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,
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 data has 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,
Outline of the presentation

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

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.

- `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 describes 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

```r
> 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

```r
> pmean(Grunfeld$inv, Grunfeld$firm)
> pmean(Grunfeld$inv)
> Grunfeld$inv - pmean(Grunfeld$inv)
> pmean(Grunfeld$inv, effect = "time")
```
### Typology of panel models

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

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

### Basic usage of the `plm` function

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

```r
Call:
  lm(formula = pmean(inv) ~ pmean(value) + pmean(capital),
     data = Grunfeld)
Coefficients:
              Value
(Intercept) -8.52711
pmean(value)  0.13465
pmean(capital)  0.03203
```

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

```r
Call:
  lm(formula = pmean(inv) ~ pmean(value) + pmean(capital),
     data = Grunfeld)
Coefficients:
              Value
(Intercept) -8.52711
pmean(value)  0.13465
pmean(capital)  0.03203
```

### Example of use: the between model

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

```r
Call:
  lm(formula = pmean(inv) ~ pmean(value) + pmean(capital),
     data = Grunfeld)
Coefficients:
              Value
(Intercept) -8.52711
pmean(value)  0.13465
pmean(capital)  0.03203
```
Time and two-ways effects

```r
> 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
```

Different random effect models

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

```r
> 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

```r
> phtest(gow)

Hausman Test

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

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