

# Package ‘pglm’

July 2, 2014

**Version** 0.1-2

**Date** 2013-12-27

**Title** panel generalized linear model

**Depends** R (>= 2.10), maxLik, plm

**Imports** statmod

**Suggests** lmtest, car

**Description** Estimation of panel models for glm-like models: this includes binomial models (logit and probit) count models (poisson and negbin) and ordered models (logit and probit)

**License** GPL (>= 2)

**URL** <http://www.r-project.org>

**Author** Yves Croissant [aut, cre]

**Maintainer** Yves Croissant <yves.croissant@univ-reunion.fr>

**Repository** CRAN

**Repository/R-Forge/Project** pglm

**Repository/R-Forge/Revision** 8

**Repository/R-Forge/DateTimeStamp** 2013-12-28 07:37:40

**Date/Publication** 2013-12-28 16:02:08

**NeedsCompilation** no

## R topics documented:

Fairness	2
HealthIns	3
PatsRD	4
pglm	5
Unions	7
<b>Index</b>	<b>8</b>

---

Fairness

*Perceived Fairness of Different Rules for Allocating Seats in Trains and Parking Spaces*

---

### Description

401 individuals who answer 14 questions

*number of observations* : 5614

*observation* : Individuals

*country* : France

### Usage

`data(Fairness)`

### Format

A dataframe containing :

**id** the individual index,

**answer** a factor with levels 0 (very unfair), 1 (essentially unfair), 2 (essentially fair) and 3 (very fair),

**good** one of ' tgv ' (French fast train) and 'Parking',

**rule** the allocation rule, a factor with levels 'peak', 'admin', 'lottery', 'addsupply', 'queuing', 'moral' and 'compensation',

**driving** does the individual has the driving license ?

**education** does the individual has a diploma ?

**recurring** does the allocation problem is reccuring ?

### Source

Raux C., Souche S. and Croissant Y., 2009, How fair is pricing to be? An empirical study, *Public Choice*, vol. 139, 1, pp.227-240.

### References

Raux C., Souche S. and Croissant Y., 2009, How fair is pricing to be? An empirical study, *Public Choice*, vol. 139, 1, pp.227-240.

---

HealthIns

*Health Insurance and Doctor Visits*

---

### Description

an unbalanced panel of 5908 individuals for 1 to 5 years

*number of observations* : 20186

*observation* : Individuals

*country* : United States

### Usage

```
data(HealthIns)
```

### Format

A dataframe containing :

**id** the individual index,

**year** the year,

**med** medical expenses,

**mdu** number of face-to face medical visits,

**coins** coinsurance rate,

**disease** count of chronic diseases,

**sex** a factor with level 'male' and 'female',

**age** the age,

**size** the size of the family,

**child** a factor with levels 'no' and 'yes'.

### Source

Rand Health Insurance Experiment

### References

Manning W. G., Newhouse J. P., Duan N., Keeler E. B. and Leibowitz A. (1987) : “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment”, *American Economic Review*, 77(3), pp. 251-277.

Deb P. and P. K. Trivedi (2002) : “The structure of demand for medical care: latent class versus two-part models”, *Journal of Health Economics*, 21, pp. 601-625.

Cameron C. and P. K. Trivedi (2008) : *Microeconometrics Using Stata*, Stata Press.

---

PatsRD

*Dynamic Relation Between Patents and R&D*


---

### Description

a panel of 346 observations from 1970 to 1979

*number of observations* : 3460

*observation* : production units

*country* : United States

### Usage

data(PatsRD)

### Format

A dataframe containing :

**cusip** Compustat's identifying number for the firm (Committee on Uniform Security Identification Procedures number)

**year** year

**ardsic** a two-digit code for the applied R&D industrial classification (roughly that in Bound, Cummins, Griliches, Hall, and Jaffe, in the Griliches R&D, Patents, and Productivity volume)

**scisect** is the firm in the scientific sector ?

**logk** the logarithm of the book value of capital in 1972.

**sumpat** the sum of patents applied for between 1972-1979.

**logr** the logarithm of R&D spending during the year (in 1972 dollars)

**pat** the number of patents applied for during the year that were eventually granted

### Source

Hall, Bronwyn , Zvi Griliches and Jerry Hausman (1986) "Patents and R&D: Is There a Lag?", *International Economic Review*, **27**, 265-283.

### References

Cameron, A.C. and Trivedi P.K. (1998) *Regression analysis of count data*, Cambridge University Press, <http://cameron.econ.ucdavis.edu/racd/racddata.html>, chapter 9.

Cameron, A.C. and P.K. Trivedi (2005) *Microeconometrics : methods and applications*, Cambridge, pp. 792-5.

**Description**

Estimation by maximum likelihood of glm (binomial and Poisson) and 'glm-like' models (Negbin and ordered) on longitudinal data

**Usage**

```
pglm(formula, data, subset, na.action,
      effect = c("individual", "time", "twoways"),
      model = c("random", "pooling", "within", "between"),
      family, other = NULL, index = NULL, start = NULL, R = 20, ...)
```

**Arguments**

formula	a symbolic description of the model to be estimated,
data	the data: a pdata.frame object or an ordinary data.frame,
subset	an optional vector specifying a subset of observations,
na.action	a function which indicates what should happen when the data contains 'NA's,
effect	the effects introduced in the model, one of "individual", "time" or "twoways",
model	one of "pooling", "within", "between", "random",,
family	the distribution to be used,
other	for developer's use only,
index	the index,
start	a vector of starting values,
R	the number of function evaluation for the gaussian quadrature method used,
...	further arguments.

**Value**

An object of class "pglm", a list with elements:

coefficients	the named vector of coefficients,
logLik	the value of the log-likelihood,
hessian	the hessian of the log-likelihood at convergence,
gradient	the gradient of the log-likelihood at convergence,
call	the matched call,
est.stat	some information about the estimation (time used, optimisation method),
freq	the frequency of choice,
residuals	the residuals,

fitted.values    the fitted values,  
 formula        the formula (a mFormula object),  
 expanded.formula  
                  the formula (a formula object),  
 model          the model frame used,  
 index          the index of the choice and of the alternatives.

### Author(s)

Yves Croissant

### Examples

```
## an ordered probit example
data('Fairness', package = 'pglm')
Parking <- subset(Fairness, good == 'parking')
op <- pglm(as.numeric(answer) ~ education + rule,
           Parking[1:105, ],
           family = ordinal('probit'), R = 5, print.level = 3,
           method = 'bfgs', index = 'id', model = "random")

## a binomial (logit) example
data('Unions', package = 'pglm')
anb <- pglm(union ~ wage + exper + rural, Unions, family = binomial('probit'),
           model = "pooling", method = "bfgs", print.level = 3, R = 5)

## a gaussian example on unbalanced panel data
data(Hedonic, package = "plm")
ra <- pglm(mv ~ crim + zn + indus + nox + age + rm, Hedonic, family = gaussian,
           model = "random", print.level = 3, method = "nr", index = "townid")

## some count data models
data("PatsRD", package="pglm")
la <- pglm(pat ~ lag(logr, 0:5) + scisect + logk + factor(year), PatsRD,
           family = negbin, model = "within", print.level=3, method="nr",
           index=c('cusip', 'year'))
la <- pglm(pat ~ lag(logr, 0:5) + scisect + logk + factor(year), PatsRD,
           family = poisson, model = "pooling", index = c("cusip", "year"),
           print.level = 0, method="nr")

## a tobit example
data("HealthIns", package="pglm")
HealthIns$med2 <- HealthIns$med / 1000
HealthIns2 <- HealthIns[-2209, ]
set.seed(2)
subs <- sample(1:20186, 200, replace = FALSE)
HealthIns2 <- HealthIns2[subs, ]
la <- pglm(med ~ mdu + disease + age, HealthIns2,
           model = 'random', family = 'tobit', print.level = 0,
           method = 'nr', R = 5)
```

---

Unions	<i>Unions and Wages</i>
--------	-------------------------

---

**Description**

a balanced panel of 545 individuals for 8 years 1980 to 1987

*number of observations* : 4360

*observation* : Individuals

*country* : United States

**Usage**

data(Unions)

**Format**

A dataframe containing :

**id** the individual index,

**year** the year,

**age** the age,

**exper** the experience, computed as age - 6 - schooling,

**union** does the wage is set by collective bargaining,

**married** is the individual married ?

**ethn** one of 'black', 'hisp' and 'other',

**disability** does the individual has health disability ?

**rural** does the individual lives in a rural area ?

**region** the region, one of 'NorthEast', 'NothernCentral', 'South' and 'other',

**wage** hourly wage in US dollars,

**sector** one of 'agricultural', 'mining', 'construction', 'trade', 'transportation', 'finance', 'businessrepair', 'personalservice', 'entertainment', 'manufacturing', 'pro.rel.service', 'pub.admin',

**occ** one of 'proftech', 'manoffpro', 'sales', 'clerical', 'craftfor', 'operative', 'laborfarm', 'farmlabor', 'service'.

**Source**

National Longitudinal Survey (Youth Sample)

**References**

Vella F. and Verbeek M. (1998) Whose wages do unions raise ? A dynamic model of unionism and wage rate determination for young men, *Journal of Applied Economics*, 13, pp.163-183.

# Index

## \*Topic **datasets**

Fairness, [2](#)

HealthIns, [3](#)

PatsRD, [4](#)

Unions, [7](#)

## \*Topic **regression**

pglm, [5](#)

Fairness, [2](#)

HealthIns, [3](#)

negbin (pglm), [5](#)

ordinal (pglm), [5](#)

PatsRD, [4](#)

pglm, [5](#)

Unions, [7](#)