Practical Questions

YULIA

3/3/2018

PRACTICAL QUESTIONS Gun Control.

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

# clear the global workspace

rm(list=ls())
library(AER)
```

```
## Loading required package: car

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

## Loading required package: sandwich

## Loading required package: survival
```

```
library(plm)
```

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

```
data("Guns")
```

3.1 PART 1

For this part we will focus on the coefficients and not the standard errors
(you may use the default OLS standard errors assuming homoskedasticity for this part only).

1. Make a table that summarizes the basic properties of all the variables in the data set.

```r
mydata <- pdata.frame(Guns,index = c("state","year"), drop.index = FALSE)
#table(index(mydata), useNA = "ifany")
summary(mydata)
```

```
##       year        violent          murder          robbery
##  1977   : 51   Min.   :  47.0   Min.   : 0.200   Min.   :   6.4
##  1978   : 51   1st Qu.: 283.1   1st Qu.: 3.700   1st Qu.:  71.1
##  1979   : 51   Median : 443.0   Median : 6.400   Median : 124.1
##  1980   : 51   Mean   : 503.1   Mean   : 7.665   Mean   : 161.8
##  1981   : 51   3rd Qu.: 650.9   3rd Qu.: 9.800   3rd Qu.: 192.7
##  1982   : 51   Max.   :2921.8   Max.   :80.600   Max.   :1635.1
##  (Other):867
##    prisoners          afam              cauc            male
##  Min.   :  19.0   Min.   : 0.2482   Min.   :21.78   Min.   :12.21
##  1st Qu.: 114.0   1st Qu.: 2.2022   1st Qu.:59.94   1st Qu.:14.65
##  Median : 187.0   Median : 4.0262   Median :65.06   Median :15.90
##  Mean   : 226.6   Mean   : 5.3362   Mean   :62.95   Mean   :16.08
##  3rd Qu.: 291.0   3rd Qu.: 6.8507   3rd Qu.:69.20   3rd Qu.:17.53
##  Max.   :1913.0   Max.   :26.9796   Max.   :76.53   Max.   :22.35
##
##    population          income         density               state
##  Min.   : 0.4027   Min.   : 8555   Min.   : 0.000707   Alabama   :  23
##  1st Qu.: 1.1877   1st Qu.:11935   1st Qu.: 0.031911   Alaska    :  23
##  Median : 3.2713   Median :13402   Median : 0.081569   Arizona   :  23
##  Mean   : 4.8163   Mean   :13725   Mean   : 0.352038   Arkansas  :  23
##  3rd Qu.: 5.6856   3rd Qu.:15271   3rd Qu.: 0.177718   California:  23
##  Max.   :33.1451   Max.   :23647   Max.   :11.102120   Colorado  :  23
##                                                        (Other)   :1035
##   law
##  no :888
##  yes:285
##
##
##
##
##
```

2. Perform a pooled OLS analysis, ignoring the panel properties of the data set and just treating all the data as iid. Report the results.

Our main explanatory variable of interest is whether a state has a shall carry law, that is, a law that allows basically anyone to get a license to carry a gun around with them. So these are laws that encourage more guns on the street, not fewer. For every regression we will include the shall-carry law dummy variable as well as controls for the gender and racial makeup of the state's population, the population density, the state average income, and the states's prison population.

```r
linear_model <- violent ~ law  + prisoners + male + density + cauc + afam + population + income
summary(linear_model)
```

```
##  Length    Class     Mode
##       3 formula     call
```

```r
#linear_model <- lm(violent ~. - murder -robbery - state - year, data = mydata)
#summary(linear_model)

naive_ols <- plm(linear_model, mydata, model ="pooling", index = c("state", "year"))
summary(naive_ols)
```

```
## Pooling Model
##
## Call:
## plm(formula = linear_model, data = mydata, model = "pooling",
##     index = c("state", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -1040.0  -109.0   -27.6   101.0   659.0
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept) -1.7878e+02  2.2401e+02 -0.7981   0.42499
## lawyes      -9.2721e+01  1.3426e+01 -6.9062 8.168e-12 ***
## prisoners    8.1559e-01  4.4174e-02 18.4633 < 2.2e-16 ***
## male         9.3165e+00  4.4414e+00  2.0977   0.03615 *
## density      9.4669e+01  5.4284e+00 17.4394 < 2.2e-16 ***
## cauc         2.7024e+00  3.4536e+00  0.7825   0.43409
## afam         1.1292e+01  6.8644e+00  1.6450   0.10024
## population   1.8506e+01  1.0549e+00 17.5433 < 2.2e-16 ***
## income       1.2348e-03  3.2073e-03  0.3850   0.70032
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    130960000
## Residual Sum of Squares: 36184000
## R-Squared:      0.7237
## Adj. R-Squared: 0.7218
## F-statistic: 381.103 on 8 and 1164 DF, p-value: < 2.22e-16
```

3. Now perform the Feasible Generalized Least Squares analysis of this panel with "classical" random effects at the state level.

```r
RE_fit <- plm(linear_model, data = mydata, model ="random", index = c("state", "year"))
summary(RE_fit)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = linear_model, data = mydata, model = "random",
##     index = c("state", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
```

```
##
## Effects:
##                    var  std.dev share
## idiosyncratic  9703.24    98.51 0.354
## individual    17694.57   133.02 0.646
## theta: 0.8474
##
## Residuals:
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -609.00  -52.60   -7.29   44.60  843.00
##
## Coefficients:
##                 Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept) -54.7031402 231.1279019 -0.2367  0.812947
## lawyes      -37.1934528  11.6024998 -3.2056  0.001384 **
## prisoners     0.3931534   0.0397895  9.8808 < 2.2e-16 ***
## male         -9.6405919   3.6232652 -2.6607  0.007904 **
## density     104.7165122  15.9784049  6.5536 8.417e-11 ***
## cauc          8.0265665   3.0930812  2.5950  0.009578 **
## afam         20.6108259   7.1965011  2.8640  0.004258 **
## population   13.1585156   3.0747605  4.2795 2.027e-05 ***
## income       -0.0060252   0.0035148 -1.7142  0.086752 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    16310000
## Residual Sum of Squares: 12145000
## R-Squared:      0.25536
## Adj. R-Squared: 0.25024
## F-statistic: 49.8972 on 8 and 1164 DF, p-value: < 2.22e-16
```

4. Now perform a fixed effects analysis of the panel with fixed effects for the state, estimated by both the "within" estimator and the "first differences" estimator.
Do you see any major difference in the two estimators here?
Give some intuition for what you see.

Fixed Effects

```r
fixed_effects_fit <- plm(linear_model, mydata, model="within", effect = "individual", index = c("state"
summary(fixed_effects_fit)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = linear_model, data = mydata, effect = "individual",
##     model = "within", index = c("state", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -786.00  -47.70   -1.42   44.90  773.00
##
## Coefficients:
```

```
##               Estimate  Std. Error t-value  Pr(>|t|)
## lawyes     -1.8575e+01  1.1563e+01 -1.6064 0.1084713
## prisoners   9.6788e-02  5.7367e-02  1.6872 0.0918520 .
## male       -2.3864e+01  3.9248e+00 -6.0803 1.646e-09 ***
## density    -1.5595e+02  5.2118e+01 -2.9923 0.0028294 **
## cauc        1.0323e+01  3.1101e+00  3.3193 0.0009317 ***
## afam        1.2808e+01  1.0883e+01  1.1769 0.2394727
## population  1.2244e+01  5.3468e+00  2.2900 0.0222075 *
## income     -4.0667e-03  3.6211e-03 -1.1230 0.2616622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    13576000
## Residual Sum of Squares: 10809000
## R-Squared:      0.20381
## Adj. R-Squared: 0.16235
## F-statistic: 35.6447 on 8 and 1114 DF, p-value: < 2.22e-16
```

First Differences

```
FD_fit <- plm(linear_model, mydata, model="fd", index = c("state", "year"))
summary(FD_fit)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = linear_model, data = mydata, model = "fd", index = c("state",
##     "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
## Observations used in estimation: 1122
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -510.00  -23.40   -2.48   -2.25   20.00  345.00
##
## Coefficients:
##               Estimate  Std. Error t-value  Pr(>|t|)
## lawyes     -1.3044e+01  1.0453e+01 -1.2479 0.2123167
## prisoners  -2.1542e-01  6.9047e-02 -3.1199 0.0018556 **
## male       -3.9103e+01  8.0786e+00 -4.8404 1.479e-06 ***
## density    -3.3427e+02  6.8199e+01 -4.9014 1.093e-06 ***
## cauc        1.7752e+01  3.4819e+00  5.0983 4.024e-07 ***
## afam        6.2655e+01  1.8468e+01  3.3926 0.0007167 ***
## population  9.3195e+00  1.6610e+01  0.5611 0.5748676
## income     -1.2067e-02  4.2975e-03 -2.8079 0.0050736 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3168700
## Residual Sum of Squares: 2966900
## R-Squared:      0.065805
## Adj. R-Squared: 0.059935
## F-statistic: 10.8296 on 7 and 1114 DF, p-value: 2.887e-13
```

We can observe that R-Squared of the FE regression is larger since it preserves more information. The coefficients on male and population differ a lot in FE and FD.

5. Now perform fixed effects analysis of the panel using the "within" estimator with state and year effects.
Perform first differences analysis adding in dummies for the year. Do you see any major difference in the two procedures? Explain.

Fixed Effects

```
twoways_fixed_effects_fit <- plm(linear_model, mydata , model="within", effect = "twoways", index = c("s
summary(twoways_fixed_effects_fit)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = linear_model, data = mydata, effect = "twoways",
##     model = "within", index = c("state", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -636.000  -42.500   -0.099   41.500  642.000
##
## Coefficients:
##              Estimate  Std. Error t-value  Pr(>|t|)
## lawyes      -0.8391042  10.4030673 -0.0807 0.9357277
## prisoners    0.1915713   0.0547407  3.4996 0.0004847 ***
## male        47.1710601   9.4669657  4.9827 7.286e-07 ***
## density    -95.2038223  46.2509986 -2.0584 0.0397870 *
## cauc       -10.9364414   4.7666752 -2.2944 0.0219592 *
## afam       -36.0927015  13.7585404 -2.6233 0.0088298 **
## population   2.0577721   4.7702000  0.4314 0.6662767
## income       0.0068879   0.0039016  1.7654 0.0777774 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    8871100
## Residual Sum of Squares: 7871200
## R-Squared:      0.11272
## Adj. R-Squared: 0.047716
## F-statistic: 17.3406 on 8 and 1092 DF, p-value: < 2.22e-16
```

First Differences

```
FD_fit_year <- plm(violent ~ law  + prisoners + male + density + cauc + afam + population + income + fac
summary(FD_fit_year)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = violent ~ law + prisoners + male + density + cauc +
##     afam + population + income + factor(year), data = mydata,
```

```
##     model = "fd", index = c("state", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
## Observations used in estimation: 1122
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -464.000  -20.700    0.159   21.400  323.000
##
## Coefficients:
##                    Estimate  Std. Error t-value  Pr(>|t|)
## lawyes          -3.4082e+00  9.5820e+00 -0.3557 0.7221393
## prisoners       -2.6817e-02  6.5768e-02 -0.4077 0.6835427
## male             2.5903e+01  1.6825e+01  1.5396 0.1239515
## density         -3.2036e+02  6.2113e+01 -5.1578 2.966e-07 ***
## cauc            -1.0584e+01  7.2833e+00 -1.4532 0.1464474
## afam             1.4819e+01  2.3417e+01  0.6328 0.5269851
## population       3.2734e+00  1.5253e+01  0.2146 0.8301150
## income           1.1825e-02  5.5392e-03  2.1347 0.0330062 *
## factor(year)1978  1.5337e+01  7.7813e+00  1.9711 0.0489671 *
## factor(year)1979  5.9977e+01  1.1377e+01  5.2718 1.627e-07 ***
## factor(year)1980  9.2022e+01  1.4467e+01  6.3609 2.945e-10 ***
## factor(year)1981  9.8849e+01  1.7697e+01  5.5857 2.938e-08 ***
## factor(year)1982  8.4494e+01  2.1648e+01  3.9031 0.0001008 ***
## factor(year)1983  5.5575e+01  2.6115e+01  2.1281 0.0335519 *
## factor(year)1984  5.2689e+01  3.1537e+01  1.6707 0.0950707 .
## factor(year)1985  6.6106e+01  3.6742e+01  1.7992 0.0722653 .
## factor(year)1986  9.6393e+01  4.2141e+01  2.2874 0.0223637 *
## factor(year)1987  8.7685e+01  4.7470e+01  1.8472 0.0649931 .
## factor(year)1988  1.1560e+02  5.2916e+01  2.1846 0.0291276 *
## factor(year)1989  1.3655e+02  5.7932e+01  2.3570 0.0185980 *
## factor(year)1990  2.1600e+02  6.7284e+01  3.2103 0.0013648 **
## factor(year)1991  2.4508e+02  7.1293e+01  3.4376 0.0006090 ***
## factor(year)1992  2.5545e+02  7.5434e+01  3.3864 0.0007334 ***
## factor(year)1993  2.6436e+02  7.8805e+01  3.3547 0.0008219 ***
## factor(year)1994  2.4419e+02  8.2520e+01  2.9591 0.0031518 **
## factor(year)1995  2.3070e+02  8.5646e+01  2.6936 0.0071764 **
## factor(year)1996  1.9186e+02  8.8441e+01  2.1693 0.0302732 *
## factor(year)1997  1.6654e+02  9.1095e+01  1.8282 0.0677996 .
## factor(year)1998  1.2694e+02  9.4085e+01  1.3492 0.1775605
## factor(year)1999  9.2039e+01  9.6889e+01  0.9499 0.3423505
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3168700
## Residual Sum of Squares: 2333100
## R-Squared:      0.2637
## Adj. R-Squared: 0.24415
## F-statistic: 13.4862 on 29 and 1092 DF, p-value: < 2.22e-16
```

Still large difference in coefficients is observed.

3.2 PART 2

1. Compute the canonical White standard errors for this model and report the results.

```r
HC_coefs <- vcovHC(fixed_effects_fit, method = "white1")
white_ses_state <- sqrt(diag(HC_coefs))
print(white_ses_state)
```

```
##       lawyes    prisoners         male      density         cauc
## 9.960723e+00 9.807655e-02 4.430832e+00 1.685339e+02 2.731201e+00
##         afam   population       income
## 1.646692e+01 7.520902e+00 4.010081e-03
```

2. Now compute the canonical "clustered" standard errors with clustering at the state level (as in Arrelano 1987).

```r
HCV_coefs_state <- vcovHC(fixed_effects_fit, method = "arellano", cluster = "group")
clustered_ses_state <- sqrt(diag(HCV_coefs_state))
print(clustered_ses_state)
```

```
##       lawyes    prisoners         male      density         cauc
## 1.922120e+01 1.546612e-01 9.327876e+00 1.119639e+02 6.067522e+00
##         afam   population       income
## 2.613087e+01 9.434884e+00 6.193339e-03
```

3. Now compute the canonical "clustered" standard errors with clustering at the time unit (as in Arrelano 1987 but for T not N).

```r
HCV_coefs_time <- vcovHC(fixed_effects_fit, method = "arellano", cluster = "time")
clustered_ses_time <- sqrt(diag(HCV_coefs_time))
print(clustered_ses_time)
```

```
##       lawyes    prisoners         male      density         cauc
##  12.58309410   0.10333974   7.05731407 141.49055513   3.34585934
##         afam   population       income
##  17.93822669   7.73829461   0.00733827
```

4. Now compute the Newey-West standard errors for panel data.

```r
NW_coefs <- vcovNW(fixed_effects_fit)
NW_ses_state <- sqrt(diag(NW_coefs))
print(NW_ses_state)
```

```
##       lawyes    prisoners         male      density         cauc
## 1.390541e+01 1.354396e-01 6.494327e+00 2.336544e+02 3.960799e+00
##         afam   population       income
## 2.407151e+01 1.070802e+01 5.762757e-03
```

5. Comment on the relative magnitude of these errors for the ?? on our key variable, the shall-carry law. Give some intuition for what you see.
Arellano approach gives highest standard errors and NW lowest.