Lab11a-Panel data

Amir Jafarzadeh

08/11/2021

```r
setwd("U:/econometrics/lab11a")

library(tidyverse) # Modern data science library
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## Warning: package 'ggplot2' was built under R version 4.1.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(plm)       # Panel data analysis library
```

```
## Warning: package 'plm' was built under R version 4.1.1
##
## Attaching package: 'plm'
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```r
library(car)       # Companion to applied regression
```

```
## 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
```

```r
library(gplots)    # Various programing tools for plotting data
```

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

```
##
##     lowess
```

```r
library(tseries)   # For timeseries analysis
```

```
## Warning: package 'tseries' was built under R version 4.1.1

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

```r
library(lmtest)    # For hetoroskedasticity analysis
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

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

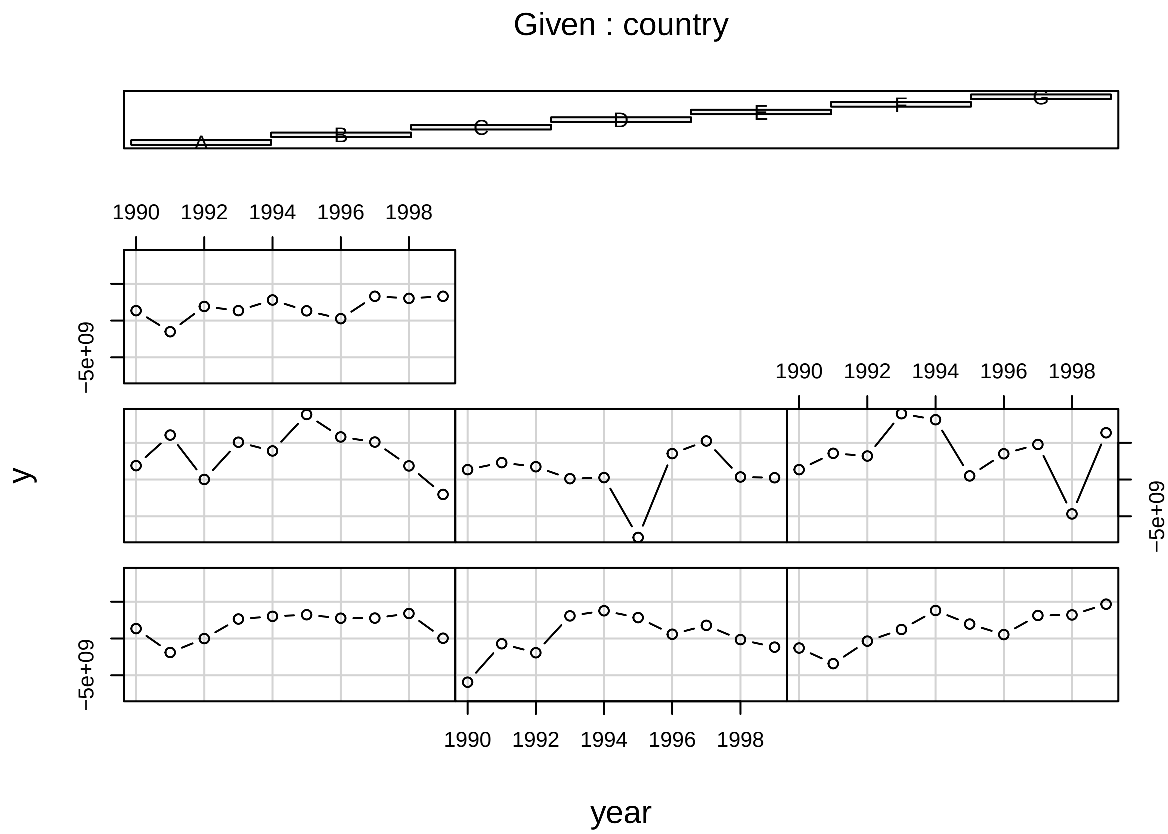
```r
dataPanel101 <- read_csv("https://github.com/ds777/sample-datasets/blob/master/dataPanel101.csv?raw=tru
```

```
##
## -- 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()
## )
```

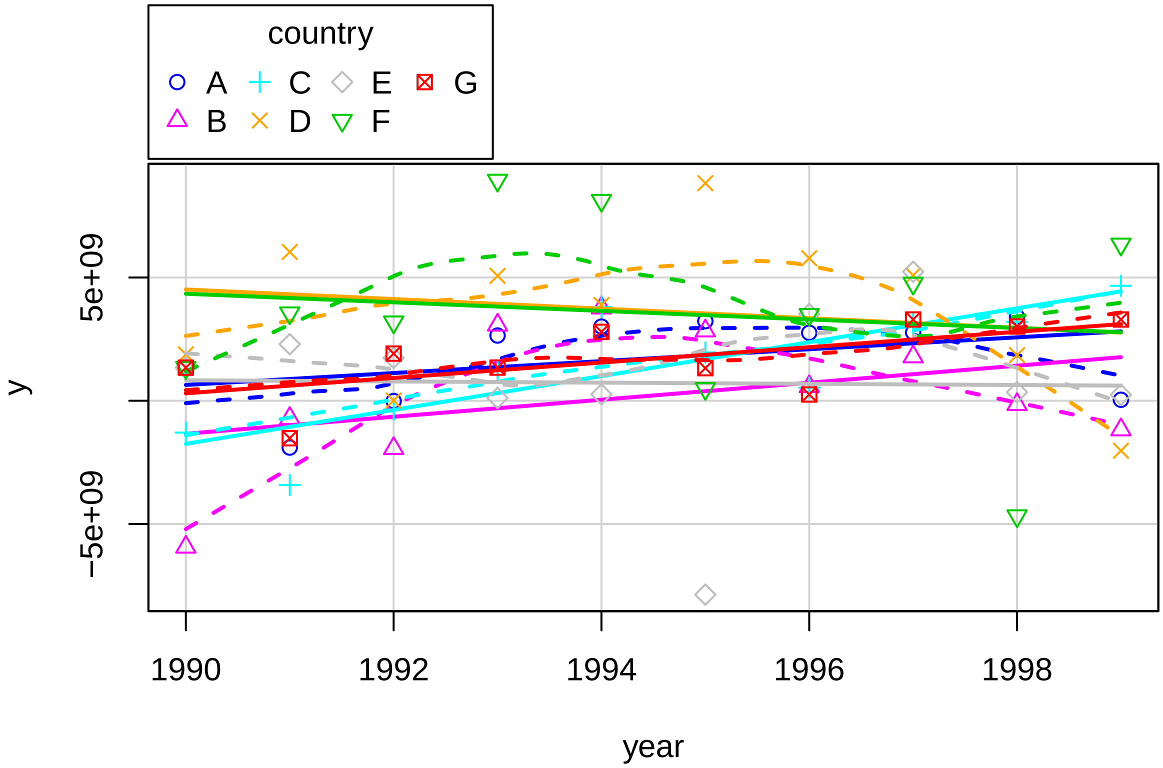
```r
dataPanel101
```

```
## # A tibble: 70 x 8
##    country  year           y y_bin     x1     x2      x3 opinion
##    <chr>   <dbl>       <dbl> <dbl>  <dbl>  <dbl>   <dbl> <chr>
##  1 A        1990  1342787840     1 0.278  -1.11   0.283  Str agree
##  2 A        1991 -1899660544     0 0.321  -0.949  0.493  Disag
##  3 A        1992   -11234363     0 0.363  -0.789  0.703  Disag
##  4 A        1993  2645775360     1 0.246  -0.886 -0.0944 Disag
##  5 A        1994  3008334848     1 0.425  -0.730  0.946  Disag
##  6 A        1995  3229574144     1 0.477  -0.723  1.03   Str agree
##  7 A        1996  2756754176     1 0.500  -0.782  1.09   Disag
##  8 A        1997  2771810560     1 0.0516 -0.705  1.42   Str agree
##  9 A        1998  3397338880     1 0.366  -0.698  1.55   Disag
## 10 A        1999    39770336     1 0.396  -0.643  1.79   Str disag
## # ... with 60 more rows
```

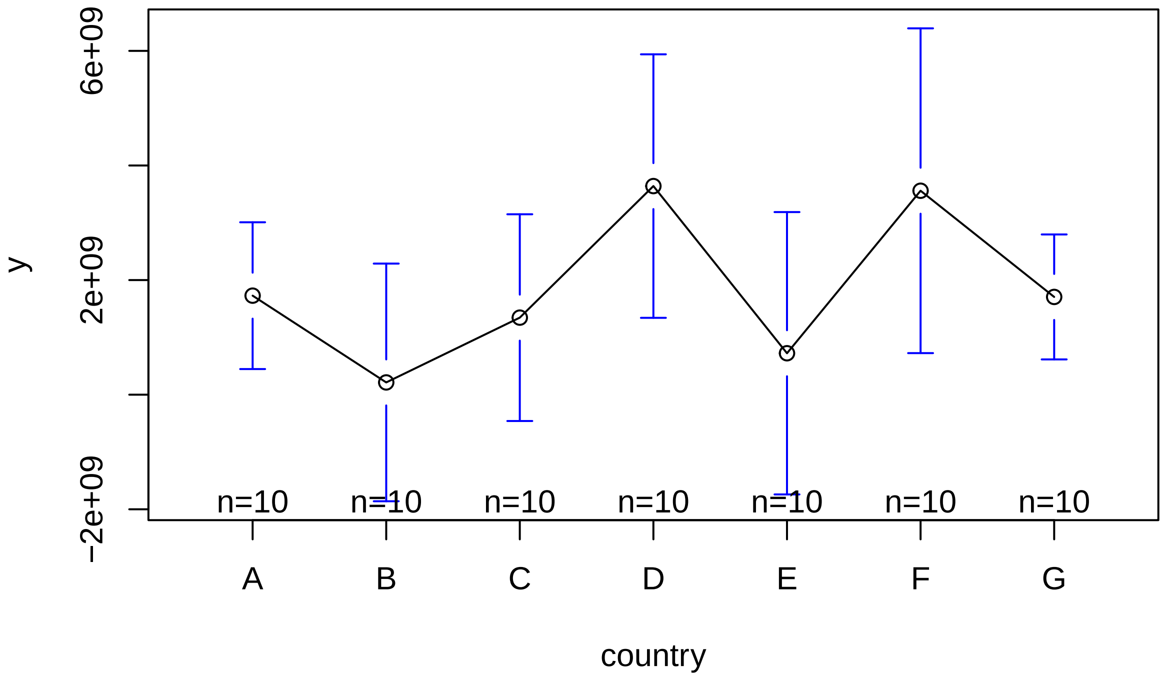
```r
coplot(y ~ year|country, type="b", data=dataPanel101)
```



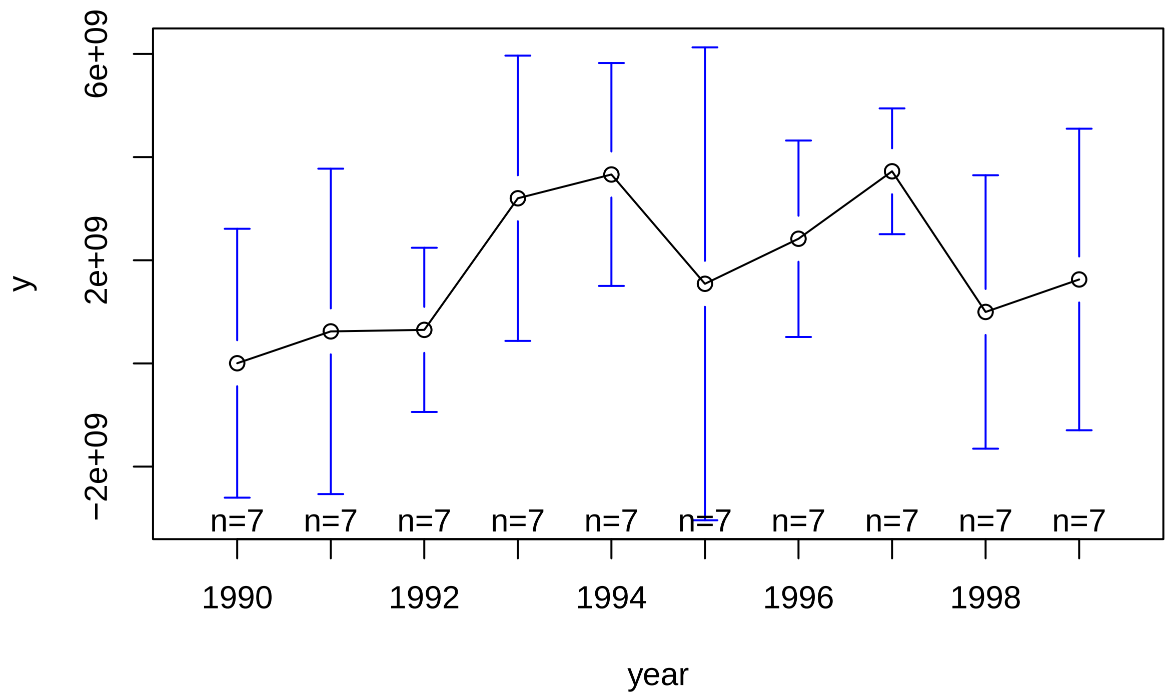

```
scatterplot(y~year|country, data=dataPanel101)
```




```
plotmeans(y ~ country, data = dataPanel101)
```




```
plotmeans(y ~ year, data = dataPanel101)
```

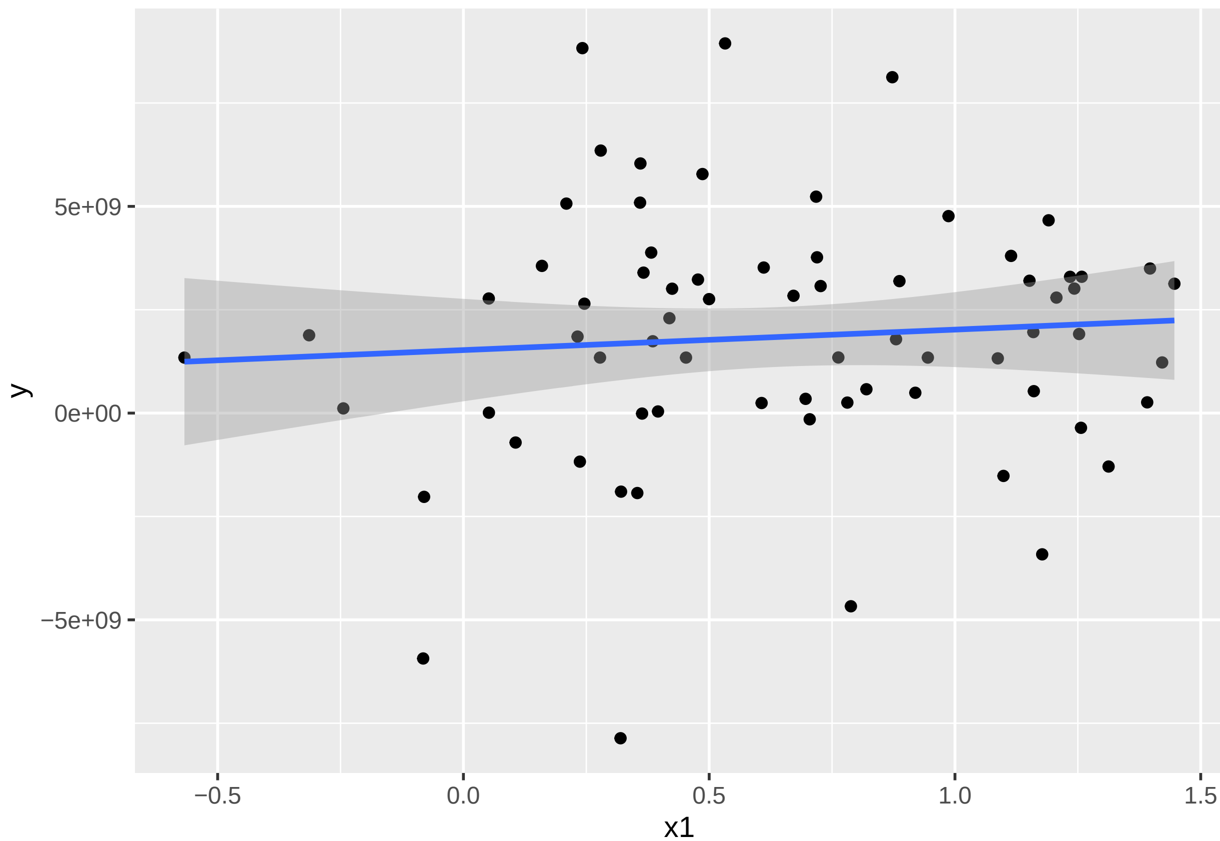



```
ols <-lm(y ~ x1, data = dataPanel101)
summary(ols)
```

```
##
## 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
```

```
yhat <- ols$fitted
ggplot(dataPanel101, aes(x = x1, y = y))+
  geom_point() +
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

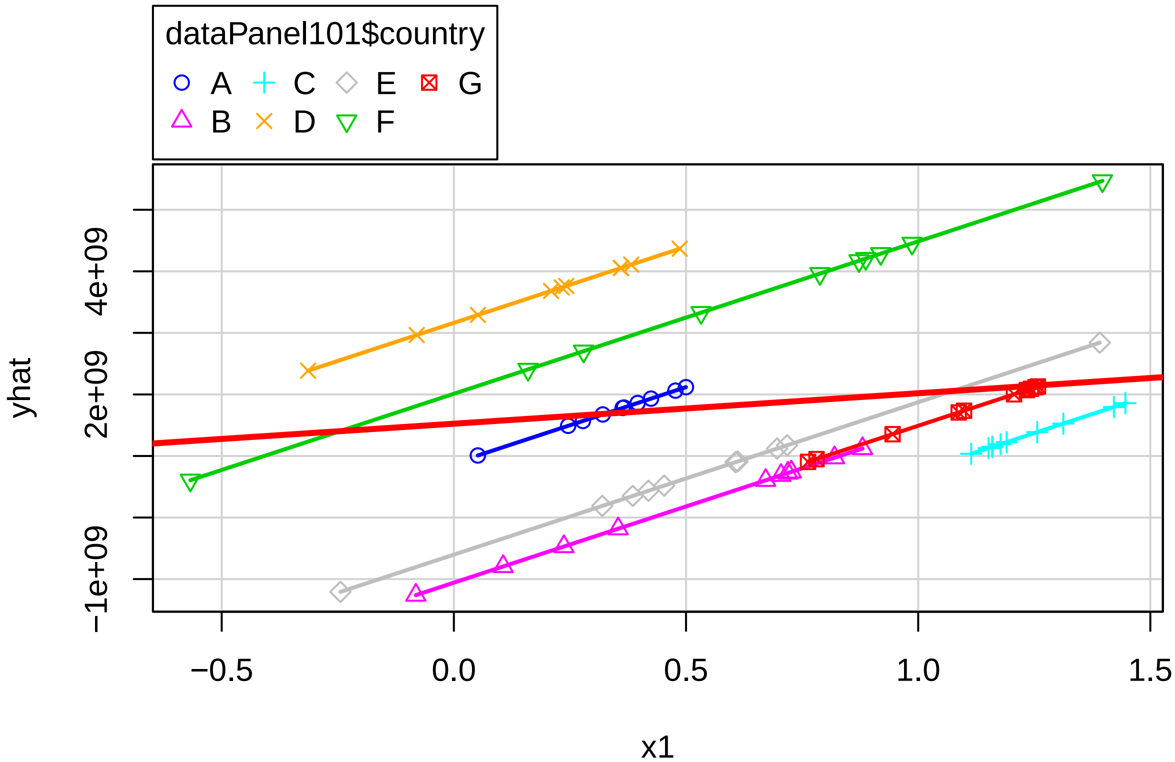



```r
fixed.dum <-lm(y ~ x1 + factor(country) - 1, data = dataPanel101)
summary(fixed.dum)
```

```
##
## 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
```

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




```
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
```

```
# 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
```

```
##Random Effect
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
```

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

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

```r
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
```

```r
# 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
```

```r
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
```

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

```
## Pooling Model
##
## Call:
## plm(formula = y ~ x1, data = dataPanel101, model = "pooling")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -9.55e+09 -1.58e+09  1.55e+08  0.00e+00  1.42e+09  7.18e+09
##
## Coefficients:
##               Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 1524319101  621072623  2.4543  0.01668 *
## x1           494988866  778861258  0.6355  0.52722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    6.2729e+20
## Residual Sum of Squares: 6.2359e+20
## R-Squared:      0.0059046
## Adj. R-Squared: -0.0087145
## F-statistic: 0.403897 on 1 and 68 DF, p-value: 0.52722
```

```r
# Breusch-Pagan Lagrange Multiplier for random effects. Null is no panel effect (i.e. OLS better).
plmtest(pool, type=c("bp"))
```

```
##
##  Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels
##
## data:  y ~ x1
## chisq = 2.6692, df = 1, p-value = 0.1023
## alternative hypothesis: significant effects
```

```r
##H0) The null is that there is not cross-sectional dependence
fixed <- plm(y ~ x1, data=dataPanel101, model="within")
pcdtest(fixed, test = c("lm"))
```

```
##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  y ~ x1
## chisq = 28.914, df = 21, p-value = 0.1161
```

```
## alternative hypothesis: cross-sectional dependence
```

```r
pcdtest(fixed, test = c("cd"))
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  y ~ x1
## z = 1.1554, p-value = 0.2479
## alternative hypothesis: cross-sectional dependence
```

```r
##H0) The null is that there is not serial correlation.
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
```

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

```r
##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
```

```r
# 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
```

```r
# 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
```

```r
# Heteroskedasticity consistent coefficients, type 3
coeftest(random, vcovHC(random, type = "HC3"))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1037014329  943438278  1.0992   0.2756
## x1          1247001710  867137595  1.4381   0.1550
```

```r
# The following shows the HC standard errors of the coefficients
t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(random, type = x)))))
```

```
##     (Intercept)        x1
## HC0   907983024 828970258
## HC1   921238952 841072654
## HC2   925403814 847733484
## HC3   943438278 867137595
## HC4   941376025 866024042
```

```r
# 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
```

```r
# Heteroskedasticity consistent coefficients
coeftest(fixed, vcovHC)
```

```
##
## 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
```

```r
# 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
```

```r
# Heteroskedasticity consistent coefficients, type 3
coeftest(fixed, vcovHC(fixed, type = "HC3"))
```

```
##
## t test of coefficients:
```

```
##
##      Estimate Std. Error t value Pr(>|t|)
## x1 2475617742 1439083498  1.7203  0.09037 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# The following shows the HC standard errors of the coefficients
t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(fixed, type = x)))))
```

```
##          HC0.x1     HC1.x1     HC2.x1     HC3.x1     HC4.x1
## [1,] 1358388924 1368196913 1397037348 1439083498 1522166001
```