plm

Linear models for panel data

Yves Croissant

LET Université de Lyon 2

UseR! 2006, Vienna

Introduction
Reading the data
Transforming the data
Estimation
Test
Conclusion

Motivations

- panel data is one of the main fields in modern econometrics,

Yves Croissant plm

Introduction
Reading the data
Transforming the data
Estimation
Test
Conclusion

Motivations

- panel data is one of the main fields in modern econometrics,
- all econometric softwares have implemented at least basic panel data estimation and test procedures,

Yves Croissant plm

Motivations

- panel data is one of the main fields in modern econometrics,
- all econometric softwares have implemented at least basic panel data estimation and test procedures,
- no R's package provide the tools to estimate panel data model in a straightforward way.

Main characteristics

- panel datas have specific structure. A new class (`pdata.frame`) is provided, which add to `data.frame` objects some attributes useful for panel data.

Main characteristics

- panel datas have specific structure. A new class (`pdata.frame`) is provided, which add to `data.frame` objects some attributes useful for panel data.
- panel data estimation consist on different estimators. The main estimation function provided, called `plm` returns a list of several models, and not just one model,

Main characteristics

- panel datas have specific structure. A new class (`pdata.frame`) is provided, which add to `data.frame` objects some attributes useful for panel data.
- panel data estimation consist on different estimators. The main estimation function provided, called `plm` returns a list of several models, and not just one model,
- all the basic panel models are linear models (ordinary or generalized least squares models) and can therefore be estimated with the `lm` function.

Outline of the presentation

- reading the data,
- transformation functions,
- estimation of the model,
- tests.

pdata.frame

The `pdata.frame` function takes as arguments a `data.frame` and two variables which represents the individual and time index. It returns a `pdata.frame` object, *i.e.* a `data.frame` with additional attributes.

The pdata.frame attributes

- `pvar` : a list of two logical vectors which indicates if each variable has an individual or a time variation.

The pdata.frame attributes

- `pvar` : a list of two logical vectors which indicates if each variable has an individual or a time variation.
- `pdim` : a list which describe the dimension of the panel, *i.e.* the number of time observations for each individual,

The pdata.frame attributes

- pvar : a list of two logical vectors which indicates if each variable has an individual or a time variation.
- pdim : a list which describe the dimension of the panel, *i.e.* the number of time observations for each individual,
- indexes : the name of the index variables.

Introduction
Reading the data
Transforming the data
Estimation
Test
Conclusion
pdata.frame
The pdata.frame attributes
An example

The pdata.frame attributes

- pvar : a list of two logical vectors which indicates if each variable has an individual or a time variation.
- pdim : a list which describe the dimension of the panel, *i.e.* the number of time observations for each individual,
- indexes : the name of the index variables.
- data : a data attribute for each variable in the pdata.frame.

An example

```
> library(plm)
> library(Ecdat)
> data(Grunfeld)
> Grunfeld = pdata.frame(Grunfeld, firm, year)
> Grunfeld = pdata.frame(Grunfeld, firm)
> Grunfeld[1:2, ]
  firm year   inv  value capital time
1    1 1935 317.6 3078.5     2.8    1
2    1 1936 391.8 4661.7    52.6    2

> Grunfeld = pdata.frame(Grunfeld, 10)
> Grunfeld[1:2, ]
  firm year   inv  value capital time id
1    1 1935 317.6 3078.5     2.8    1  1
2    1 1936 391.8 4661.7    52.6    2  1
```


The pmean function

```
> pmean(Grunfeld$inv, Grunfeld$firm)
> pmean(Grunfeld$inv)
> Grunfeld$inv - pmean(Grunfeld$inv)
> pmean(Grunfeld$inv, effect = "time")
```

plag and pdiff functions

```
> plag(Grunfeld$inv)
> plag(Grunfeld$inv, order = 2)
> pdiff(Grunfeld$inv)
> pdiff(Grunfeld$inv, lag = 1)
```


Example of use : the between model

```
> lm(pmean(inv) ~ pmean(value) + pmean(capital),
      data = Grunfeld)

Call:
lm(formula = pmean(inv) ~ pmean(value) + pmean(capital),
    data = Grunfeld)

Coefficients:
   (Intercept)    pmean(value)  pmean(capital)
      -8.52711         0.13465         0.03203
```


Typology of panel models

- balanced *vs* unbalanced panel,
- one–way *vs* two–ways effects,
- instrumental variables or not,

Basic usage of the plm function

```
> form = inv ~ value + capital
> gow = plm(form, data = Grunfeld)
> summary(gow)
```

	between	bse	within	wse	random	rs
(intercept)	-8.5271	47.5153	.	.	-57.8344	28.9
value	0.1346	0.0287	0.1101	0.0119	0.1098	0.0
capital	0.0320	0.1909	0.3101	0.0174	0.3081	0.0

```
> summary(gow$within)

Model formula: inv ~ value + capital

Residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.84e+02 -1.76e+01  5.63e-01 -2.12e-15  1.92e+01  2.51e+02

        Estimate Std. Error z-value Pr(>|z|)
value     0.1101     0.0119    9.29   <2e-16 ***
capital   0.3101     0.0174   17.87   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```


Time and two–ways effects

```
> gowt = plm(inv ~ value + capital, effect = "time",
+      data = Grunfeld)
> gtw = plm(inv ~ value + capital, effect = "double",
+      data = Grunfeld)
```


Different random effect models

```
> gow = plm(inv ~ value + capital,data = Grunfeld,
       theta="amemiya")
> gow = plm(inv ~ value + capital,data = Grunfeld,
       theta="nerlove")
> gow = plm(inv ~ value + capital,data = Grunfeld,
       theta="walhus")
> gow = plm(inv ~ value + capital,data = Grunfeld,
       theta="swar")
```


Instrumental Variable estimator

```
> data(Crime)
> Crime = pdata.frame(Crime, county, year)
> form = log(crmrte) ~ log(prbarr) + log(polpc) + log(prbconv
+     log(pctymle) + log(pctmin) + region + smsa
> inst = ~log(prbconv) + log(pctymle) + log(pctmin) + region
+     smsa + log(taxpc) + log(mix)
> cr = plm(form, inst, data = Crime)
```

Four main tests for panel data

- test of poolability
- tests of the presence of individual effects,
- tests of correlation of the effects with explanatory variables,
- tests of autocorrelation and heteroscedasticity.

Tests of poolability

The pooltest function is just a Chow test of stability applied to panel data.

```
> gow = plm(form, data = Grunfeld, np = T)
> pooltest(gow)
```

```
F statistic

data:  gow
F = 27.7486, p-value < 2.2e-16
alternative hypothesis: true is stability
```


ols residuals : lagrange multiplier test

```
> plmtest(gow)
```

```
   Lagrange Multiplier Test - individual effects
    ( Breush-Pagan )

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


within fixed effects (F test)

```
> pFtest(gow)
```

```
F statistic

data:  data.name
F = 49.1766, p-value < 2.2e-16
alternative hypothesis: true is null.value
```

Hausman test

```
> phtest(gow)

Hausman Test

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


- estimation of systems of equations,

- estimation of systems of equations,
- estimation of robust covariance matrices,

- estimation of systems of equations,
- estimation of robust covariance matrices,
- estimation of dynamic panel models,