Introduction to plm

Yves Croissant

October 30, 2006

1 Introduction

The aim of package `plm` is to provide an easy way to estimate panel models. Some panel models may be estimated with package `nlme` (*non–linear mixed effect models*), but not in an intuitive way for an econometrician. `plm` provides methods to read panel data, to estimate a wide range of models and to make some tests. This library is loaded using :

```
> library(plm)
```

This document illustrates the features of `plm`, using data available in package `Ecdat`.

```
> library(Ecdat)
```

These data are used in Baltagi (2001).

2 Reading data

With `plm`, data are stored in an object of class `pdata.frame`, which is a `data.frame` with additional attributes describing the structure of the data set. A `pdata.frame` may be created from an ordinary `data.frame` using the `pdata.frame` function or from a text file using the `pread.table` function.

2.1 Reading the data from a data.frame

We illustrate the use of the `pdata.frame` function with the `Produc` data :

```
> data(Produc)
> pdata.frame(Produc, "state", "year", "pprod")
```

The `pdata.frame` function has 4 arguments :

- the name of the `data.frame`,
- `id` : the individual index,

- `time` : the time index,
- `name` : the name under which the `pdata.frame` will be stored.

Observations are assumed to be sorted by individuals first, and by period. The third argument is optional, if `NULL` a new variable called `time` is added. The fourth argument is also optional, if `NULL` the `pdata.frame` is stored under the same name as the `data.frame`.

```
> data(Hedonic)
> pdata.frame(Hedonic, "townid")
```

In case of a balanced panel, the `id` may be the number of individuals. In this case, two new variables (called `id` and `time`) are added.

```
> data(Wages)
> pdata.frame(Wages, 595)
```

A description of the data is obtained using the `summary` method :

```
> summary(Hedonic)
```

```
---------------------------------------------------------------------------------
----------------------------------- Indexes --------------------------------------
Individual index : townid
Time index       : time
---------------------------------------------------------------------------------
------------------------------- Panel Dimensions ---------------------------------
Unbalanced Panel
Number of Individuals        :  92
Number of Time Obserbations  :  from 1 to 30
Total Number of Observations :  506
---------------------------------------------------------------------------------
--------------------------- Time/Individual Variation ----------------------------
no time variation   :  zn indus rad tax ptratio
---------------------------------------------------------------------------------
---------------------------- Descriptive Statistics ------------------------------
       mv               crim                zn               indus        chas
 Min.   : 8.517   Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   no :471
 1st Qu.: 9.742   1st Qu.: 0.08205   1st Qu.:  0.00   1st Qu.: 5.19   yes: 35
 Median : 9.962   Median : 0.25651   Median :  0.00   Median : 9.69
 Mean   : 9.942   Mean   : 3.61352   Mean   : 11.36   Mean   :11.14
 3rd Qu.:10.127   3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10
 Max.   :10.820   Max.   :88.97620   Max.   :100.00   Max.   :27.74

      nox              rm             age              dis
 Min.   :14.82   Min.   :12.68   Min.   :  2.90   Min.   :0.1219
 1st Qu.:20.16   1st Qu.:34.64   1st Qu.: 45.02   1st Qu.:0.7420
```

```
 Median :28.94   Median :38.55   Median : 77.50   Median :1.1655  
 Mean   :32.11   Mean   :39.99   Mean   : 68.57   Mean   :1.1880  
 3rd Qu.:38.94   3rd Qu.:43.87   3rd Qu.: 94.07   3rd Qu.:1.6464  
 Max.   :75.86   Max.   :77.09   Max.   :100.00   Max.   :2.4954  

      rad             tax           ptratio          blacks       
 Min.   :0.000   Min.   :187.0   Min.   :12.60   Min.   :0.00032  
 1st Qu.:1.386   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:0.37538  
 Median :1.609   Median :330.0   Median :19.05   Median :0.39144  
 Mean   :1.868   Mean   :408.2   Mean   :18.46   Mean   :0.35667  
 3rd Qu.:3.178   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:0.39623  
 Max.   :3.178   Max.   :711.0   Max.   :22.00   Max.   :0.39690  

     lstat              townid          time      
 Min.   :-4.0582   29      : 30   1      : 92  
 1st Qu.:-2.6659   84      : 23   2      : 75  
 Median :-2.1747   5       : 22   3      : 60  
 Mean   :-2.2342   83      : 19   4      : 50  
 3rd Qu.:-1.7744   41      : 18   5      : 39  
 Max.   :-0.9684   28      : 15   6      : 33  
                   (Other):379   (Other):157  
```

The printing consists on four sections :

- `indexes` indicates the names of the index variables,
- `panel dimensions` gives information about the dimension of the panel,
- `Time/individual variation` indicates whether some variables have only individual or time variation,
- `Descriptive statistics` gives descriptive statistics about the variables.

2.2 Reading the data from a text file

`pread.table` reads panel data from a text file, with the following syntax :

```
pread.table("c:/mes documents/essai/mydata.txt",
         "firm","year","dataname",header=T,sep=";",dec=",")
```

The arguments of `pread.table` are :

- the text file,
- `id` : the individual index,
- `time` : the time index,

- `name` : the name under which the `pdata.frame` will be stored (if `NULL`, the name of the `pdata.frame` is the name of the file without the path and the extension),
- further arguments that will be passed to `read.table`.

3 Model estimation

A panel model is estimated with the `plm` function.

3.1 Basic use of plm

There are two ways to use `plm` : the first one is to estimate a list of models (the default behavior), the second to estimate just one model. In the first case, the estimated models are :

- the fixed effects model (`within`),
- the pooling model (`pooling`),
- the between model (`between`),
- the error components model (`random`).

The basic use of `plm` is to indicate the model formula and the `pdata.frame` [1] :

```
> zz <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
+     data = pprod)
```

The result of the estimation is stored in a `plms` object which is a list of 4 estimated models, each of them being objects of class `plm`. Each individual model can be easily extracted :

```
> zzwith <- zz$within
```

A particular model to be estimated may also be indicated by filling the `model` argument of `plm`.

```
> zzra <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
+     data = pprod, model = "random")
```

Objects of class `plm` and `plms` have a `print` method.

```
> print(zzra)
```

[1] The following example is from Baltagi (2001), pp. 25–28.

```
Model Formula: log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp

Coefficients:
(intercept)   log(pcap)     log(pc)    log(emp)       unemp
  2.1354110   0.0044386   0.3105484   0.7296705  -0.0061725
```

There is also a `summary` method :

- for `plms` objects, coefficients and standard errors of the fixed effects and the error components models are printed,
- for `plm` object, the table of coefficients and some statistics are printed.

```
> summary(zz)

----------------------------------------------------------------------
_________________________ Model Description __________________________
Oneway (individual) effect

Model Formula        : log(gsp) ~ log(pcap) + log(pc) + log(emp) +
                             unemp
----------------------------------------------------------------------
__________________________ Panel Dimensions __________________________
Balanced Panel
Number of Individuals        :  48
Number of Time Obserbations  :  17
Total Number of Observations :  816
----------------------------------------------------------------------
____________________________ Coefficients ____________________________
                 within         wse      random    rse
(intercept)           .           .  2.13541100 0.1335
log(pcap)   -0.02614965  0.02900158  0.00443859 0.0234
log(pc)      0.29200693  0.02511967  0.31054843 0.0198
log(emp)     0.76815947  0.03009174  0.72967053 0.0249
unemp       -0.00529774  0.00098873 -0.00617247 0.0009
----------------------------------------------------------------------
________________________________ Tests ________________________________
Hausman Test                  : chi2(4) = 9.525416 (p.value=0.04922762)
F Test                        : F(47,764) = 75.8204 (p.value=0)
Lagrange Multiplier Test      : chi2(1) = 4134.961 (p.value=0)
----------------------------------------------------------------------

> summary(zzra)

----------------------------------------------------------------------
_________________________ Model Description __________________________
Oneway (individual) effect
```

```
Random Effect Model (Swamy-Arora's transformation)
Model Formula            : log(gsp) ~ log(pcap) + log(pc) +
                                log(emp) + unemp
----------------------------------------------------------------------
_________________________ Panel Dimensions ___________________________
Balanced Panel
Number of Individuals        :  48
Number of Time Obserbations  :  17
Total Number of Observations :  816
----------------------------------------------------------------------
_____________________________ Effects ________________________________
                    var   std.dev  share
idiosyncratic 0.0014544 0.0381371 0.1754
individual    0.0068377 0.0826905 0.8246
theta   :  0.88884
----------------------------------------------------------------------
____________________________ Residuals _______________________________
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.07e-01 -2.46e-02 -2.37e-03 -9.93e-19  2.17e-02  2.00e-01
----------------------------------------------------------------------
___________________________ Coefficients _____________________________
               Estimate  Std. Error z-value  Pr(>|z|)
(intercept)  2.13541100  0.13346149 16.0002 < 2.2e-16 ***
log(pcap)    0.00443859  0.02341732  0.1895    0.8497
log(pc)      0.31054843  0.01980475 15.6805 < 2.2e-16 ***
log(emp)     0.72967053  0.02492022 29.2803 < 2.2e-16 ***
unemp       -0.00617247  0.00090728 -6.8033 1.023e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
----------------------------------------------------------------------
________________________ Overall Statistics __________________________
Total Sum of Squares       : 29.209
Sum of Squares Residuals   : 1.1879
Rsq                        : 0.95933
F                          : 4782.77
P(F>0)                     : 8.76231e-08
----------------------------------------------------------------------
```

For a `random` model, the `summary` method gives information about the variance of the components of the errors.

`plm`'s can be updated using the `update` method :

```
> zzwithmod <- update(zzwith, . ~ . - unemp - log(emp) + emp)
> zzmod <- update(zz, . ~ . - unemp - log(emp) + emp)
> summary(zzwithmod)
```

```
----------------------------------------------------------------------
_________________________ Model Description ___________________________
Oneway (individual) effect

Model Formula        : log(gsp) ~ log(pcap) + log(pc) + emp
----------------------------------------------------------------------
__________________________ Panel Dimensions ___________________________
Balanced Panel
Number of Individuals        :  48
Number of Time Obserbations  :  17
Total Number of Observations :  816
----------------------------------------------------------------------
____________________________ Coefficients ____________________________
               within        wse     random       rse
(intercept)         .          . 7.1982e-01    0.1846
log(pcap)   1.7888e-01 4.0690e-02 3.4357e-01    0.0322
log(pc)     6.9975e-01 2.9154e-02 6.0369e-01    0.0256
emp         3.7909e-05 8.7824e-06 5.0924e-05 8.218e-06
----------------------------------------------------------------------
_______________________________ Tests ________________________________
Hausman Test                 : chi2(3) = 80.35868 (p.value=0)
F Test                       : F(47,765) = 101.9109 (p.value=0)
Lagrange Multiplier Test     : chi2(1) = 4355.292 (p.value=0)
----------------------------------------------------------------------
```

Fixed effects may be extracted easily from a `plms` or a `plm` object using `FE` :

```
> FE(zzmod)
       ALABAMA        ARIZONA       ARKANSAS     CALIFORNIA       COLORADO
     1.1717531      1.3062389      1.1877004      1.6191982      1.4582149
   CONNECTICUT       DELAWARE        FLORIDA        GEORGIA          IDAHO
     1.7060341      1.2035746      1.5564969      1.4460171      1.1002049
      ILLINOIS        INDIANA           IOWA         KANSAS       KENTUCKY
     1.5496106      1.3451714      1.2323038      1.1735476      1.3492604
     LOUISIANA          MAINE       MARYLAND  MASSACHUSETTS       MICHIGAN
     1.1652834      1.2659480      1.6011871      1.7384231      1.5290312
     MINNESOTA    MISSISSIPPI       MISSOURI        MONTANA       NEBRASKA
     1.3654287      1.1545345      1.4809262      0.7960951      1.0905033
        NEVADA  NEW_HAMPSHIRE     NEW_JERSEY     NEW_MEXICO       NEW_YORK
     1.0627992      1.4138235      1.7420589      1.0925399      1.6694387
NORTH_CAROLINA   NORTH_DAKOTA           OHIO       OKLAHOMA         OREGON
     1.5048751      0.7663694      1.4985974      1.2784660      1.3345094
  PENNSYLVANIA   RHODE_ISLAND SOUTH_CAROLINA   SOUTH_DAKOTA       TENNESSE
     1.4972243      1.5948140      1.2344011      0.8705826      1.3123010
         TEXAS           UTAH        VERMONT       VIRGINIA     WASHINGTON
     1.3230328      1.2464927      1.1804339      1.6175357      1.3492922
```

```
WEST_VIRGINIA      WISCONSIN        WYOMING
    1.0129871      1.4860561      0.7842841
```

3.2 Options for the random effect model

The random effect model is obtained as a linear estimation on quasi–differentiated data. The parameter of this transformation is obtained using preliminary estimations. Four estimators of this parameter are available, depending on the value of the argument `theta.method` :

- `swar` : from Swamy and Arora (1972), the default value,
- `walhus` : from Wallace and Hussain (1969),
- `amemiya` : from Amemiyia (1971),
- `nerlove` : from Nerlove (1971).

For exemple, to use the `amemiya` estimator :

```
> zzra <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
+     data = pprod, model = "random", theta.method = "amemiya")
```

3.3 Choosing the effects

The default behavior of `plm` is to introduce individual effects. Using the `effect` argument, one may also introduce :

- time effects (`effect="time"`),
- individual and time effects (`effect="twoways"`).

For example, to estimate a two–ways effect model for the `Grunfeld` data :

```
> data(Grunfeld)
> pdata.frame(Grunfeld, "firm", "year")
> z <- plm(inv ~ value + capital, data = Grunfeld, effect = "twoways",
+     theta.method = "amemiya")
> summary(z$random)

______________________________________________________________________
_________________________ Model Description ___________________________
Twoways effects
Random Effect Model (Swamy-Arora's transformation)
Model Formula             : inv ~ value + capital
______________________________________________________________________
__________________________ Panel Dimensions ___________________________
Balanced Panel
Number of Individuals        :  10
```

```
Number of Time Obserbations  :  20
Total Number of Observations :  200
----------------------------------------------------------------------
______________________________ Effects _______________________________
                   var  std.dev  share
idiosyncratic 2675.426   51.725 0.2738
individual    7095.252   84.233 0.7262
time             0.000    0.000 0.0000
theta  : 0.86397 (id) 0 (time) 0 (total)
----------------------------------------------------------------------
_____________________________ Residuals ______________________________
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.77e+02 -1.98e+01  4.60e+00  8.77e-16  1.95e+01  2.53e+02
----------------------------------------------------------------------
____________________________ Coefficients ____________________________
              Estimate Std. Error z-value Pr(>|z|)
(intercept) -57.865377  29.393359 -1.9687  0.04899 *
value         0.109790   0.010528 10.4285  < 2e-16 ***
capital       0.308190   0.017171 17.9483  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
----------------------------------------------------------------------
_________________________ Overall Statistics _________________________
Total Sum of Squares       : 2376000
Sum of Squares Residuals   : 547910
Rsq                        : 0.7694
F                          : 328.647
P(F>0)                     : 0.0030381
----------------------------------------------------------------------
```

In the "effects" section of the result is printed now the variance of the three elements of the error term and the three parameters used in the transformation.

The two–ways effect model is for the moment only available for balanced panels.

3.4 Hausman–Taylor's model

Hausman–Taylor's model may be estimated with `plm` by equating the `model` argument to `"ht"` and filling the second argument `instruments` with a formula indicating the variables used as instruments.

```
> data(Wages)
> pdata.frame(Wages, 595)
> form = lwage ~ wks + south + smsa + married + exp + I(exp^2) +
+     bluecol + ind + union + sex + black + ed
> ht = plm(form, ~sex + black + bluecol + south + smsa + ind, data = Wages,
```

```
+     model = "ht")
> summary(ht)

----------------------------------------------------------------------
_________________________ Model Description __________________________
Oneway (individual) effect
Hausman-Taylor Model
Model Formula             : lwage ~ wks + south + smsa + married +
                                  exp + I(exp^2) + bluecol + ind +
                                  union + sex + black + ed
Instrumental Variables    : ~sex + black + bluecol + south +
                                  smsa + ind
Time--Varying Variables
    exogenous variables   :  bluecolyes,southyes,smsayes,ind
    endogenous variables  :  wks,marriedyes,exp,I(exp^2),unionyes
Time--Invariant Variables
    exogenous variables   :  sexmale,blackyes
    endogenous variables  :  ed
----------------------------------------------------------------------
__________________________ Panel Dimensions __________________________
Balanced Panel
Number of Individuals        :  595
Number of Time Obserbations  :  7
Total Number of Observations :  4165
----------------------------------------------------------------------
______________________________ Effects _______________________________
                   var  std.dev  share
idiosyncratic 0.023044 0.151803 0.0253
individual    0.886993 0.941803 0.9747
theta   :  0.93919
----------------------------------------------------------------------
_____________________________ Residuals ______________________________
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.92e+00 -7.07e-02  6.57e-03 -2.46e-17  7.97e-02  2.03e+00
----------------------------------------------------------------------
____________________________ Coefficients ____________________________
               Estimate  Std. Error z-value  Pr(>|z|)
(intercept)  2.7818e+00  3.0768e-01  9.0411 < 2.2e-16 ***
wks          8.3740e-04  5.9981e-04  1.3961   0.16268
southyes     7.4398e-03  3.1959e-02  0.2328   0.81592
smsayes     -4.1833e-02  1.8960e-02 -2.2064   0.02736 *
marriedyes  -2.9851e-02  1.8982e-02 -1.5726   0.11582
exp          1.1313e-01  2.4713e-03 45.7795 < 2.2e-16 ***
I(exp^2)    -4.1886e-04  5.4605e-05 -7.6709 1.710e-14 ***
bluecolyes  -2.0705e-02  1.3783e-02 -1.5022   0.13304
ind          1.3604e-02  1.5239e-02  0.8927   0.37202
```

```
unionyes     3.2771e-02  1.4910e-02  2.1979   0.02796 *
sexmale      1.3092e-01  1.2667e-01  1.0335   0.30135
blackyes    -2.8575e-01  1.5572e-01 -1.8350   0.06651 .
ed           1.3794e-01  2.1251e-02  6.4912 8.518e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
----------------------------------------------------------------------
_________________________ Overall Statistics _________________________
Total Sum of Squares       : 243.04
Sum of Squares Residuals   : 95.947
Rsq                        : 0.60522
F                          : 530.318
P(F>0)                     : 2.88658e-15
----------------------------------------------------------------------
```

3.5 Instrumental variables estimation

One or all of the models may be estimated using instrumental variables by indicating the list of the instrumental variables. This can be done using one of the two following techniques :

- specifying the total list of instruments (using the `instruments` argument of `plm`),
- specifying, on the one hand the external instruments in the argument `instrument` and on the other hand the variables of the model that are assumed to be endogenous in the argument `endog`.

The instrumental variables estimator used may be indicated with the `inst.method` argument :

- `bvk`, from Balestra et Varadharajan–Krishnakumar (1987), the default value,
- `baltagi`, from Baltagi (1981).

We illustrate instrumental variables estimation with the `Crime` data[2]. The same estimation is done using the first syntax (`cr1`) and the second (`cr2`). The `prbarr` and `polpc` variables are assumed to be endogenous and there are two external instruments `taxpc` and `mix` :

```
> data(Crime)
> pdata.frame(Crime, "county", "year")
> form = log(crmrte) ~ log(prbarr) + log(polpc) + log(prbconv) +
+     log(prbpris) + log(avgsen) + log(density) + log(wcon) + log(wtuc) +
+     log(wtrd) + log(wfir) + log(wser) + log(wmfg) + log(wfed) +
```

[2]See Baltagi (2001), pp.119–120.

```
+     log(wsta) + log(wloc) + log(pctymle) + log(pctmin) + region +
+     smsa + year
> inst = ~log(prbconv) + log(prbpris) + log(avgsen) + log(density) +
+     log(wcon) + log(wtuc) + log(wtrd) + log(wfir) + log(wser) +
+     log(wmfg) + log(wfed) + log(wsta) + log(wloc) + log(pctymle) +
+     log(pctmin) + region + smsa + log(taxpc) + log(mix) + year
> inst2 = ~log(taxpc) + log(mix)
> endog = ~log(prbarr) + log(polpc)
> cr = plm(form, data = Crime)
> cr1 = plm(form, inst, data = Crime)
> cr2 = plm(form, inst2, endog, data = Crime)
> summary(cr2$random)

--------------------------------------------------------------------------
_________________________ Model Description ______________________________
Oneway (individual) effect
Random Effect Model (Swamy-Arora's transformation)
Instrumental variable estimation (Balestra-Varadharajan-Krishnakumar's transformation)
Model Formula             : log(crmrte) ~ log(prbarr) + log(polpc) +
                                  log(prbconv) + log(prbpris) +
                                  log(avgsen) + log(density) +
                                  log(wcon) + log(wtuc) + log(wtrd) +
                                  log(wfir) + log(wser) + log(wmfg) +
                                  log(wfed) + log(wsta) + log(wloc) +
                                  log(pctymle) + log(pctmin) +
                                  region + smsa + year
Endogenous Variables    : ~log(prbarr) + log(polpc)
Instrumental Variables    : ~log(taxpc) + log(mix)
--------------------------------------------------------------------------
__________________________ Panel Dimensions ______________________________
Balanced Panel
Number of Individuals       :  90
Number of Time Obserbations  :  7
Total Number of Observations :  630
--------------------------------------------------------------------------
______________________________ Effects ___________________________________
                   var  std.dev share
idiosyncratic 0.022269 0.149228 0.326
individual    0.046036 0.214561 0.674
theta   :  0.74576
--------------------------------------------------------------------------
_____________________________ Residuals __________________________________
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-5.02e+00 -4.76e-01  2.73e-02  7.11e-16  5.26e-01  3.19e+00
--------------------------------------------------------------------------
____________________________ Coefficients ________________________________
```

```
               Estimate Std. Error z-value  Pr(>|z|)
(intercept)   -0.4538241  1.7029840 -0.2665  0.789864
log(prbarr)   -0.4141200  0.2210540 -1.8734  0.061015 .
log(polpc)     0.5049285  0.2277811  2.2167  0.026642 *
log(prbconv)  -0.3432383  0.1324679 -2.5911  0.009567 **
log(prbpris)  -0.1900437  0.0733420 -2.5912  0.009564 **
log(avgsen)   -0.0064374  0.0289406 -0.2224  0.823977
log(density)   0.4343519  0.0711528  6.1045 1.031e-09 ***
log(wcon)     -0.0042963  0.0414225 -0.1037  0.917392
log(wtuc)      0.0444572  0.0215449  2.0635  0.039068 *
log(wtrd)     -0.0085626  0.0419822 -0.2040  0.838387
log(wfir)     -0.0040302  0.0294565 -0.1368  0.891175
log(wser)      0.0105604  0.0215822  0.4893  0.624620
log(wmfg)     -0.2017917  0.0839423 -2.4039  0.016220 *
log(wfed)     -0.2134634  0.2151074 -0.9924  0.321023
log(wsta)     -0.0601083  0.1203146 -0.4996  0.617362
log(wloc)      0.1835137  0.1396721  1.3139  0.188884
log(pctymle)  -0.1458448  0.2268137 -0.6430  0.520214
log(pctmin)    0.1948760  0.0459409  4.2419 2.217e-05 ***
regionwest    -0.2281780  0.1010317 -2.2585  0.023916 *
regioncentral -0.1987675  0.0607510 -3.2718  0.001068 **
smsayes       -0.2595423  0.1499780 -1.7305  0.083535 .
year82         0.0132140  0.0299923  0.4406  0.659518
year83        -0.0847676  0.0320008 -2.6489  0.008075 **
year84        -0.1062004  0.0387893 -2.7379  0.006184 **
year85        -0.0977398  0.0511685 -1.9102  0.056113 .
year86        -0.0719390  0.0605821 -1.1875  0.235045
year87        -0.0396520  0.0758537 -0.5227  0.601153
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
----------------------------------------------------------------------
_________________________ Overall Statistics _________________________
Total Sum of Squares       : 1354.7
Sum of Squares Residuals   : 557.64
Rsq                        : 0.58836
F                          : 33.1494
P(F>0)                     : 7.77156e-16
----------------------------------------------------------------------
```

3.6 Variable coefficients model

If there is enough time observations for each individual, the model may be estimate for each individual. The `nopool` function provide this kind of estimation. It can be done using :

- directly the `nopool` function,

- plm with the argument np fixed to TRUE.

With the Grunfeld data, we get :

```
> znp <- nopool(inv ~ value + capital, data = Grunfeld)
```

or

```
> z <- plm(inv ~ value + capital, data = Grunfeld, np = TRUE)
> znp <- z$nopool
> print(znp)
```

```
    (intercept)       value     capital
1  -149.7824533 0.119280833 0.371444807
2   -49.1983219 0.174856015 0.389641889
3    -9.9563065 0.026551189 0.151693870
4    -6.1899605 0.077947821 0.315718185
5    22.7071160 0.162377704 0.003101737
6    -8.6855434 0.131454842 0.085374274
7    -4.4995344 0.087527198 0.123781407
8    -0.5093902 0.052894126 0.092406492
9    -7.7228371 0.075387943 0.082103558
10    0.1615186 0.004573432 0.437369190
```

```
> summary(znp)
```

```
  (intercept)           value               capital
 Min.   :-149.782   Min.   :0.004573   Min.   :0.003102
 1st Qu.:  -9.639   1st Qu.:0.058518   1st Qu.:0.087132
 Median :  -6.956   Median :0.082738   Median :0.137738
 Mean   : -21.368   Mean   :0.091285   Mean   :0.205264
 3rd Qu.:  -1.507   3rd Qu.:0.128411   3rd Qu.:0.357513
 Max.   :  22.707   Max.   :0.174856   Max.   :0.437369
```

The result is an object of class nopool. The print method presents the coefficients estimated for each individual. The summary method gives descriptive statistics for these coefficients.

3.7 Unbalanced panel

plm offers limited support for unbalanced panels. The following example is based on the Hedonic data[3]:

```
> form = mv ~ crim + zn + indus + chas + nox + rm + age + dis +
+     rad + tax + ptratio + blacks + lstat
> ba = plm(form, data = Hedonic)
> summary(ba$random)
```

[3]See Baltagi (2001), p. 174.

```
----------------------------------------------------------------------
________________________ Model Description ___________________________
Oneway (individual) effect
Random Effect Model (Swamy-Arora's transformation)
Model Formula            : mv ~ crim + zn + indus + chas + nox +
                                rm + age + dis + rad + tax +
                                ptratio + blacks + lstat
----------------------------------------------------------------------
_________________________ Panel Dimensions ___________________________
Unbalanced Panel
Number of Individuals        :  92
Number of Time Obserbations  :  from 1 to 30
Total Number of Observations :  506
----------------------------------------------------------------------
_____________________________ Effects ________________________________
                   var  std.dev share
idiosyncratic 0.016965 0.130249 0.502
individual    0.016832 0.129738 0.498
theta  :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.2915  0.5904  0.6655  0.6499  0.7447  0.8197
----------------------------------------------------------------------
____________________________ Residuals _______________________________
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.641000 -0.066100 -0.000519 -0.001990  0.069800  0.527000
----------------------------------------------------------------------
___________________________ Coefficients _____________________________
               Estimate  Std. Error  z-value  Pr(>|z|)
(intercept)  9.6778e+00  2.0714e-01  46.7207 < 2.2e-16 ***
crim        -7.2338e-03  1.0346e-03  -6.9921 2.707e-12 ***
zn           3.9575e-05  6.8778e-04   0.0575 0.9541153
indus        2.0794e-03  4.3403e-03   0.4791 0.6318706
chasyes     -1.0591e-02  2.8960e-02  -0.3657 0.7145720
nox         -5.8630e-03  1.2455e-03  -4.7074 2.509e-06 ***
rm           9.1773e-03  1.1792e-03   7.7828 7.105e-15 ***
age         -9.2715e-04  4.6468e-04  -1.9952 0.0460159 *
dis         -1.3288e-01  4.5683e-02  -2.9088 0.0036279 **
rad          9.6863e-02  2.8350e-02   3.4168 0.0006337 ***
tax         -3.7472e-04  1.8902e-04  -1.9824 0.0474298 *
ptratio     -2.9723e-02  9.7538e-03  -3.0473 0.0023089 **
blacks       5.7506e-01  1.0103e-01   5.6920 1.256e-08 ***
lstat       -2.8514e-01  2.3855e-02 -11.9533 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
----------------------------------------------------------------------
________________________ Overall Statistics __________________________
```

```
Total Sum of Squares       : 893.08
Sum of Squares Residuals   : 8.6843
Rsq                        : 0.99028
F                          : 3854.18
P(F>0)                     : 0
----------------------------------------------------------------------
```

4 Tests

4.1 Tests of poolability

`pooltest` tests the hypothesis that the same coefficients apply to each individual. It is a standard F test, based on the comparison of a model obtained for the full sample and a model based on the estimation of an equation for each individual. The main argument of `pooltest` is a `plm` object. If the model has been estimated with the argument `np=F`, one has to indicate a second argument of class `nopool`. A third argument `effect` should be fixed to `FALSE` if the intercepts are assumed to be identical (the default value) or `TRUE` if not[4].

```
> form = inv ~ value + capital
> znp = nopool(form, data = Grunfeld)
> zplm = plm(form, data = Grunfeld)
> pooltest(zplm, znp)

        F statistic

data:  zplm
F = 27.7486, df1 = 27, df2 = 170, p-value < 2.2e-16

> pooltest(zplm, znp, effect = T)

        F statistic

data:  zplm
F = 5.7805, df1 = 18, df2 = 170, p-value = 1.219e-10

> z = plm(form, data = Grunfeld, effect = "time", np = TRUE)
> pooltest(z, effect = F)

        F statistic

data:  z
F = 1.1204, df1 = 57, df2 = 140, p-value = 0.2928
```

[4] The following examples are from Baltagi (2001), pp. 57–58.

4.2 Tests for individual and time effects

4.2.1 Lagrange multiplier tests

`plmtest` implements tests of individual or/and time effects based on the results of the pooling model. It's main argument is a `plm` object (the result of a pooling model) or a `plms` object.

Two additional arguments can be added to indicate the kind of test to be computed. The argument `type` is whether :

- `bp` : Breusch–Pagan (1980), the default value,
- `honda` : Honda (1985),
- `kw` : King and Wu (1997).

The effects tested are indicated with the `effect` argument :

- `individual` for individual effects (the default value),
- `time` for time effects,
- `twoways` for individuals and time effects.

Some examples of the use of `plmtest` are shown below[5]:

```
> library(Ecdat)
> g <- plm(inv ~ value + capital, data = Grunfeld)
> plmtest(g)

        Lagrange Multiplier Test - individual effects (Breush-Pagan)

data:  g
chi2 = 798.1615, df = 1, p-value < 2.2e-16

> plmtest(g, effect = "time")

        Lagrange Multiplier Test - time effects (Breush-Pagan)

data:  g
chi2 = 6.4539, df = 1, p-value = 0.01107

> plmtest(g, type = "honda")

        Lagrange Multiplier Test - individual effects (Honda)

data:  g
normal = 28.2518, p-value < 2.2e-16
```

[5]See Baltagi (2001), p. 65.

```
> plmtest(g, type = "ghm", effect = "twoways")

        Lagrange Multiplier Test - two-ways effects (Gourierroux, Holly and
        Monfort)

data:  g
chi2 = 798.1615, df = 2, p-value < 2.2e-16

> plmtest(g, type = "kw", effect = "twoways")

        Lagrange Multiplier Test - two-ways effects (King and Wu)

data:  g
normal = 21.8322, df = 2, p-value < 2.2e-16
```

4.2.2 F tests

`pFtest` computes F tests of effects based on the comparison of the `within` and the `pooling` models. Its arguments are whether a `plms` object or two `plm` objects (the results of a `pooling` and a `within` model). Some examples of the use of `pFtest` are shown below[6]:

```
> library(Ecdat)
> gi <- plm(inv ~ value + capital, data = Grunfeld)
> gt <- plm(inv ~ value + capital, data = Grunfeld, effect = "time")
> gd <- plm(inv ~ value + capital, data = Grunfeld, effect = "twoways")
> pFtest(gi)

        F test for effects

data:  gi
F = 49.1766, df1 = 9, df2 = 188, p-value < 2.2e-16

> pFtest(gi$within, gi$pooling)

        F test for effects

data:  gi$within and gi$pooling
F = 49.1766, df1 = 9, df2 = 188, p-value < 2.2e-16

> pFtest(gt)

        F test for effects

data:  gt
F = 0.5229, df1 = 9, df2 = 188, p-value = 0.8569
```

[6] Voir Baltagi (2001), p. 65.

```
> pFtest(gd)

        F test for effects

data:  gd
F = 17.4031, df1 = 28, df2 = 169, p-value < 2.2e-16
```

4.3 Hausman's test

`phtest` computes the HAUSMAN's test which is based on the comparison of two models. It's main argument may be :

- a `plms` object. In this case, the two models used in the test are the `within` and the `random` models (the most usual case with panel data),
- two `plm` objects.

Some examples of the use of `phtest` are shown below [7]:

```
> g <- plm(inv ~ value + capital, data = Grunfeld)
> phtest(g)

        Hausman Test

data:  g
chi2 = 2.3304, df = 2, p-value = 0.3119

> phtest(g$between, g$random)

        Hausman Test

data:  g$between and g$random
chi2 = 2.1314, df = 3, p-value = 0.5456
```

5 Bibiographie

Amemiyia, T. (1971), The estimation of the variances in a variance–components model, *International Economic Review*, 12, pp.1–13.

Balestra, P. and J. Varadharajan–Krishnakumar (1987), Full information estimations of a system of simultaneous equations with error components structure, *Econometric Theory*, 3, pp.223–246.

Baltagi, B.H. (1981), Simultaneous equations with error components, *Journal of econometrics*, 17, pp.21–49.

[7] See BALTAGI (2001), p. 71.

Baltagi, B.H. (2001) *Econometric Analysis of Panel Data.* John Wiley and sons. ltd.

Breusch, T.S. and A.R. Pagan (1980), The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies*, 47, pp.239–253.

Gourieroux, C., A. Holly and A. Monfort (1982), Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica*, 50, pp.63–80.

Hausman, G. (1978), Specification tests in econometrics, *Econometrica*, 46, pp.1251–1271.

Hausman, J.A. and W.E. Taylor (1981), Panel data and unobservable individual effects, *Econometrica*, 49, pp.1377–1398.

Honda, Y. (1985), Testing the error components model with non–normal disturbances, *Review of Economic Studies*, 52, pp.681–690.

King, M.L. and P.X. Wu (1997), Locally optimal one–sided tests for multiparameter hypotheses, *Econometric Reviews*, 33, pp.523–529.

Nerlove, M. (1971), Further evidence on the estimation of dynamic economic relations from a time–series of cross–sections, *Econometrica*, 39, pp.359–382.

Swamy, P.A.V.B. and S.S. Arora (1972), The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica*, 40, pp.261–275.

Wallace, T.D. and A. Hussain (1969), The use of error components models in combining cross section with time series data, *Econometrica*, 37(1), pp.55–72.