CLASS QUESTIONS FOR WEEK 5 of EC402, LENT SEMESTER

YULIA

2/6/2018

CLASS QUESTIONS FOR WEEK 5 of EC402, LENT SEMESTER

Today we are going to look at the determinants of crime in North Carolina. If Gary Becker is right about crime, economic variables and the ease of commiting crimes should be able to predict how much crime happens. We will consider factors such as population density, percentage of minorities, wages and tax revenues… The problem is that places with higher crime on average are probably systematically different to places with lower crime on average, in ways correlated to these variables –> OVB Let's use FEs to deal with that problem.

We highly suggest you use the "plm" package rather than trying to do this manually. We will use the "Crime" dataset from the plm package. Please look up the documentation of the data here https://cran.r-project.org/web/packages/plm/plm.pdf (Search for "Crime" in the text to see the dataset)

clear the global workspace

```
rm(list=ls())
```

Load and install packages (toggles are there for future reference if running code on remote servers etc):

```
#installation_needed  <- TRUE
#loading_needed <- TRUE

#package_list <- c('foreign', 'xtable', 'plm','gmm', 'AER','stargazer','readstata13')
#if(installation_needed){install.packages(package_list, repos='http://cran.us.r-project.org')}
#if(loading_needed){lapply(package_list, require, character.only = TRUE)}
```

Load the data

```
library(plm)
```

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

```
data("Crime")
```

Tell R that this is panel data

```
data <- pdata.frame(Crime, index = c("county","year"), drop.index = FALSE)
attach(data)
summary(data)
```

```
##      county     year        crmrte              prbarr
##  1      :  7   81:90   Min.   :0.001812   Min.   :0.05882
##  3      :  7   82:90   1st Qu.:0.018352   1st Qu.:0.21790
```

```
## 5      :  7   83:90   Median :0.028441   Median :0.27824
## 7      :  7   84:90   Mean   :0.031588   Mean   :0.30737
## 9      :  7   85:90   3rd Qu.:0.038406   3rd Qu.:0.35252
## 11     :  7   86:90   Max.   :0.163835   Max.   :2.75000
## (Other):588   87:90
##    prbconv           prbpris           avgsen           polpc
## Min.   : 0.06838   Min.   :0.1489   Min.   : 4.220   Min.   :0.0004585
## 1st Qu.: 0.34769   1st Qu.:0.3744   1st Qu.: 7.160   1st Qu.:0.0011913
## Median : 0.47437   Median :0.4286   Median : 8.495   Median :0.0014506
## Mean   : 0.68862   Mean   :0.4255   Mean   : 8.955   Mean   :0.0019168
## 3rd Qu.: 0.63560   3rd Qu.:0.4832   3rd Qu.:10.197   3rd Qu.:0.0018033
## Max.   :37.00000   Max.   :0.6786   Max.   :25.830   Max.   :0.0355781
##
##    density           taxpc           region     smsa
## Min.   :0.1977   Min.   : 14.30   other  :245   no :574
## 1st Qu.:0.5329   1st Qu.: 23.43   west   :147   yes: 56
## Median :0.9526   Median : 27.79   central:238
## Mean   :1.3861   Mean   : 30.24
## 3rd Qu.:1.5078   3rd Qu.: 33.27
## Max.   :8.8277   Max.   :119.76
##
##     pctmin           wcon              wtuc              wtrd
## Min.   : 1.284   Min.   :  65.62   Min.   :  28.86   Min.   :  16.87
## 1st Qu.:10.005   1st Qu.: 201.66   1st Qu.: 317.60   1st Qu.: 168.05
## Median :24.852   Median : 236.46   Median : 358.20   Median : 185.48
## Mean   :25.713   Mean   : 245.67   Mean   : 406.10   Mean   : 192.82
## 3rd Qu.:38.223   3rd Qu.: 269.69   3rd Qu.: 411.02   3rd Qu.: 204.82
## Max.   :64.348   Max.   :2324.60   Max.   :3041.96   Max.   :2242.75
##
##      wfir              wser              wmfg            wfed
## Min.   :  3.516   Min.   :   1.844   Min.   :101.8   Min.   :255.4
## 1st Qu.:235.705   1st Qu.: 191.319   1st Qu.:234.0   1st Qu.:361.5
## Median :264.423   Median : 216.475   Median :271.6   Median :404.0
## Mean   :272.059   Mean   : 224.671   Mean   :285.2   Mean   :403.9
## 3rd Qu.:302.440   3rd Qu.: 247.155   3rd Qu.:320.0   3rd Qu.:444.6
## Max.   :509.466   Max.   :2177.068   Max.   :646.9   Max.   :598.0
##
##      wsta            wloc            mix               pctymle
## Min.   :173.0   Min.   :163.6   Min.   :0.002457   Min.   :0.06216
## 1st Qu.:258.2   1st Qu.:226.8   1st Qu.:0.075324   1st Qu.:0.07859
## Median :289.4   Median :253.1   Median :0.102089   Median :0.08316
## Mean   :296.9   Mean   :258.0   Mean   :0.139396   Mean   :0.08897
## 3rd Qu.:331.5   3rd Qu.:289.3   3rd Qu.:0.149009   3rd Qu.:0.08919
## Max.   :548.0   Max.   :388.1   Max.   :4.000000   Max.   :0.27436
##
```

Define the linear model we will use throughout

```
linear_model_crime <- crmrte ~ density + taxpc + wcon + pctmin
```

Q1: Fit a naive OLS regression to the data, assuming all observations are iid.

```r
naive_ols <- plm(linear_model_crime, data)
summary(naive_ols)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = linear_model_crime, data = data)
##
## Balanced Panel: n = 90, T = 7, N = 630
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.042300 -0.002580 -0.000317  0.002230  0.093800
##
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)
## density -1.4823e-03  5.2269e-03 -0.2836  0.77683
## taxpc    8.8219e-05  4.5217e-05  1.9510  0.05157 .
## wcon    -2.0597e-06  2.6270e-06 -0.7841  0.43335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.026722
## Residual Sum of Squares: 0.026491
## R-Squared:      0.0086426
## Adj. R-Squared: -0.1612
## F-statistic: 1.5605 on 3 and 537 DF, p-value: 0.19799
```

Q2: Fit a classical Random Effects model, implementing FGLS.

```r
RE_model <- plm(linear_model_crime, data, model ="random")
summary(RE_model)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = linear_model_crime, data = data, model = "random")
##
## Balanced Panel: n = 90, T = 7, N = 630
##
## Effects:
##                     var   std.dev share
## idiosyncratic 4.933e-05 7.024e-03 0.313
## individual    1.080e-04 1.039e-02 0.687
## theta: 0.7526
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.031500 -0.003180 -0.000673  0.002380  0.105000
##
## Coefficients:
```

```
##                  Estimate  Std. Error t-value  Pr(>|t|)    
## (Intercept)  1.2570e-02  2.5927e-03  4.8482 1.574e-06 ***
## density      8.6440e-03  7.8850e-04 10.9627 < 2.2e-16 ***
## taxpc        5.5558e-05  3.7754e-05  1.4715 0.1416460    
## wcon        -2.4950e-06  2.6012e-06 -0.9592 0.3378491    
## pctmin       2.3215e-04  6.7316e-05  3.4487 0.0006012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Total Sum of Squares:    0.037732
## Residual Sum of Squares: 0.030934
## R-Squared:      0.18018
## Adj. R-Squared: 0.17494
## F-statistic: 34.3411 on 4 and 625 DF, p-value: < 2.22e-16
```

Q3: Fit a Fixed Effects Model with entity fixed effects using the within transformation. Use variotion within each cross-sectional observation

NB!
ai is time-invariant, individual specific, unobserved effect on the level of yit.
ai is reffered to as fixed effect ?V fixed over time.
ai is reffered to as unobserved heterogeneity, or individual heterogeneity.

```r
fixed_effects_fit <- plm(linear_model_crime, data, model="within", effect = "individual", index = c("co
summary(fixed_effects_fit)
```

```
## Oneway (individual) effect Within Model
## 
## Call:
## plm(formula = linear_model_crime, data = data, effect = "individual", 
##     model = "within", index = c("county", "year"))
## 
## Balanced Panel: n = 90, T = 7, N = 630
## 
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max. 
## -0.042300 -0.002580 -0.000317  0.002230  0.093800 
## 
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)  
## density -1.4823e-03  5.2269e-03 -0.2836  0.77683  
## taxpc    8.8219e-05  4.5217e-05  1.9510  0.05157 .
## wcon    -2.0597e-06  2.6270e-06 -0.7841  0.43335  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Total Sum of Squares:    0.026722
## Residual Sum of Squares: 0.026491
## R-Squared:      0.0086426
## Adj. R-Squared: -0.1612
## F-statistic: 1.5605 on 3 and 537 DF, p-value: 0.19799
```

Q4: Comment on what happened to the variable "pcmin". R deleted it, but why?

We have transformed the model to the demeaned model by subtracting means from all observations. The mean of pcmin which is constant over time is just itself so the constant disappears after the transformation. Pcmin is constant within entity, not varying over time, when you subtract the mean it gets cancelled out.

Q5: Fit a Fixed Effects Model with entity and time fixed effects.

```
twoways_fixed_effects_fit <- plm(linear_model_crime, data, model="within", effect = "twoways", index = c
summary(twoways_fixed_effects_fit)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = linear_model_crime, data = data, effect = "twoways",
##     model = "within", index = c("county", "year"))
##
## Balanced Panel: n = 90, T = 7, N = 630
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -4.37e-02 -2.16e-03 -9.29e-05  1.94e-03  9.35e-02
##
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)
## density  9.0867e-04  5.3391e-03  0.1702  0.86492
## taxpc    1.0375e-04  5.2807e-05  1.9647  0.04997 *
## wcon    -7.1573e-07  2.6270e-06 -0.2724  0.78538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.025341
## Residual Sum of Squares: 0.025131
## R-Squared:      0.0082799
## Adj. R-Squared: -0.17475
## F-statistic: 1.47778 on 3 and 531 DF, p-value: 0.21962
```

Q6: Fit a First Differences Model.

```
FD_model <- plm(linear_model_crime, data, model="fd", index = c("county", "year"))
summary(FD_model)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = linear_model_crime, data = data, model = "fd",
##     index = c("county", "year"))
##
## Balanced Panel: n = 90, T = 7, N = 630
## Observations used in estimation: 540
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.132000 -0.002480 -0.000131 -0.000055  0.002370  0.103000
##
## Coefficients:
```

```
##            Estimate  Std. Error t-value Pr(>|t|)
## density  8.1304e-03  1.2227e-02  0.6650   0.5064
## taxpc    2.5369e-05  5.4849e-05  0.4625   0.6439
## wcon    -9.9526e-07  2.2455e-06 -0.4432   0.6578
##
## Total Sum of Squares:    0.043448
## Residual Sum of Squares: 0.043375
## R-Squared:      0.0017212
## Adj. R-Squared: -0.0019967
## F-statistic: 0.450121 on 2 and 537 DF, p-value: 0.63779
```

Intro to Standard Errors!

Q7: Compute White-style standard errors which are robust to heteroskedasticity within and across entities, but no autocorrelation. (Warning: this is probably bad, we will see why in the next lecture).

```r
HC_coefs <- vcovHC(fixed_effects_fit, method = "white1")

white_ses_state <- sqrt(diag(HC_coefs))

print(white_ses_state)
```

```
##      density        taxpc         wcon
## 4.580978e-03 5.275769e-05 5.555445e-07
```

Q8: Compute White-style robust to heteroskedasticity across entities only but no autocorrelation. (Warning: this is probably bad, we will see why in lectures.)

```r
HC_coefs <- vcovHC(fixed_effects_fit, method = "white2")

white_ses_state <- sqrt(diag(HC_coefs))

print(white_ses_state)
```

```
##      density        taxpc         wcon
## 4.015131e-03 4.923166e-05 7.088393e-07
```

Q9. Compute Clustered White-style standard errors, clustering at the entity level and allowing for arbitrary serial correlation, as in Arellano 1987.

```r
HCV_coefs <- vcovHC(fixed_effects_fit, method = "arellano", cluster = "group")

clustered_ses_state <- sqrt(diag(HCV_coefs))

print(clustered_ses_state)
```

```
##      density        taxpc         wcon
## 6.387680e-03 8.658943e-05 5.275221e-07
```

Q10. Compare the relative magnitudes of the standard errors and think about what's going on here.
Different strcutre on error terms gives different significance and wrong strucuture may produce wrong inference.

Notice how the different assumptions and structures have lead to different standard errors!