ANALYSIS OF PANEL DATA

Fixed-Effect and Random-Effect Models

datascience@berkeley

Fixed-Effect Model

An Example: The Effect of Job Training on Firm Scrap Rates

The plm Package

```
plm(formula, data, subset, na.action, effect = c("individual", "time", "twoways"),
    model = c("within", "random", "ht", "between", "pooling", "fd"),
    random.method = c("swar", "walhus", "amemiya", "nerlove", "kinla"),
    random.dfcor = NULL,
    inst.method = c("bvk", "baltagi", "am", "bmc"), restrict.matrix = NULL,
    restrict.rhs = NULL, index = NULL, ...)
```

plm is a general function for the estimation of linear panel models. It supports the following estimation methods: pooled OLS (model = "pooling"), fixed effects ("within"), random effects ("random"), first–differences ("fd"), and between ("between"). It supports unbalanced panels and two–way effects (although not with all methods).

For random effects models, four estimators of the transformation parameter are available by setting random.method to one of "swar" (Swamy and Arora (1972)) (default), "amemiya" (Amemiya (1971)), "walhus" (Wallace and Hussain (1969)), or "nerlove" (Nerlove (1971)).

Instrumental variables estimation is obtained using two–part formulas, the second part indicating the instrumental variables used. This can be a complete list of instrumental variables or an update of the first part. If, for example, the model is y ~ x1 + x2 + x3, with x1 and x2 endogenous and z1 and z2 external instruments, the model can be estimated with:

- formula=y~x1+x2+x3 | x3+z1+z2,
- formula=y~x1+x2+x3 | .-x1-x2+z1+z2.

- To structure a panel data, we use each row of the data for a specific individual in a particular time period.
- In $ data.frame can be used, but it includes an argument called *index* to indicate the structure of the data. This can be:
- NULL (the default value), it is then assumed that the first two columns contain the individual and the time index and that observations are ordered by individual and by time period
- a character string, which should be the name of the individual index
- a character vector of length two containing the names of the individual and the time index
- an integer which is the number of individuals (only in case of a balanced panel with observations ordered by individual)

- The pdata.frame function is then called internally, which returns a pdata.frame which is a data.frame with an attribute called index.
- The **plm** package is very rich and provides four estimation function. In this course, we will only use the *plm* function, as it already provides estimation for both fixed effect and random effect models.
- A nice feature of these functions is that is share the same interface as that of the *lm()* function. Their first two arguments are **formula** and **data**
- Of the other arguments, I want to highlight **index**, which we discussed above, and **$**, which is use to indicate the kind of effects to be included in the model. That is, *individual effect, time effect*, or both.
- Data Transformation: The *within* transformation, $Q = I_{nT} - P$, where $P = \frac{1}{T} I_n xjj'$ returns a vector containing individual means and I_z is a $z \times z$ identity matrix.
- The **within** function can be used to perform the within transformation

Value

An object of class `c("plm","panelmodel")`.

A `"plm"` object has the following elements :

`coefficients`	the vector of coefficients,
`vcov`	the covariance matrix of the coefficients,
`residuals`	the vector of residuals,
`df.residual`	degrees of freedom of the residuals,
`formula`	an object of class `'pFormula'` describing the model,
`model`	a data.frame of class `'pdata.frame'` containing the variables used for the estimation: the response is in first column and the two indexes in the two last columns,
`ercomp`	an object of class `'ercomp'` providing the estimation of the components of the errors (for random effects models only),
`call`	the call.

Effect of Job Training Program on Scrap Rates Revisit

```
# Use the plm.data() function to transform a data frame in a format suitable
for using with the estimation functions of plm.
jtrain.panel<-plm.data(jtrain, c("fcode","year"))
summary(jtrain.panel)
```

Description

This function transforms a data frame in a format suitable for using with the estimation functions of plm.

Usage

```
plm.data(x, indexes = NULL)
```

Arguments

x	a data.frame,
indexes	a vector (of length one or two) indicating the (individual and time) indexes.

indexes can be:

- a character string which is the name of the individual index variable, in this case a new variable called "time" containing the time index is added,
- an integer, the number of individuals in the case of balanced panel, in this case two new variables "time" and "id" containing the individual and the time indexes are added,
- a vector of two character strings which contains the names of the individual and of the time indexes.

```
> summary(jtrain.panel)
     fcode          year          employ            sales             avgsal
 410032 :  3   1987:157   Min.   :  4.00   Min.   :  110000   Min.   : 4237
 410440 :  3   1988:157   1st Qu.: 15.00   1st Qu.: 1550000   1st Qu.:14102
 410495 :  3   1989:157   Median : 30.00   Median : 3000000   Median :17773
 410500 :  3              Mean   : 59.32   Mean   : 6116037   Mean   :18873
 410501 :  3              3rd Qu.: 72.00   3rd Qu.: 7700000   3rd Qu.:22360
 410509 :  3              Max.   :525.00   Max.   :54000000   Max.   :42583
 (Other):453              NA's   :31       NA's   :98         NA's   :65
     scrap            rework           tothrs           union            grant
Min.   : 0.0100   Min.   : 0.000   Min.   :  0.0   Min.   :0.0000   Min.   :0.0000
1st Qu.: 0.5925   1st Qu.: 0.350   1st Qu.:  0.0   1st Qu.:0.0000   1st Qu.:0.0000
Median : 1.4150   Median : 1.160   Median : 12.0   Median :0.0000   Median :0.0000
Mean   : 3.8436   Mean   : 3.474   Mean   : 29.2   Mean   :0.1975   Mean   :0.1401
3rd Qu.: 4.0000   3rd Qu.: 4.000   3rd Qu.: 40.0   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :30.0000   Max.   :40.000   Max.   :320.0   Max.   :1.0000   Max.   :1.0000
NA's   :309       NA's   :348      NA's   :56
      d89              d88             totrain          hrsemp           lscrap
Min.   :0.0000   Min.   :0.0000   Min.   :  0.00   Min.   :  0.000   Min.   :-4.6052
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:  0.00   1st Qu.:  0.000   1st Qu.:-0.5234
Median :0.0000   Median :0.0000   Median :  8.00   Median :  3.308   Median : 0.3471
Mean   :0.3333   Mean   :0.3333   Mean   : 23.09   Mean   : 14.968   Mean   : 0.3937
3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 25.00   3rd Qu.: 18.663   3rd Qu.: 1.3863
Max.   :1.0000   Max.   :1.0000   Max.   :350.00   Max.   :163.917   Max.   : 3.4012
                                  NA's   :6        NA's   :81        NA's   :309
    lemploy          lsales          lrework          lhrsemp          lscrap_1
Min.   :1.386   Min.   :11.61   Min.   :-4.6052   Min.   :0.000   Min.   :-4.6052
1st Qu.:2.708   1st Qu.:14.25   1st Qu.:-0.9163   1st Qu.:0.000   1st Qu.:-0.2675
Median :3.401   Median :14.91   Median : 0.1823   Median :1.460   Median : 0.4414
Mean   :3.531   Mean   :15.03   Mean   : 0.1642   Mean   :1.650   Mean   : 0.5129
3rd Qu.:4.277   3rd Qu.:15.86   3rd Qu.: 1.3863   3rd Qu.:2.979   3rd Qu.: 1.6094
Max.   :6.263   Max.   :17.80   Max.   : 3.6889   Max.   :5.105   Max.   : 3.4012
NA's   :31      NA's   :98      NA's   :350       NA's   :81      NA's   :363
```

First-Difference Method

```
> jtrain.fd <- plm(lscrap ~ hrsemp+lsales+lemploy,data=jtrain.panel, model="fd")
> summary(jtrain.fd)
Oneway (individual) effect First-Difference Model

Call:
plm(formula = lscrap ~ hrsemp + lsales + lemploy, data = jtrain.panel,
    model = "fd")

Unbalanced Panel: n=47, T=1-3, N=135

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-3.0600 -0.1110  0.0838  0.2770  2.6500

Coefficients :
             Estimate Std. Error t-value Pr(>|t|)
(intercept) -0.2112469  0.0734771 -2.8750 0.005118 **
hrsemp      -0.0012945  0.0022429 -0.5771 0.565393
lsales      -0.3475582  0.3436699 -1.0113 0.314770
lemploy      0.2589719  0.4105471  0.6308 0.529886
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    34.021
Residual Sum of Squares: 33.353
R-Squared:      0.019647
Adj. R-Squared: 0.018754
F-statistic: 0.561147 on 3 and 84 DF, p-value: 0.64214
```

Fixed-Effect Method

```
> jtrain.fe2 <- plm(lscrap ~ hrsemp+lsales+lemploy,data=jtrain.panel, model="within")
> summary(jtrain.fe2)
Oneway (individual) effect Within Model

Call:
plm(formula = lscrap ~ hrsemp + lsales + lemploy, data = jtrain.panel,
    model = "within")

Unbalanced Panel: n=47, T=1-3, N=135

Residuals :
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.800000 -0.124000  0.000966  0.136000  1.610000

Coefficients :
          Estimate Std. Error t-value Pr(>|t|)
hrsemp  -0.0041585  0.0024653 -1.6868  0.09532 .
lsales  -0.5616153  0.3611430 -1.5551  0.12364
lemploy  0.3655257  0.4389659  0.8327  0.40735
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    27.287
Residual Sum of Squares: 24.889
R-Squared:      0.087879
Adj. R-Squared: 0.055331
F-statistic: 2.7298 on 3 and 85 DF, p-value: 0.048909
```

Berkeley SCHOOL OF INFORMATION