.286741  -9.143  < 2e-16 ***
## factor(state)FL -1.484055   0.165968  -8.942  < 2e-16 ***
## factor(state)GA -0.419331   0.174534  -2.403 0.016922 *
## factor(state)ID -0.475246   0.220535  -2.155 0.032005 *
## factor(state)IL -2.662620   0.268618  -9.912  < 2e-16 ***
## factor(state)IN -1.551228   0.232698  -6.666 1.36e-10 ***
## factor(state)IA -1.512083   0.213137  -7.094 1.04e-11 ***
## factor(state)KS -1.425295   0.206859  -6.890 3.59e-11 ***
## factor(state)KY -1.008736   0.246350  -4.095 5.52e-05 ***
```

```
## factor(state)LA -0.990443   0.165592  -5.981 6.64e-09 ***
## factor(state)ME -1.728957   0.173953  -9.939  < 2e-16 ***
## factor(state)MD -2.971806   0.277647 -10.704  < 2e-16 ***
## factor(state)MA -3.542156   0.279881 -12.656  < 2e-16 ***
## factor(state)MI -2.039022   0.220826  -9.234  < 2e-16 ***
## factor(state)MN -2.726702   0.242581 -11.240  < 2e-16 ***
## factor(state)MS  0.162334   0.130321   1.246 0.213920
## factor(state)MO -1.473051   0.227292  -6.481 4.02e-10 ***
## factor(state)MT -0.600398   0.229280  -2.619 0.009302 **
## factor(state)NE -1.812884   0.212276  -8.540 8.27e-16 ***
## factor(state)NV -3.433805   0.384945  -8.920  < 2e-16 ***
## factor(state)NH -4.226736   0.340310 -12.420  < 2e-16 ***
## factor(state)NJ -3.425003   0.306210 -11.185  < 2e-16 ***
## factor(state)NM  0.453740   0.218171   2.080 0.038446 *
## factor(state)NY -3.146312   0.279708 -11.249  < 2e-16 ***
## factor(state)NC -0.638759   0.105299  -6.066 4.17e-09 ***
## factor(state)ND -2.124115   0.224007  -9.482  < 2e-16 ***
## factor(state)OH -1.689753   0.217130  -7.782 1.33e-13 ***
## factor(state)OK -0.684285   0.142350  -4.807 2.49e-06 ***
## factor(state)OR -1.388066   0.246994  -5.620 4.56e-08 ***
## factor(state)PA -1.920440   0.236910  -8.106 1.57e-14 ***
## factor(state)RI -3.097237   0.265756 -11.654  < 2e-16 ***
## factor(state)SC -0.025195   0.105572  -0.239 0.811547
## factor(state)SD -1.345773   0.185628  -7.250 3.97e-12 ***
## factor(state)TN -0.780764   0.228633  -3.415 0.000731 ***
## factor(state)TX -1.117104   0.209679  -5.328 2.03e-07 ***
## factor(state)UT -0.766329   0.171935  -4.457 1.20e-05 ***
## factor(state)VT -1.844306   0.199658  -9.237  < 2e-16 ***
## factor(state)VA -1.891904   0.182297 -10.378  < 2e-16 ***
## factor(state)WA -2.176212   0.254450  -8.553 7.59e-16 ***
## factor(state)WV -0.214221   0.217052  -0.987 0.324504
## factor(state)WI -2.416260   0.263501  -9.170  < 2e-16 ***
## factor(state)WY -0.718353   0.274840  -2.614 0.009434 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1568 on 284 degrees of freedom
## Multiple R-squared:  0.9359, Adjusted R-squared:  0.9244
## F-statistic:  81.3 on 51 and 284 DF,  p-value: < 2.2e-16
```

1.2.3 Fixed effects: nentity-specific intercepts (using plm)

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

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

```
fixed <-plm(fatal ~ beertax+spircons+unrate+perinck, data=Panel, index=c("state", "year
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = fatal ~ beertax + spircons + unrate + perinck,
##     data = Panel, model = "within", index = c("state", "year"))
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.44378892 -0.07922880  0.00078846  0.06761301  0.56861719
##
## Coefficients:
##            Estimate Std. Error t-value  Pr(>|t|)
## beertax  -0.4840728  0.1625106 -2.9787  0.003145 **
## spircons  0.8169652  0.0792118 10.3137 < 2.2e-16 ***
## unrate   -0.0290499  0.0090274 -3.2180  0.001441 **
```

```
## perinck   0.1047103  0.0205986  5.0834 6.738e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10.785
## Residual Sum of Squares: 6.9816
## R-Squared:      0.35265
## Adj. R-Squared: 0.2364
## F-statistic: 38.6774 on 4 and 284 DF, p-value: < 2.22e-16
```

1.2.4 Test: fixed or ols?

```
pFtest(fixed, ols)
```

```
##
##  F test for individual effects
##
## data:  fatal ~ beertax + spircons + unrate + perinck
## F = 59.768, df1 = 47, df2 = 284, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

2 Difference in differences

2.1 Getting sample data

```
library("foreign")
mydata <-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assist
```

- Create an interaction between time and treated. We will call this interaction *did*.
- $did = t : treated$

```r
didreg <- lm(fte ~ treated + t + t:treated, data = mydata)
summary(didreg)
```

```
##
## Call:
## lm(formula = fte ~ treated + t + t:treated, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.573  -6.292  -1.323   4.427  62.935
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.949      1.019  19.569   <2e-16 ***
## treated       -2.884      1.135  -2.541   0.0112 *
## t             -2.407      1.446  -1.664   0.0965 .
## treated:t      2.914      1.611   1.809   0.0708 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.003 on 797 degrees of freedom
## Multiple R-squared:  0.008046,   Adjusted R-squared:  0.004312
## F-statistic: 2.155 on 3 and 797 DF,  p-value: 0.09195
```

- The coefficient for *did* is the differences-in-differences estimator.The effect is significant at 10% with the treatment having a negative effect.