Lab13

## Setup

setwd("C:/Users/22700/Desktop")  
library(data.table)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.4 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::between() masks data.table::between()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks data.table::first()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks data.table::last()  
## x purrr::transpose() masks data.table::transpose()

library(plm)

##   
## Attaching package: 'plm'

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

## The following object is masked from 'package:data.table':  
##   
## between

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)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

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")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## country = col\_character(),  
## year = col\_double(),  
## y = col\_double(),  
## y\_bin = col\_double(),  
## x1 = col\_double(),  
## x2 = col\_double(),  
## x3 = col\_double(),  
## opinion = col\_character()  
## )

dataPanel101 <- plm.data(dataPanel101, index=c("country","year"))

## Warning: use of 'plm.data' is discouraged, better use 'pdata.frame' instead

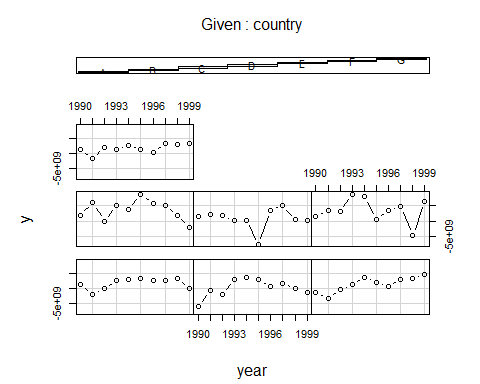
## View Datat panel

dataPanel101

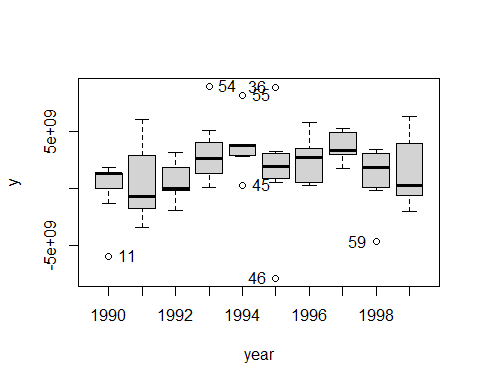
## country year y y\_bin x1 x2 x3 opinion  
## 1 A 1990 1342787840 1 0.27790365 -1.1079559 0.28255358 Str agree  
## 2 A 1991 -1899660544 0 0.32068470 -0.9487200 0.49253848 Disag  
## 3 A 1992 -11234363 0 0.36346573 -0.7894840 0.70252335 Disag  
## 4 A 1993 2645775360 1 0.24614404 -0.8855330 -0.09439092 Disag  
## 5 A 1994 3008334848 1 0.42462304 -0.7297683 0.94613063 Disag  
## 6 A 1995 3229574144 1 0.47721413 -0.7232460 1.02968040 Str agree  
## 7 A 1996 2756754176 1 0.49980500 -0.7815716 1.09228810 Disag  
## 8 A 1997 2771810560 1 0.05162839 -0.7048455 1.41590080 Str agree  
## 9 A 1998 3397338880 1 0.36641079 -0.6983712 1.54872270 Disag  
## 10 A 1999 39770336 1 0.39584252 -0.6431540 1.79419800 Str disag  
## 11 B 1990 -5934699520 0 -0.08184998 1.4251202 0.02342812 Agree  
## 12 B 1991 -711623744 0 0.10616001 1.6496018 0.26036251 Str agree  
## 13 B 1992 -1933116160 0 0.35378519 1.5937191 -0.23439877 Agree  
## 14 B 1993 3072741632 1 0.72677696 1.6917576 0.25622433 Str disag  
## 15 B 1994 3768078848 1 0.71939486 1.7414261 0.41174951 Disag  
## 16 B 1995 2837581312 1 0.67154658 1.7083139 0.53584301 Str disag  
## 17 B 1996 577199360 1 0.81985730 1.5324961 -0.49964902 Str agree  
## 18 B 1997 1786851584 1 0.88016719 1.5021962 -0.57626772 Disag  
## 19 B 1998 -149072048 0 0.70451611 1.4236463 -0.44841924 Agree  
## 20 B 1999 -1174480128 0 0.23696731 1.4545859 -0.04936399 Str disag  
## 21 C 1990 -1292379264 0 1.31256070 -1.2931356 0.20408297 Agree  
## 22 C 1991 -3415966464 0 1.17748360 -1.3442180 0.28397188 Str agree  
## 23 C 1992 -355804672 0 1.25640800 -1.2599510 0.37339270 Agree  
## 24 C 1993 1225180032 1 1.42154460 -1.3117452 -0.37596563 Disag  
## 25 C 1994 3802287616 1 1.11419310 -1.2849948 0.56046754 Str disag  
## 26 C 1995 1959696640 1 1.15948390 -1.2188276 0.69540799 Agree  
## 27 C 1996 530576672 1 1.16045430 -1.2350063 0.81689382 Agree  
## 28 C 1997 3128852224 1 1.44641160 -1.3275964 -0.14206907 Str disag  
## 29 C 1998 3201045760 1 1.15162670 -1.2061129 1.19458140 Str agree  
## 30 C 1999 4663067648 1 1.19054410 -1.1266172 1.67016040 Disag  
## 31 D 1990 1883025152 1 -0.31391269 1.7366557 0.64663702 Disag  
## 32 D 1991 6037768704 1 0.36009100 2.1318641 1.09994170 Disag  
## 33 D 1992 10244189 1 0.05188770 1.6816775 0.96976823 Str agree  
## 34 D 1993 5067265024 1 0.20944354 1.6149769 -0.21257821 Str agree  
## 35 D 1994 3882478336 1 0.38207000 1.5683011 -1.16538670 Disag  
## 36 D 1995 8827006976 1 0.24208580 1.5412215 -0.18413101 Agree  
## 37 D 1996 5782000128 1 0.48636678 1.7423391 -0.03731453 Str disag  
## 38 D 1997 5090524160 1 0.35942599 1.8742865 0.08786795 Str agree  
## 39 D 1998 1850565248 1 0.23220351 1.5953021 0.07247547 Disag  
## 40 D 1999 -2025476864 0 -0.07998896 1.7047973 0.55843300 Str agree  
## 41 E 1990 1342787840 1 0.45286715 1.7284026 0.59705788 Str disag  
## 42 E 1991 2296009472 1 0.41904032 1.7068400 0.79313534 Str agree  
## 43 E 1992 1737627776 1 0.38521346 1.6852775 0.98921281 Agree  
## 44 E 1993 113973136 1 -0.24428773 1.6492835 1.22413280 Str agree  
## 45 E 1994 260098048 1 1.39114000 2.5302765 -0.52620137 Str disag  
## 46 E 1995 -7863482880 0 0.31968558 1.1890552 -0.48425370 Agree  
## 47 E 1996 3520491520 1 0.61097682 1.4845277 -0.97895509 Agree  
## 48 E 1997 5234565120 1 0.71761495 1.5544620 -0.98863661 Str disag  
## 49 E 1998 344746176 1 0.69613826 1.7010406 -0.08965246 Disag  
## 50 E 1999 243920688 1 0.60662067 1.6119040 -0.08929884 Str disag  
## 51 F 1990 1342787840 1 -0.56757486 -0.3466710 1.25841930 Str agree  
## 52 F 1991 3560401920 1 0.15974578 -0.4641182 0.32665297 Str disag  
## 53 F 1992 3192281088 1 0.88706642 -0.5815655 -0.60511333 Agree  
## 54 F 1993 8941232128 1 0.53241795 -0.7553238 -0.51157588 Agree  
## 55 F 1994 8124504576 1 0.87260014 -0.7114431 0.20570269 Str agree  
## 56 F 1995 491740096 1 0.91935229 -0.3697441 -0.01292755 Str agree  
## 57 F 1996 3497164544 1 1.39689230 -0.3601406 0.67867643 Str agree  
## 58 F 1997 4764803072 1 0.98688608 -0.3590902 0.24226174 Str agree  
## 59 F 1998 -4671723520 0 0.78830910 -0.7556524 0.73347801 Agree  
## 60 F 1999 6349319168 1 0.27938697 -0.4601679 1.17317200 Disag  
## 61 G 1990 1342787840 1 0.94488174 -1.5150151 1.45265730 Str disag  
## 62 G 1991 -1518985728 0 1.09872830 -1.4614717 1.43964470 Agree  
## 63 G 1992 1912769920 1 1.25257490 -1.4079282 1.42663200 Str agree  
## 64 G 1993 1345690240 1 0.76276451 -1.3519315 1.85448630 Str disag  
## 65 G 1994 2793515008 1 1.20645560 -1.3252175 2.23653030 Str disag  
## 66 G 1995 1323696384 1 1.08718650 -1.4098167 2.82980850 Str disag  
## 67 G 1996 254524176 1 0.78107548 -1.3279996 4.27822400 Str agree  
## 68 G 1997 3297033216 1 1.25787950 -1.5773667 4.58732560 Disag  
## 69 G 1998 3011820800 1 1.24277660 -1.6012177 6.11376190 Disag  
## 70 G 1999 3296283392 1 1.23420020 -1.6217614 7.16892190 Disag

## explanatory data analysis

coplot(y ~ year|country, type="b", data=dataPanel101)

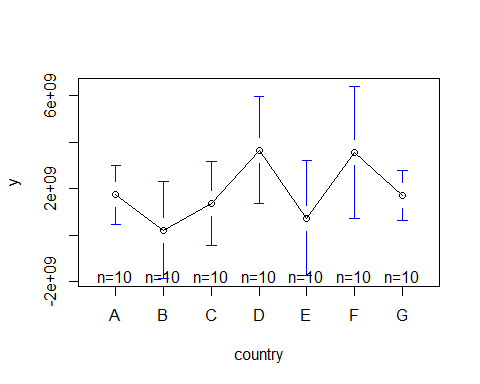


scatterplot(y~year|country, data=dataPanel101)

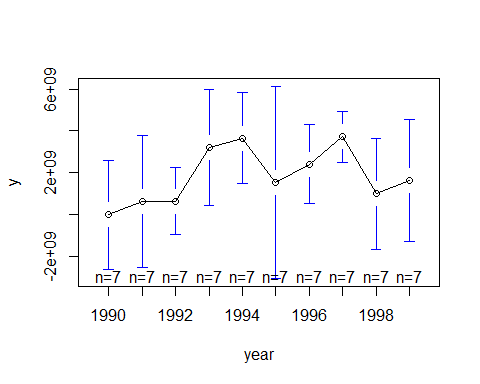


## [1] "11" "54" "45" "55" "46" "36" "59"

plotmeans(y ~ country, data = dataPanel101)

 # Heterogeneity across years

plotmeans(y ~ year, data = dataPanel101)



## Panel Data Modeling

# we start with the basic OLS

#The problem is that the basic OLS regression model does not consider heterogeneity across countries or across years

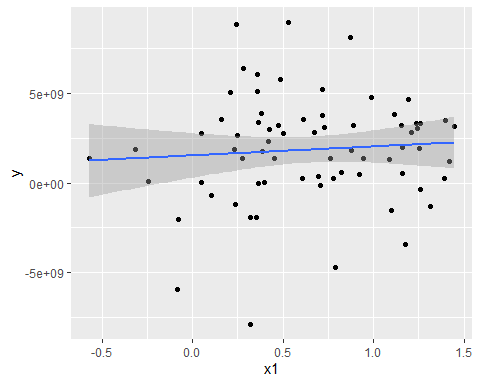
ols <-lm(y ~ x1, data = dataPanel101)  
summary(ols)

##   
## 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

# FItted values

yhat <- ols$fitted  
ggplot(dataPanel101, aes(x = x1, y = y))+  
 geom\_point() +  
 geom\_smooth(method=lm)

## `geom\_smooth()` using formula 'y ~ x'

 ## 4.2 Fixed Effects Model

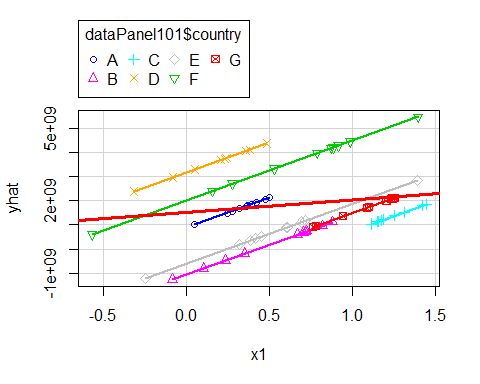
# 4.2.1 Country-Specific Fixed Effects using Dummy Variables (LSDV Model)

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$country, xlab ="x1", ylab ="yhat", boxplots = FALSE,smooth = FALSE)  
abline(lm(dataPanel101$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 x1 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

# 4.3 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.

## 4.4 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

# 4.5 Regression Diagnostics

## 4.5.1 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 1389050209 1319849569 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 2848156371 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

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

## 4.5.5 Unit roots/stationarity testing

# The Dickey-Fuller test to check for stochastic trends.

# H0) The null hypothesis is that the series has a unit root (i.e. non-stationary)

# If unit root is present you can take the first difference of the variable.

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

# Because p-value < 0.05, we conclude that the series does NOT have unit root. In other words, the series is stationary

## Heteroskedasticity testing

# H0) The null hypothesis for the Breusch-Pagan test is homoskedasticity

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

# Because p-value < 0.05, we detect hetersokedasticity

# => If hetersokedasticity is detected we need to use a robust covariance matrix (Sandwich estimator) to account for it

#4.5.6.1 Controlling for heteroskedasticity: Random effects #The -vcovHC- function estimates three heteroskedasticity-consistent covariance estimators:

# “white1” - for general heteroskedasticity but no serial correlation. Recommended for random effects.

# “white2” - is “white1” restricted to a common variance within groups. Recommended for random effects.

# arellano" - both heteroskedasticity and serial correlation. Recommended for fixed effects.

# Original coefficients

coeftest(random)

##   
## 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

# Heteroskedasticity consistent coefficients

coeftest(random, vcovHC)

##   
## 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

## 4.5.6.2 Controlling for heteroskedasticity: Fixed effects

# Original coefficients

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

# Heteroskedasticity consistent coefficients (Arellano)

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