```
# Panel Data Models in R
# Copyright 2013 by Ani Katchova

# install.packages("plm")
library(plm)

mydata<- read.csv("C:/Econometrics/Data/panel_wage.csv")
attach(mydata)

Y <- cbind(lwage)
X <- cbind(exp, exp2, wks, ed)

# Set data as panel data
pdata <- plm.data(mydata, index=c("id","t"))

# Descriptive statistics
summary(Y)
summary(X)

# Pooled OLS estimator
pooling <- plm(Y ~ X, data=pdata, model= "pooling")
summary(pooling)

# Between estimator
between <- plm(Y ~ X, data=pdata, model= "between")
summary(between)

# First differences estimator
firstdiff <- plm(Y ~ X, data=pdata, model= "fd")
summary(firstdiff)

# Fixed effects or within estimator
fixed <- plm(Y ~ X, data=pdata, model= "within")
summary(fixed)

# Random effects estimator
random <- plm(Y ~ X, data=pdata, model= "random")
summary(random)

# LM test for random effects versus OLS
plmtest(pooling)

# LM test for fixed effects versus OLS
pFtest(fixed, pooling)

# Hausman test for fixed versus random effects model
phtest(random, fixed)
```

```
> # Panel Data Models in R
> # Copyright 2013 by Ani Katchova
>
> # install.packages("plm")
> library(plm)
Loading required package: bdsmatrix

Attaching package: ‘bdsmatrix’

The following object(s) are masked from ‘package:base’:

    backsolve

Loading required package: nlme
Loading required package: Formula
Loading required package: MASS
Loading required package: sandwich
Loading required package: zoo

Attaching package: ‘zoo’

The following object(s) are masked from ‘package:base’:

    as.Date, as.Date.numeric

>
> mydata<- read.csv("C:/Econometrics/Data/panel_wage.csv")
> attach(mydata)
>
> Y <- cbind(lwage)
> X <- cbind(exp, exp2, wks, ed)
>
> # Set data as panel data
> pdata <- plm.data(mydata, index=c("id","t"))
>
> # Descriptive statistics
> summary(Y)
     lwage
 Min.   :4.605
 1st Qu.:6.395
 Median :6.685
 Mean   :6.676
 3rd Qu.:6.953
 Max.   :8.537
> summary(X)
      exp             exp2             wks              ed
 Min.   : 1.00   Min.   :   1.0   Min.   : 5.00   Min.   : 4.00
 1st Qu.:11.00   1st Qu.: 121.0   1st Qu.:46.00   1st Qu.:12.00
 Median :18.00   Median : 324.0   Median :48.00   Median :12.00
 Mean   :19.85   Mean   : 514.4   Mean   :46.81   Mean   :12.85
 3rd Qu.:29.00   3rd Qu.: 841.0   3rd Qu.:50.00   3rd Qu.:16.00
 Max.   :51.00   Max.   :2601.0   Max.   :52.00   Max.   :17.00
>
> # Pooled OLS estimator
> pooling <- plm(Y ~ X, data=pdata, model= "pooling")
```

```
> summary(pooling)
Oneway (individual) effect Pooling Model

Call:
plm(formula = Y ~ X, data = pdata, model = "pooling")

Balanced Panel: n=595, T=7, N=4165

Residuals :
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.160000 -0.250000  0.000273  0.268000  2.130000

Coefficients :
               Estimate  Std. Error  t-value  Pr(>|t|)
(Intercept)  4.9080e+00  6.7330e-02  72.8945 < 2.2e-16 ***
Xexp         4.4675e-02  2.3929e-03  18.6701 < 2.2e-16 ***
Xexp2       -7.1563e-04  5.2794e-05 -13.5552 < 2.2e-16 ***
Xwks         5.8270e-03  1.1826e-03   4.9271 8.673e-07 ***
Xed          7.6041e-02  2.2266e-03  34.1511 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    886.9
Residual Sum of Squares: 635.41
R-Squared      :  0.28356
      Adj. R-Squared :  0.28322
F-statistic: 411.624 on 4 and 4160 DF, p-value: < 2.22e-16
>
> # Between estimator
> between <- plm(Y ~ X, data=pdata, model= "between")
> summary(between)
Oneway (individual) effect Between Model

Call:
plm(formula = Y ~ X, data = pdata, model = "between")

Balanced Panel: n=595, T=7, N=4165

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-0.9780 -0.2200  0.0366  0.2500  0.9860

Coefficients :
               Estimate  Std. Error t-value  Pr(>|t|)
(Intercept)  4.68303917  0.21009890 22.2897 < 2.2e-16 ***
Xexp         0.03815295  0.00569666  6.6974 4.953e-11 ***
Xexp2       -0.00063127  0.00012568 -5.0228 6.757e-07 ***
Xwks         0.01309028  0.00406592  3.2195  0.001355 **
Xed          0.07378378  0.00489848 15.0626 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    92.322
Residual Sum of Squares: 62.187
R-Squared      :  0.32641
```

```
      Adj. R-Squared :  0.32367
F-statistic: 71.4768 on 4 and 590 DF, p-value: < 2.22e-16
>
> # First differences estimator
> firstdiff <- plm(Y ~ X, data=pdata, model= "fd")
> summary(firstdiff)
Oneway (individual) effect First-Difference Model

Call:
plm(formula = Y ~ X, data = pdata, model = "fd")

Balanced Panel: n=595, T=7, N=4165

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-2.11000 -0.06550 -0.00958  0.04840  2.33000

Coefficients :
               Estimate  Std. Error t-value  Pr(>|t|)
(intercept)  0.11706540  0.00631057 18.5507 < 2.2e-16 ***
Xexp2       -0.00053212  0.00013927 -3.8207 0.0001354 ***
Xwks        -0.00026826  0.00056483 -0.4749 0.6348525
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    118.06
Residual Sum of Squares: 117.58
R-Squared      :  0.004108
      Adj. R-Squared :  0.0041046
F-statistic: 7.35691 on 2 and 3567 DF, p-value: 0.0006479
>
> # Fixed effects or within estimator
> fixed <- plm(Y ~ X, data=pdata, model= "within")
> summary(fixed)
Oneway (individual) effect Within Model

Call:
plm(formula = Y ~ X, data = pdata, model = "within")

Balanced Panel: n=595, T=7, N=4165

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-1.81000 -0.05110  0.00371  0.06140  1.94000

Coefficients :
         Estimate  Std. Error t-value  Pr(>|t|)
Xexp   1.1379e-01  2.4689e-03 46.0888 < 2.2e-16 ***
Xexp2 -4.2437e-04  5.4632e-05 -7.7678 1.036e-14 ***
Xwks   8.3588e-04  5.9967e-04  1.3939    0.1634
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    240.65
Residual Sum of Squares: 82.632
```

```
R-Squared      :  0.65663
      Adj. R-Squared :  0.56235
F-statistic: 2273.74 on 3 and 3567 DF, p-value: < 2.22e-16
>
> # Random effects estimator
> random <- plm(Y ~ X, data=pdata, model= "random")
> summary(random)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = Y ~ X, data = pdata, model = "random")

Balanced Panel: n=595, T=7, N=4165

Effects:
                  var std.dev share
idiosyncratic 0.02317 0.15220 0.185
individual    0.10209 0.31952 0.815
theta:  0.8228

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-2.0400 -0.1060  0.0071  0.1150  2.0900

Coefficients :
               Estimate  Std. Error  t-value Pr(>|t|)
(Intercept)  3.8294e+00  9.3634e-02  40.8974   <2e-16 ***
Xexp         8.8861e-02  2.8178e-03  31.5360   <2e-16 ***
Xexp2       -7.7257e-04  6.2262e-05 -12.4083   <2e-16 ***
Xwks         9.6577e-04  7.4329e-04   1.2993   0.1939
Xed          1.1171e-01  6.0572e-03  18.4426   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    260.94
Residual Sum of Squares: 151.35
R-Squared      :  0.42
      Adj. R-Squared :  0.4195
F-statistic: 753.113 on 4 and 4160 DF, p-value: < 2.22e-16
>
> # LM test for random effects versus OLS
> plmtest(pooling)

       Lagrange Multiplier Test - (Honda)

data:  Y ~ X
normal = 72.0564, p-value < 2.2e-16
alternative hypothesis: significant effects

>
> # LM test for fixed effects versus OLS
> pFtest(fixed, pooling)

       F test for individual effects
```

```
data:  Y ~ X
F = 40.2394, df1 = 593, df2 = 3567, p-value < 2.2e-16
alternative hypothesis: significant effects

>
> # Hausman test for fixed versus random effects model
> phtest(random, fixed)

       Hausman Test

data:  Y ~ X
chisq = 6191.428, df = 3, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```