Lab 13

AML

09/11/2020

Setup

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version [illegible]
```

```
## ── Attaching packages ──────────────────────── tidyverse [illegible] ──
```

```
## ✓ ggplot2 [illegible]     ✓ purrr   [illegible]
## ✓ tibble  [illegible]     ✓ dplyr   [illegible]
## ✓ tidyr   [illegible]     ✓ stringr [illegible]
## ✓ readr   [illegible]     ✓ forcats [illegible]
```

```
## Warning: package 'ggplot2' was built under R version [illegible]
```

```
## Warning: package 'forcats' was built under R version [illegible]
```

```
## ── Conflicts ──────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(plm)
```

```
##
## Attaching package: 'plm'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```
library(car)
```

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

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(gplots)
```

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

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

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version [illegible]
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

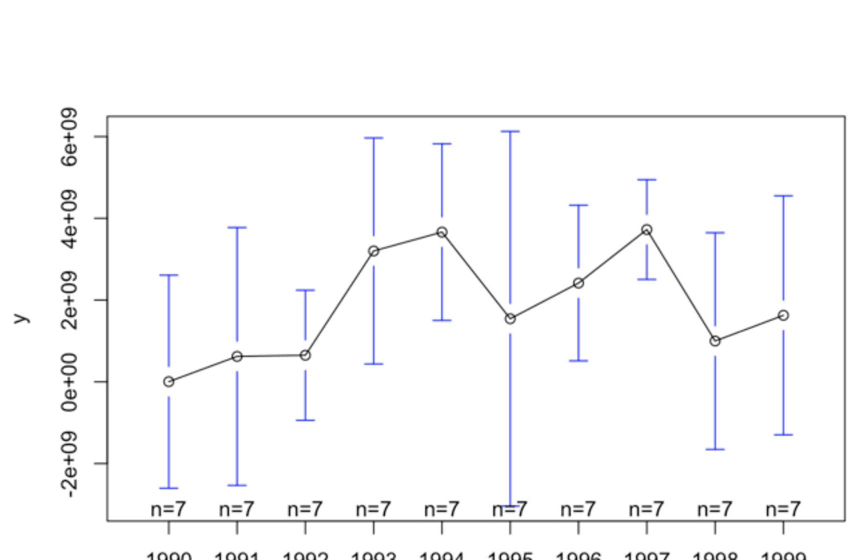
```
##
## Attaching package: 'zoo'
```

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

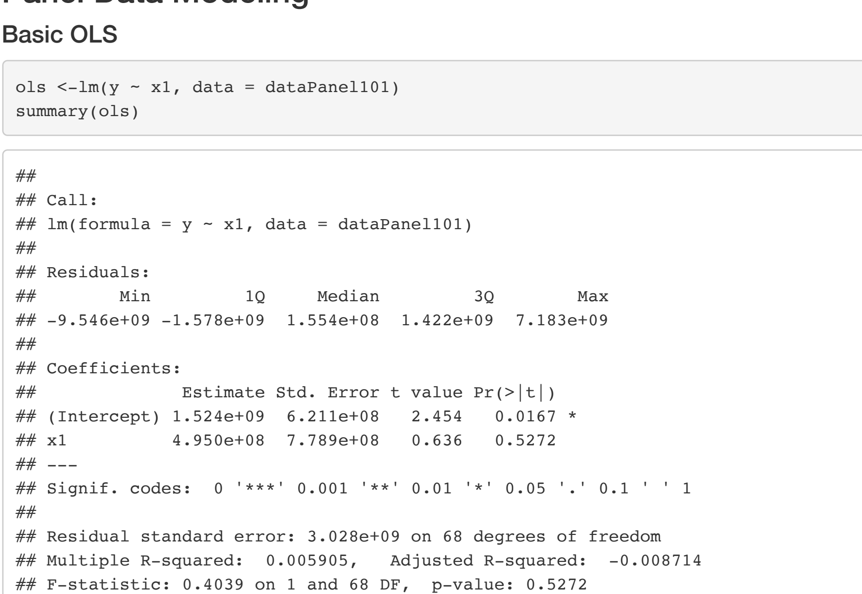
```
dataPanel101 <-read.csv("https://github.com/ds777/sample-datasets/blob/master/dataPanel101.csv?raw=true")
dataPanel101 <- pdata.frame(dataPanel101, index=c("country","year"))
#dataPanel101 #see data
```

Explanatory Data Analysis

```
#coplot(y ~ year|country, type="l", data=dataPanel101)
plotmeans(y ~ country, data = dataPanel101)
```



```
plotmeans(y ~ year, data = dataPanel101) #Heterogeineity across years
```



Panel Data Modeling

Basic OLS

```
ols <-lm(y ~ x1, data = dataPanel101)
summary(ols)
```

```
##
## Call:
## lm(formula = y ~ x1, data = dataPanel101)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -9.546e+09 -1.578e+09  1.554e+08  1.422e+09  7.183e+09
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.524e+09  6.211e+08   2.454   0.0167 *
## x1          4.950e+08  7.789e+08   0.636   0.5272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.028e+09 on 68 degrees of freedom
## Multiple R-squared:  0.005905,   Adjusted R-squared:  -0.008714
## F-statistic: 0.4039 on 1 and 68 DF,  p-value: 0.5272
```

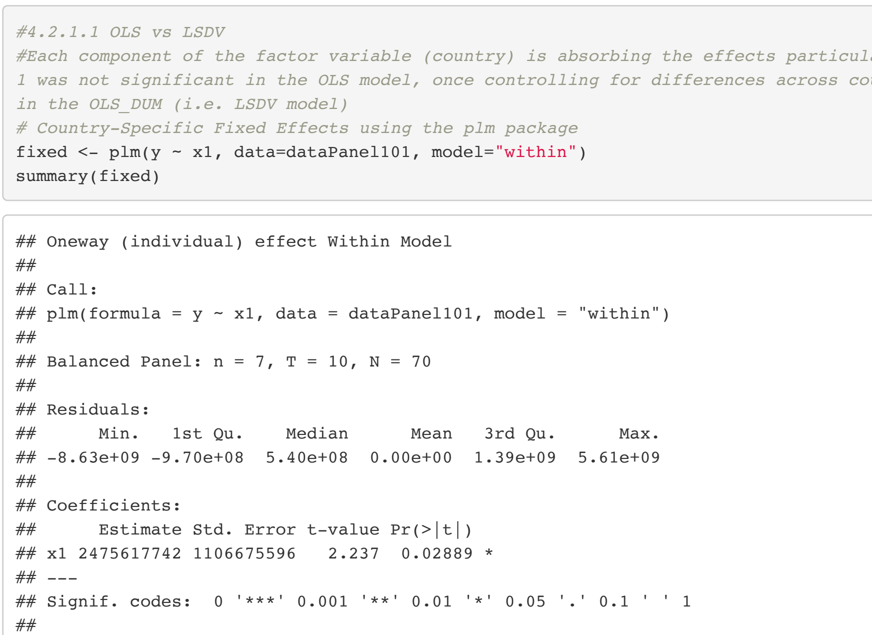
Fixed Effects Model

#4.2.1 Country-Specific Fixed Effects using Dummy Variables (LSDV Model)

```
fixed.dum <-lm(y ~ x1 + factor(country) - 1, data = dataPanel101)
summary(fixed.dum)
```

```
##
## Call:
## lm(formula = y ~ x1 + factor(country) - 1, data = dataPanel101)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -8.634e+09 -9.697e+08  5.405e+08  1.386e+09  5.612e+09
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## x1                 2.476e+09  1.107e+09   2.237  0.02889 *
## factor(country)A  8.805e+08  9.618e+08   0.916  0.36347
## factor(country)B -1.058e+09  1.051e+09  -1.006  0.31811
## factor(country)C -1.723e+09  1.632e+09  -1.056  0.29508
## factor(country)D  3.163e+09  9.095e+08   3.478  0.00093 ***
## factor(country)E -6.026e+08  1.064e+09  -0.566  0.57329
## factor(country)F  2.011e+09  1.123e+09   1.791  0.07821 .
## factor(country)G -9.847e+08  1.493e+09  -0.660  0.51190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.796e+09 on 62 degrees of freedom
## Multiple R-squared:  0.4402, Adjusted R-squared:  0.368
## F-statistic: 6.095 on 8 and 62 DF,  p-value: 8.892e-06
```

```
#Fit
yhat <- fixed.dum$fitted
scatterplot(yhat ~ dataPanel101$x1 | dataPanel101$country, xlab ="x1", ylab ="yhat", boxplots = FALSE,smooth = FA
LSE)
abline(lm(dataPanel101$y~dataPanel101$x1),lwd=3, col="red")
```



```
#4.2.1.1 OLS vs LSDV
#Each component of the factor variable (country) is absorbing the effects particular to each country. Predictor x
1 was not significant in the OLS model, once controlling for differences across countries, x1 became significant
in the OLS_DUM (i.e. LSDV model)
# Country-Specific Fixed Effects using the plm package
fixed <- plm(y ~ x1, data=dataPanel101, model="within")
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = y ~ x1, data = dataPanel101, model = "within")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -8.63e+09 -9.70e+08  5.40e+08  0.00e+00  1.39e+09  5.61e+09
##
## Coefficients:
##      Estimate Std. Error t-value Pr(>|t|)
## x1 2475617742 1106675596   2.237  0.02889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5.2364e+20
## Residual Sum of Squares: 4.8454e+20
## R-Squared:      0.074684
## Adj. R-Squared: -0.029788
## F-statistic: 5.00411 on 1 and 62 DF, p-value: 0.028892
```

```
# Display the fixed effects (constants for each country)
fixef(fixed)
```

```
##           A           B           C           D           E           F
##   880542434 -1057858320 -1722810680  3162826916  -602621958  2010731852
##           G
##  -984717393
```

```
###The coeff of x1 indicates how much Y changes overtime, on average per country, when X increases by one unit.

#4.2.2.1 Fixed Effects vs OLS
# Testing for fixed effects, null: OLS better than fixed
pFtest(fixed, ols)
```

```
##
##  F test for individual effects
##
## data:  y ~ x1
## F = 2.9655, df1 = 6, df2 = 62, p-value = 0.01307
## alternative hypothesis: significant effects
```

If the p-value is < 0.05 then the fixed effects model is a better choice.

Random Effects Model

```
random <- plm(y ~ x1, data=dataPanel101, model="random")
summary(random)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = y ~ x1, data = dataPanel101, model = "random")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Effects:
##                     var   std.dev share
## idiosyncratic 7.815e+18 2.796e+09 0.873
## individual    1.133e+18 1.065e+09 0.127
## theta: 0.3611
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -8.94e+09 -1.51e+09  2.82e+08  0.00e+00  1.56e+09  6.63e+09
##
## Coefficients:
##               Estimate Std. Error z-value Pr(>|z|)
## (Intercept) 1037014329  790626206  1.3116   0.1896
## x1          1247001710  902145599  1.3823   0.1669
##
## Total Sum of Squares:    5.6595e+20
## Residual Sum of Squares: 5.5048e+20
## R-Squared:      0.02733
## Adj. R-Squared: 0.013026
## Chisq: 1.91065 on 1 DF, p-value: 0.16689
```

Interpretation of the coefficients is tricky since they include both the within-entity and between-entity effects. In the case of TSCS data represents the average effect of X over Y when X changes across time and between countries by one unit.
Also remember that the Random Effects assumptions are much stronger.

Fixed vs. Random

To decide between fixed or random effects you can run a Hausman test where the null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects (see Green, 2008, chapter 9). It basically tests whether the unique errors are correlated with the regressors, the null hypothesis is they are not. If the p-value is significant (for example <0.05) then use fixed effects, if not use random effects.

```
phtest(fixed, random)
```

```
##
##  Hausman Test
##
## data:  y ~ x1
## chisq = 3.674, df = 1, p-value = 0.05527
## alternative hypothesis: one model is inconsistent
```

We should use the random effects model.

Regression Diagnostics

Time-fixed effects testing

```
fixed.time <- plm(y ~ x1 + factor(year), data=dataPanel101, model="within")
summary(fixed.time)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = y ~ x1 + factor(year), data = dataPanel101, model = "within")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -7.92e+09 -1.05e+09 -1.40e+08  0.00e+00  1.63e+09  5.49e+09
##
## Coefficients:
##                    Estimate Std. Error t-value Pr(>|t|)
## x1               1389050208 1319849568  1.0524  0.29738
## factor(year)1991  296381592 1503368532  0.1971  0.84447
## factor(year)1992  145369724 1547226550  0.0940  0.92550
## factor(year)1993 2874386825 1503862558  1.9113  0.06138 .
## factor(year)1994 2848156370 1661498931  1.7142  0.09233 .
## factor(year)1995  973941363 1567245752  0.6214  0.53698
## factor(year)1996 1672812635 1631539257  1.0253  0.30988
## factor(year)1997 2991770146 1627062033  1.8388  0.07156 .
## factor(year)1998  367463673 1587924443  0.2314  0.81789
## factor(year)1999 1258751990 1512397631  0.8323  0.40898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5.2364e+20
## Residual Sum of Squares: 4.0201e+20
## R-Squared:      0.23229
## Adj. R-Squared: 0.0005285
## F-statistic: 1.60365 on 10 and 53 DF, p-value: 0.13113
```

```
# Testing time-fixed effects. The null is that no time-fixed effects are needed
pFtest(fixed.time, fixed)
```

```
##
##  F test for individual effects
##
## data:  y ~ x1 + factor(year)
## F = 1.209, df1 = 9, df2 = 53, p-value = 0.3094
## alternative hypothesis: significant effects
```

```
plmtest(fixed, c("time"), type=("bp"))
```

```
##
##  Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced
##  panels
##
## data:  y ~ x1
## chisq = 0.16532, df = 1, p-value = 0.6843
## alternative hypothesis: significant effects
```

If the p value < 0.05 then use time-fixed effects. In this example, no need to use time-fixed effects

Serial correlation testing

```
pbgtest(fixed)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  y ~ x1
## chisq = 14.137, df = 10, p-value = 0.1668
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
#Because p-value > 0.05, we conclude that there is NO serial correlation
```

Unit roots/stationarity testing

```
adf.test(dataPanel101$y, k=2)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  dataPanel101$y
## Dickey-Fuller = -3.9051, Lag order = 2, p-value = 0.0191
## alternative hypothesis: stationary
```

BP test for heteroskedasticity testing

```
bptest(y ~ x1 + factor(country), data = dataPanel101, studentize=F)
```

```
##
##  Breusch-Pagan test
##
## data:  y ~ x1 + factor(country)
## BP = 14.606, df = 7, p-value = 0.04139
```

```
coeftest(random) # original coefficients
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1037014329  790626206  1.3116   0.1941
## x1          1247001710  902145599  1.3823   0.1714
```

```
coeftest(random, vcovHC) #heteroskedasticy consistent coeffs.
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1037014329  907983024  1.1421   0.2574
## x1          1247001710  828970258  1.5043   0.1371
```

Controlling for heteroskedasticity: Fixed effects

```
coeftest(fixed)
```

```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## x1 2475617742 1106675596   2.237  0.02889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(fixed, vcovHC(fixed, method = "arellano"))
```

```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## x1 2475617742 1358388924  1.8225  0.07321 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```