

# Package ‘plm’

November 7, 2017

**Version** 1.6-6

**Date** 2017-10-31

**Title** Linear Models for Panel Data

**Depends** R (>= 3.1.0), Formula (>= 0.2-0), stats

**Imports** MASS, bdsmatrix, zoo, nlme, sandwich, lattice, lmtest, maxLik

**Suggests** AER, car, pcse, clusterSEs, clubSandwich, pglm, spdep, splm,  
statmod, Ecdat, pder, stargazer, texreg, foreign

**Description** A set of estimators and tests for panel data econometrics.

**License** GPL (>= 2)

**URL** <https://cran.r-project.org/package=plm>,  
<https://r-forge.r-project.org/projects/plm/>

**Author** Yves Croissant [aut, cre],  
Giovanni Millo [aut],  
Kevin Tappe [aut],  
Ott Toomet [ctb],  
Christian Kleiber [ctb],  
Achim Zeileis [ctb],  
Arne Henningsen [ctb],  
Liviu Andronic [ctb],  
Nina Schoenfelder [ctb]

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

**Repository** CRAN

**Repository/R-Forge/Project** plm

**Repository/R-Forge/Revision** 644

**Repository/R-Forge/DateTimeStamp** 2017-11-07 03:41:50

**Date/Publication** 2017-11-07 17:22:35 UTC

**NeedsCompilation** no

**R topics documented:**

plm-package . . . . .	3
aneweytest . . . . .	4
Cigar . . . . .	5
cipstest . . . . .	6
cortab . . . . .	8
Crime . . . . .	8
detect_lin_dep . . . . .	10
dynformula . . . . .	13
EmplUK . . . . .	14
ercomp . . . . .	15
fixef.plm . . . . .	17
Gasoline . . . . .	19
Grunfeld . . . . .	20
Hedonic . . . . .	21
index . . . . .	23
is.pbalanced . . . . .	24
is.pconsecutive . . . . .	25
LaborSupply . . . . .	28
lag . . . . .	29
make.pbalanced . . . . .	31
make.pconsecutive . . . . .	33
Males . . . . .	36
model.frame.pFormula . . . . .	37
mttest . . . . .	39
nobs . . . . .	40
Parity . . . . .	41
pbgtest . . . . .	42
pbltest . . . . .	44
pbsytest . . . . .	45
pcce . . . . .	48
pcdtest . . . . .	49
pdata.frame . . . . .	52
pdim . . . . .	55
pdwtest . . . . .	56
pFormula . . . . .	58
pFtest . . . . .	59
pggls . . . . .	60
pgmm . . . . .	62
pgrangertest . . . . .	65
pht . . . . .	67
phtest . . . . .	68
piest . . . . .	70
pldv . . . . .	71
plm . . . . .	72
plm.data . . . . .	76
plmtest . . . . .	78

pmg	80
pmodel.response	82
pooltest	83
Produc	84
pseries	85
punbalancedness	87
purtest	90
pvar	93
pvcmm	94
pwaldtest	96
pwartest	98
pwfdtest	100
pwtest	101
r.squared	103
ranef.plm	104
RiceFarms	105
sargan	106
Snmesp	107
SumHes	108
summary.plm	109
vcovBK	111
vcovDC	113
vcovG	115
vcovHC	116
vcovNW	119
vcovSCC	121
Wages	123
within_intercept	124
<b>Index</b>	<b>126</b>

---

plm-package

*plm package: linear models for panel data*

---

## Description

plm is a package for R which intends to make the estimation of linear panel models straightforward. plm provides functions to estimate a wide variety of models and to make (robust) inference.

## Details

For a gentle and comprehensive introduction to the package, please see the package's vignette.

The main functions to estimate models are:

`plm` panel data estimators using `lm` on transformed data,

`pgmm` generalized method of moments (GMM) estimation for panel data,

pvcmm variable coefficients models for panel data,

pmg mean groups (MG), demeaned MG and common correlated effects (CCEMG) estimators.

Next to the model estimation functions, the package offers several functions for statistical tests related to panel data/models.

Multiple functions for (robust) variance–covariance matrices are at hand as well.

The package also provides data sets to demonstrate functions and to replicate some text book/paper results. Use `data(package="plm")` to view a list of available data sets in the package.

### Examples

```
data("Produc", package = "plm")
zz <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
          data = Produc, index = c("state", "year"))
summary(zz)

# replicates some results from Baltagi (2013), table 3.1
data("Grunfeld", package = "plm")
p <- plm(inv ~ value + capital,
         data = Grunfeld, model="pooling")

wi <- plm(inv ~ value + capital,
          data = Grunfeld, model="within", effect = "twoways")

swar <- plm(inv ~ value + capital,
            data = Grunfeld, model="random", effect = "twoways")

amemiya <- plm(inv ~ value + capital,
               data = Grunfeld, model = "random", random.method = "amemiya",
               effect = "twoways")

walhus <- plm(inv ~ value + capital,
              data = Grunfeld, model = "random", random.method = "walhus",
              effect = "twoways")
```

---

aneweytest

*Chamberlain estimator and test for fixed effects*

---

### Description

Angrist and Newey's version of the Chamberlain test

### Usage

```
aneweytest(formula, data, subset, na.action, index = NULL, ...)
```

**Arguments**

formula            a symbolic description for the model to be estimated,  
 data                a `data.frame`,  
 subset             see `lm`,  
 na.action         see `lm`,  
 index              the indexes,  
 ...                 further arguments.

**Details**

Angrist and Newey's test is based on the results of the artifactual regression of the within residuals on the covariates for all the period.

**Value**

An object of class "htest".

**Author(s)**

Yves Croissant

**References**

Angrist, J.D. and Newey, W.K. (1991). Over-identification tests in earnings functions with fixed effects, *Journal of Business & Economic Statistics*, **9(3)**, pp. 317–323.

**Examples**

```
data("RiceFarms", package = "plm")
aneweytest(log(goutput) ~ log(seed) + log(totlabor) + log(size), RiceFarms, index = "id")
```

---

Cigar

*Cigarette Consumption*

---

**Description**

a panel of 46 observations from 1963 to 1992

*total number of observations* : 1380

*observation* : regional

*country* : United States

**Usage**

```
data(Cigar)
```

**Format**

A data frame containing :

**state** state abbreviation

**year** the year

**price** price per pack of cigarettes

**pop** population

**pop16** population above the age of 16

**cpi** consumer price index (1983=100)

**ndi** per capita disposable income

**sales** cigarette sales in packs per capita

**pimin** minimum price in adjoining states per pack of cigarettes

**Source**

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

**References**

Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.

Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.

Baltagi, B.H. and D. Levin (1992) "Cigarette taxation: raising revenues and reducing consumption", *Structural Changes and Economic Dynamics*, **3**(2), pp. 321–335.

Baltagi, B.H., J.M. Griffin and W. Xiong (2000) "To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand", *Review of Economics and Statistics*, **82**(1), pp. 117–126.

---

cipstest

*Cross-sectionally Augmented IPS Test for Unit Roots in Panel Models*

---

**Description**

Cross-sectionally augmented Im, Pesaran and Shin (IPS) test for unit roots in panel models.

**Usage**

```
cipstest(x, lags = 2, type = c("trend", "drift", "none"),
        model = c("cmg", "mg", "dmg"), truncated = FALSE, ...)
```

**Arguments**

x	an object of class "pseries",
lags	lag order for Dickey-Fuller augmentation,
type	one of "trend", "drift", "none",
model	one of "cmg", "mg", "dmg",
truncated	logical specifying whether to calculate the truncated version of the test,
...	further arguments passed to critvals.

**Details**

This cross-sectionally augmented version of the IPS unit root test ( $H_0$ : the pseries has a unit root) is a so-called second-generation panel unit root test: it is in fact robust against cross-sectional dependence, provided that the default `type="cmg"` is calculated. Else one can obtain the standard (`model="mg"`) or cross-sectionally demeaned (`model="dmg"`) versions of the IPS test.

**Value**

An object of class "htest".

**Author(s)**

Giovanni Millo

**References**

Pesaran, M.H. (2007) A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics*, **22**(2), pp. 265–312.

**See Also**

[purtest](#)

**Examples**

```
data("Produc", package = "plm")
Produc <- pdata.frame(Produc, index=c("state", "year"))
## check whether the gross state product (gsp) is trend-stationary
cipstest(Produc$gsp, type = "trend")
```

---

cortab	<i>Cross-sectional correlation matrix</i>
--------	---

---

**Description**

Computes the cross-sectional correlation matrix

**Usage**

```
cortab(x, grouping, groupnames = NULL, value = "statistic", ...)
```

**Arguments**

x	an object of class pseries
grouping	grouping variable,
groupnames	a character vector of group names,
value	to complete
...	further arguments

**Value**

A matrix

---

Crime	<i>Crime in North Carolina</i>
-------	--------------------------------

---

**Description**

a panel of 90 observational units (counties) from 1981 to 1987

*total number of observations* : 630

*observation* : regional

*country* : United States

**Usage**

```
data(Crime)
```



**Format**

A data frame containing :

**county** county identifier

**year** year from 1981 to 1987

**crmrte** crimes committed per person

**prbarr** 'probability' of arrest

**prbconv** 'probability' of conviction

**prbpris** 'probability' of prison sentence

**avgsen** average sentence, days

**polpc** police per capita

**density** people per square mile

**taxpc** tax revenue per capita

**region** factor. One of 'other', 'west' or 'central'.

**smsa** factor. (Also called "urban".) Does the individual reside in a SMSA (standard metropolitan statistical area)?

**pctmin** percentage minority in 1980

**wcon** weekly wage in construction

**wtuc** weekly wage in transportation, utilities, communications

**wtrd** weekly wage in wholesale and retail trade

**wfir** weekly wage in finance, insurance and real estate

**wser** weekly wage in service industry

**wmfg** weekly wage in manufacturing

**wfed** weekly wage in federal government

**wsta** weekly wage in state government

**wloc** weekly wage in local government

**mix** offence mix: face-to-face/other

**pctymle** percentage of young males (between ages 15 to 24)

**Source**

Journal of Applied Econometrics Data Archive (complements Baltagi (2006)):

<http://qed.econ.queensu.ca/jae/2006-v21.4/baltagi/>

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

See also Journal of Applied Econometrics data archive entry for Baltagi (2006) at <http://qed.econ.queensu.ca/jae/2006-v21.4/baltagi/>.

## References

- Cornwell, C. and W.N. Trumbull (1994) “Estimating the economic model of crime with panel data”, *Review of Economics and Statistics*, **76**(2), pp. 360–366.
- Baltagi, B. H. (2006) “Estimating an economic model of crime using panel data from North Carolina”, *Journal of Applied Econometrics*, **21**(4), pp. 543–547.
- Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.
- Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.

---

detect\_lin\_dep                      *Functions to detect linear dependence*

---

## Description

Little helper functions to aid users to detect linear dependent columns in a two-dimensional data structure, especially in a (transformed) model matrix - typically useful in interactive mode during model building phase.

## Usage

```
detect_lin_dep(object, ...)
## S3 method for class 'matrix'
detect_lin_dep(object, suppressPrint = FALSE, ...)
## S3 method for class 'data.frame'
detect_lin_dep(object, suppressPrint = FALSE, ...)
## S3 method for class 'plm'
detect_lin_dep(object, suppressPrint = FALSE, ...)

## S3 method for class 'plm'
alias(object, ...)
## S3 method for class 'pFormula'
alias(object, data,
       model = c("pooling", "within", "Between", "between", "mean", "random", "fd"),
       effect = c("individual", "time", "twoways"), ...)
```

## Arguments

object	for detect_lin_dep: an object which should be checked for linear dependence (of class "matrix", "data.frame", or "plm"); for alias: either an estimated model of class "plm" or a "pFormula". Usually, one wants to input a model matrix here or check an already estimated plm model,
suppressPrint	for detect_lin_dep only: logical indicating whether a message shall be printed; defaults to printing the message, i. e. to suppressPrint = FALSE,
data	for alias.pFormula only: model frame (created by plm's model.frame method),
model, effect	for alias.pFormula only: model and effect to specify the data transformation, see <a href="#">plm</a> ,
...	further arguments.

## Details

Linear dependence of columns/variables is (usually) readily avoided when building one's model. However, linear dependence is sometimes not obvious and harder to detect for less experienced applied statisticians. The so called "dummy variable trap" is a common and probably the best-known fallacy of this kind (see e. g. Wooldridge (2016), sec. 7-2.). When building linear models with `lm` or `plm`'s pooling model, linear dependence in one's model is easily detected, at times post hoc.

However, linear dependence might also occur after some transformations of the data, albeit it is not present in the untransformed data. The within transformation (also called fixed effect transformation) used in the "within" model can result in such linear dependence and this is harder to come to mind when building a model. See **Examples** for two examples of linear dependent columns after the within transformation: ex. 1) the transformed variables have the opposite sign of one another; ex. 2) the transformed variables are identical.

During `plm`'s model estimation, linear dependent columns and their corresponding coefficients in the resulting object are silently dropped, while the corresponding model frame and model matrix still contain the affected columns. The `plm` object contains an element `aliased` which indicates any such aliased coefficients by a named logical.

Both functions, `detect_lin_dep` and `alias`, help to detect linear dependence and accomplish almost the same: `detect_lin_dep` is a stand alone implementation while `alias` is a wrapper around `alias.lm`, extending the `alias` generic to classes "plm" and "pFormula". `alias` hinges on the availability of the package **MASS** on the system. Not all arguments of `alias.lm` are supported. Output of `alias` is more informative as it gives the linear combination of dependent columns (after data transformations, i. e. after (quasi-)demeaning) while `detect_lin_dep` only gives columns involved in the linear dependence in a simple format (thus being more suited for automatic post-processing of the information).

## Value

For `detect_lin_dep`: A named numeric vector containing column numbers of the linear dependent columns in the object after data transformation, if any are present. `NULL` if no linear dependent columns are detected.

For `alias`: return value of `alias.lm` run on the (quasi-)demeaned model, i. e. the information outputted applies to the transformed model matrix, not the original data.

## Author(s)

Kevin Tappe

## References

Wooldridge, J.M. (2016) *Introductory Econometrics: A Modern Approach*, 6th ed., Cengage Learning, Boston, sec. 7-2, pp. 206–211.

## See Also

`alias`, `model.matrix` and especially `plm`'s `model.matrix` for (transformed) model matrices, `model.frame`.

## Examples

```

### Example 1 ###
# prepare the data
data("Cigar" , package = "plm")
Cigar[ , "fact1"] <- c(0,1)
Cigar[ , "fact2"] <- c(1,0)
Cigar.p <- pdata.frame(Cigar)

# setup a pFormula and a model frame
pform <- pFormula(price ~ 0 + cpi + fact1 + fact2)
mf <- model.frame(pform, data = Cigar.p)

# no linear dependence in the pooling model's model matrix
# (with intercept in the formula, there would be linear dependence)
detect_lin_dep(model.matrix(pform, data = mf, model = "pooling"))

# linear dependence present in the FE transformed model matrix
modmat_FE <- model.matrix(pform, data = mf, model = "within")
detect_lin_dep(modmat_FE)
mod_FE <- plm(pform, data = Cigar.p, model = "within")
detect_lin_dep(mod_FE)
alias(mod_FE) # => fact1 == -1*fact2
plm(pform, data = mf, model = "within")$aliased # "fact2" indicated as aliased

# look at the data: after FE transformation fact1 == -1*fact2
head(modmat_FE)
all.equal(modmat_FE[ , "fact1"], -1*modmat_FE[ , "fact2"])

### Example 2 ###
# Setup the data:
# Assume CEOs stay with the firms of the Grunfeld data
# for the firm's entire lifetime and assume some fictional
# data about CEO tenure and age in year 1935 (first observation
# in the data set) to be at 1 to 10 years and 38 to 55 years, respectively.
# => CEO tenure and CEO age increase by same value (+1 year per year).
data(Grunfeld, package = "plm")
set.seed(42)
# add fictional data
Grunfeld$CEOtenure <- c(replicate(10, seq(from=s<-sample(1:10, 1), to=s+19, by=1)))
Grunfeld$CEOage <- c(replicate(10, seq(from=s<-sample(38:65, 1), to=s+19, by=1)))

# look at the data
head(Grunfeld, 50)

pform <- pFormula(inv ~ value + capital + CEOtenure + CEOage)
mf <- model.frame(pform, data=pdata.frame(Grunfeld))

# no linear dependent columns in original data/pooling model
modmat_pool <- model.matrix(pform, data = mf, model="pooling")
detect_lin_dep(modmat_pool)
mod_pool <- plm(pform, data = Grunfeld, model = "pooling")
alias(mod_pool)

```

```

# CEOtenure and CEOage are linear dependent after FE transformation
# (demeaning per individual)
modmat_FE <- model.matrix(pform, data = mf, model="within")
detect_lin_dep(modmat_FE)
mod_FE <- plm(pform, data = Grunfeld, model = "within")
detect_lin_dep(mod_FE)
alias(mod_FE)

# look at the transformed data: after FE transformation CEOtenure == 1*CEOage
head(modmat_FE, 50)
all.equal(modmat_FE[, "CEOtenure"], modmat_FE[, "CEOage"])

```

---

dynformula

*Dynamic Formula*


---

## Description

A function to easily create a formula with lags and differences

## Usage

```
dynformula(formula, lag.form = NULL, diff.form = NULL, log.form = NULL)
```

## Arguments

formula	a formula,
lag.form	a list containing the lag structure of each variable in the formula,
diff.form	a vector (or a list) of logical values indicating whether variables should be differenced,
log.form	a vector (or a list) of logical values indicating whether variables should be in logarithms.

## Details

lag.form is a list, diff.form and log.form are vectors (or lists) that should be of length equal to the total number of variables. Each element of these lists/vectors is:

- either a vector of length 2 (c(1,4) means lags 1,2,3 and 4) or a scalar (3 means lags 0,1,2,3 except for the left-hand side variable for which it is 1,2,3) for lag.form.
- logical values for diff.form and log.form.

It can also be an incomplete named list/vector (for example, to apply the transformation for only some variables) with eventually an unnamed element which then is the default value.

**Value**

An object of class `c("dynformula", "formula")`, which is a formula with four additional attributes: `var`, the names of the variables in the formula, `lag`, `diff`, and `log`, which store the information about lags, differences and logs, respectively.

A `formula` method coerces the `dynformula` object to a standard formula.

**Author(s)**

Yves Croissant

**Examples**

```
# all variables in log, x1, x2 and x3 lagged twice, y lagged once and x3 differenced
z <- dynformula(y ~ x1 + x2 + x3, lag.form = list(2, y = 1),
               diff.form = c(x3 = TRUE), log.form = TRUE)
formula(z)
```

---

EmplUK

*Employment and Wages in the United Kingdom*

---

**Description**

An unbalanced panel of 140 observations from 1976 to 1984

*total number of observations* : 1031

*observation* : firms

*country* : United Kingdom

**Usage**

```
data(EmplUK)
```

**Format**

A data frame containing :

**firm** firm index

**year** year

**sector** the sector of activity

**emp** employment

**wage** wages

**capital** capital

**output** output

**Source**

Arellano, M. and Bond, S. (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies*, **58**(2), 1991, pp. 227–297.

ercomp

*Estimation of the error components***Description**

This function enables the estimation of the variance components of a panel model.

**Usage**

```
ercomp(object, ...)
## S3 method for class 'formula'
ercomp(object, data,
        effect = c("individual", "time", "twoways", "nested"),
        method = NULL, models = NULL,
        dfcor = NULL, index = NULL, ...)
## S3 method for class 'plm'
ercomp(object, ...)
## S3 method for class 'ercomp'
print(x, digits= max(3, getOption("digits") - 3), ...)
```

**Arguments**

object	a formula or a plm object,
data	a data.frame,
effect	the effects introduced in the model, see <a href="#">plm</a> for details,
method	method of estimation for the variance components, see <a href="#">plm</a> for details,
models	the models used to estimate the variance components (an alternative to the previous argument),
dfcor	a numeric vector of length 2 indicating which degree of freedom should be used,
index	the indexes,
x	an ercomp object,
digits	digits,
...	further arguments.

**Value**

An object of class "ercomp": a list containing

- sigma2 a named numeric with estimates of the variance components,
- theta contains the parameter(s) used for the transformation of the variables: For a one-way model, a numeric corresponding to the selected effect (individual or time); for a two-ways model a list of length 3 with the parameters. In case of a balanced model, the numeric has length 1 while for an unbalanced model, the numerics' length equal the number of observations.

**Author(s)**

Yves Croissant

**References**

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

Nerlove, M. (1971) Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica*, **39**(2), pp. 359–382.

Swamy, P.A.V.B. and Arora, S.S. (1972) The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica*, **40**(2), pp. 261–275.

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

**See Also**

[plm](#) where the estimates of the variance components are used if a random effects model is estimated

**Examples**

```
data("Produc", package = "plm")
# an example of the formula method
ercomp(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc,
       method = "walhus", effect = "time")
# same with the plm method
z <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
        data = Produc, random.method = "walhus",
        effect = "time", model = "random")
ercomp(z)
# a two-ways model
ercomp(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc,
       method = "amemiya", effect = "twoways")
```



---

fixef.plm	<i>Extract the Fixed Effects</i>
-----------	----------------------------------

---

## Description

Function to extract the fixed effects from a plm object and associated summary method.

## Usage

```
## S3 method for class 'plm'
fixef(object, effect = NULL,
      type = c("level", "dfirst", "dmean"), vcov = NULL, ...)
## S3 method for class 'fixef'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'fixef'
summary(object, ...)
## S3 method for class 'summary.fixef'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
```

## Arguments

x,object	an object of class "plm", an object of class "fixef" for the print and the summary method,
effect	one of "individual" or "time", only relevant in case of two-ways effects models,
vcov	a variance-covariance matrix furnished by the user or a function to calculate one (see <b>Examples</b> ),
type	one of "level", "dfirst", or "dmean",
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

## Details

Function `fixef` calculates the fixed effects and returns an object of class `c("fixef", "numeric")`. By setting the `type` argument, the fixed effects may be returned in levels ("level"), as deviations from the first value of the index ("dfirst"), or as deviations from the overall mean ("dmean"). If the argument `vcov` was specified, the standard errors (stored as attribute "se" in the return value) are the respective robust standard errors.

The associated summary method returns an extended object of class `c("summary.fixef", "matrix")` with more information (see sections **Value** and **Examples**).

References with formulae (except for the two-ways unbalanced case) are, e.g., Greene (2012), Ch. 11.4.4, p. 364, formulae (11-25); Wooldridge (2010), Ch. 10.5.3, pp. 308-309, formula (10.58).

**Value**

For function `fixef` an object of class `c("fixef", "numeric")` is returned:

It is a numeric vector containing the fixed effects with attribute `se` which contains the standard errors. There are two further attributes: attribute `type` contains the chosen type (the value of argument `type` as a character); attribute `df.residual` holds the residual degrees of freedom (integer) from the fixed effects model (plm object) on which `fixef` was run.

For function `summary.fixef` an object of class `c("summary.fixef", "matrix")` is returned:

It is a matrix with four columns in this order: the estimated fixed effects, their standard errors and associated `t`-values and `p`-values. The type of the fixed effects and the standard errors in the `summary.fixef` objects correspond to was requested in the `fixef` function by arguments `type` and `vcov`, respectively.

**Author(s)**

Yves Croissant

**References**

Greene, W.H. (2012), *Econometric Analysis*, 7th ed., Prentice Hall, Ch. 11.4.4.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Ch. 10.5.3.

**See Also**

[within\\_intercept](#) for the overall intercept of fixed effect models along its standard error, [plm](#) for plm objects and within models (= fixed effects models) in general. See [ranef](#) to extract the random effects from a random effects model.

**Examples**

```
data("Grunfeld", package = "plm")
gi <- plm(inv ~ value + capital, data = Grunfeld, model = "within")
fixef(gi)
summary(fixef(gi))
summary(fixef(gi))[ , c("Estimate", "Pr(>|t|)")] # only estimates and p-values

# relationship of type = "dmean" and "level" and overall intercept
fx_level <- fixef(gi, type = "level")
fx_dmean <- fixef(gi, type = "dmean")
overallint <- within_intercept(gi)
all.equal(overallint + fx_dmean, fx_level, check.attributes = FALSE) # TRUE

# extract time effects in a twoways effects model
gi_tw <- plm(inv ~ value + capital, data = Grunfeld,
             model = "within", effect = "twoways")
fixef(gi_tw, effect = "time")

# with supplied variance-covariance matrix as matrix, function,
# and function with additional arguments
fx_level_robust1 <- fixef(gi, vcov = vcovHC(gi))
```

```

fx_level_robust2 <- fixef(gi, vcov = vcovHC)
fx_level_robust3 <- fixef(gi, vcov = function(x) vcovHC(x, method = "white2"))
summary(fx_level_robust1) # gives fixed effects, robust SEs, t- and p-values

# calc. fitted values of oneway within model:
fixefs <- fixef(gi)[index(gi, which = "id")]
fitted_by_hand <- fixefs + gi$coefficients["value"] * gi$model$value +
                    gi$coefficients["capital"] * gi$model$capital

```

---

Gasoline

*Gasoline Consumption*

---

### Description

A panel of 18 observations from 1960 to 1978

*total number of observations* : 342

*observation* : country

*country* : OECD

### Usage

```
data(Gasoline)
```

### Format

A data frame containing :

**country** a factor with 18 levels

**year** the year

**lgaspcar** logarithm of motor gasoline consumption per car

**lincomep** logarithm of real per-capita income

**lrpmg** logarithm of real motor gasoline price

**lcarpcap** logarithm of the stock of cars per capita

### Source

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

## References

- Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.
- Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.
- Baltagi, B.H. and J.M. Griffin (1983) “Gasoline demand in the OECD: An application of pooling and testing procedures”, *European Economic Review*, **22**(2), pp. 117–137.

---

Grunfeld

*Grunfeld's Investment Data*

---

## Description

A balanced panel of 10 observational units (firms) from 1935 to 1954

*total number of observations* : 200

*observation* : production units

*country* : United States

## Usage

```
data(Grunfeld)
```

## Format

A data frame containing :

**firm** observation

**year** date

**inv** gross Investment

**value** value of the firm

**capital** stock of plant and equipment

## Note

The Grunfeld data as provided in package `plm` is the same data as used in Baltagi (2001), see **Examples** below.

NB:

Various versions of the Grunfeld data circulate online. Also, various text books (and also varying among editions) and papers use different subsets of the original Grunfeld data, some of which contain errors in a few data points compared to the original data used by Grunfeld (1958) in his PhD thesis. See Kleiber/Zeileis (2010) and its accompanying website for a comparison of various Grunfeld data sets in use.

**Source**

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/> <http://www.wiley.com/legacy/wileychi/baltagi/supp/Grunfeld.fil>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

**References**

Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons; accompanying website <http://www.wiley.com/legacy/wileychi/baltagi/>.

Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons; accompanying website <http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>.

Grunfeld, Yehuda (1958) *The Determinants of Corporate Investment*, Ph.D. thesis, Department of Economics, University of Chicago.

Kleiber, C./Zeileis, A. (2010) “The Grunfeld Data at 50”, *German Economic Review*, **11**(4), pp. 404–417, <http://dx.doi.org/10.1111/j.1468-0475.2010.00513.x>; website accompanying the paper with various variants of the Grunfeld data: <http://statmath.wu-wien.ac.at/~zeileis/grunfeld/>.

**See Also**

For the complete Grunfeld data (11 firms), see [Grunfeld](#), in the AER package.

**Examples**

```
## Not run:
# Compare plm's Grunfeld data to Baltagi's (2001) Grunfeld data:
data("Grunfeld", package="plm")
Grunfeld_baltagi2001 <- read.csv("http://www.wiley.com/legacy/wileychi/
  baltagi/supp/Grunfeld.fil", sep=";", header = FALSE)
library(compare)
compare::compare(Grunfeld, Grunfeld_baltagi2001, allowAll = T) # same data set

## End(Not run)
```

---

Hedonic

*Hedonic Prices of Census Tracts in the Boston Area*

---

**Description**

A cross-section

*number of observations* : 506

*observation* : regional

*country* : United States

**Usage**

```
data(Hedonic)
```

**Format**

A dataframe containing:

**mv** median value of owner-occupied homes

**crim** crime rate

**zn** proportion of 25,000 square feet residential lots

**indus** proportion of no-retail business acres

**chas** is the tract bounds the Charles River?

**nox** annual average nitrogen oxide concentration in parts per hundred million

**rm** average number of rooms

**age** proportion of owner units built prior to 1940

**dis** weighted distances to five employment centers in the Boston area

**rad** index of accessibility to radial highways

**tax** full value property tax rate (\$/\$10,000)

**ptratio** pupil/teacher ratio

**blacks** proportion of blacks in the population

**lstat** proportion of population that is lower status

**townid** town identifier

**Source**

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

**References**

Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.

Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.

Belsley, D.A., E. Kuh and R. E. Welsch (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley, New York.

Harrison, D. and D.L. Rubinfeld (1978): "Hedonic housing prices and the demand for clean air", *Journal of Environmental Economics and Management*, **5**(1), pp. 81–102.

---

index	<i>Extract the indexes of panel data</i>
-------	--

---

### Description

This function extracts the information about the structure of the individual and time dimensions of panel data. Grouping information can also be extracted if the panel data were created with a grouping variable.

### Usage

```
## S3 method for class 'pindex'  
index(x, which = NULL, ...)  
## S3 method for class 'pdata.frame'  
index(x, which = NULL, ...)  
## S3 method for class 'pseries'  
index(x, which = NULL, ...)  
## S3 method for class 'panelmodel'  
index(x, which = NULL, ...)
```

### Arguments

x	an object of class "pindex", "pdata.frame", "pseries" or "panelmodel",
which	the index(es) to be extracted (see details),
...	further arguments.

### Details

Panel data are stored in a "pdata.frame" which has an "index" attribute. Fitted models in "plm" have a "model" element which is also a "pdata.frame" and therefore also has an "index" attribute. Finally, each series, once extracted from a "pdata.frame", becomes of class "pseries", which also has this "index" attribute. "index" methods are available for all these objects. The argument "which" indicates which index should be extracted. If which = NULL, all indexes are extracted. "which" can also be a vector of length 1, 2, or 3 (3 only if the pdata frame was constructed with an additional group index) containing either characters (the names of the individual variable and/or of the time variable and/or the group variable or "id" and "time") and "group" or integers (1 for the individual index, 2 for the time index, and 3 for the group index (the latter only if the pdata frame was constructed with such).)

### Value

A vector or an object of class c("pindex", "data.frame") containing either one index, individual and time index, or (any combination of) individual, time and group indexes.

### Author(s)

Yves Croissant

**See Also**[pdata.frame](#), [plm](#)**Examples**

```

data("Grunfeld", package = "plm")
Gr <- pdata.frame(Grunfeld, index = c("firm", "year"))
m <- plm(inv ~ value + capital, data = Gr)
index(Gr, "firm")
index(Gr, "time")
index(Gr$inv, c(2, 1))
index(m, "id")

# with additional group index
data("Produc", package = "plm")
pProduc <- pdata.frame(Produc, index = c("state", "year", "region"))
index(pProduc, 3)
index(pProduc, "region")
index(pProduc, "group")

```

is.pbalanced

*Check if data are balanced***Description**

This function checks if the data are balanced, i.e. if each individual has the same time periods

**Usage**

```

## S3 method for class 'pdata.frame'
is.pbalanced(x, ...)
## S3 method for class 'data.frame'
is.pbalanced(x, index = NULL, ...)
## S3 method for class 'pseries'
is.pbalanced(x, ...)
## S3 method for class 'panelmodel'
is.pbalanced(x, ...)
## S3 method for class 'pgmm'
is.pbalanced(x, ...)

```

**Arguments**

x	an object of class <code>pdata.frame</code> , <code>data.frame</code> , <code>pseries</code> , <code>panelmodel</code> , <code>pgmm</code> ;
index	only relevant for <code>data.frame</code> interface; if <code>NULL</code> , the first two columns of the <code>data.frame</code> are assumed to be the index variables; if not <code>NULL</code> , both dimensions ('individual', 'time') need to be specified by <code>index</code> as character of length 2 for data frames, for further details see <a href="#">pdata.frame</a> ,
...	further arguments.



**Details**

Balanced data are data for which each individual has the same time periods. The returned values of the `is.pbalanced(object)` methods are identical to `pdim(object)$balanced`. `is.pbalanced` is provided as a short cut and is faster than `pdim(object)$balanced` because it avoids those computations performed by `pdim` which are unnecessary to determine the balancedness of the data.

**Value**

A logical indicating whether the data associated with object `x` are balanced (TRUE) or not (FALSE).

**See Also**

[punbalancedness](#) for two measures of unbalancedness, [make.pbalanced](#) to make data balanced; [is.pconsecutive](#) to check if data are consecutive; [make.pconsecutive](#) to make data consecutive (and, optionally, also balanced).

[pdim](#) to check the dimensions of a 'pdata.frame' (and other objects), [pvar](#) to check for individual and time variation of a 'pdata.frame' (and other objects), [pseries](#), [data.frame](#), [pdata.frame](#).

**Examples**

```
# take balanced data and make it unbalanced
# by deletion of 2nd row (2nd time period for first individual)
data("Grunfeld", package = "plm")
Grunfeld_missing_period <- Grunfeld[-2, ]
is.pbalanced(Grunfeld_missing_period)      # check if balanced: FALSE
pdim(Grunfeld_missing_period)$balanced     # same

# pdata.frame interface
pGrunfeld_missing_period <- pdata.frame(Grunfeld_missing_period)
is.pbalanced(Grunfeld_missing_period)

# pseries interface
is.pbalanced(pGrunfeld_missing_period$inv)
```

---

<code>is.pconsecutive</code>	<i>Check if time periods are consecutive</i>
------------------------------	--

---

**Description**

This function checks for each individual if its associated time periods are consecutive (no "gaps" in time dimension per individual)

**Usage**

```
## S3 method for class 'pdata.frame'
is.pconsecutive(x, na.rm.tindex = FALSE, ...)
## S3 method for class 'panelmodel'
is.pconsecutive(x, na.rm.tindex = FALSE, ...)
```

```
## S3 method for class 'pseries'
is.pconsecutive(x, na.rm.tindex = FALSE, ...)
## S3 method for class 'data.frame'
is.pconsecutive(x, index = NULL, na.rm.tindex = FALSE, ...)
## Default S3 method:
is.pconsecutive(x, id, time, na.rm.tindex = FALSE, ...)
```

## Arguments

<code>x</code>	usually, an object of class <code>pdata.frame</code> , <code>data.frame</code> , <code>pseries</code> , or an estimated <code>panelmodel</code> ; for the default method <code>x</code> can also be an arbitrary vector or <code>NULL</code> , see <b>Details</b> ,
<code>na.rm.tindex</code>	logical indicating whether any NA values in the time index are removed before consecutiveness is evaluated (defaults to <code>FALSE</code> ),
<code>index</code>	only relevant for <code>data.frame</code> interface; if <code>NULL</code> , the first two columns of the <code>data.frame</code> are assumed to be the index variables; if not <code>NULL</code> , both dimensions ('individual', 'time') need to be specified by <code>index</code> for <code>is.pconsecutive</code> on <code>data.frames</code> , for further details see <a href="#">pdata.frame</a> ,
<code>id, time</code>	only relevant for default method: vectors specifying the id and time dimensions, i. e. a sequence of individual and time identifiers, each as stacked time series,
<code>...</code>	further arguments.

## Details

(p)`data.frame`, `pseries` and estimated `panelmodel` objects can be tested if their time periods are consecutive per individual. For evaluation of consecutiveness, the time dimension is interpreted to be numeric, and the data are tested for being a regularly spaced sequence with distance 1 between the time periods for each individual (for each individual the time dimension can be interpreted as sequence  $t, t+1, t+2, \dots$  where  $t$  is an integer). As such, the "numerical content" of the time index variable is considered for consecutiveness, not the "physical position" of the various observations for an individuals in the (p)`data.frame`/`pseries` (it is not about "neighbouring" rows).

The default method also works for argument `x` being an arbitrary vector (see **Examples**), provided one can supply arguments `id` and `time`, which need to ordered as stacked time series. As only `id` and `time` are really necessary for the default method to evaluate the consecutiveness, `x = NULL` is also possible. However, if the vector `x` is also supplied, additional input checking for equality of the lengths of `x`, `id` and `time` is performed, which is safer.

For the `data.frame` interface, the data is ordered in the appropriate way (stacked time series) before the consecutiveness is evaluated. For the `pdata.frame` and `pseries` interface, ordering is not performed because both data types are already ordered in the appropriate way when created.

Note: Only the presence of the time period itself in the object is tested, not if there are any other variables. NA values in individual index are not examined but silently dropped - In this case, it is not clear which individual is meant by `id` value NA, thus no statement about consecutiveness of time periods for those "NA-individuals" is possible.

## Value

A named logical vector (names are those of the individuals). The  $i$ -th element of the returned vector corresponds to the  $i$ -th individual. The values of the  $i$ -th element can be:

TRUE	if the i-th individual has consecutive time periods,
FALSE	if the i-th individual has non-consecutive time periods,
NA	if there are any NA values in time index of the i-th the individual; see also argument <code>na.rm.tindex</code> to remove those.

**Author(s)**

Kevin Tappe

**See Also**

[make.pconsecutive](#) to make data consecutive (and, as an option, balanced at the same time) and [make.pbalanced](#) to make data balanced.

[pdim](#) to check the dimensions of a 'pdata.frame' (and other objects), [pvar](#) to check for individual and time variation of a 'pdata.frame' (and other objects), [lag](#) for lagged (and leading) values of a 'pseries' object.

[pseries](#), [data.frame](#), [pdata.frame](#), for class 'panelmodel' see [plm](#) and [pgmm](#).

**Examples**

```
data("Grunfeld", package = "plm")
is.pconsecutive(Grunfeld)
is.pconsecutive(Grunfeld, index=c("firm", "year"))

# delete 2nd row (2nd time period for first individual)
# -> non consecutive
Grunfeld_missing_period <- Grunfeld[-2, ]
is.pconsecutive(Grunfeld_missing_period)
all(is.pconsecutive(Grunfeld_missing_period)) # FALSE

# delete rows 1 and 2 (1st and 2nd time period for first individual)
# -> consecutive
Grunfeld_missing_period_other <- Grunfeld[-c(1,2), ]
is.pconsecutive(Grunfeld_missing_period_other) # all TRUE

# delete year 1937 (3rd period) for _all_ individuals
Grunfeld_wo_1937 <- Grunfeld[Grunfeld$year != 1937, ]
is.pconsecutive(Grunfeld_wo_1937) # all FALSE

# pdata.frame interface
pGrunfeld <- pdata.frame(Grunfeld)
pGrunfeld_missing_period <- pdata.frame(Grunfeld_missing_period)
is.pconsecutive(pGrunfeld) # all TRUE
is.pconsecutive(pGrunfeld_missing_period) # first FALSE, others TRUE

# panelmodel interface (first, estimate some models)
mod_pGrunfeld <- plm(inv ~ value + capital, data = Grunfeld)
mod_pGrunfeld_missing_period <- plm(inv ~ value + capital, data = Grunfeld_missing_period)
```

```

is.pconsecutive(mod_pGrunfeld)
is.pconsecutive(mod_pGrunfeld_missing_period)

nobs(mod_pGrunfeld) # 200
nobs(mod_pGrunfeld_missing_period) # 199

# pseries interface
pinv <- pGrunfeld$inv
pinv_missing_period <- pGrunfeld_missing_period$inv

is.pconsecutive(pinv)
is.pconsecutive(pinv_missing_period)

# default method for arbitrary vectors or NULL
inv <- Grunfeld$inv
inv_missing_period <- Grunfeld_missing_period$inv
is.pconsecutive(inv, id = Grunfeld$firm, time = Grunfeld$year)
is.pconsecutive(inv_missing_period, id = Grunfeld_missing_period$firm,
                 time = Grunfeld_missing_period$year)

# (not run) demonstrate mismatch lengths of x, id, time
# is.pconsecutive(x = inv_missing_period, id = Grunfeld$firm, time = Grunfeld$year)

# only id and time are needed for evaluation
is.pconsecutive(NULL, id = Grunfeld$firm, time = Grunfeld$year)

```

---

LaborSupply

*Wages and Hours Worked*


---

### Description

A panel of 532 observations from 1979 to 1988  
*number of observations : 5320*

### Usage

```
data(LaborSupply)
```

### Format

A data frame containing :

**lnhr** log of annual hours worked

**lnwg** log of hourly wage

**kids** number of children

**age** age

**disab** bad health

**id** id

**year** year

**Source**

Online complements to Ziliak (1997).

Journal of Business Economics and Statistics web site: <http://amstat.tandfonline.com/loi/ubes20/>.

**References**

Cameron, A.C. and P.K. Trivedi (2005) *Microeconometrics : methods and applications*, Cambridge, pp. 708–15, 754–6.

Ziliak, J.P. (1997) “Efficient Estimation With Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators”, *Journal of Business and Economic Statistics*, **15**(4), pp. 419–431.

---

lag *lag, lead, and diff for panel data*

---

**Description**

lag, lead, and diff functions for class pseries.

**Usage**

```
## S3 method for class 'pseries'
lag(x, k = 1, ...)
## S3 method for class 'pseries'
lead(x, k = 1, ...)
## S3 method for class 'pseries'
diff(x, lag = 1, ...)
```

**Arguments**

x	a pseries object,
k	an integer vector, the number of lags for the lag and lead methods (can also be negative). For the lag method, a positive (negative) k gives lagged (leading) values. For the lead method, a positive (negative) k gives leading (lagged) values, thus, lag(x, k = -1) yields the same as lead(x, k = 1). If k is an integer vector with length > 1 (k = c(k1, k2, ...)) a matrix with multiple lagged pseries is returned,
lag	the number of lags for the diff method, can also be a vector of length > 1 (see argument k) (only non-negative values in argument lag are allowed for diff),
...	further arguments.

**Value**

- \* An object of class `pseries`, if the argument specifying the lag has length 1 (argument `k` in functions `lag` and `lead`, argument `lag` in function `diff`).
- \* A matrix containing the various series in its columns, if the argument specifying the lag has length  $> 1$ .

**Note**

The sign of `k` in `lag.pseries` results in inverse behavior compared to [lag](#) and [lag.zoo](#).

**Author(s)**

Yves Croissant

**See Also**

For further function for 'pseries' objects: [between](#), [Between](#), [Within](#), [summary.pseries](#), [print.summary.pseries](#), [as.matrix.pseries](#). To check if the time periods are consecutive per individual, see [is.pconsecutive](#).

**Examples**

```
# First, create a pdata.frame
data("EmplUK", package = "plm")
Em <- pdata.frame(EmplUK)

# Then extract a series, which becomes additionally a pseries
z <- Em$output
class(z)

# compute the first and third lag, and the difference lagged twice
lag(z)
lag(z, 3)
diff(z, 2)

# compute negative lags (= leading values)
lag(z, -1)
lead(z, 1) # same as line above
identical(lead(z, 1), lag(z, -1)) # TRUE

# compute more than one lag and diff at once
# (each returns matrix)
lag(z, c(1,2))
diff(z, c(1,2))
```

---

make.pbalanced	<i>Make data balanced</i>
----------------	---------------------------

---

## Description

This function makes the data balanced, i.e. each individual has the same time periods, by filling in or dropping observations

## Usage

```
## S3 method for class 'pdata.frame'
make.pbalanced(x,
               balance.type = c("fill", "shared.times", "shared.individuals"), ...)
## S3 method for class 'pseries'
make.pbalanced(x,
               balance.type = c("fill", "shared.times", "shared.individuals"), ...)
## S3 method for class 'data.frame'
make.pbalanced(x,
               balance.type = c("fill", "shared.times", "shared.individuals"),
               index = NULL, ...)
```

## Arguments

x	an object of class <code>pdata.frame</code> , <code>data.frame</code> , or <code>pseries</code> ;
balance.type	character, one of "fill", "shared.times", or "shared.individuals", see <b>Details</b> ,
index	only relevant for <code>data.frame</code> interface; if <code>NULL</code> , the first two columns of the <code>data.frame</code> are assumed to be the index variables; if not <code>NULL</code> , both dimensions ('individual', 'time') need to be specified by <code>index</code> as character of length 2 for <code>data.frames</code> , for further details see <a href="#">pdata.frame</a> ,
...	further arguments.

## Details

(p)`data.frame` and `pseries` objects are made balanced, meaning each individual has the same time periods. Depending on the value of `balance.type`, the balancing is done in different ways:

- `balance.type = "fill"` (default): The union of available time periods over all individuals is taken (w/o NA values). Missing time periods for an individual are identified and corresponding rows (elements for `pseries`) are inserted and filled with NA for the non-index variables (elements for a `pseries`). This means, only time periods present for at least one individual are inserted, if missing.
- `balance.type = "shared.times"`: The intersect of available time periods over all individuals is taken (w/o NA values). Thus, time periods not available for all individuals are discarded, i. e., only time periods shared by all individuals are left in the result).

- `balance.type = "shared.individuals"`: All available time periods are kept and those individuals are dropped for which not all time periods are available, i. e., only individuals shared by all time periods are left in the result (symmetric to `"shared.times"`).

The data are not necessarily made consecutive (regular time series with distance 1), because balancedness does not imply consecutiveness. For making the data consecutive, use `make.pconsecutive` (and, optionally, set argument `balanced = TRUE` to make consecutive and balanced, see also **Examples** for a comparison of the two functions).

Note: Rows of (p)data.frames (elements for pseries) with NA values in individual or time index are not examined but silently dropped before the data are made balanced. In this case, it cannot be inferred which individual or time period is meant by the missing value(s) (see also **Examples**). Especially, this means: NA values in the first/last position of the original time periods for an individual are dropped, which are usually meant to depict the beginning and ending of the time series for that individual. Thus, one might want to check if there are any NA values in the index variables before applying `make.pbalanced`, and especially check for NA values in the first and last position for each individual in original data and, if so, maybe set those to some meaningful begin/end value for the time series.

### Value

An object of the same class as the input `x`, i.e. a `pdata.frame`, `data.frame` or a `pseries` which is made balanced based on the index variables. The returned data are sorted as a stacked time series.

### Author(s)

Kevin Tappe

### See Also

`is.pbalanced` to check if data are balanced; `is.pconsecutive` to check if data are consecutive; `make.pconsecutive` to make data consecutive (and, optionally, also balanced). `punbalancedness` for two measures of unbalancedness, `pdim` to check the dimensions of a 'pdata.frame' (and other objects), `pvar` to check for individual and time variation of a 'pdata.frame' (and other objects), `lag` for lagging (and leading) values of a 'pseries' object. `pseries`, `data.frame`, `pdata.frame`.

### Examples

```
# take data and make it unbalanced
# by deletion of 2nd row (2nd time period for first individual)
data("Grunfeld", package = "plm")
nrow(Grunfeld) # 200 rows
Grunfeld_missing_period <- Grunfeld[-2, ]
pdim(Grunfeld_missing_period)$balanced # check if balanced: FALSE
make.pbalanced(Grunfeld_missing_period) # make it balanced (by filling)
make.pbalanced(Grunfeld_missing_period, balance.type = "shared.times") # (shared periods)
nrow(make.pbalanced(Grunfeld_missing_period))
nrow(make.pbalanced(Grunfeld_missing_period, balance.type = "shared.times"))

# more complex data:
```



```

# First, make data unbalanced (and non-consecutive)
# by deletion of 2nd time period (year 1936) for all individuals
# and more time periods for first individual only
Grunfeld_unbalanced <- Grunfeld[Grunfeld$year != 1936, ]
Grunfeld_unbalanced <- Grunfeld_unbalanced[-c(1,4), ]
pdim(Grunfeld_unbalanced)$balanced      # FALSE
all(is.pconsecutive(Grunfeld_unbalanced)) # FALSE

g_bal <- make.pbalanced(Grunfeld_unbalanced)
pdim(g_bal)$balanced                    # TRUE
unique(g_bal$year)                       # all years but 1936
nrow(g_bal)                              # 190 rows
head(g_bal)                             # 1st individual: years 1935, 1939 are NA

# NA in 1st, 3rd time period (years 1935, 1937) for first individual
Grunfeld_NA <- Grunfeld
Grunfeld_NA[c(1, 3), "year"] <- NA
g_bal_NA <- make.pbalanced(Grunfeld_NA)
head(g_bal_NA)                          # years 1935, 1937: NA for non-index vars
nrow(g_bal_NA)                          # 200

# pdata.frame interface
pGrunfeld_missing_period <- pdata.frame(Grunfeld_missing_period)
make.pbalanced(Grunfeld_missing_period)

# pseries interface
make.pbalanced(pGrunfeld_missing_period$inv)

# comparison to make.pconsecutive
g_consec <- make.pconsecutive(Grunfeld_unbalanced)
all(is.pconsecutive(g_consec)) # TRUE
pdim(g_consec)$balanced        # FALSE
head(g_consec, 22)            # 1st individual: no years 1935/6; 1939 is NA;
                             # other individuals: years 1935-1954, 1936 is NA
nrow(g_consec)                # 198 rows

g_consec_bal <- make.pconsecutive(Grunfeld_unbalanced, balanced = TRUE)
all(is.pconsecutive(g_consec_bal)) # TRUE
pdim(g_consec_bal)$balanced        # TRUE
head(g_consec_bal)                # year 1936 is NA for all individuals
nrow(g_consec_bal)                # 200 rows

head(g_bal)                      # no year 1936 at all
nrow(g_bal)                       # 190 rows

```

---

make.pconsecutive

*Make data consecutive (and, optionally, also balanced)*


---

## Description

This function makes the data consecutive for each individual (no "gaps" in time dimension per individual) and, optionally, also balanced

## Usage

```
## S3 method for class 'pdata.frame'
make.pconsecutive(x, balanced = FALSE, ...)
## S3 method for class 'pseries'
make.pconsecutive(x, balanced = FALSE, ...)
## S3 method for class 'data.frame'
make.pconsecutive(x, balanced = FALSE, index = NULL, ...)
```

## Arguments

<code>x</code>	an object of class <code>pdata.frame</code> , <code>data.frame</code> , or <code>pseries</code> ,
<code>balanced</code>	logical, indicating whether the data should <code>_additionally_</code> be made balanced (default: <code>FALSE</code> ),
<code>index</code>	only relevant for <code>data.frame</code> interface; if <code>NULL</code> , the first two columns of the <code>data.frame</code> are assumed to be the index variables; if not <code>NULL</code> , both dimensions ('individual', 'time') need to be specified by <code>index</code> as character of length 2 for data frames, for further details see <a href="#">pdata.frame</a> ,
<code>...</code>	further arguments.

## Details

(p)`data.frame` and `pseries` objects are made consecutive, meaning their time periods are made consecutive per individual. For consecutiveness, the time dimension is interpreted to be numeric, and the data are extended to a regularly spaced sequence with distance 1 between the time periods for each individual (for each individual the time dimension become a sequence  $t, t+1, t+2, \dots$  where  $t$  is an integer). Non-index variables are filled with NA for the inserted elements (rows for (p)`data.frames`, vector elements for `pseries`).

With argument `balanced = TRUE`, additionally to be made consecutive, the data also can be made a balanced panel/`pseries`. Note: This means consecutive AND balanced; balancedness does not imply consecutiveness. In the result, each individual will have the same time periods in their time dimension by taking the min and max of the time index variable over all individuals (w/o NA values) and inserting the missing time periods. Looking at the number of rows of the resulting (p)`data.frame` (elements for `pseries`), this results in `nrow(make.pconsecutive, balanced = FALSE) <= nrow(make.pconsecutive, balanced = TRUE)`. For making the data only balanced, i.e. not demanding consecutiveness at the same time, use [make.pbalanced](#) (see **Examples** for a comparison).

Note: rows of (p)`data.frames` (elements for `pseries`) with NA values in individual or time index are not examined but silently dropped before the data are made consecutive. In this case, it is not clear which individual or time period is meant by the missing value(s). Especially, this means: If there are NA values in the first/last position of the original time periods for an individual, which usually depicts the beginning and ending of the time series for that individual, the beginning/end of the resulting time series is taken to be the min and max (w/o NA values) of the original time series for

that individual, see also **Examples**. Thus, one might want to check if there are any NA values in the index variables before applying `make.pconsecutive`, and especially check for NA values in the first and last position for each individual in original data and, if so, maybe set those to some meaningful begin/end value for the time series.

### Value

An object of the same class as the input `x`, i.e. a `pdata.frame`, `data.frame` or a `pseries` which is made time-consecutive based on the index variables. The returned data are sorted as a stacked time series.

### Author(s)

Kevin Tappe

### See Also

[is.pconsecutive](#) to check if data are consecutive; [make.pbalanced](#) to make data only balanced (not consecutive).  
[punbalancedness](#) for two measures of unbalancedness, [pdim](#) to check the dimensions of a 'pdata.frame' (and other objects), [pvar](#) to check for individual and time variation of a 'pdata.frame' (and other objects), [lag](#) for lagged (and leading) values of a 'pseries' object.  
[pseries](#), [data.frame](#), [pdata.frame](#).

### Examples

```
# take data and make it non-consecutive
# by deletion of 2nd row (2nd time period for first individual)
data("Grunfeld", package = "plm")
nrow(Grunfeld) # 200 rows
Grunfeld_missing_period <- Grunfeld[-2, ]
is.pconsecutive(Grunfeld_missing_period) # check for consecutiveness
make.pconsecutive(Grunfeld_missing_period) # make it consecutiveness

# argument balanced:
# First, make data non-consecutive and unbalanced
# by deletion of 2nd time period (year 1936) for all individuals
# and more time periods for first individual only
Grunfeld_unbalanced <- Grunfeld[Grunfeld$year != 1936, ]
Grunfeld_unbalanced <- Grunfeld_unbalanced[-c(1,4), ]
all(is.pconsecutive(Grunfeld_unbalanced)) # FALSE
pdim(Grunfeld_unbalanced)$balanced # FALSE

g_consec_bal <- make.pconsecutive(Grunfeld_unbalanced, balanced = TRUE)
all(is.pconsecutive(g_consec_bal)) # TRUE
pdim(g_consec_bal)$balanced # TRUE
nrow(g_consec_bal) # 200 rows
head(g_consec_bal) # 1st individual: years 1935, 1936, 1939 are NA

g_consec <- make.pconsecutive(Grunfeld_unbalanced) # default: balanced = FALSE
all(is.pconsecutive(g_consec)) # TRUE
pdim(g_consec)$balanced # FALSE
```

```

nrow(g_consec)          # 198 rows
head(g_consec)         # 1st individual: years 1935, 1936 dropped, 1939 is NA

# NA in 1st, 3rd time period (years 1935, 1937) for first individual
Grunfeld_NA <- Grunfeld
Grunfeld_NA[c(1, 3), "year"] <- NA
g_NA <- make.pconsecutive(Grunfeld_NA)
head(g_NA)             # 1936 is begin for 1st individual, 1937: NA for non-index vars
nrow(g_NA)             # 199, year 1935 from original data is dropped

# pdata.frame interface
pGrunfeld_missing_period <- pdata.frame(Grunfeld_missing_period)
make.pconsecutive(Grunfeld_missing_period)

# pseries interface
make.pconsecutive(pGrunfeld_missing_period$inv)

# comparison to make.pbalanced (makes the data only balanced, not consecutive)
g_bal <- make.pbalanced(Grunfeld_unbalanced)
all(is.pconsecutive(g_bal)) # FALSE
pdim(g_bal)$balanced      # TRUE
nrow(g_bal) # 190 rows

```

---

Males

*Wages and Education of Young Males*


---

### Description

A panel of 545 observations from 1980 to 1987  
*total number of observations* : 4360  
*observation* : individuals  
*country* : United States

### Usage

```
data(Males)
```

### Format

A data frame containing :

- nr** identifier
- year** year
- school** years of schooling

**exper** years of experience (computed as age-6-school)  
**union** wage set by collective bargaining ?  
**ethn** a factor with levels black, hisp, other  
**married** married?  
**health** health problem ?  
**wage** log of hourly wage  
**industry** a factor with 12 levels  
**occupation** a factor with 9 levels  
**residence** a factor with levels rural area, north east, northern central, south

### Source

Journal of Applied Econometrics data archive <http://qed.econ.queensu.ca/jae/1998-v13.2/vella-verbeek/>.

### References

Vella, F. and M. Verbeek (1998) "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men", *Journal of Applied Econometrics*, **13**(2), pp. 163–183.  
 Verbeek, M. (2004) *A Guide to Modern Econometrics*, John Wiley and Sons, chapter 10.

---

model.frame.pFormula *model.frame and model.matrix for panel data*

---

### Description

Methods to create model frame and model matrix for panel data.

### Usage

```
## S3 method for class 'pFormula'
model.frame(formula, data, ..., lhs = NULL, rhs = NULL)

## S3 method for class 'pFormula'
model.matrix(object, data,
             model = c("pooling", "within", "Between", "Sum",
                       "between", "mean", "random", "fd"),
             effect = c("individual", "time", "twoways", "nested"),
             rhs = 1,
             theta = NULL,
             cstcovar.rm = NULL,
             ...)

## S3 method for class 'plm'
model.matrix(object, ...)
```

**Arguments**

object, formula	an object of class "pFormula" or an estimated model object of class "plm",
data	a pdata.frame, see <b>Details</b> ,
effect	the effects introduced in the model, one of "individual", "time", "twoways" or "nested",
model	one of "pooling", "within", "Sum", "Between", "between", "random", "fd" and "ht",
theta	the parameter for the transformation if model = "random",
cstcovar.rm	remove the constant columns, one of "none", "intercept", "covariates", "all"),
lhs	inherited from package <a href="#">Formula</a> (see there),
rhs	inherited from package <a href="#">Formula</a> (see there),
...	further arguments.

**Details**

The lhs and rhs arguments are inherited from [Formula](#), see there for more details.

The model.frame methods return a pdata.frame object suitable as an input to plm's model.matrix.

The model.matrix methods builds a model matrix with transformations performed as specified by the model and effect arguments (and theta if model = "random" is requested), in this case the supplied data argument should be a model frame created by plm's model.frame method. If not, it is tried to construct the model frame from the data. Constructing the model frame first ensures proper NA handling, see **Examples**.

**Value**

The model.frame methods return a pdata.frame.

The model.matrix methods return a matrix.

**Author(s)**

Yves Croissant

**See Also**

[pmodel.response](#) for (transformed) response variable.

[pFormula](#), especially for coercing a formula to a pFormula.

[Formula](#) from package [Formula](#), especially for the lhs and rhs arguments.

**Examples**

```
# First, make a pdata.frame
data("Grunfeld", package = "plm")
pGrunfeld <- pdata.frame(Grunfeld)

# then make a model frame from a pFormula and a pdata.frame
pform <- pFormula(inv ~ value + capital)
```

```

mf <- model.frame(pform, data = pGrunfeld)

# then construct the (transformed) model matrix (design matrix)
# from formula and model frame
modmat <- model.matrix(pform, data = mf, model = "within")

## retrieve model frame and model matrix from an estimated plm object
fe_model <- plm(pform, data = pGrunfeld, model = "within")
model.frame(fe_model)
model.matrix(fe_model)

# same as constructed before
all.equal(mf, model.frame(fe_model), check.attributes = FALSE) # TRUE
all.equal(modmat, model.matrix(fe_model), check.attributes = FALSE) # TRUE

```

---

mtest

*Arellano–Bond test of Serial Correlation*


---

### Description

Test of serial correlation for models estimated by GMM

### Usage

```
mtest(object, order = 1, vcov = NULL)
```

### Arguments

object	an object of class "pgmm",
order	the order of the serial correlation (1 or 2),
vcov	a matrix of covariance for the coefficients or a function to compute it.

### Details

The Arellano–Bond test is a test of correlation based on the residuals of the estimation. By default, the computation is done with the standard covariance matrix of the coefficients. A robust estimator of this covariance matrix can be supplied with the `vcov` argument.

### Value

An object of class "htest".

### Author(s)

Yves Croissant

## References

Arellano, M. and Bond, S. (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies*, **58**(2), 1991, pp. 227–297.

## See Also

[pgmm](#)

## Examples

```
data("EmplUK", package = "plm")
ar <- pgmm(log(emp) ~ lag(log(emp), 1:2) + lag(log(wage), 0:1) +
           lag(log(capital), 0:2) + lag(log(output), 0:2) | lag(log(emp), 2:99),
           data = EmplUK, effect = "twoways", model = "twosteps")
mtest(ar, order = 1)
mtest(ar, order = 2, vcov = vcovHC)
```

---

nobs

*Extract Total Number of Observations Used in Estimated Panelmodel*

---

## Description

This function extracts the total number of 'observations' from a fitted panel model.

## Usage

```
## S3 method for class 'panelmodel'
nobs(object, ...)
```

## Arguments

object	a panelmodel object for which the number of total observations is to be extracted,
...	further arguments.

## Details

The number of observations is usually the length of the residuals vector. Thus, nobs gives the number of observations actually used by the estimation procedure. It is not necessarily the number of observations of the model frame (number of rows in the model frame), because sometimes the model frame is further reduced by the estimation procedure. This is e.g. the case for first-difference models estimated by `plm(..., model = "fd")` where the model frame does not yet contain the differences (see also **Examples**).

## Value

A single number, normally an integer.



**See Also**[pdim](#)**Examples**

```
# estimate a panelmodel
data("Produc", package = "plm")
z <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp,data=Produc,
        model="random", subset = gsp > 5000)

nobs(z)      # total observations used in estimation
pdim(z)$nT$N # same information
pdim(z)      # more information about the dimensions (no. of individuals and time periods)

# illustrate difference between nobs and pdim for first-difference model
data("Grunfeld", package = "plm")
fdmod <- plm(inv ~ value + capital, data = Grunfeld, model = "fd")
nobs(fdmod)  # 190
pdim(fdmod)$nT$N # 200
```

Parity

*Purchasing Power Parity and other parity relationships***Description**

A panel of 104 quarterly observations from 1973Q1 to 1998Q4  
*total number of observations* : 1768  
*observation* : country  
*country* : OECD

**Usage**

```
data(Parity)
```

**Format**

A data frame containing :

**country** country codes: a factor with 17 levels  
**time** the quarter index, 1973Q1-1998Q4  
**ls** log spot exchange rate vs. USD  
**lp** log price level  
**is** short term interest rate  
**il** long term interest rate  
**ld** log price differential vs. USA  
**uis** U.S. short term interest rate  
**uil** U.S. long term interest rate

**Source**

Coakley, J., Fuertes, A. M., and Smith, R. (2006) “Unobserved heterogeneity in panel time series models”, *Computational Statistics & Data Analysis*, **50**(9), 2361–2380.

**References**

Coakley, J., Fuertes, A. M., and Smith, R. (2006) “Unobserved heterogeneity in panel time series models”, *Computational Statistics & Data Analysis*, **50**(9), 2361–2380.

Driscoll, J. C., and Kraay, A. C. (1998). “Consistent covariance matrix estimation with spatially dependent panel data”, *Review of Economics and Statistics*, **80**(4), 549–560.

---

pbgtest

*Breusch–Godfrey Test for Panel Models*

---

**Description**

Test of serial correlation for (the idiosyncratic component of) the errors in panel models.

**Usage**

```
pbgtest(x, ...)
## S3 method for class 'panelmodel'
pbgtest(x, order = NULL, type = c("Chisq", "F"), ...)
## S3 method for class 'formula'
pbgtest(x, order = NULL, type = c("Chisq", "F"),
        data, model=c("pooling", "random", "within"), ...)
```

**Arguments**

x	an object of class "panelmodel" or of class "formula",
order	an integer indicating the order of serial correlation to be tested for. NULL (default) uses the minimum number of observations over the time dimension (see also section <b>Details</b> below),
type	type of test statistic to be calculated; either "Chisq" (default) for the Chi-squared test statistic or "F" for the F test statistic,
data	only relevant for formula interface: data set for which the respective panel model (see model) is to be evaluated,
model	only relevant for formula interface: compute test statistic for model pooling (default), random, or within. When model is used, the data argument needs to be passed as well,
...	further arguments (see <a href="#">bgtest</a> ).

## Details

This Lagrange multiplier test uses the auxiliary model on (quasi-)demeaned data taken from a model of class `plm` which may be a pooling (default for formula interface), random or within model. It performs a Breusch–Godfrey test (using `bgtest` from package **lmtest**) on the residuals of the (quasi-)demeaned model, which should be serially uncorrelated under the null of no serial correlation in idiosyncratic errors, as illustrated in Wooldridge (2002/2010). The function takes the demeaned data, estimates the model and calls `bgtest`.

Unlike most other tests for serial correlation in panels, this one allows to choose the order of correlation to test for.

## Value

An object of class "htest".

## Note

The argument `order` defaults to the minimum number of observations over the time dimension, while for `lmtest::bgtest` it defaults to 1.

## Author(s)

Giovanni Millo

## References

Breusch, T.S. (1978), "Testing for autocorrelation in dynamic linear models", *Australian Economic Papers*, **17**(31), pp. 334–355.

Godfrey, L.G. (1978), "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables", *Econometrica*, **46**(6), pp. 1293–1301.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, pp. 288–291.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, pp. 328–334.

Wooldridge, J.M. (2013) *Introductory Econometrics: A Modern Approach*, 5th ed., South-Western (Cengage Learning), Sec. 12.2, pp. 421–422.

## See Also

For the original test in package **lmtest** see `bgtest`. See `pdwtest` for the analogous panel Durbin–Watson test. See `pbltest`, `pbsytest`, `pwartest` and `pwfdtest` for other serial correlation tests for panel models.

## Examples

```
data("Grunfeld", package = "plm")
g <- plm(inv ~ value + capital, data = Grunfeld, model = "random")

# panelmodel interface
```

```

pbgtest(g)
pbgtest(g, order = 4)

# formula interface
pbgtest(inv ~ value + capital, data = Grunfeld, model = "random")

# F test statistic (instead of default type="Chisq")
pbgtest(g, type="F")
pbgtest(inv ~ value + capital, data = Grunfeld, model = "random", type = "F")

```

---

pbltest

*Baltagi and Li Serial Dependence Test For Random Effects Models*


---

### Description

Baltagi and Li (1995)'s Lagrange multiplier test for AR(1) or MA(1) idiosyncratic errors in panel models with random effects.

### Usage

```

## S3 method for class 'formula'
pbltest(x, data,
        alternative = c("twosided", "onesided"), index = NULL, ...)
## S3 method for class 'plm'
pbltest(x, alternative = c("twosided", "onesided"), ...)

```

### Arguments

x	a model formula or an estimated random-effects model of class plm ,
data	for the formula interface only: a data.frame,
alternative	one of "twosided", "onesided". Selects either $H_A : \rho \neq 0$ or $H_A : \rho = 0$ (i.e., the Normal or the Chi-squared version of the test),
index	the index of the data.frame,
...	further arguments.

### Details

This is a Lagrange multiplier test for the null of no serial correlation, against the alternative of either an AR(1) or an MA(1) process, in the idiosyncratic component of the error term in a random effects panel model (as the analytical expression of the test turns out to be the same under both alternatives, see Baltagi and Li (1995, 1997)). The alternative argument, defaulting to twosided, allows testing for positive serial correlation only, if set to onesided.

### Value

An object of class "hctest".

**Author(s)**

Giovanni Millo

**References**

Baltagi, B.H. and Li, Q. (1995) Testing AR(1) against MA(1) disturbances in an error component model, *Journal of Econometrics* **68**(1), pp. 133–151.

Baltagi, B.H. and Li, Q. (1997) Monte Carlo results on pure and pretest estimators of an error component model with autocorrelated disturbances, *Annales d'économie et de statistique* **48**, pp. 69–82.

**See Also**

[pdwtest](#), [bgtest](#), [pbsytest](#), [pwartest](#) and [pwfdtest](#) for other serial correlation tests for panel models.

**Examples**

```
data("Grunfeld", package = "plm")

# formula interface
pbltest(inv ~ value + capital, data = Grunfeld)

# plm interface
re_mod <- plm(inv ~ value + capital, data = Grunfeld, model = "random")
pbltest(re_mod)
pbltest(re_mod, alternative = "onesided")
```

---

pbsytest

*Bera, Sosa-Escudero and Yoon Locally–Robust Lagrange Multiplier Tests for Panel Models and Joint Test by Baltagi and Li*

---

**Description**

Test for residual serial correlation (or individual random effects) locally robust vs. individual random effects (serial correlation) for panel models and joint test by Baltagi and Li.

**Usage**

```
pbsytest(x, ...)
## S3 method for class 'panelmodel'
pbsytest(x, test = c("ar", "re", "j"),
         re.normal = if (test == "re") TRUE else NULL, ...)
## S3 method for class 'formula'
pbsytest(x, data, ..., test = c("ar", "re", "j"),
         re.normal = if (test == "re") TRUE else NULL)
```

## Arguments

<code>x</code>	an object of class <code>formula</code> or of class <code>panelmodel</code> ,
<code>data</code>	a <code>data.frame</code> ,
<code>test</code>	a character string indicating which test to perform: first-order serial correlation (" <code>ar</code> "), random effects (" <code>re</code> ") or joint test for either of them (" <code>j</code> "),
<code>re.normal</code>	logical, only relevant for <code>test = "re"</code> : TRUE (default) computes the one-sided " <code>re</code> " test, FALSE the two-sided test (see also Details); not relevant for other values of <code>test</code> and, thus, should be NULL,
<code>...</code>	further arguments.

## Details

These Lagrange multiplier tests are robust vs. local misspecification of the alternative hypothesis, i.e. they test the null of serially uncorrelated residuals against AR(1) residuals in a pooling model, allowing for local departures from the assumption of no random effects; or they test the null of no random effects allowing for local departures from the assumption of no serial correlation in residuals. They use only the residuals of the pooled OLS model and correct for local misspecification as outlined in Bera *et al.* (2001).

For `test = "re"`, the default (`re.normal = TRUE`) is to compute a one-sided test which is expected to lead to a more powerful test (asymptotically  $N(0,1)$  distributed). Setting `re.normal = FALSE` gives the two-sided test (asymptotically chi-squared(2) distributed). Argument `re.normal` is irrelevant for all other values of `test`.

The joint test (`test = "j"`) is due to Baltagi and Li (1991) (also mentioned in Baltagi and Li (1995), pp. 135–136) and is added for convenience under this same function.

The unbalanced version of all tests are derived in Sosa-Escudero and Bera (2008). The functions implemented are suitable for balanced as well as unbalanced panel data sets.

A concise treatment of the statistics for only balanced panels is given in Baltagi (2013), p. 108.

Here is an overview of how the various values of the `test` argument relate to the literature:

- `test = "ar"`:
  - $RS*_\rho$  in Bera *et al.* (2001), p. 9 (balanced)
  - $LM*_\rho$  in Baltagi (2013), p. 108 (balanced)
  - $RS*_\lambda$  in Sosa-Escudero/Bera (2008), p. 73 (unbalanced)
- `test = "re"`, `re.normal = TRUE` (default) (one-sided test, asymptotically  $N(0,1)$  distributed):
  - $RSO*_\mu$  in Bera *et al.* (2001), p. 11 (balanced)
  - $RSO*_\mu$  in Sosa-Escudero/Bera (2008), p. 75 (unbalanced)
- `test = "re"`, `re.normal = FALSE` (two-sided test, asymptotically chi-squared(2) distributed):
  - $RS*_\mu$  in Bera *et al.* (2001), p. 7 (balanced)
  - $LM*_\mu$  in Baltagi (2013), p. 108 (balanced)
  - $RS*_\mu$  in Sosa-Escudero/Bera (2008), p. 73 (unbalanced)
- `test = "j"`:

- $RS_{\mu\rho}$  in Bera et al. (2001), p. 10 (balanced)
- $LM$  in Baltagi/Li (2001), p. 279 (balanced)
- $LM_1$  in Baltagi and Li (1995), pp. 135–136 (balanced)
- $LM1$  in Baltagi (2013), p. 108 (balanced)
- $RS_{\lambda\rho}$  in Sosa-Escudero/Bera (2008), p. 74 (unbalanced)

### Value

An object of class "htest".

### Author(s)

Giovanni Millo (initial implementation) & Kevin Tappe (extension to unbalanced panels)

### References

- Bera, A.K., Sosa-Escudero, W. and Yoon, M. (2001), Tests for the error component model in the presence of local misspecification, *Journal of Econometrics*, **101**(1), pp. 1–23.
- Baltagi, B. H. (2013) *Econometric Analysis of Panel Data*, 5th edition, Wiley.
- Baltagi, B. H. and Li, Q. (1991), A joint test for serial correlation and random individual effects, *Statistics and Probability Letters*, **11**(3), pp. 277–280.
- Baltagi, B. H. and Li, Q. (1995), Testing AR(1) against MA(1) disturbances in an error component model, *Journal of Econometrics*, **68**(1), pp. 133–151.
- Sosa-Escudero, W. and Bera, A.K. (2008), Tests for unbalanced error–components models under local misspecification, *The Stata Journal*, **8**(1), pp. 68–78.

### See Also

[plmtest](#) for individual and/or time random effects tests based on a correctly specified model; [pbltest](#), [pbgtest](#) and [pdwtest](#) for serial correlation tests in random effects models.

### Examples

```
## Bera et al. (2001), p. 13, table 1 use a subset of the original Grunfeld
## data which contains three errors -> construct this subset:
data("Grunfeld", package = "plm")
Grunsubset <- rbind(Grunfeld[1:80, ], Grunfeld[141:160, ])
Grunsubset[Grunsubset$firm == 2 & Grunsubset$year %in% c(1940, 1952), ][["inv"]] <- c(261.6, 645.2)
Grunsubset[Grunsubset$firm == 2 & Grunsubset$year == 1946, ][["capital"]] <- 232.6

## default is AR testing (formula interface)
pbsytest(inv ~ value + capital, data = Grunsubset, index = c("firm", "year"))
pbsytest(inv ~ value + capital, data = Grunsubset, index = c("firm", "year"), test = "re")
pbsytest(inv ~ value + capital, data = Grunsubset, index = c("firm", "year"),
  test = "re", re.normal = FALSE)
pbsytest(inv ~ value + capital, data = Grunsubset, index = c("firm", "year"), test = "j")

## plm interface
mod <- plm(inv ~ value + capital, data = Grunsubset, model = "pooling")
pbsytest(mod)
```

pcce

*Common Correlated Effects estimators***Description**

Common Correlated Effects Mean Groups (CCEMG) and Pooled (CCEP) estimators for panel data with common factors (balanced or unbalanced)

**Usage**

```
pcce(formula, data, subset, na.action,
      model=c("mg", "p"),
      index = NULL, trend = FALSE, ...)
## S3 method for class 'pcce'
summary(object, ...)
## S3 method for class 'summary.pcce'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
```

**Arguments**

formula	a symbolic description of the model to be estimated,
object, x	an object of class "pcce",
data	a data.frame,
subset	see <code>lm</code> ,
na.action	see <code>lm</code> ,
model	one of "mg", "p", selects Mean Groups vs. Pooled CCE model,
index	the indexes, see <a href="#">pdata.frame</a> ,
trend	logical specifying whether an individual-specific trend has to be included,
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

**Details**

pcce is a function for the estimation of linear panel models by the Common Correlated Effects Mean Groups or Pooled estimator, consistent under the hypothesis of unobserved common factors and idiosyncratic factor loadings. The CCE estimator works by augmenting the model by cross-sectional averages of the dependent variable and regressors in order to account for the common factors, and adding individual intercepts and possibly trends.



**Value**

An object of class `c("pcce", "panelmodel")` containing:

<code>coefficients</code>	the vector of coefficients,
<code>residuals</code>	the vector of (defactored) residuals,
<code>stdres</code>	the vector of (raw) residuals,
<code>tr.model</code>	the transformed data after projection on H,
<code>fitted.values</code>	the vector of fitted values,
<code>vcov</code>	the covariance matrix of the coefficients,
<code>df.residual</code>	degrees of freedom of the residuals,
<code>model</code>	a data.frame containing the variables used for the estimation,
<code>call</code>	the call,
<code>sigma</code>	always NULL, sigma is here only for compatibility reasons (to allow using the same summary and print methods as <code>pggls</code> ),
<code>indcoef</code>	the matrix of individual coefficients from separate time series regressions.

**Author(s)**

Giovanni Millo

**References**

G. Kapetanios, M. Hashem Pesaran, T. Yamagata (2011), Panels with non-stationary multifactor error structures, *Journal of Econometrics*, **160**(2), pp. 326–348.

**Examples**

```
data("Produc", package = "plm")
ccepmod <- pcce(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc, model="p")
summary(ccepmod)
```

---

pcdtest

*Tests of cross-section dependence for panel models*

---

**Description**

Pesaran's CD or Breusch–Pagan's LM (local or global) tests for cross sectional dependence in panel models

**Usage**

```
pcdtest(x, ...)
## S3 method for class 'formula'
pcdtest(x, data, index = NULL, model = NULL,
        test = c("cd", "sclm", "lm", "rho", "absrho"),
        w = NULL, ...)
## S3 method for class 'panelmodel'
pcdtest(x, test = c("cd", "sclm", "lm", "rho", "absrho"),
        w = NULL, ...)
## S3 method for class 'pseries'
pcdtest(x, test = c("cd", "sclm", "lm", "rho", "absrho"),
        w = NULL, ...)
```

**Arguments**

<code>x</code>	an object of class <code>formula</code> , <code>panelmodel</code> , or <code>pseries</code> (depending on the respective interface) describing the model to be tested,
<code>data</code>	a <code>data.frame</code> ,
<code>index</code>	an optional numerical index, if <code>NULL</code> , the first two columns of the <code>data.frame</code> provided in argument <code>data</code> are assumed to be the index variables; for further details see <a href="#">pdata.frame</a> ,
<code>model</code>	an optional character string indicating which type of model to estimate; if left to <code>NULL</code> , the original heterogeneous specification of Pesaran is used,
<code>test</code>	the type of test statistic to be returned. One of <ul style="list-style-type: none"> <li>• "cd" for Pesaran's CD statistic,</li> <li>• "lm" for Breusch and Pagan's original LM statistic,</li> <li>• "sclm" for the scaled version of Breusch and Pagan's LM statistic,</li> <li>• "rho" for the average correlation coefficient,</li> <li>• "absrho" for the average absolute correlation coefficient,</li> </ul>
<code>w</code>	either <code>NULL</code> (default) for the global tests or – for the local versions of the statistics – a $n \times n$ matrix describing proximity between individuals, with $w_{ij} = a$ where $a$ is any number such that <code>as.logical(a)==TRUE</code> , if $i, j$ are neighbours, 0 or any number $b$ such that <code>as.logical(b)==FALSE</code> elsewhere. Only the lower triangular part (without diagonal) of $w$ after coercing by <code>as.logical()</code> is evaluated for neighbouring information (but $w$ can be symmetric). See also <b>Details</b> and <b>Examples</b> ,
<code>...</code>	further arguments to be passed on to <code>plm</code> , such as <code>effect</code> or <code>random.method</code> .

**Details**

These tests are originally meant to use the residuals of separate estimation of one time–series regression for each cross-sectional unit in order to check for cross–sectional dependence. If a different model specification (`within`, `random`, ...) is assumed consistent, one can resort to its residuals for testing (which is common, e.g., when the time dimension's length is insufficient for estimating the heterogeneous model). If the time dimension is insufficient and `model=NULL`, the function defaults to estimation of a `within` model and issues a warning. The main argument of this function may be

either a model of class `panelmodel` or a formula and `dataframe`; in the second case, unless `model` is set to `NULL`, all usual parameters relative to the estimation of a `plm` model may be passed on. The test is compatible with any consistent `panelmodel` for the data at hand, with any specification of effect. E.g., specifying `effect="time"` or `effect="twoways"` allows to test for residual cross-sectional dependence after the introduction of time fixed effects to account for common shocks.

A *local* version of either test can be computed by supplying a proximity matrix (elements coercible to `logical`) with argument `w` which provides information on whether any pair of individuals are neighbours or not. If `w` is supplied, only neighbouring pairs will be used in computing the test; else, `w` will default to `NULL` and all observations will be used. The matrix need not be binary, so commonly used “row-standardized” matrices can be employed as well. `nb` objects from **spdep** must instead be transformed into matrices by **spdep**’s function `nb2mat` before using.

The methods implemented are suitable also for unbalanced panels.

Pesaran’s CD test (`test="cd"`), Breusch and Pagan’s LM test (`test="lm"`), and its scaled version (`test="sclm"`) are all described in Pesaran (2004) (and complemented by Pesaran (2015)). Breusch/Pagan (1980) is the original source for the LM test.

The test on a `pseries` is the same as a test on a pooled regression model of that variable on a constant, i.e. `pcdtest(some_pseries)` is equivalent to `pcdtest(plm(some_var ~ 1, data = some_pdata.frame, model = "pooled"))` and also equivalent to `pcdtest(some_var ~ 1, data = some_data)`, where `some_var` is the variable name in the data which corresponds to `some_pseries`.

## Value

An object of class `"hctest"`.

## References

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**(1), pp. 239–253.

Pesaran, M.H. (2004), General Diagnostic Tests for Cross Section Dependence in Panels, *CESifo Working Paper* 1229.

Pesaran, M.H. (2015), Testing Weak Cross-Sectional Dependence in Large Panels, *Econometric Reviews*, **34**(6-10), pp. 1089–1117.

## Examples

```
data("Grunfeld", package = "plm")
## test on heterogeneous model (separate time series regressions)
pcdtest(inv ~ value + capital, data = Grunfeld,
        index = c("firm", "year"))

## test on two-way fixed effects homogeneous model
pcdtest(inv ~ value + capital, data = Grunfeld, model = "within",
        effect = "twoways", index = c("firm", "year"))

## test on panelmodel object
g <- plm(inv ~ value + capital, data = Grunfeld, index = c("firm", "year"))
pcdtest(g)
```

```
## scaled LM test
pcdtest(g, test = "sclm")

## test on pseries
pGrunfeld <- pdata.frame(Grunfeld)
pcdtest(pGrunfeld$value)

## local test
## define neighbours for individual 2: 1, 3, 4, 5 in lower triangular matrix
w <- matrix(0, ncol= 10, nrow=10)
w[2,1] <- w[3,2] <- w[4,2] <- w[5,2] <- 1
pcdtest(g, w = w)
```

---

`pdata.frame`

*data.frame for panel data*

---

## Description

An object of class 'pdata.frame' is a data.frame with an index attribute that describes its individual and time dimensions.

## Usage

```
pdata.frame(x, index = NULL, drop.index = FALSE, row.names = TRUE,
            stringsAsFactors = default.stringsAsFactors(),
            replace.non.finite = FALSE,
            drop.NA.series = FALSE, drop.const.series = FALSE,
            drop.unused.levels = FALSE)

## S3 method for class 'pdata.frame'
x[i, j, drop]
## S3 method for class 'pdata.frame'
x[[y]]
## S3 method for class 'pdata.frame'
x$y
## S3 method for class 'pdata.frame'
print(x, ...)
## S3 method for class 'pdata.frame'
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
## S3 method for class 'pdata.frame'
as.list(x, keep.attributes = FALSE, ...)
```

## Arguments

<code>x</code>	a data.frame for the <code>pdata.frame</code> function and a <code>pdata.frame</code> for the methods,
<code>i</code>	see <a href="#">Extract</a> ,
<code>j</code>	see <a href="#">Extract</a> ,

y	one of the columns of the data.frame,
index	this argument indicates the individual and time indexes. See <b>Details</b> ,
drop	see <a href="#">Extract</a> ,
drop.index	logical, indicates whether the indexes are to be excluded from the resulting pdata.frame,
optional	see <a href="#">as.data.frame</a> ,
row.names	NULL or logical, indicates whether “fancy” row names (a combination of individual index and time index) are to be added to the returned (p)data.frame (NULL and FALSE have the same meaning),
stringsAsFactors	logical, indicating whether character vectors are to be converted to factors,
replace.non.finite	logical, indicating whether values for which <code>is.finite()</code> yields TRUE are to be replaced by NA values, except for character variables (defaults to FALSE),
drop.NA.series	logical, indicating whether all-NA columns are to be removed from the pdata.frame (defaults to FALSE),
drop.const.series	logical, indicating whether constant columns are to be removed from the pdata.frame (defaults to FALSE),
drop.unused.levels	logical, indicating whether unused levels of factors are to be dropped (defaults to FALSE) (unused levels are always dropped from variables serving to construct the index variables),
keep.attributes	logical, only for <code>as.list</code> , indicating the elements of the returned list should have the pdata.frame’s attributes added (default: FALSE),
...	further arguments

## Details

The `index` argument indicates the dimensions of the panel. It can be:

- a vector of two character strings which contains the names of the individual and of the time indexes,
- a character string which is the name of the individual index variable. In this case, the time index is created automatically and a new variable called “time” is added, assuming consecutive and ascending time periods in the order of the original data,
- an integer, the number of individuals. In this case, the data need to be a balanced panel and be organized as a stacked time series (successive blocks of individuals, each block being a time series for the respective individual) assuming consecutive and ascending time periods in the order of the original data. Two new variables are added: “id” and “time” which contain the individual and the time indexes.

The “[[” and “\$” extract a series from the `pdata.frame`. The “index” attribute is then added to the series and a class attribute “pseries” is added. The “[” method behaves as for `data.frame`,

except that the extraction is also applied to the index attribute. A safe way to extract the index attribute is to use the function `index` for 'pdata.frames' (and other objects).

`as.data.frame` removes the index from the `pdata.frame` and adds it to each column.

`as.list` behaves by default identical to `as.list.data.frame` which means it drops the attributes specific to a `pdata.frame`; if a list of `pseries` is wanted, the attribute `keep.attributes` can be set to `TRUE`. This also makes `lapply` work as expected on a `pdata.frame` (see also **Examples**).

## Value

a `pdata.frame` object: this is a `data.frame` with an index attribute which is a `data.frame` with two variables, the individual and the time indexes, both being factors. The resulting `pdata.frame` is sorted by the individual index, then by the time index.

## Author(s)

Yves Croissant

## See Also

`index` to extract the index variables from a 'pdata.frame' (and other objects), `pdim` to check the dimensions of a 'pdata.frame' (and other objects), `pvar` to check for each variable if it varies cross-sectionally and over time. To check if the time periods are consecutive per individual, see `is.pconsecutive`.

## Examples

```
# Gasoline contains two variables which are individual and time indexes
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)

# Hedonic is an unbalanced panel, townid is the individual index
data("Hedonic", package = "plm")
Hed <- pdata.frame(Hedonic, index = "townid", row.names = FALSE)

# In case of balanced panel, it is sufficient to give number of individuals
# data set 'Wages' is organized as a stacked time series
data("Wages", package = "plm")
Wag <- pdata.frame(Wages, 595)

# lapply on a pdata.frame by making it a list of pseries first
lapply(as.list(Wag[, c("ed", "lwage")]), keep.attributes = TRUE, lag)
```

pdim

*Check for the Dimensions of the Panel***Description**

This function checks the number of individuals and time observations in the panel and whether it is balanced or not.

**Usage**

```
pdim(x, ...)
## S3 method for class 'data.frame'
pdim(x, index = NULL, ...)
## S3 method for class 'panelmodel'
pdim(x, ...)
## S3 method for class 'pseries'
pdim(x, ...)
## S3 method for class 'pdata.frame'
pdim(x, ...)
## S3 method for class 'pgmm'
pdim(x, ...)
```

**Arguments**

x	a data.frame, a pdata.frame, a pseries, a panelmodel, or a pgmm object,
index	see <a href="#">pdata.frame</a> ,
...	further arguments.

**Details**

pdim is called by the estimation functions and can be also used stand-alone.

**Value**

An object of class pdim containing the following elements:

nT	a list containing n, the number of individuals, T, the number of time observations, N the total number of observations,
Tint	a list containing two vectors (of type integer): Ti gives the number of observations for each individual and nt gives the number of individuals observed for each period,
balanced	a logical value: TRUE for a balanced panel, FALSE for an unbalanced panel,
panel.names	a list of character vectors: id.names contains the names of each individual and time.names contains the names of each period.

**Note**

Calling `pdim` on an estimated `panelmodel` object and on the corresponding `(p)data.frame` used for this estimation does not necessarily yield the same result. When called on an estimated `panelmodel`, the number of observations (individual, time) actually used for model estimation are taken into account. When called on a `(p)data.frame`, the rows in the `(p)data.frame` are considered, disregarding any NA values in the dependent or independent variable(s) which would be dropped during model estimation.

**Author(s)**

Yves Croissant

**See Also**

[is.pbalanced](#) to just determine balancedness of data (slightly faster than `pdim`),  
[punbalancedness](#) for measures of unbalancedness,  
[nobs](#), [pdata.frame](#),  
[pvar](#) to check for each variable if it varies cross-sectionally and over time.

**Examples**

```
# There are 595 individuals
data("Wages", package = "plm")
pdim(Wages, 595)

# Gasoline contains two variables which are individual and time indexes
# and are the first two variables
data("Gasoline", package="plm")
pdim(Gasoline)

# Hedonic is an unbalanced panel, townid is the individual index
data("Hedonic", package = "plm")
pdim(Hedonic, "townid")

# An example of the panelmodel method
data("Produc", package = "plm")
z <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp,data=Produc,
         model="random", subset = gsp > 5000)
pdim(z)
```

---

pdwtest

*Durbin–Watson Test for Panel Models*

---

**Description**

Test of serial correlation for (the idiosyncratic component of) the errors in panel models.



**Usage**

```
pdwtest(x, ...)
## S3 method for class 'panelmodel'
pdwtest(x, ...)
## S3 method for class 'formula'
pdwtest(x, data, ...)
```

**Arguments**

x	an object of class "panelmodel" or of class "formula",
data	a data.frame,
...	further arguments to be passed on to dwtest, e.g. alternative, see <a href="#">dwtest</a> for further details.

**Details**

This Durbin–Watson test uses the auxiliary model on (quasi-)demeaned data taken from a model of class `plm` which may be a pooling (the default), random or within model. It performs a Durbin–Watson test (using `dwtest` from package **lmtest**) on the residuals of the (quasi-)demeaned model, which should be serially uncorrelated under the null of no serial correlation in idiosyncratic errors. The function takes the demeaned data, estimates the model and calls `dwtest`.

**Value**

An object of class "hctest".

**Author(s)**

Giovanni Millo

**References**

Baltagi, B.H. (2005) *Econometric Analysis of Panel Data*, 3rd. ed., Wiley, p. 98.  
 Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, p. 288.

**See Also**

[dwtest](#) for the Durbin–Watson test in **lmtest**, [pbgtest](#) for the analogous Breusch–Godfrey test for panel models, [bgtest](#) for the Breusch–Godfrey test for serial correlation in the linear model. [pbltest](#), [pbsytest](#), [pwartest](#) and [pwfdtest](#) for other serial correlation tests for panel models.

**Examples**

```
data("Grunfeld", package = "plm")
g <- plm(inv ~ value + capital, data = Grunfeld, model="random")
pdwtest(g)
pdwtest(g, alternative="two.sided")
## formula interface
pdwtest(inv ~ value + capital, data=Grunfeld, model="random")
```

---

pFormula

*pFormula: An extended Formula interface for panel data*


---

### Description

pFormula is a Formula object used in the plm package.

### Usage

```
pFormula(object)
## S3 method for class 'pFormula'
as.Formula(x, ...)
```

### Arguments

object	an object of class "formula", the formula to be coerced to class "pFormula",
x	an object of class "pFormula", to be coerced to class{"Formula"},
...	further arguments.

### Value

For pFormula, the return value is an object of class c("pFormula", "Formula", "formula").  
 For as.Formula, the return value is an object of class c("Formula", "formula").

### Author(s)

Yves Croissant

### See Also

plm's [model.frame](#) and [model.matrix](#) to create a model frame for panel data and a model matrix with data transformations applied, respectively.  
 plm's [pmodel.response](#) for (transformed) response variable.  
[Formula](#) from package Formula.

### Examples

```
# First, make a pdata.frame
data("Grunfeld", package = "plm")
pGrunfeld <- pdata.frame(Grunfeld)

# then make a model frame from a pFormula and a pdata.frame
pform <- pFormula(inv ~ value + capital)
mf <- model.frame(pform, data = pGrunfeld)

# then construct the (transformed) model matrix (design matrix)
```

```
# from formula and model frame
modmat <- model.matrix(pform, data = mf, model = "within")
```

---

pFtest *F Test for Individual and/or Time Effects*

---

### Description

Test of individual and/or time effects based on the comparison of the within and the pooling model.

### Usage

```
pFtest(x, ...)
## S3 method for class 'plm'
pFtest(x, z, ...)
## S3 method for class 'formula'
pFtest(x, data, ...)
```

### Arguments

x	an object of class "plm" or of class "formula",
z	an object of class "plm",
data	a data.frame,
...	further arguments.

### Details

For the plm method, the argument of this function is two plm objects, the first being a within model, the second a pooling model. The effects tested are either individual, time or twoways, depending on the effects introduced in the within model.

### Value

An object of class "htest".

### Author(s)

Yves Croissant

### See Also

[plmtest](#) for Lagrange multiplier tests of individuals and/or time effects.

**Examples**

```

data("Grunfeld", package="plm")
gp <- plm(inv ~ value + capital, data = Grunfeld, model = "pooling")
gi <- plm(inv ~ value + capital, data = Grunfeld,
          effect = "individual", model = "within")
gt <- plm(inv ~ value + capital, data = Grunfeld,
          effect = "time", model = "within")
gd <- plm(inv ~ value + capital, data = Grunfeld,
          effect = "twoways", model = "within")
pFtest(gi, gp)
pFtest(gt, gp)
pFtest(gd, gp)
pFtest(inv ~ value + capital, data = Grunfeld, effect = "twoways")

```

pggls

*General FGLS Estimators***Description**

General FGLS estimators for panel data (balanced or unbalanced)

**Usage**

```

pggls(formula, data, subset, na.action, effect = c("individual", "time"),
      model = c("within", "random", "pooling", "fd"),
      index = NULL, ...)
## S3 method for class 'pggls'
summary(object, ...)
## S3 method for class 'summary.pggls'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)

```

**Arguments**

formula	a symbolic description of the model to be estimated,
object, x	an object of class pggls,
data	a data.frame,
subset	see <a href="#">lm</a> ,
na.action	see <a href="#">lm</a> ,
effect	the effects introduced in the model, one of "individual" or "time",
model	one of "within", "pooling", "random" or "fd",
index	the indexes, see <a href="#">pdata.frame</a> ,
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

## Details

`pggls` is a function for the estimation of linear panel models by general feasible generalized least squares, either with or without fixed effects. General FGLS is based on a two-step estimation process: first a model is estimated by OLS (`model = "pooling"`), fixed effects (`model = "within"`) or first differences (`model = "fd"`), then its residuals are used to estimate an error covariance matrix for use in a feasible-GLS analysis. This framework allows the error covariance structure inside every group (if `effect = "individual"`, else symmetric) of observations to be fully unrestricted and is therefore robust against any type of intragroup heteroskedasticity and serial correlation. Conversely, this structure is assumed identical across groups and thus general FGLS estimation is inefficient under groupwise heteroskedasticity. Note also that this method requires estimation of  $T(T + 1)/2$  variance parameters, thus efficiency requires  $N \gg T$  (if `effect = "individual"`, else the opposite). Setting `model = "random"` or `model = "pooling"`, both produce an unrestricted FGLS model as in Wooldridge, Ch. 10, although the former is deprecated and included only for retro-compatibility reasons. If `model = "within"` (the default) then a FEGLS (fixed effects GLS, see *ibid.*) is estimated; if `model = "fd"` a FDGLS (first-difference GLS).

## Value

An object of class `c("pggls", "panelmodel")` containing:

<code>coefficients</code>	the vector of coefficients,
<code>residuals</code>	the vector of residuals,
<code>fitted.values</code>	the vector of fitted values,
<code>vcov</code>	the covariance matrix of the coefficients,
<code>df.residual</code>	degrees of freedom of the residuals,
<code>model</code>	a data.frame containing the variables used for the estimation,
<code>call</code>	the call,
<code>sigma</code>	the estimated intragroup (or cross-sectional, if <code>effect = "time"</code> ) covariance of errors,

## Author(s)

Giovanni Millo

## References

- Kiefer, N. M. (1980) Estimation of Fixed Effects Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance, *Journal of Econometrics*, **14**(2), pp. 195–202.
- Im, K. S. and Ahn, S. C. and Schmidt, P. and Wooldridge, J. M. (1999) Efficient Estimation of Panel Data Models with Strictly Exogenous Explanatory Variables, *Journal of Econometrics*, **93**(1), pp. 177–201.
- Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press.

## Examples

```
data("Produc", package = "plm")
zz_wi <- pggls(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
              data = Produc, model = "within")
summary(zz_wi)

zz_pool <- pggls(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
                data = Produc, model = "pooling")
summary(zz_pool)

zz_fd <- pggls(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
              data = Produc, model = "fd")
summary(zz_fd)
```

---

pgmm

*Generalized Method of Moments (GMM) Estimation for Panel Data*

---

## Description

Generalized method of moments estimation for static or dynamic models with panel data.

## Usage

```
pgmm(formula, data, subset, na.action,
      effect = c("twoways", "individual"),
      model = c("onestep", "twosteps"),
      collapse = FALSE,
      lost.ts = NULL,
      transformation = c("d", "ld"),
      fsm = NULL, index = NULL, ...)
## S3 method for class 'pgmm'
summary(object, robust, time.dummies = FALSE, ...)
## S3 method for class 'summary.pgmm'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
```

## Arguments

formula	a symbolic description for the model to be estimated. The preferred interface is now to indicate a multi-part formula, the first two parts describing the covariates and the GMM instruments and, if any, the third part the 'normal' instruments,
object, x	an object of class "pgmm",
data	a data.frame (neither factors nor character vectors will be accepted in data.frame),
subset	see <a href="#">lm</a> ,
na.action	see <a href="#">lm</a> ,

effect	the effects introduced in the model, one of "twoways" (the default) or "individual",
model	one of "onestep" (the default) or "twosteps",
collapse	if TRUE, the GMM instruments are collapsed,
lost.ts	the number of lost time series: if NULL, this is automatically computed. Otherwise, it can be defined by the user as a numeric vector of length 1 or 2. The first element is the number of lost time series in the model in difference, the second one in the model in level. If the second element is missing, it is set to the first one minus one,
transformation	the kind of transformation to apply to the model: either "d" (the default value) for the "difference GMM" model or "ld" for the "system GMM",
fsm	the matrix for the one step estimator: one of "I" (identity matrix) or "G" (= $D'D$ where $D$ is the first-difference operator) if transformation="d", one of "GI" or "full" if transformation="ld",
index	the indexes,
digits	digits,
width	the maximum length of the lines in the print output,
robust	if TRUE, robust inference is performed in the summary,
time.dummies	if TRUE, the estimated coefficients of time dummies are present in the table of coefficients,
...	further arguments.

## Details

pgmm estimates a model for panel data with a generalized method of moments (GMM) estimator. The description of the model to estimate is provided with a multi-part formula which is (or which is coerced to) a Formula object. The first right-hand side part describes the covariates. The second one, which is mandatory, describes the GMM instruments. The third one, which is optional, describes the 'normal' instruments. By default, all the variables of the model which are not used as GMM instruments are used as normal instruments with the same lag structure as the one specified in the model.

$y \sim \text{lag}(y, 1:2) + \text{lag}(x1, 0:1) + \text{lag}(x2, 0:2) \mid \text{lag}(y, 2:99)$  is similar to  $y \sim \text{lag}(y, 1:2) + \text{lag}(x1, 0:1) + \text{lag}(x2, 0:2) + \text{lag}(x1, 0:1) + \text{lag}(x2, 0:2)$  and indicates that all lags from 2 of  $y$  are used as GMM instruments.

transformation indicates how the model should be transformed for the estimation. "d" gives the "difference GMM" model (see Arellano and Bond (1991)), "ld" the "system GMM" model (see Blundell and Bond (1998)).

pgmm is an attempt to adapt GMM estimators available within the DPD library for GAUSS (see Arellano and Bond (1998)) and Ox (see Doornik, Arellano and Bond (2006)) and within the xtabond2 library for Stata (see Roodman (2009)).

## Value

An object of class c("pgmm", "panelmodel"), which has the following elements:

coefficients	the vector (or the list for fixed effects) of coefficients,
residuals	the vector of residuals,

<code>vcov</code>	the covariance matrix of the coefficients,
<code>fitted.values</code>	the vector of fitted values,
<code>df.residual</code>	degrees of freedom of the residuals,
<code>model</code>	a list containing the variables used for the estimation for each individual,
<code>W</code>	a list containing the instruments for each individual (two lists in case of “sys-GMM”),
<code>A1</code>	the weighting matrix for the one-step estimator,
<code>A2</code>	the weighting matrix for the two-steps estimator,
<code>call</code>	the call.

It has `print`, `summary` and `print.summary` methods.

### Author(s)

Yves Croissant

### References

Arellano, M. and Bond, S. (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies*, **58**(2), 1991, pp. 227–297.

Arellano, M. and Bond, S. (1998) Dynamic Panel Data Estimation Using DPD98 for GAUSS: A Guide for Users.

Blundell, R. and Bond, S. (1998) Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, vol. **87**(1), pp. 115–143.

Doornik, J., Arellano, M. and Bond, S. (2006) Panel Data Estimation using DPD for Ox. [http://www.doornik.com/download/oxmetrics7/Ox\\_Packages/dpd.pdf](http://www.doornik.com/download/oxmetrics7/Ox_Packages/dpd.pdf)

Roodman, D. (2009) How to do xtabond2: An introduction to difference and system GMM in Stata, *Stata Journal*, vol. **9**(1), pp. 86–136. <http://www.stata-journal.com/article.html?article=st0159>.

### See Also

[sargan](#) for Sargan tests and [mtest](#) for Arellano–Bond’s tests of serial correlation. [dynformula](#) for dynamic formulas (deprecated).

### Examples

```
data("EmplUK", package = "plm")

## Arellano and Bond (1991), table 4 col. b
z1 <- pgmm(log(emp) ~ lag(log(emp), 1:2) + lag(log(wage), 0:1)
          + log(capital) + lag(log(output), 0:1) | lag(log(emp), 2:99),
          data = EmplUK, effect = "twoways", model = "twosteps")
summary(z1, robust = FALSE)

## Blundell and Bond (1998) table 4 (cf. DPD for OX p. 12 col. 4)
```



```

z2 <- pgmm(log(emp) ~ lag(log(emp), 1)+ lag(log(wage), 0:1) +
           lag(log(capital), 0:1) | lag(log(emp), 2:99) +
           lag(log(wage), 2:99) + lag(log(capital), 2:99),
           data = EmplUK, effect = "twoways", model = "onestep",
           transformation = "ld")
summary(z2, robust = TRUE)

## Not run:
## Same with the old formula or dynformula interface
## Arellano and Bond (1991), table 4, col. b
z1 <- pgmm(log(emp) ~ log(wage) + log(capital) + log(output),
           lag.form = list(2,1,0,1), data = EmplUK,
           effect = "twoways", model = "twosteps",
           gmm.inst = ~log(emp), lag.gmm = list(c(2,99)))
summary(z1, robust = FALSE)

## Blundell and Bond (1998) table 4 (cf DPD for OX p. 12 col. 4)
z2 <- pgmm(dynformula(log(emp) ~ log(wage) + log(capital), list(1,1,1)),
           data = EmplUK, effect = "twoways", model = "onestep",
           gmm.inst = ~log(emp) + log(wage) + log(capital),
           lag.gmm = c(2,99), transformation = "ld")
summary(z2, robust = TRUE)

## End(Not run)

```

pgrangertest

*Panel Granger Causality Test (Dumitrescu/Hurlin (2012))***Description**

Test for Granger causality in panel data.

**Usage**

```

pgrangertest(formula, data, test = c("Ztilde", "Zbar", "Wbar"),
             order = 1L, index = NULL)

```

**Arguments**

formula	a formula object to describe the direction of the hypothesized Granger causation,
data	a <code>pdata.frame</code> or a <code>data.frame</code> ,
test	a character to request the statistic to be returned, either "Ztilde" (default), or "Zbar",
order	integer(s) giving the number of lags to include in the test's auxiliary regressions, the length of order must be either 1 (same lag order for all individuals) or equal to the number of individuals (to specify a lag order per individual),
index	only relevant if data is <code>data.frame</code> and not a <code>pdata.frame</code> ; if NULL, the first two columns of the <code>data.frame</code> are assumed to be the index variables, for further details see <a href="#">pdata.frame</a> .

## Details

The panel Granger causality test is a combination of Granger tests (Granger (1969)) performed per individual. The test is developed by Dumitrescu/Hurlin (2012), a shorter exposition is given in Lopez/Weber (2017).

The formula  $y \sim x$  describes the direction of the (panel) Granger causation where  $y \sim x$  means "x (panel) Granger causes y".

By setting argument `test` to either "Ztilde" (default) or "Zbar", two different statistics can be requested. "Ztilde" gives the standardised statistic recommended by Dumitrescu/Hurlin (2012) for fixed T samples. If set to "Wbar", the intermediate Wbar statistic (average of individual Granger chi-square statistics) is given which is used to derive the other two.

The Zbar statistic is not suitable for unbalanced panels. For the Wbar statistic, no p-value is available.

The implementation uses `grangertest` from package `lmtest` to perform the individual Granger tests.

## Value

An object of class `c("pgrangertest", "htest")`. Besides the usual elements of a `htest` object, it contains element `indgranger` which carries the Granger test statistics along their p-values, degrees of freedom and the specified lag order per individual as a data frame.

## Author(s)

Kevin Tappe

## References

Dumitrescu, E.-I./Hurlin, C. (2012), Testing for Granger non-causality in heterogeneous panels, *Economic Modelling*, **29**(4), pp. 1450–460.

Granger, C. W. J. (1969), Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, **37**(3), pp. 424–438.

Lopez, L./Weber, S. (2017), Testing for Granger causality in panel data, *IRENE Working paper 17-03*, September 11, 2017.

## See Also

`grangertest` for the Granger causality test in `lmtest`,

## Examples

```
## not meaningful, just to demonstrate usage
## H0: 'value' Granger causes 'inv' for all individuals

data("Grunfeld", package = "plm")
pgrangertest(inv ~ value, data = Grunfeld)
pgrangertest(inv ~ value, data = Grunfeld, order = 2L)
pgrangertest(inv ~ value, data = Grunfeld, order = 2L, test = "Zbar")
```

```
# varying lag order (last individual lag order 3, others lag order 2)
pgrangertest(inv ~ value, data = Grunfeld, order = c(rep(2L, 9), 3L))
```

pht

*Hausman–Taylor Estimator for Panel Data***Description**

The Hausman–Taylor estimator is an instrumental variable estimator without external instruments.

**Usage**

```
pht(formula, data, subset, na.action, model = c("ht", "am", "bms"), index = NULL, ...)
## S3 method for class 'pht'
summary(object, ...)
## S3 method for class 'summary.pht'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), subset = NULL, ...)
```

**Arguments**

formula	a symbolic description for the model to be estimated,
object, x	an object of class "plm",
data	a data.frame,
subset	see <a href="#">lm</a> for "plm", a character or numeric vector indicating a subset of the table of coefficient to be printed for "print.summary.plm",
na.action	see <a href="#">lm</a> ,
model	one of "ht" for Hausman–Taylor, "am" for Amemiya–MaCurdy and "bms" for Breusch–Mizon–Schmidt,
index	the indexes,
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

**Details**

pht estimates panels models using the Hausman–Taylor estimator, Amemiya–MaCurdy estimator, or Breusch–Mizon–Schmidt estimator, depending on the argument model. The model is specified as a two–part formula, the second part containing the exogenous variables.

**Value**

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

A "pht" object contains the same elements as plm, with a further argument called varlist which describes the typology of the variables. It has summary and print.summary methods.

**Author(s)**

Yves Croissant

**References**

Amemiya, T. and MaCurdy, T.E. (1986) Instrumental–variable estimation of an error components model, *Econometrica*, **54**(4), pp. 869–880.

Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.

Breusch, T.S., Mizon, G.E. and Schmidt, P. (1989) Efficient estimation using panel data, *Econometrica*, **57**(3), pp. 695–700.

Hausman, J.A. and Taylor W.E. (1981) Panel data and unobservable individual effects, *Econometrica*, **49**(6), pp. 1377–1398.

**Examples**

```
# replicates Baltagi (2005, 2013), table 7.4
data("Wages", package = "plm")
ht <- pht(lwage ~ wks + south + smsa + married + exp + I(exp^2) +
         bluecol + ind + union + sex + black + ed |
         sex + black + bluecol + south + smsa + ind,
         data = Wages, model = "ht", index = 595)
summary(ht)
```

---

phtest

*Hausman Test for Panel Models*


---

**Description**

Specification test for panel models.

**Usage**

```
phtest(x, ...)
## S3 method for class 'panelmodel'
phtest(x, x2, ...)
## S3 method for class 'formula'
phtest(x, data, model = c("within", "random"),
       method = c("chisq", "aux"),
       index = NULL, vcov = NULL, ...)
```

**Arguments**

x                    an object of class "panelmodel" or "formula",

x2                    an object of class "panelmodel",

model                a character vector containing the names of two models (length(model) must be 2),

data	a data.frame,
method	one of "chisq" or "aux",
index	an optional vector of index variables,
vcov	an optional covariance function,
...	further arguments to be passed on. For the formula method, place argument effect here to compare e.g. twoway models (effect = "twoways") Note: Argument effect is not respected in the panelmodel method.

### Details

The Hausman test (sometimes also called Durbin–Wu–Hausman test) is based on the difference of the vectors of coefficients of two different models. The `panelmodel` method computes the original version of the test based on a quadratic form (Hausman (1978)). The `formula` method, if `method="chisq"` (default), computes the original version of the test based on a quadratic form; if `method="aux"` then the auxiliary-regression-based version in Wooldridge (2010, Sec. 10.7.3.) is computed instead. Only the latter can be robustified by specifying a robust covariance estimator as a function through the argument `vcov` (see **Examples**).

The equivalent tests in the **one-way** case using a between model (either "within vs. between" or "random vs. between") (see Hausman/Taylor (1981) or Baltagi (2013), Sec. 4.3) can also be performed by `phtest`, but only for `test = "chisq"`, not for the regression-based test. NB: These equivalent tests using the between model do not extend to the two-ways case. (There are, however, some other equivalent tests, see Kang (1985) or Baltagi (2013), Sec. 4.3.7), but those are unsupported by `phtest`.)

### Value

An object of class "htest".

### Author(s)

Yves Croissant, Giovanni Millo

### References

- Hausman, J.A. (1978), Specification tests in econometrics, *Econometrica*, **46**(6), pp. 1251–1271.
- Hausman, J.A./Taylor, W.E. (1981), Panel data and unobservable individual effects, *Econometrica*, **49**(6), pp. 1377–1398.
- Kang, Suk (1985), A note on the equivalence of specification tests in the two-factor multivariate variance components model, *Journal of Econometrics*, **28**(2), pp. 193–203.
- Wooldridge, Jeffrey M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Sec. 10.7.3., pp. 328–334.
- Baltagi, Badi H. (2013), *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons., Sec. 4.3.

**Examples**

```

data("Gasoline", package = "plm")
form <- lgaspcar ~ lincomep + lrpmpg + lcarpcap
wi <- plm(form, data = Gasoline, model = "within")
re <- plm(form, data = Gasoline, model = "random")
phtest(wi, re)
phtest(form, data = Gasoline)
phtest(form, data = Gasoline, method = "aux")

# robust Hausman test (regression-based)
phtest(form, data = Gasoline, method = "aux", vcov = vcovHC)

# robust Hausman test with vcov supplied as a
# function and additional parameters
phtest(form, data = Gasoline, method = "aux",
        vcov = function(x) vcovHC(x, method="white2", type="HC3"))

```

---

piest

*Chamberlain estimator and test for fixed effects*


---

**Description**

General estimator useful for testing the within specification

**Usage**

```

piest(formula, data, subset, na.action, index = NULL, robust = TRUE, ...)
## S3 method for class 'piest'
summary(object, ...)
## S3 method for class 'summary.piest'
print(x, ...)

```

**Arguments**

formula	a symbolic description for the model to be estimated,
object, x	an object of class "plm",
data	a data.frame,
subset	see <a href="#">lm</a> ,
na.action	see <a href="#">lm</a> ,
index	the indexes,
robust	if FALSE, the error as assumed to be spherical, otherwise, a robust estimation of the covariance matrix is computed,
...	further arguments.

**Details**

The Chamberlain method consists on using the covariates of all the periods as regressors. It allows to test the within specification.

**Value**

An object of class "piest".

**Author(s)**

Yves Croissant

**References**

Chamberlain, G. (1982) Multivariate regression for panel data, *Journal of Econometrics*, **18**(1), pp. 5–46.

**Examples**

```
data("RiceFarms", package = "plm")
pirice <- piest(log(goutput) ~ log(seed) + log(totlabor) + log(size), RiceFarms, index = "id")
summary(pirice)
```

---

pldv

*Panel estimators for limited dependent variables*

---

**Description**

Fixed and random effects estimators for truncated or censored limited dependent variable

**Usage**

```
pldv(formula, data, subset, weights, na.action,
      model = c("fd", "random", "pooling"), index = NULL,
      R = 20, start = NULL, lower = 0, upper = + Inf,
      objfun = c("lsq", "lad"), sample = c("cens", "trunc"), ...)
```

**Arguments**

formula	a symbolic description for the model to be estimated,
data	a data.frame,
subset	see lm,
weights	see lm,
na.action	see lm,
model	one of "fd", "random" or "pooling",
index	the indexes, see <a href="#">pdata.frame</a> ,

R	the number of points for the gaussian quadrature,
start	a vector of starting values,
lower	the lower bound for the censored/truncated dependent variable,
upper	the upper bound for the censored/truncated dependent variable,
objfun	the objective function for the fixed effect model, one of "lsq" for least squares and "lad" for least absolute deviations,
sample	"cens" for a censored (tobit-like) sample, "trunc" for a truncated sample,
...	further arguments.

### Details

pldv computes two kinds of models : maximum likelihood estimator with an assumed normal distribution for the individual effects and a LSQ/LAD estimator for the first-difference model.

### Value

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

### Author(s)

Yves Croissant

### References

Honoré, Bo (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, *Econometrica*, **60**(3), pp. 533–565.

---

plm

*Panel Data Estimators*

---

### Description

Linear models for panel data estimated using the `lm` function on transformed data.

### Usage

```
plm(formula, data, subset, weights, na.action,
     effect = c("individual", "time", "twoways", "nested"),
     model = c("within", "random", "ht", "between", "pooling", "fd"),
     random.method = NULL,
     random.models = NULL,
     random.dfcor = NULL,
     inst.method = c("bvk", "baltagi", "am", "bms"),
     restrict.matrix = NULL,
     restrict.rhs = NULL,
     index = NULL, ...)
```



```

## S3 method for class 'panelmodel'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'plm'
plot(x, dx = 0.2, N = NULL, seed = 1,
     within = TRUE, pooling = TRUE, between = FALSE, random = FALSE, ...)

```

### Arguments

formula	a symbolic description for the model to be estimated,
x	an object of class "plm",
data	a data.frame,
subset	see <a href="#">lm</a> ,
weights	see <a href="#">lm</a> ,
na.action	see <a href="#">lm</a> ; currently, not fully supported,
effect	the effects introduced in the model, one of "individual", "time", "twoways", or "nested",
model	one of "pooling", "within", "between", "random", "fd", or "ht",
random.method	method of estimation for the variance components in the random effects model, one of "swar" (default), "amemiya", "walhus", or "nerlove",
random.models	an alternative to the previous argument, the models used to compute the variance components estimations are indicated,
random.dfcor	a numeric vector of length 2 indicating which degree of freedom should be used,
inst.method	the instrumental variable transformation: one of "bvk", "baltagi", "am", or "bms" (see also Details),
index	the indexes,
restrict.matrix	a matrix which defines linear restrictions on the coefficients,
restrict.rhs	the right hand side vector of the linear restrictions on the coefficients,
digits	number of digits for printed output,
width	the maximum length of the lines in the printed output,
dx	the half-length of the individual lines for the plot method (relative to x range),
N	the number of individual to plot,
seed	the seed which will lead to individual selection,
within	if TRUE, the within model is plotted,
pooling	if TRUE, the pooling model is plotted,
between	if TRUE, the between model is plotted,
random	if TRUE, the random effect model is plotted,
...	further arguments.

## Details

`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)), `"wallhus"` (Wallace and Hussain (1969)), or `"nerlove"` (Nerlove (1971)).

For first-difference models, the intercept is maintained (which from a specification viewpoint amounts to allowing for a trend in the levels model). The user can exclude it from the estimated specification the usual way by adding `"-1"` to the model formula.

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  $y \sim x_1 + x_2 + x_3$ , with  $x_1$  and  $x_2$  endogenous and  $z_1$  and  $z_2$  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.`

If an instrument variable estimation is requested, argument `inst.method` selects the instrument variable transformation method:

- `"bvk"` (default) for Balestra and Varadharajan-Krishnakumar (1987),
- `"baltagi"` for Baltagi (1981),
- `"am"` for Amemiya and MaCurdy (1986),
- `"bms"` for Breusch, Mizon, and Schmidt (1989).

The Hausman–Taylor estimator (Hausman and Taylor (1981)) is computed if `model = "ht"`.

## Value

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

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

<code>coefficients</code>	the vector of coefficients,
<code>vcov</code>	the variance–covariance matrix of the coefficients,
<code>residuals</code>	the vector of residuals (these are the residuals of the (quasi-)demeaned model),
<code>df.residual</code>	degrees of freedom of the residuals,
<code>formula</code>	an object of class <code>"pFormula"</code> describing the model,
<code>model</code>	the model frame as a <code>"pdata.frame"</code> containing the variables used for estimation: the response is in first column followed by the other variables, the individual and time indexes are in the <code>'index'</code> attribute of <code>model</code> ,
<code>ercomp</code>	an object of class <code>"ercomp"</code> providing the estimation of the components of the errors (for random effects models only),

aliased	named logical vector indicating any aliased coefficients which are silently dropped by plm due to linearly dependent terms (see also <a href="#">detect_lin_dep</a> ),
call	the call.

It has `print`, `summary` and `print.summary` methods. The `summary` method creates an object of class "summary.plm" that extends the object it is run on with information about (inter alia) F statistic and (adjusted) R-squared of model, standard errors, t-values, and p-values of coefficients, (if supplied) the furnished vcov, see [summary.plm](#) for further details.

### Author(s)

Yves Croissant

### References

- Amemiya, T. (1971) The estimation of the variances in a variance-components model, *International Economic Review*, **12**(1), pp. 1–13.
- Amemiya, T. and MaCurdy, T.E. (1986) Instrumental-variable estimation of an error components model, *Econometrica*, **54**(4), pp. 869–880.
- Balestra, P. and Varadharajan-Krishnakumar, J. (1987) Full information estimations of a system of simultaneous equations with error components structure, *Econometric Theory*, **3**(2), pp. 223–246.
- Baltagi, B.H. (1981) Simultaneous equations with error components, *Journal of Econometrics*, **17**(2), pp. 189–200.
- Baltagi, B.H.; Song, S.H.; Jung, B.C. (2001), "The unbalanced nested error component regression model", *Journal of Econometrics*, **101**(2), pp. 357–381.
- Baltagi, B.H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.
- Breusch, T.S., Mizon, G.E. and Schmidt, P. (1989) Efficient estimation using panel data, *Econometrica*, **57**(3), pp. 695–700.
- Hausman, J.A. and Taylor, W.E. (1981) Panel data and unobservable individual effects, *Econometrica*, **49**(6), pp. 1377–1398.
- Nerlove, M. (1971) Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica*, **39**(2), pp. 359–382.
- Swamy, P.A.V.B. and Arora, S.S. (1972) The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica*, **40**(2), pp. 261–275.
- Wallace, T.D. and Hussain, A. (1969) The use of error components models in combining cross section with time series data, *Econometrica*, **37**(1), pp. 55–72.

### See Also

[summary.plm](#) for further details about the associated `summary` method and the "summary.plm" object both of which provide some model tests and tests of coefficients. [fixef](#) to compute the fixed effects for "within" models (=fixed effects models).

## Examples

```

data("Produc", package = "plm")
zz <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
          data = Produc, index = c("state", "year"))
summary(zz)

# replicates some results from Baltagi (2013), table 3.1
data("Grunfeld", package = "plm")
p <- plm(inv ~ value + capital,
          data = Grunfeld, model = "pooling")

wi <- plm(inv ~ value + capital,
           data = Grunfeld, model = "within", effect = "twoways")

swar <- plm(inv ~ value + capital,
             data = Grunfeld, model = "random", effect = "twoways")

amemiya <- plm(inv ~ value + capital,
                data = Grunfeld, model = "random", random.method = "amemiya",
                effect = "twoways")

walhus <- plm(inv ~ value + capital,
               data = Grunfeld, model = "random", random.method = "walhus",
               effect = "twoways")

# summary and summary with a furnished vcov (passed as matrix,
# as function, and as function with additional argument)
summary(wi)
summary(wi, vcov = vcovHC(wi))
summary(wi, vcov = vcovHC)
summary(wi, vcov = function(x) vcovHC(x, method = "white2"))

# nested random effect model
# replicate Baltagi/Song/Jung (2001), p. 378 (table 6), columns SA, WH
# == Baltagi (2013), pp. 204-205
data("Produc", package = "plm")
pProduc <- pdata.frame(Produc, index = c("state", "year", "region"))
form <- log(gsp) ~ log(pc) + log(emp) + log(hwy) + log(water) + log(util) + unemp
summary(plm(form, data = pProduc, model = "random", effect = "nested"))
summary(plm(form, data = pProduc, model = "random", effect = "nested",
            random.method = "walhus"))

```

---

plm.data

*Data Frame Special Format for Panel Data (Deprecated)*

---

## Description

This function was once used to transform a data frame in a format suitable for using with the estimation functions of `plm`. New code should rather use the more versatile function `pdata.frame`.

**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 (see Details).

**Details**

This function is kept due to backward compatibility of old code. New code should use the function `pdata.frame` instead.

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.

**Value**

An object of class `c("plm.dim", "data.frame")`.

**Author(s)**

Yves Croissant

**Examples**

```
# There are 595 individuals
data("Wages", package = "plm")
Wages <- plm.data(Wages, 595)

# Gasoline contains two variables which are individual and time indexes
# The pdata.frame is called gas
data("Gasoline", package = "plm")
Gasoline <- plm.data(Gasoline, c("country","year"))
summary(Gasoline)

# Hedonic is an unbalanced panel, townid is the individual index
data("Hedonic", package = "plm")
Hedonic <- plm.data(Hedonic, "townid")
```

**Description**

Test of individual and/or time effects for panel models.

**Usage**

```
plmtest(x, ...)
## S3 method for class 'plm'
plmtest(x, effect = c("individual", "time", "twoways"),
        type = c("honda", "bp", "ghm", "kw"), ...)
## S3 method for class 'formula'
plmtest(x, data, ..., effect = c("individual", "time", "twoways"),
        type = c("honda", "bp", "ghm", "kw"))
```

**Arguments**

x	an object of class "plm" or a formula of class "formula",
data	a data.frame,
effect	a character string indicating which effects are tested: individual effects ("individual"), time effects ("time") or both ("twoways"),
type	a character string indicating the test to be performed: <ul style="list-style-type: none"> <li>• "honda" (default) for Honda (1985),</li> <li>• "bp" for Breusch/Pagan (1980),</li> <li>• "kw" for King/Wu (1997), or</li> <li>• "ghm" for Gouriéroux/Holly/Monfort (1982);</li> </ul> For unbalanced panel data sets, the respective unbalanced version of the tests are computed,
...	further arguments passed to plmtest.

**Details**

These Lagrange multiplier tests use only the residuals of the pooling model. The first argument of this function may be either a pooling model of class `plm` or an object of class `formula` describing the model. For inputted within (fixed effects) or random effects models, the corresponding pooling model is calculated internally first as the tests are based on the residuals of the pooling model.

The "bp" test for unbalanced panels was derived in Baltagi/Li (1990), the "kw" test for unbalanced panels in Baltagi/Chang/Li (1998). The "ghm" test and the "kw" test were extended to two-way effects in Baltagi/Chang/Li (1992).

For a concise overview of all these statistics see Baltagi (2013), Sec. 4.2, pp. 68–76 (for balanced panels) and Sec. 9.5, pp. 200–203 (for unbalanced panels).

**Value**

An object of class "htest".

**Note**

For the King-Wu statistics ("kw"), the oneway statistics ("individual" and "time") coincide with the respective Honda statistics ("honda"); twoway statistics of "kw" and "honda" differ.

**Author(s)**

Yves Croissant (initial implementation), Kevin Tappe (generalization to unbalanced panels)

**References**

- Baltagi, B. H. (2013) *Econometric Analysis of Panel Data*, 5th edition, Sec. 4.2, pp. 68–76 and Sec. 9.5, pp. 200–203.
- Baltagi, B. H./Li, Q. (1990) A lagrange multiplier test for the error components model with incomplete panels, *Econometric Reviews*, **9**(1), pp. 103–107.
- Baltagi B. H./Chang, Y./Li, Q. (1992) Monte Carlo results on several new and existing tests for the error component model, *Journal of Econometrics*, **54**(1-3), pp. 95–120.
- Baltagi B. H./Chang, Y./Li, Q. (1998) Testing for random individual and time effects using unbalanced panel data, *Advances in Econometrics*, **13**, pp. 1–20.
- Breusch, T. S./Pagan, A. R. (1980) The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies*, **47**(1), pp. 239–253.
- Gourieroux, C./Holly, A./Monfort, A. (1982) Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica*, **50**(1), pp. 63–80.
- Honda, Y. (1985) Testing the error components model with non–normal disturbances, *Review of Economic Studies*, **52**(4), pp. 681–690.
- King, M. L./Wu, P. X. (1997) Locally optimal one–sided tests for multiparameter hypotheses, *Econometric Reviews*, **16**(2), pp. 131–156.

**See Also**

[pFtest](#) for individual and/or time effects tests based on the within model.

**Examples**

```
data("Grunfeld", package = "plm")
g <- plm(inv ~ value + capital, data = Grunfeld, model = "pooling")
plmtest(g)
plmtest(g, effect="time")
plmtest(inv ~ value + capital, data = Grunfeld, type = "honda")
plmtest(inv ~ value + capital, data = Grunfeld, type = "bp")
plmtest(inv ~ value + capital, data = Grunfeld, type = "bp", effect = "twoways")
plmtest(inv ~ value + capital, data = Grunfeld, type = "ghm", effect = "twoways")
plmtest(inv ~ value + capital, data = Grunfeld, type = "kw", effect = "twoways")
```

```

Grunfeld_unbal <- Grunfeld[1:(nrow(Grunfeld)-1), ] # create an unbalanced panel data set
g_unbal <- plm(inv ~ value + capital, data = Grunfeld_unbal, model = "pooling")
plmtest(g_unbal) # unbalanced version of test is indicated in output

```

pmg

*Mean Groups (MG), Demeaned MG and CCE MG estimators***Description**

Mean Groups (MG), Demeaned MG (DMG) and Common Correlated Effects MG (CCEMG) estimators for heterogeneous panel models, possibly with common factors (CCEMG)

**Usage**

```

pmg(formula, data, subset, na.action,
     model = c("mg", "cmg", "dmg"),
     index = NULL, trend = FALSE, ...)
## S3 method for class 'pmg'
summary(object, ...)
## S3 method for class 'summary.pmg'
print(x, digits = max(3, getOption("digits") -
2), width = getOption("width"), ...)

```

**Arguments**

formula	a symbolic description of the model to be estimated,
object, x	an object of class pmg,
data	a data.frame,
subset	see <a href="#">lm</a> ,
na.action	see <a href="#">lm</a> ,
model	one of c("mg", "cmg", "dmg"),
index	the indexes, see <a href="#">pdata.frame</a> ,
trend	logical specifying whether an individual-specific trend has to be included,
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

**Details**

pmg is a function for the estimation of linear panel models with heterogeneous coefficients by the Mean Groups estimator. `model = "mg"` specifies the standard Mean Groups estimator, based on the average of individual time series regressions. If `model = "dmg"` the data are demeaned cross-sectionally, which is believed to reduce the influence of common factors (and is akin to what is done in homogeneous panels when `model = "within"` and `effect = "time"`). Lastly, if



model = "cmg" the CCEMG estimator is employed: this latter is consistent under the hypothesis of unobserved common factors and idiosyncratic factor loadings; it works by augmenting the model by cross-sectional averages of the dependent variable and regressors in order to account for the common factors, and adding individual intercepts and possibly trends.

### Value

An object of class c("pmg", "panelmodel") containing:

coefficients	the vector of coefficients,
residuals	the vector of residuals,
fitted.values	the vector of fitted values,
vcov	the covariance matrix of the coefficients,
df.residual	degrees of freedom of the residuals,
model	a data.frame containing the variables used for the estimation,
call	the call,
sigma	always NULL, sigma is here only for compatibility reasons (to allow using the same summary and print methods as pggls),
indcoef	the matrix of individual coefficients from separate time series regressions.

### Author(s)

Giovanni Millo

### References

M. Hashem Pesaran (2006), Estimation and Inference in Large Heterogeneous Panels with a Multi-factor Error Structure, *Econometrica*, **74**(4), pp. 967–1012.

### Examples

```
data("Produc", package = "plm")
## Mean Groups estimator
mgmod <- pmg(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc)
summary(mgmod)

## demeaned Mean Groups
dmgmod <- pmg(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
              data = Produc, model = "dmg")
summary(dmgmod)

## Common Correlated Effects Mean Groups
ccemgmod <- pmg(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
               data = Produc, model = "cmg")
summary(ccemgmod)
```

pmodel.response      *A function to extract the model.response*

---

### Description

pmodel.response has several methods to conveniently extract the response of several objects.

### Usage

```
pmodel.response(object, ...)  
## S3 method for class 'data.frame'  
pmodel.response(object, ...)  
## S3 method for class 'formula'  
pmodel.response(object, data, ...)  
## S3 method for class 'plm'  
pmodel.response(object, ...)
```

### Arguments

object	an object of class "plm", or a formula of class "pFormula",
data	a data.frame
...	further arguments.

### Details

The model response is extracted from a pdata.frame (where the response must reside in the first column; this is the case for a model frame), a pFormula + data or a plm object, and the transformation specified by effect and model is applied to it.

Constructing the model frame first ensures proper NA handling and the response being placed in the first column, see also **Examples** for usage.

### Value

A numeric vector.

### Author(s)

Yves Croissant

### See Also

plm's [model.matrix](#) for (transformed) model matrix and the corresponding [model.frame](#) method to construct a model frame.

**Examples**

```

# First, make a pdata.frame
data(Grunfeld)
pGrunfeld <- pdata.frame(Grunfeld)

# then make a model frame from a pFormula and a pdata.frame
pform <- pFormula(inv ~ value + capital)
mf <- model.frame(pform, data = pGrunfeld)

# construct (transformed) response of the within model
resp <- pmodel.response(pform, data = mf, model = "within", effect = "individual")

# retrieve (transformed) response directly from model frame
resp_mf <- pmodel.response(mf, model = "within", effect = "individual")

# retrieve (transformed) response from a plm object, i.e. an estimated model
fe_model <- plm(pform, data = pGrunfeld, model = "within")
pmodel.response(fe_model)

# same as constructed before
all.equal(resp, pmodel.response(fe_model), check.attributes = FALSE) # TRUE

```

---

pooltest

*Test of Poolability*

---

**Description**

A Chow test for the poolability of the data.

**Usage**

```

pooltest(x, ...)
## S3 method for class 'plm'
pooltest(x, z, ...)
## S3 method for class 'formula'
pooltest(x, data, ...)

```

**Arguments**

x	an object of class "plm" for the plm method; an object of class "formula" for the formula interface,
z	an object of class "pvcv" obtained with model="within",
data	a data.frame,
...	further arguments passed to plm.

**Details**

`pooltest` is a  $F$  test of stability (or Chow test) for the coefficients of a panel model. For argument `x`, the estimated `plm` object should be a "pooling" model or a "within" model (the default); intercepts are assumed to be identical in the first case and different in the second case.

**Value**

An object of class "htest".

**Author(s)**

Yves Croissant

**Examples**

```
data("Gasoline", package = "plm")
form <- lgaspcar ~ lincomep + lrpmg + lcarpcap
gasw <- plm(form, data = Gasoline, model = "within")
gasp <- plm(form, data = Gasoline, model = "pooling")
gasnp <- pvcn(form, data = Gasoline, model = "within")
pooltest(gasw, gasnp)
pooltest(gasp, gasnp)

pooltest(form, data = Gasoline, effect = "individual", model = "within")
pooltest(form, data = Gasoline, effect = "individual", model = "pooling")
```

---

Produc

*US States Production*

---

**Description**

A panel of 48 observations from 1970 to 1986

*total number of observations* : 816

*observation* : regional

*country* : United States

**Usage**

```
data(Produc)
```

**Format**

A data frame containing :

**state** the state

**year** the year

**region** the region

**pcap** public capital stock  
**hwy** highway and streets  
**water** water and sewer facilities  
**util** other public buildings and structures  
**pc** private capital stock  
**gsp** gross state product  
**emp** labor input measured by the employment in non-agricultural payrolls  
**unemp** state unemployment rate

### Source

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

### References

Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.

Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.

Baltagi, B. H. and N. Pinnoi (1995) "Public capital stock and state productivity growth: further evidence", *Empirical Economics*, **20**(2), pp. 351–359.

Munnell, A.H. (1990) "Why has productivity growth declined? Productivity and public investment", *New England Economic Review*, pp. 3–22.

---

pseries

*panel series*

---

### Description

A class for panel series for which several useful computations and data transformations are available.

### Usage

```
between(x, ...)
Between(x, ...)
Within(x, ...)
## S3 method for class 'pseries'
between(x, effect = c("individual", "time", "group"), ...)
## S3 method for class 'pseries'
Between(x, effect = c("individual", "time", "group"), ...)
```

```
## S3 method for class 'pseries'
Within(x, effect = c("individual", "time", "group", "twoways"), ...)
## S3 method for class 'pseries'
summary(object, ...)
## S3 method for class 'summary.pseries'
print(x, ...)
## S3 method for class 'pseries'
as.matrix(x, idbyrow = TRUE, ...)
```

### Arguments

<code>x</code> , <code>object</code>	a <code>pseries</code> or a <code>summary.pseries</code> object,
<code>effect</code>	character string indicating the "individual" or "time" effect,
<code>idbyrow</code>	if TRUE in the <code>as.matrix</code> method, the lines of the matrix are the individuals,
<code>...</code>	further arguments, e. g. <code>na.rm = TRUE</code> for transformation functions like <code>between</code> , see <b>Details</b> and <b>Examples</b> .

### Details

The functions `between`, `Between`, and `Within` perform specific data transformations, i. e. the `between` and `within` transformation.

`between` returns a vector containing the individual means (over time) with the length of the vector equal to the number of individuals (if `effect = "individual"` (default); if `effect = "time"`, it returns the time means (over individuals)). `Between` duplicates the values and returns a vector which length is the number of total observations. `Within` returns a vector containing the values in deviation from the individual means (if `effect = "individual"`, from time means if `effect = "time"`), the so called demeaned data.

For `between`, `Between`, and `Within` in presence of NA values it can be useful to supply `na.rm = TRUE` as an additional argument to keep as many observations as possible in the resulting transformation, see also **Examples**.

### Value

All these functions return an object of class `pseries`, except: `between`, which returns a numeric vector, `as.matrix`, which returns a matrix.

### Author(s)

Yves Croissant

### See Also

For more functions on class 'pseries' see [lag](#), [lead](#), [diff](#) for lagging values, leading values (negative lags) and differencing.

**Examples**

```

# First, create a pdata.frame
data("EmplUK", package = "plm")
Em <- pdata.frame(EmplUK)

# Then extract a series, which becomes additionally a pseries
z <- Em$output
class(z)

# obtain the matrix representation
as.matrix(z)

# compute the between and within transformations
between(z)
Within(z)

# Between replicates the values for each time observation
Between(z)

# between, Between, and Within transformations on other dimension
between(z, effect = "time")
Between(z, effect = "time")
Within(z, effect = "time")

# NA treatment for between, Between, and Within
z2 <- z
z2[length(z2)] <- NA # set last value to NA
between(z2, na.rm = TRUE) # non-NA value for last individual
Between(z2, na.rm = TRUE) # only the NA observation is lost
Within(z2, na.rm = TRUE) # only the NA observation is lost

sum(is.na(Between(z2))) # 9 observations lost due to one NA value
sum(is.na(Between(z2, na.rm = TRUE))) # only the NA observation is lost
sum(is.na(Within(z2))) # 9 observations lost due to one NA value
sum(is.na(Within(z2, na.rm = TRUE))) # only the NA observation is lost

```

---

punbalancedness

*Measures for Unbalancedness of Panel Data*


---

**Description**

This function reports unbalancedness measures for panel data as defined in Ahrens/Pincus (1981) and Baltagi/Song/Jung (2001).

**Usage**

```

punbalancedness(x, ...)
## S3 method for class 'panelmodel'
punbalancedness(x, ...)

```

```
## S3 method for class 'data.frame'
punbalancedness(x, index = NULL, ...)
## S3 method for class 'pdata.frame'
punbalancedness(x, ...)
```

### Arguments

`x` a `panelmodel`, a `data.frame`, or a `pdata.frame` object,  
`index` only relevant for `data.frame` interface, for details see [pdata.frame](#),  
`...` further arguments.

### Details

`punbalancedness` returns measures for the unbalancedness of a panel data set.

- For two-dimensional data:  
 The two measures of Ahrens/Pincus (1981) are calculated, called "gamma" ( $\gamma$ ) and "nu" ( $\nu$ ). If the panel data are balanced, both measures equal 1. The more "unbalanced" the panel data, the lower the measures (but  $> 0$ ). The upper and lower bounds as given in Ahrens/Pincus (1981) are:  
 $0 < \gamma, \nu \leq 1$ , and for  $\nu$  more precisely  $\frac{1}{n} < \nu \leq 1$ , with  $n$  being the number of individuals (as in `pdim(x)$nT$n`).
- For nested panel data (meaning including a grouping variable):  
 The extension of the above measures by Baltagi/Song/Jung (2001), p. 368, are calculated:
  - `c1`: measure of subgroup (individual) unbalancedness,
  - `c2`: measure of time unbalancedness,
  - `c3`: measure of group unbalancedness due to each group size.

Values are 1 if the data are balanced and become smaller as the data become more unbalanced.

An application of the measure "gamma" is found in e. g. Baltagi/Song/Jung (2002), pp. 488-491, and Baltagi/Chang (1994), pp. 78-87, where it is used to measure the unbalancedness of various unbalanced data sets used for Monte Carlo simulation studies. Measures `c1`, `c2`, `c3` are used for similar purposes in Baltagi/Song/Jung (2001).

In the two-dimensional case, `punbalancedness` uses output of `pdim` to calculate the two unbalancedness measures, so inputs to `punbalancedness` can be whatever `pdim` works on. `pdim` returns detailed information about the number of individuals and time observations (see [pdim](#)).

### Value

A named numeric containing either two or three entries, depending on the panel structure inputted:

- \* For the two-dimensional panel structure, the entries are called `gamma` and `nu`,
- \* For a nested panel structure, the entries are called `c1`, `c2`, `c3`.



**Note**

Calling `punbalancedness` on an estimated `panelmodel` object and on the corresponding `(p)data.frame` used for this estimation does not necessarily yield the same result (true also for `pdim`). When called on an estimated `panelmodel`, the number of observations (individual, time) actually used for model estimation are taken into account. When called on a `(p)data.frame`, the rows in the `(p)data.frame` are considered, disregarding any NA values in the dependent or independent variable(s) which would be dropped during model estimation.

**Author(s)**

Kevin Tappe

**References**

- Ahrens, H.; Pincus, R. (1981), "On two measures of unbalancedness in a one-way model and their relation to efficiency", *Biometrical Journal*, **23**(3), pp. 227–235.
- Baltagi, Badi H.; Chang, Young-Jae (1994), "Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model", *Journal of Econometrics*, **62**(2), pp. 67–89.
- Baltagi, Badi H.; Song, Seuck Heun; Jung, Byoung Cheol (2001), "The unbalanced nested error component regression model", *Journal of Econometrics*, **101**(2), pp. 357–381.
- Baltagi, Badi H.; Song, Seuck H.; Jung, Byoung C. (2002), "A comparative study of alternative estimators for the unbalanced two-way error component regression model", *Econometrics Journal*, **5**(2), pp. 480–493.

**See Also**

[nobs](#), [pdim](#), [pdata.frame](#)

**Examples**

```
# Grunfeld is a balanced panel, Hedonic is an unbalanced panel
data(list=c("Grunfeld", "Hedonic"), package="plm")

# Grunfeld has individual and time index in first two columns
punbalancedness(Grunfeld) # c(1,1) indicates balanced panel
pdim(Grunfeld)$balanced  # TRUE

# Hedonic has individual index in column "townid" (in last column)
punbalancedness(Hedonic, index="townid") # c(0.472, 0.519)
pdim(Hedonic, index="townid")$balanced  # FALSE

# punbalancedness on estimated models
plm_mod_pool <- plm(inv ~ value + capital, data = Grunfeld)
punbalancedness(plm_mod_pool)

plm_mod_fe <- plm(inv ~ value + capital, data = Grunfeld[1:99, ], model = "within")
punbalancedness(plm_mod_fe)
```

```
# replicate results for panel data design no. 1 in Ahrens/Pincus (1981), p. 234
ind_d1 <- c(1,1,1,2,2,2,3,3,3,3,3,4,4,4,4,4,4,4,4,4,5,5,5,5,5,5,5)
time_d1 <- c(1,2,3,1,2,3,1,2,3,4,5,1,2,3,4,5,6,7,1,2,3,4,5,6,7)
df_d1 <- data.frame(individual = ind_d1, time = time_d1)
punbalancedness(df_d1) # c(0.868, 0.887)

# example for a nested panel structure with a third index variable
# specifying a group (states are grouped by region) and without grouping
data("Produc", package = "plm")
punbalancedness(Produc, index = c("state", "year", "region"))
punbalancedness(Produc, index = c("state", "year"))
```

---

purtest

*Unit root tests for panel data*

---

## Description

purtest implements several testing procedures that have been proposed to test unit root hypotheses with panel data.

## Usage

```
purtest(object, data = NULL, index = NULL,
        test = c("levinlin", "ips", "madwu", "Pm", "invnormal", "logit", "hadri"),
        exo = c("none", "intercept", "trend"),
        lags = c("SIC", "AIC", "Hall"), pmax = 10, Hcons = TRUE,
        q = NULL, dfcor = FALSE, fixedT = TRUE, ...)
## S3 method for class 'purtest'
print(x, ...)
## S3 method for class 'purtest'
summary(object, ...)
```

## Arguments

object, x	Either a "data.frame" or a matrix containing the time series, a "pseries" object, a formula, or the name of a column of a "data.frame", or a "pdata.frame" on which the test has to be computed; a "purtest" object for the print and summary methods,
data	a "data.frame" or a "pdata.frame" object,
index	the indexes,
test	the test to be computed: one of "levinlin" for Levin, Lin and Chu (2002), "ips" for Im, Pesaran and Shin (2003), "madwu" for Maddala and Wu (1999), "Pm", "invnormal", or "logit" for various tests as in Choi (2001), or "hadri" for Hadri (2000), see Details,

exo	the exogenous variables to introduce in the augmented Dickey–Fuller (ADF) regressions, one of: no exogenous variables ("none"), individual intercepts ("intercept"), or individual intercepts and trends ("trend"), but see Details,
lags	the number of lags to be used for the augmented Dickey-Fuller regressions: either an integer (the number of lags for all time series), a vector of integers (one for each time series), or a character string for an automatic computation of the number of lags, based on either the AIC ("AIC"), the SIC ("SIC"), or on the method by Hall (1994) ("Hall"),
pmax	maximum number of lags,
Hcons	logical, only relevant for test = "hadri", indicating whether the heteroskedasticity-consistent test of Hadri (2000) should be computed,
q	the bandwidth for the estimation of the long-run variance,
dfcor	logical, indicating whether the standard deviation of the regressions is to be computed using a degrees-of-freedom correction,
fixedT	logical, indicating whether the different ADF regressions are to be computed using the same number of observations,
...	further arguments.

## Details

All these tests except "hadri" are based on the estimation of augmented Dickey-Fuller (ADF) regressions for each time series. A statistic is then computed using the t-statistics associated with the lagged variable. The Hadri residual-based LM statistic is the cross-sectional average of the individual KPSS statistics (Kwiatkowski/Phillips/Schmidt/Shin (1992)), standardized by their asymptotic mean and standard deviation.

Several Fisher-type tests that combine p-values from tests based on ADF regressions per individual are available:

- "madwu" is the inverse chi-squared test (Maddala and Wu (1999)), also called P test by Choi (2001),
- "Pm" is the modified P test proposed by Choi (2001) for large N,
- "invnormal" is the inverse normal test by Choi (2001), and
- "logit" is the logit test by Choi (2001).

The individual p-values for the Fisher-type tests are approximated as described in MacKinnon (1994).

The kind of test to be computed can be specified in several ways, depending on how the data is handed over to the function:

- For the formula/data interface (if data is a data.frame, an additional index argument should be specified); the formula should be of the form:  $y \sim 0$ ,  $y \sim 1$ , or  $y \sim \text{trend}$  for a test with no exogenous variables, with an intercept, or with individual intercepts and time trend, respectively. The exo argument is ignored in this case.
- For the data.frame, matrix, and pseries interfaces: in these cases, the exogenous variables are specified using the exo argument.

With the associated summary and print methods, additional information can be extracted/displayed (see also Value).

**Value**

An object of class "purtest": a list with the elements "statistic" (a "htest" object), "call", "args", "idres" (containing results from the individual regressions), and "adjval" (containing the simulated means and variances needed to compute the statistic).

**Author(s)**

Yves Croissant

**References**

- Choi, I. (2001). "Unit root tests for panel data", *Journal of International Money and Finance*, **20**(2), pp. 249–272.
- Hadri K. (2000). "Testing for Stationarity in Heterogeneous Panel Data", *The Econometrics Journal*, **3**(2), pp. 148–161.
- Hall A. (1994). "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection", *Journal of Business & Economic Statistics*, **12**(1), pp. 461–470.
- Im K.S., Pesaran M.H. and Shin Y. (2003). "Testing for Unit Roots in Heterogeneous Panels", *Journal of Econometrics*, **115**(1), pp. 53–74.
- Kwiatkowski D., Phillips P. C. B., Schmidt P. and Shin Y. (1992). "Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", *Journal of Econometrics*, **54**(1–3), pp. 159–178.
- Levin A., Lin C.-F. and Chu C.-S.J. (2002). "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties", *Journal of Econometrics*, **108**(1), pp. 1–24.
- MacKinnon, J.G. (1994). "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests", *Journal of Business & Economic Statistics*, **12**(2), pp. 167–176.
- Maddala G.S. and Wu S. (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test", *Oxford Bulletin of Economics and Statistics*, **61**, Supplement 1, pp. 631–652.

**See Also**

[cipstest](#)

**Examples**

```
data("Grunfeld", package = "plm")
y <- data.frame(split(Grunfeld$inv, Grunfeld$firm))

purtest(y, pmax = 4, exo = "intercept", test = "madwu")

## same via formula interface
purtest(inv ~ 1, data = Grunfeld, index = c("firm", "year"), pmax = 4, test = "madwu")
```

**Description**

This function checks for each variable of a panel if it varies cross-sectionally and over time.

**Usage**

```
pvar(x, ...)
## S3 method for class 'data.frame'
pvar(x, index = NULL, ...)
## S3 method for class 'pdata.frame'
pvar(x, ...)
## S3 method for class 'matrix'
pvar(x, index = NULL, ...)
## S3 method for class 'pvar'
print(x, ...)
```

**Arguments**

x                    a (p)data.frame or a matrix,  
 index                see [pdata.frame](#),  
 ...                   further arguments.

**Details**

For (p)data.frame and matrix interface: All-NA columns are removed prior to calculation of variation due to coercing to pdata.frame first.

**Value**

An object of class pvar containing the following elements:

id.variation        a logical vector with TRUE values if the variable has individual variation, FALSE if not,  
 time.variation    a logical vector with TRUE values if the variable has time variation, FALSE if not,  
 id.variation\_anyNA    a logical vector with TRUE values if the variable has at least one individual-time combination with all NA values in the individual dimension for at least one time period, FALSE if not,  
 time.variation\_anyNA    a logical vector with TRUE values if the variable has at least one individual-time combination with all NA values in the time dimension for at least one individual, FALSE if not.

**Note**

pvar can be time consuming for “big” panels.

**Author(s)**

Yves Croissant

**See Also**

[pdim](#) to check the dimensions of a 'pdata.frame' (and other objects),

**Examples**

```
# Gasoline contains two variables which are individual and time indexes
# and are the first two variables
data("Gasoline", package = "plm")
pvar(Gasoline)

# Hedonic is an unbalanced panel, townid is the individual index;
# the drop.index argument is passed to pdata.frame
data("Hedonic", package = "plm")
pvar(Hedonic, "townid", drop.index = TRUE)

# same using pdata.frame
Hed <- pdata.frame(Hedonic, "townid", drop.index = TRUE)
pvar(Hed)

# Gasoline with pvar's matrix interface
Gasoline_mat <- as.matrix(Gasoline)
pvar(Gasoline_mat)
pvar(Gasoline_mat, index=c("country", "year"))
```

---

pvcm

*Variable Coefficients Models for Panel Data*

---

**Description**

Estimators for random and fixed effects models with variable coefficients.

**Usage**

```
pvcm(formula, data, subset, na.action, effect = c("individual", "time"),
      model = c("within", "random"), index = NULL, ...)
## S3 method for class 'pvcm'
summary(object, ...)
## S3 method for class 'summary.pvcm'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
```

**Arguments**

formula	a symbolic description for the model to be estimated,
object, x	an object of class "pvcn",
data	a data.frame,
subset	see lm,
na.action	see lm,
effect	the effects introduced in the model: one of "individual", "time",
model	one of "within", "random",
index	the indexes, see <a href="#">pdata.frame</a> ,
digits	digits,
width	the maximum length of the lines in the print output,
...	further arguments.

**Details**

pvcn estimates variable coefficients models. Time or individual effects are introduced, respectively, if `effect = "time"` or `effect = "individual"` (the default value).

Coefficients are assumed to be fixed if `model = "within"` and random if `model = "random"`. In the first case, a different model is estimated for each individual (or time period). In the second case, the Swamy (1970) model is estimated. It is a generalized least squares model which uses the results of the previous model.

**Value**

An object of class `c("pvcn", "panelmodel")`, which has the following elements:

coefficients	the vector (or the list for fixed effects) of coefficients,
residuals	the vector of residuals,
fitted.values	the vector of fitted values,
vcov	the covariance matrix of the coefficients,
df.residual	degrees of freedom of the residuals,
model	a data.frame containing the variables used for the estimation,
call	the call,
Delta	the estimation of the covariance matrix of the coefficients (random effect models only),
std.error	the standard errors for all the coefficients for each individual (within models only).

pvcn objects have `print`, `summary` and `print.summary` methods.

**Author(s)**

Yves Croissant

## References

Swamy, P.A.V.B. (1970). Efficient Inference in a Random Coefficient Regression Model, *Econometrica*, **38**(2), pp. 311–323.

## Examples

```
data("Produc", package = "plm")
zw <- pvcm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc, model = "within")
zr <- pvcm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc, model = "random")

## replicate Greene (2012), p. 419, table 11.14
summary(pvcm(log(gsp) ~ log(pc) + log(hwy) + log(water) + log(util) + log(emp) + unemp,
            data = Produc, model = "random"))
```

---

pwaldtest

*Wald-style Chi-square Test and F Test*

---

## Description

Wald-style chi-square test and F test of slope coefficients being zero jointly, including robust versions of the tests.

## Usage

```
pwaldtest(x, ...)
## S3 method for class 'plm'
pwaldtest(x, test = c("Chisq", "F"), vcov = NULL,
          df2adj = (test == "F" && !is.null(vcov) && missing(.df2)),
          .df1, .df2, ...)
```

## Arguments

<code>x</code>	an estimated model of which the coefficients should be tested (usually of class "plm"),
<code>test</code>	a character, indicating the test to be performed, may be either "Chisq" or "F" for the Wald-style Chi-square test or F test, respectively,
<code>vcov</code>	NULL by default; a matrix giving a variance–covariance matrix or a function which computes such; if supplied (non NULL), the test is carried out using the variance–covariance matrix indicated resulting in a robust test,
<code>df2adj</code>	logical, only relevant for <code>test = "F"</code> , indicating whether the adjustment for clustered standard errors for the second degrees of freedom parameter should be performed (see <b>Details</b> , also for further requirements regarding the variance–covariance matrix in <code>vcov</code> for the adjustment to be performed),
<code>.df1</code>	a numeric, used if one wants to overwrite the first degrees of freedom parameter in the performed test (usually not used),



.df2            a numeric, used if one wants to overwrite the second degrees of freedom parameter for the F test (usually not used),

...            further arguments (currently none).

## Details

pwaldtest can be used stand-alone with a plm object. It is also used in `summary.plm` to produce the F statistic.

pwaldtest performs the test if the slope coefficients of a panel regression are jointly zero. It does not perform general purpose Wald-style tests (for those, see `waldtest` (from package `lmtest`) or `linearHypothesis` (from `car`)).

If a user specified variance-covariance matrix/function is given in argument `vcov`, the robust version of the tests are carried out. In that case, if the F test is requested (`test = "F"`) and no overwriting of the second degrees of freedom parameter is given (by supplying argument (`.df2`)), the adjustment of the second degrees of freedom parameter is performed by default. The second degrees of freedom parameter is adjusted to be the number of unique elements of the cluster variable - 1, e. g. the number of individuals - 1. For the degrees of freedom adjustment of the F test in general, see e. g. Cameron/Miller (2015), section VII; Andress/Golsch/Schmidt (2013), pp. 126, footnote 4.

The degrees of freedom adjustment requires the `vcov` object supplied or created by a supplied function to carry an attribute called "cluster" with a known clustering described as a character (for now this could be either "group" or "time"). The `vcovXX` functions of the package `plm` provide such an attribute for their returned variance-covariance matrices. No adjustment is done for unknown descriptions given in the attribute "cluster" or when the attribute "cluster" is not present. Robust `vcov` objects/functions from package `clubSandwich` work as inputs to pwaldtest's F test because they are translated internally to match the needs described above.

## Value

An object of class "hctest".

## Author(s)

Yves Croissant (initial implementation) and Kevin Tappe (extensions: `vcov` argument and F test's `df2` adjustment)

## References

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Sec. 4.2.3, pp. 60–62.

Andress, H.-J./Golsch, K./Schmidt, A. (2013), *Applied Panel Data Analysis for Economic and Social Surveys*, Springer, Heidelberg et al.

Cameron, A. C./Miller, D. L. (2015), "A Practitioner's Guide to Cluster-Robust Inference", *Journal of Human Resources*, 2015, **50**(2), pp. 317–373; see also the supplements under <http://cameron.econ.ucdavis.edu/research/papers.html>.

**See Also**

[vcovHC](#) for an example of the `vcovXX` functions, a robust estimation for the variance–covariance matrix; [summary.plm](#)

**Examples**

```
data("Grunfeld", package = "plm")
mod_fe <- plm(inv ~ value + capital, data = Grunfeld, model = "within")
mod_re <- plm(inv ~ value + capital, data = Grunfeld, model = "random")
pwaldtest(mod_fe, test = "F")
pwaldtest(mod_re, test = "Chisq")

# with robust vcov (matrix, function)
pwaldtest(mod_fe, vcov = vcovHC(mod_fe))
pwaldtest(mod_fe, vcov = function(x) vcovHC(x, type = "HC3"))

pwaldtest(mod_fe, vcov = vcovHC(mod_fe), df2adj = FALSE) # w/o df2 adjustment

# example without attribute "cluster" in the vcov
vcov_mat <- vcovHC(mod_fe)
attr(vcov_mat, "cluster") <- NULL # remove attribute
pwaldtest(mod_fe, vcov = vcov_mat) # no df2 adjustment performed
```

---

pwartest

*Wooldridge Test for AR(1) Errors in FE Panel Models*


---

**Description**

Test of serial correlation for (the idiosyncratic component of) the errors in fixed–effects panel models.

**Usage**

```
pwartest(x, ...)
## S3 method for class 'panelmodel'
pwaldtest(x, ...)
## S3 method for class 'formula'
pwaldtest(x, data, ...)
```

**Arguments**

`x` an object of class `formula` or of class `panelmodel`,  
`data` a `data.frame`,  
`...` further arguments to be passed on to `vcovHC` (see [Details](#) and [Examples](#)).

## Details

As Wooldridge (2003/2010, Sec. 10.5.4) observes, under the null of no serial correlation in the errors, the residuals of a FE model must be negatively serially correlated, with  $cor(\hat{u}_{it}, \hat{u}_{is}) = -1/(T-1)$  for each  $t, s$ . He suggests basing a test for this null hypothesis on a pooled regression of FE residuals on their first lag:  $\hat{u}_{i,t} = \alpha + \delta\hat{u}_{i,t-1} + \eta_{i,t}$ . Rejecting the restriction  $\delta = -1/(T-1)$  makes us conclude against the original null of no serial correlation.

`pwartest` estimates the within model and retrieves residuals, then estimates an AR(1) pooling model on them. The test statistic is obtained by applying a F test to the latter model to test the above restriction on  $\delta$ , setting the covariance matrix to `vcovHC` with the option `method="arellano"` to control for serial correlation.

Unlike the `pbgtest` and `pdwtest`, this test does not rely on large- $T$  asymptotics and has therefore good properties in “short” panels. Furthermore, it is robust to general heteroskedasticity.

## Value

An object of class “`htest`”.

## Author(s)

Giovanni Millo

## References

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Sec. 10.5.4, pp. 274–276.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Sec. 10.5.4, pp. 310–312.

## See Also

[pwfdtest](#), [pdwtest](#), [pbgtest](#), [pbltest](#), [pbsytest](#).

## Examples

```
data("EmplUK", package = "plm")
pwartest(log(emp) ~ log(wage) + log(capital), data = EmplUK)

# pass argument 'type' to vcovHC used in test
pwartest(log(emp) ~ log(wage) + log(capital), data = EmplUK, type = "HC3")
```

---

pwfdtest	<i>Wooldridge first-difference-based test for AR(1) errors in levels or first-differenced panel models</i>
----------	--

---

### Description

First-differencing-based test of serial correlation for (the idiosyncratic component of) the errors in either levels or first-differenced panel models.

### Usage

```
pwfdtest(x, ...)
## S3 method for class 'panelmodel'
pwfdtest(x, ..., h0 = c("fd", "fe"))
## S3 method for class 'formula'
pwfdtest(x, data, ..., h0 = c("fd", "fe"))
```

### Arguments

x	an object of class formula or a "fd"-model (plm object),
data	a data.frame,
h0	the null hypothesis: one of "fd", "fe",
...	further arguments to be passed on to vcovHC (see Details and Examples).

### Details

As Wooldridge (2003/2010, Sec. 10.6.3) observes, if the idiosyncratic errors in the model in levels are uncorrelated (which we label hypothesis "fe"), then the errors of the model in first differences (FD) must be serially correlated with  $cor(\hat{e}_{it}, \hat{e}_{is}) = -0.5$  for each  $t, s$ . If on the contrary the levels model's errors are a random walk, then there must be no serial correlation in the FD errors (hypothesis "fd"). Both the fixed effects (FE) and the first-differenced (FD) estimators remain consistent under either assumption, but the relative efficiency changes: FE is more efficient under "fe", FD under "fd".

Wooldridge (ibid.) suggests basing a test for either hypothesis on a pooled regression of FD residuals on their first lag:  $\hat{e}_{i,t} = \alpha + \rho \hat{e}_{i,t-1} + \eta_{i,t}$ . Rejecting the restriction  $\rho = -0.5$  makes us conclude against the null of no serial correlation in errors of the levels equation ("fe"). The null hypothesis of no serial correlation in differenced errors ("fd") is tested in a similar way, but based on the zero restriction on  $\rho$  ( $\rho = 0$ ). Rejecting "fe" favours the use of the first-differences estimator and the contrary, although it is possible that both be rejected.

pwfdtest estimates the fd model (or takes an fd model as input for the panelmodel interface) and retrieves its residuals, then estimates an AR(1) pooling model on them. The test statistic is obtained by applying a F test to the latter model to test the relevant restriction on  $\rho$ , setting the covariance matrix to vcovHC with the option method="arellano" to control for serial correlation.

Unlike the pbgttest and pdwtest, this test does not rely on large-T asymptotics and has therefore good properties in "short" panels. Furthermore, it is robust to general heteroskedasticity. The

"fe" version can be used to test for error autocorrelation regardless of whether the maintained specification has fixed or random effects (see Drukker (2003)).

### Value

An object of class "htest".

### Author(s)

Giovanni Millo

### References

Drukker, D.M. (2003) Testing for serial correlation in linear panel–data models, *The Stata Journal*, **3(2)**, pp. 168–177.

Wooldridge, J.M. (2003) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Sec. 10.6.3, pp. 282–283.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Sec. 10.6.3, pp. 319–320.

### See Also

pdwtest, pbgtest, pwartest,

### Examples

```
data("EmplUK" , package = "plm")
pwfdtest(log(emp) ~ log(wage) + log(capital), data = EmplUK)
pwfdtest(log(emp) ~ log(wage) + log(capital), data = EmplUK, h0 = "fe")

# pass argument 'type' to vcovHC used in test
pwfdtest(log(emp) ~ log(wage) + log(capital), data = EmplUK, type = "HC3", h0 = "fe")

# same with panelmodel interface
mod <- plm(log(emp) ~ log(wage) + log(capital), data = EmplUK, model = "fd")
pwfdtest(mod)
pwfdtest(mod, h0 = "fe")
pwfdtest(mod, type = "HC3", h0 = "fe")
```

---

pwtest

*Wooldridge's Test for Unobserved Effects in Panel Models*

---

### Description

Semi-parametric test for the presence of (individual or time) unobserved effects in panel models.

**Usage**

```
pwtest(x, ...)
## S3 method for class 'formula'
pwtest(x, data, effect = c("individual", "time"), ...)
## S3 method for class 'panelmodel'
pwtest(x, effect = c("individual", "time"), ...)
```

**Arguments**

x	an object of class "formula", or an estimated model of class panelmodel,
effect	the effect to be tested for, one of "individual" (default) or "time",
data	a data.frame,
...	further arguments passed to plm.

**Details**

This semi-parametric test checks the null hypothesis of zero correlation between errors of the same group. Therefore, it has power both against individual effects and, more generally, any kind of serial correlation.

The test relies on large-N asymptotics. It is valid under error heteroskedasticity and departures from normality.

The above is valid if effect="individual", which is the most likely usage. If effect="time", symmetrically, the test relies on large-T asymptotics and has power against time effects and, more generally, against cross-sectional correlation.

If the panelmodel interface is used, the inputted model must be a pooling model.

**Value**

An object of class "htest".

**Author(s)**

Giovanni Millo

**References**

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Sec. 10.4.4., pp. 264–265.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Sec. 10.4.4, pp. 299–300.

**See Also**

[pbltest](#), [pbgtest](#), [pdwtest](#), [pbsytest](#), [pwartest](#), [pwfdtest](#) for tests for serial correlation in panel models. [plmtest](#) for tests for random effects.

**Examples**

```

data("Produc", package = "plm")
## formula interface
pwtest(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc)
pwtest(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp, data = Produc, effect = "time")

## panelmodel interface
# first, estimate a pooling model, than compute test statistics
form <- formula(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp)
pool_prodc <- plm(form, data = Produc, model = "pooling")
pwtest(pool_prodc) # == effect="individual"
pwtest(pool_prodc, effect="time")

```

r.squared

*R squared and adjusted R squared for panel models***Description**

This function computes R squared or adjusted R squared for plm objects. It allows to define on which transformation of the data the (adjusted) R squared is to be computed and which method for calculation is used.

**Usage**

```
r.squared(object, model = NULL, type = c("cor", "rss", "ess"), dfcor = FALSE)
```

**Arguments**

object	an object of class "plm",
model	on which transformation of the data the R-squared is to be computed. If NULL, the transformation used to estimate the model is also used for the computation of R squared,
type	indicates method which is used to compute R squared. One of "rss" (residual sum of squares), "ess" (explained sum of squares), or "cor" (coefficient of correlation between the fitted values and the response),
dfcor	if TRUE, the adjusted R squared is computed.

**Value**

A numerical value. The R squared or adjusted R squared of the model estimated on the transformed data, e. g. for the within model the so called "within R squared".

**See Also**

[plm](#) for estimation of various models; [summary.plm](#) which makes use of r.squared.

**Examples**

```
data("Grunfeld", package = "plm")
p <- plm(inv ~ value + capital, data = Grunfeld, model = "pooling")
r.squared(p)
r.squared(p, dfcor = TRUE)
```

---

ranef.plm

*Extract the Random Effects*


---

**Description**

Function to calculate the random effects from a plm object (random effects model).

**Usage**

```
## S3 method for class 'plm'
ranef(object, effect = NULL, ...)
```

**Arguments**

object	an object of class "plm", needs to be a fitted random effects model,
effect	NULL, "individual", or "time", the effects to be extracted, see <b>Details</b> ,
...	further arguments (currently not used).

**Details**

Function ranef calculates the random effects of a fitted random effects model. For one-way models, the effects of the estimated model are extracted (either individual or time effects). For two-way models, extracting the individual effects is the default (both, argument effect = NULL and effect = "individual" will give individual effects). Time effects can be extracted by setting effect = "time".

Not all random effect model types are supported (yet?).

**Value**

A named numeric with the random effects per dimension (individual or time).

**Author(s)**

Kevin Tappe

**See Also**

[fixef](#) to extract the fixed effects from a fixed effects model (within model).



**Examples**

```

data("Grunfeld", package = "plm")
m1 <- plm(inv ~ value + capital, data = Grunfeld, model = "random")
ranef(m1) # individual random effects

# compare to random effects by ML estimation via lmer from package lme4
## Not run:
library(lme4)
m2 <- lmer(inv ~ value + capital + (1 | firm), data = Grunfeld)
cbind("plm" = ranef(m1), "lmer" = unname(ranef(m2)$firm))

## End(Not run)

# two-ways RE model, calculate individual and time random effects
data("Cigar", package = "plm")
tw <- plm(sales ~ pop + price, data = Cigar, model = "random", effect = "twoways")
ranef(tw) # individual random effects
ranef(tw, effect = "time") # time random effects

```

---

RiceFarms

*Production of Rice in Indonesia*


---

**Description**

a panel of 171 observations  
*number of observations* : 1026  
*observation* : farms  
*country* : Indonesia

**Usage**

```
data(RiceFarms)
```

**Format**

A dataframe containing :

**id** the farm identifier

**size** the total area cultivated with rice, measured in hectares

**status** land status, on of 'owner' (non sharecroppers, owner operators or leaseholders or both), 'share' (sharecroppers), 'mixed' (mixed of the two previous status)

**varieties** one of 'trad' (traditional varieties), 'high' (high yielding varieties) and 'mixed' (mixed varieties)

**bimas** bIMAS is an intensification program; one of 'no' (non-bimas farmer), 'yes' (bimas farmer) or 'mixed' (part but not all of farmer's land was registered to be in the bimas program)

**seed** seed in kilogram  
**urea** urea in kilogram  
**phosphate** phosphate in kilogram  
**pesticide** pesticide cost in Rupiah  
**pseed** price of seed in Rupiah per kg  
**purea** price of urea in Rupiah per kg  
**pphosph** price of phosphate in Rupiah per kg  
**hiredlabor** hired labor in hours  
**famlabor** family labor in hours  
**totlabor** total labor (excluding harvest labor)  
**wage** labor wage in Rupiah per hour  
**goutput** gross output of rice in kg  
**noutput** net output, gross output minus harvesting cost (paid in terms of rice)  
**price** price of rough rice in Rupiah per kg  
**region** one of 'wargabinangun', 'langan', 'gunungwangi', 'malausma', 'sukaambit', 'ciwangi'

### Source

Qu Feng and William C. Horrace, (2012) “Alternative Measures of Technical Efficiency: Skew, Bias and Scale”, *Journal of Applied Econometrics*, **27(2)**, pp. 253–268.

---

sargan

*Hansen–Sargan Test of Overidentifying Restrictions*

---

### Description

A test of overidentifying restrictions for models estimated by GMM.

### Usage

```
sargan(object, weights = c("twosteps", "onestep"))
```

### Arguments

**object** an object of class "pgmm",  
**weights** the weighting matrix to be used for the computation of the test.

### Details

The Hansen–Sargan test calculates the quadratic form of the moment restrictions that is minimized while computing the GMM estimator. It follows asymptotically a chi-square distribution with number of degrees of freedom equal to the difference between the number of moment conditions and the number of coefficients.

**Value**

An object of class "htest".

**Author(s)**

Yves Croissant

**References**

Hansen, L.P. (1982), Large Sample Properties of Generalized Methods of Moments Estimators, *Econometrica*, **50**(4), pp. 1029–1054.

Sargan, J.D. (1958), The Estimation of Economic Relationships using Instrumental Variables, *Econometrica*, **26**(3), pp. 393–415.

**See Also**

[pgmm](#)

**Examples**

```
data("EmplUK", package = "plm")
ar <- pgmm(log(emp) ~ lag(log(emp), 1:2) + lag(log(wage), 0:1) +
           lag(log(capital), 0:2) + lag(log(output), 0:2) | lag(log(emp), 2:99),
           data = EmplUK, effect = "twoways", model = "twosteps")
sargan(ar)
```

---

Snmesp

*Employment and Wages in Spain*

---

**Description**

A panel of 738 observations from 1983 to 1990

*total number of observations:* 5904

*observation:* firms

*country:* Spain

**Usage**

```
data(Snmesp)
```

**Format**

A data frame containing:

**firm** firm index

**year** year

**n** log of employment

**w** log of wages

**y** log of real output

**i** log of intermediate inputs

**k** log of real capital stock

**f** real cash flow

**Source**

Journal of Business Economics and Statistics data archive:

<http://amstat.tandfonline.com/loi/ubes20/>.

**References**

Alonso-Borrego, C. and Arellano, M. (1999). Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data, *Journal of Business and Economic Statistics*, **17(1)**, pp. 36–49.

---

SumHes

*The Penn World Table, v. 5*

---

**Description**

A panel of 125 observations from 1960 to 1985

*total number of observations* : 3250

*observation* : country

*country* : World

**Usage**

data(SumHes)

**Format**

A data frame containing :

**year** the year

**country** the country name (factor)

**opec** OPEC member?

**com** communist regime?

**pop** country's population (in thousands)

**gdp** real GDP per capita (in 1985 US dollars)

**sr** saving rate (in percent)

**Source**

Online supplements to Hayashi (2000).

<http://fhayashi.fc2web.com/datasets.htm>

**References**

Hayashi, F. (2000) *Econometrics*, Princeton University Press, chapter 5, pp. 358–363.

Summers, R. and A. Heston (1991) “The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988”, *Quarterly Journal of Economics*, **106**(2), pp. 327-368.

---

summary.plm

*Summary for plm objects*

---

**Description**

The summary method for plm objects generates some more information about estimated plm models.

**Usage**

```
## S3 method for class 'plm'
summary(object, vcov = NULL, ..., .vcov = NULL)
## S3 method for class 'summary.plm'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), subset = NULL, ...)
```

**Arguments**

object	an object of class "plm",
x	an object of class "summary.plm",
subset	a character or numeric vector indicating a subset of the table of coefficients to be printed for "print.summary.plm",
vcov	a variance–covariance matrix furnished by the user or a function to calculate one (see <b>Examples</b> ),
.vcov	deprecated, use argument vcov,
digits	number of digits for printed output,
width	the maximum length of the lines in the printed output,
...	further arguments.

**Details**

The summary method for plm objects (summary.plm) creates an object of class c("summary.plm", "plm", "panelmodel") that extends the plm object it is run on with various information about the estimated model like (inferential) statistics, see **Value**. It has an associated print method (print.summary.plm).

**Value**

An object of class c("summary.plm", "plm", "panelmodel"). Some of its elements are carried over from the associated plm object and described there ([plm](#)). The following elements are new or changed relative to the elements of a plm object:

fstatistic	F statistic for joint test of significance of coefficients (class "htest") (robust F statistic in case of supplied argument vcov, see <a href="#">pwaldtest</a> for details),
coefficients	a matrix with the estimated coefficients, standard errors, t–values, and p–values, if argument vcov was set to non-NULL the standard errors (and t– and p–values) in their respective robust variant,
vcov	the "regular" variance–covariance matrix of the coefficients (class "matrix"),
rvcov	only present if argument vcov was set to non-NULL: the furnished variance–covariance matrix of the coefficients (class "matrix"),
r.squared	a named numeric containing the R-squared ("rsq") and the adjusted R-squared ("adjrsq") of the model,
df	an integer vector with 3 components, (p, n-p, p*), where p is the number of estimated (non-aliased) coefficients of the model, n-p are the residual degrees of freedom (n being number of observations), and p* is the total number of coefficients (incl. any aliased ones).

**Author(s)**

Yves Croissant

**See Also**

[plm](#) for estimation of various models; [vcovHC](#) for an example of a robust estimation of variance-covariance matrix; [r.squared](#) for the function to calculate R-squared; [print.htest](#) for some information about class "htest"; [fixef](#) to compute the fixed effects for "within" (=fixed effects) models and [within\\_intercept](#) for an "overall intercept" for such models; [pwaldtest](#)

**Examples**

```
data("Produc", package = "plm")
zz <- plm(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,
          data = Produc, index = c("state","year"))
summary(zz)

# summary with a finished vcov, passed as matrix, as function, and
# as function with additional argument
data("Grunfeld", package = "plm")
wi <- plm(inv ~ value + capital,
          data = Grunfeld, model="within", effect = "individual")
summary(wi, vcov = vcovHC(wi))
summary(wi, vcov = vcovHC)
summary(wi, vcov = function(x) vcovHC(x, method = "white2"))

# extract F statistic
wi_summary <- summary(wi)
Fstat <- wi_summary[["fstatistic"]]

# extract estimates and p-values
est <- wi_summary[["coefficients"]][ , "Estimate"]
pval <- wi_summary[["coefficients"]][ , "Pr(>|t|)"]

# print summary only for coefficient "value"
print(wi_summary, subset = "value")
```

---

vcovBK

*Beck and Katz Robust Covariance Matrix Estimators*


---

**Description**

Unconditional Robust covariance matrix estimators *a la Beck and Katz* for panel models (a.k.a. Panel Corrected Standard Errors (PCSE)).

**Usage**

```
## S3 method for class 'plm'
vcovBK(x, type = c("HC0", "HC1", "HC2", "HC3", "HC4"),
       cluster = c("group", "time"),
       diagonal = FALSE,
       ...)
```

**Arguments**

<code>x</code>	an object of class "plm"
<code>type</code>	the weighting scheme used, one of "HC0", "HC1", "HC2", "HC3", "HC4", see Details,
<code>cluster</code>	one of "group", "time"
<code>diagonal</code>	a logical value specifying whether to force nondiagonal elements to zero
<code>...</code>	further arguments.

**Details**

vcovBK is a function for estimating a robust covariance matrix of parameters for a panel model according to the Beck and Katz (1995) method, a.k.a. Panel Corrected Standard Errors (PCSE), which uses an unconditional estimate of the error covariance across time periods (groups) inside the standard formula for coefficient covariance. Observations may be clustered either by "group" to account for timewise heteroskedasticity and serial correlation or by "time" to account for cross-sectional heteroskedasticity and correlation. It must be borne in mind that the Beck and Katz formula is based on N- (T-) asymptotics and will not be appropriate elsewhere.

The `diagonal` logical argument can be used, if set to TRUE, to force to zero all nondiagonal elements in the estimated error covariances; this is appropriate if both serial and cross-sectional correlation are assumed out, and yields a timewise- (groupwise-) heteroskedasticity-consistent estimator.

Weighting schemes specified by `type` are analogous to those in `vcovHC` in package **sandwich** and are justified theoretically (although in the context of the standard linear model) by MacKinnon and White (1985) and Cribari-Neto (2004) (see Zeileis (2004)).

The main use of `vcovBK` is to be an argument to other functions, e.g. for Wald-type testing: argument `vcov.` to `coefTest()`, argument `vcov` to `waldtest()` and other methods in the **lmtest** package; and argument `vcov.` to `linearHypothesis()` in the **car** package (see the examples). Notice that the `vcov` and `vcov.` arguments allow to supply a function (which is the safest) or a matrix (see Zeileis (2004), 4.1-2 and examples below).

**Value**

An object of class "matrix" containing the estimate of the covariance matrix of coefficients.

**Author(s)**

Giovanni Millo

**References**

- Beck, N. and Katz, J. (1995) What to do (and not to do) with time-series cross-section data in comparative politics. *American Political Science Review*, **89**(3), pp. 634–647.
- Cribari-Neto, F. (2004) Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* **45**(2), pp. 215–233.
- Greene, W. H. (2003) *Econometric Analysis*, 5th ed., Prentice Hall/Pearson, Upper Saddle River, New Jersey, Sec. 13.9.3, 13.9.7.



MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* **29**(3), pp. 305–325.

Zeileis, A. (2004) Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, **11**(10), pp. 1–17. URL <http://www.jstatsoft.org/v11/i10/>.

### See Also

`vcovHC` from the **sandwich** package for weighting schemes (type argument).

### Examples

```
library(lmtest)
library(car)
data("Produc", package="plm")
zz <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp, data=Produc, model="random")
## standard coefficient significance test
coeftest(zz)
## robust significance test, cluster by group
## (robust vs. serial correlation), default arguments
coeftest(zz, vcov.=vcovBK)
## idem with parameters, pass vcov as a function argument
coeftest(zz, vcov.=function(x) vcovBK(x, type="HC1"))
## idem, cluster by time period
## (robust vs. cross-sectional correlation)
coeftest(zz, vcov.=function(x) vcovBK(x,
  type="HC1", cluster="time"))
## idem with parameters, pass vcov as a matrix argument
coeftest(zz, vcov.=vcovBK(zz, type="HC1"))
## joint restriction test
waldtest(zz, update(zz, .~-log(emp)-unemp), vcov=vcovBK)
## test of hyp.: 2*log(pc)=log(emp)
linearHypothesis(zz, "2*log(pc)=log(emp)", vcov.=vcovBK)
```

---

vcovDC

---

*Double-Clustering Robust Covariance Matrix Estimator*


---

### Description

High-level convenience wrapper for double-clustering robust covariance matrix estimators *a la* Thompson (2011) and Cameron, Gelbach and Miller (2011) for panel models.

### Usage

```
## S3 method for class 'plm'
vcovDC(x, type = c("HC0", "sss", "HC1", "HC2", "HC3", "HC4"),
  ...)
```

**Arguments**

x	an object of class "plm" or "pcce"
type	the weighting scheme used, one of "HC0", "sss", "HC1", "HC2", "HC3", "HC4", see Details,
...	further arguments.

**Details**

vcovDC is a function for estimating a robust covariance matrix of parameters for a panel model with errors clustering along both dimensions. The function is a convenience wrapper simply summing a group- and a time-clustered covariance matrix and subtracting a diagonal one *a la* White.

Weighting schemes specified by type are analogous to those in `vcovHC` in package `sandwich` and are justified theoretically (although in the context of the standard linear model) by MacKinnon and White (1985) and Cribari-Neto (2004) (see Zeileis (2004)).

The main use of `vcovDC` is to be an argument to other functions, e.g. for Wald-type testing: argument `vcov.` to `coefest()`, argument `vcov` to `waldtest()` and other methods in the `lmtest` package; and argument `vcov.` to `linearHypothesis()` in the `car` package (see the examples). Notice that the `vcov` and `vcov.` arguments allow to supply a function (which is the safest) or a matrix (see Zeileis (2004), 4.1-2 and examples below).

**Value**

An object of class "matrix" containing the estimate of the covariance matrix of coefficients.

**Author(s)**

Giovanni Millo

**References**

- Cameron, A.C., Gelbach, J.B., & Miller, D.L. (2011) Robust inference with multiway clustering, *Journal of Business and Economic Statistics* **29(2)**, pp. 238–249.
- Cribari-Neto, F. (2004) Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* **45(2)**, pp. 215–233.
- MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* **29(3)**, pp. 305–325.
- Thompson, S.B. (2011) Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* **99(1)**, pp. 1–10.
- Zeileis, A. (2004) Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, **11(10)**, pp. 1–17. URL <http://www.jstatsoft.org/v11/i10/>.

**See Also**

`vcovHC` from the `sandwich` package for weighting schemes (type argument).

**Examples**

```

library(lmtest)
library(car)
data("Produc", package="plm")
zz <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp, data=Produc, model="pooling")
## standard coefficient significance test
coeftest(zz)
## DC robust significance test, default
coeftest(zz, vcov.=vcovDC)
## idem with parameters, pass vcov as a function argument
coeftest(zz, vcov.=function(x) vcovDC(x, type="HC1", maxlag=4))
## joint restriction test
waldtest(zz, update(zz, .~-log(emp)-unemp), vcov=vcovDC)
## test of hyp.: 2*log(pc)=log(emp)
linearHypothesis(zz, "2*log(pc)=log(emp)", vcov.=vcovDC)

```

vcovG

*Generic Lego building block for Robust Covariance Matrix Estimators***Description**

Generic Lego building block for robust covariance matrix estimators of the vcovXX kind for panel models.

**Usage**

```

## S3 method for class 'plm'
vcovG(x, type = c("HC0", "sss", "HC1", "HC2", "HC3", "HC4"),
      cluster=c("group", "time"),
      l=0,
      inner=c("cluster", "white", "diagavg"),
      ...)

```

**Arguments**

x	an object of class "plm" or "pcce"
type	the weighting scheme used, one of "HC0", "sss", "HC1", "HC2", "HC3", "HC4",
cluster	one of "group", "time",
l	lagging order, defaulting to zero
inner	the function to be applied to the residuals inside the sandwich: one of "cluster" or "white" or "diagavg",
...	further arguments.

**Details**

vcovG is the generic building block for use by higher-level wrappers [vcovHC](#), [vcovSCC](#), [vcovDC](#), and [vcovNW](#). The main use of vcovG is to be used internally by the former, but it is made available in the user space for use in non-standard combinations. For more documentation, see wrapper functions mentioned.

**Value**

An object of class "matrix" containing the estimate of the covariance matrix of coefficients.

**Author(s)**

Giovanni Millo

**References**

Millo, G. (2014) Robust standard error estimators for panel models: a unifying approach *Unpublished manuscript*.

**See Also**

[vcovHC](#), [vcovSCC](#), [vcovDC](#), [vcovNW](#), and [vcovBK](#) albeit the latter does not make use of vcovG.

**Examples**

```
data("Produc", package="plm")
zz <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp, data=Produc,
model="pooling")
## reproduce Arellano's covariance matrix
vcovG(zz, cluster="group", inner="cluster", l=0)
## use in coefficient significance test
library(lmtest)
## define custom covariance function
## (in this example, same as vcovHC)
myvcov <- function(x) vcovG(x, cluster="group", inner="cluster", l=0)
## robust significance test
coeftest(zz, vcov.=myvcov)
```

---

vcovHC

*Robust Covariance Matrix Estimators*


---

**Description**

Robust covariance matrix estimators *a la White* for panel models.

**Usage**

```
## S3 method for class 'plm'
vcovHC(x, method = c("arellano", "white1", "white2"),
        type = c("HC0", "sss", "HC1", "HC2", "HC3", "HC4"),
        cluster = c("group", "time"), ...)

## S3 method for class 'pgmm'
vcovHC(x, ...)
```

**Arguments**

x	an object of class "plm" which should be the result of a random effects or a within model or a model of class "pgmm" or an object of class "pcce",
method	one of "arellano", "white1", "white2",
type	the weighting scheme used, one of "HC0", "sss", "HC1", "HC2", "HC3", "HC4", see Details,
cluster	one of "group", "time"
...	further arguments.

**Details**

vcovHC is a function for estimating a robust covariance matrix of parameters for a fixed effects or random effects panel model according to the White method (White 1980, 1984; Arellano 1987). Observations may be clustered by "group" ("time") to account for serial (cross-sectional) correlation.

All types assume no intragroup (serial) correlation between errors and allow for heteroskedasticity across groups (time periods). As for the error covariance matrix of every single group of observations, "white1" allows for general heteroskedasticity but no serial (cross-sectional) correlation; "white2" is "white1" restricted to a common variance inside every group (time period) (see Greene (2003, Sec. 13.7.1-2; 2012, Sec. 11.6.1-2) and Wooldridge (2002), Sec. 10.7.2); "arellano" (see *ibid.* and the original ref. Arellano (1987)) allows a fully general structure w.r.t. heteroskedasticity and serial (cross-sectional) correlation.

Weighting schemes specified by type are analogous to those in `vcovHC` in package **sandwich** and are justified theoretically (although in the context of the standard linear model) by MacKinnon and White (1985) and Cribari-Neto (2004) (see Zeileis (2004)).

The main use of `vcovHC` is to be an argument to other functions, e.g. for Wald-type testing: argument `vcov.` to `coefTest()`, argument `vcov` to `waldtest()` and other methods in the **lmtest** package; and argument `vcov.` to `linearHypothesis()` in the **car** package (see the examples). Notice that the `vcov` and `vcov.` arguments allow to supply a function (which is the safest) or a matrix (see Zeileis (2004), 4.1-2 and examples below).

A special procedure for `pgmm` objects, proposed by Windmeijer (2005), is also provided.

**Value**

An object of class "matrix" containing the estimate of the asymptotic covariance matrix of coefficients.



```

## idem with parameters, pass vcov as a matrix argument
coefstest(zz, vcov.=vcovHC(zz, method="arellano", type="HC1"))
## joint restriction test
waldtest(zz, update(zz, .~-log(emp)-unemp), vcov=vcovHC)
## test of hyp.: 2*log(pc)=log(emp)
linearHypothesis(zz, "2*log(pc)=log(emp)", vcov.=vcovHC)

## Robust inference for GMM models
data("EmplUK", package="plm")
ar <- pgmm(dynformula(log(emp) ~ log(wage) + log(capital) + log(output),
  list(2, 1, 2, 2)), data = EmplUK, effect = "twoways",
  model = "twosteps", gmm.inst = ~ log(emp),
  lag.gmm = list(c(2, 99)))
rv <- vcovHC(ar)
mtest(ar, order = 2, vcov = rv)

```

---

vcovNW

*Newey and West (1987) Robust Covariance Matrix Estimator*


---

## Description

Nonparametric robust covariance matrix estimators *a la Newey and West* for panel models with serial correlation.

## Usage

```

## S3 method for class 'plm'
vcovNW(x, type = c("HC0", "sss", "HC1", "HC2", "HC3", "HC4"),
  maxlag=NULL,
  wj=function(j, maxlag) 1-j/(maxlag+1),
  ...)

```

## Arguments

x	an object of class "plm" or "pcce"
type	the weighting scheme used, one of "HC0", "sss", "HC1", "HC2", "HC3", "HC4", see Details,
maxlag	either NULL or a positive integer specifying the maximum lag order before truncation
wj	weighting function to be applied to lagged terms,
...	further arguments.

## Details

vcovNW is a function for estimating a robust covariance matrix of parameters for a panel model according to the Newey and West (1987) method. The function works as a restriction of the Driscoll and Kraay (1998) covariance (see `vcovSCC`) to no cross-sectional correlation.

Weighting schemes specified by `type` are analogous to those in `vcovHC` in package `sandwich` and are justified theoretically (although in the context of the standard linear model) by MacKinnon and White (1985) and Cribari-Neto (2004) (see Zeileis (2004)).

The main use of `vcovNW` is to be an argument to other functions, e.g. for Wald-type testing: argument `vcov.` to `coefTest()`, argument `vcov` to `waldtest()` and other methods in the `lmtest` package; and argument `vcov.` to `linearHypothesis()` in the `car` package (see the examples). Notice that the `vcov` and `vcov.` arguments allow to supply a function (which is the safest) or a matrix (see Zeileis (2004), 4.1-2 and examples below).

## Value

An object of class "matrix" containing the estimate of the covariance matrix of coefficients.

## Author(s)

Giovanni Millo

## References

Cribari-Neto, F. (2004) Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* **45**(2), pp. 215–233.

Driscoll, J.C. and Kraay, A.C. (1998) Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* **80**(4), pp. 549–560.

MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* **29**(3), pp. 305–325.

Newey, W.K. & West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3), pp. 703–708.

Zeileis, A. (2004) Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, **11**(10), pp. 1–17. URL <http://www.jstatsoft.org/v11/i10/>.

## See Also

`vcovHC` from the `sandwich` package for weighting schemes (`type` argument).

## Examples

```
library(lmtest)
library(car)
data("Produc", package="plm")
zz <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp, data=Produc, model="pooling")
## standard coefficient significance test
coefTest(zz)
## NW robust significance test, default
coefTest(zz, vcov.=vcovNW)
```



```
## idem with parameters, pass vcov as a function argument
coefest(zz, vcov.=function(x) vcovNW(x, type="HC1", maxlag=4))
## joint restriction test
waldtest(zz, update(zz, .~-log(emp)-unemp), vcov=vcovNW)
## test of hyp.: 2*log(pc)=log(emp)
linearHypothesis(zz, "2*log(pc)=log(emp)", vcov.=vcovNW)
```

vcovSCC

*Driscoll and Kraay (1998) Robust Covariance Matrix Estimator***Description**

Nonparametric robust covariance matrix estimators *a la Driscoll and Kraay* for panel models with cross-sectional *and* serial correlation.

**Usage**

```
## S3 method for class 'plm'
vcovSCC(x, type = c("HC0", "sss", "HC1", "HC2", "HC3", "HC4"),
        cluster="time",
        maxlag=NULL,
        inner=c("cluster", "white", "diagavg"),
        wj=function(j, maxlag) 1-j/(maxlag+1),
        ...)
```

**Arguments**

x	an object of class "plm" or "pcc"
type	the weighting scheme used, one of "HC0", "sss", "HC1", "HC2", "HC3", "HC4", see Details,
cluster	switch for vcovG; set at "time" here,
maxlag	either NULL or a positive integer specifying the maximum lag order before truncation
inner	the function to be applied to the residuals inside the sandwich: "cluster" for SCC, "white" for Newey-West, ("diagavg" for compatibility reasons)
wj	weighting function to be applied to lagged terms,
...	further arguments.

**Details**

vcovSCC is a function for estimating a robust covariance matrix of parameters for a panel model according to the Driscoll and Kraay (1998) method, which is consistent with cross-sectional and serial correlation in a T-asymptotic setting and irrespective of the N dimension. The use with random effects models is undocumented.

Weighting schemes specified by `type` are analogous to those in `vcovHC` in package `sandwich` and are justified theoretically (although in the context of the standard linear model) by MacKinnon and White (1985) and Cribari-Neto (2004) (see Zeileis (2004)).

The main use of `vcovSCC` is to be an argument to other functions, e.g. for Wald-type testing: argument `vcov.` to `coefstest()`, argument `vcov` to `waldtest()` and other methods in the `lmtest` package; and argument `vcov.` to `linearHypothesis()` in the `car` package (see the examples). Notice that the `vcov` and `vcov.` arguments allow to supply a function (which is the safest) or a matrix (see Zeileis (2004), 4.1-2 and examples below).

### Value

An object of class "matrix" containing the estimate of the covariance matrix of coefficients.

### Author(s)

Giovanni Millo, partially ported from Daniel Hoehle's (2007) Stata code

### References

- Cribari-Neto, F. (2004) Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis* **45(2)**, pp. 215–233.
- Driscoll, J.C. and Kraay, A.C. (1998) Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* **80(4)**, pp. 549–560.
- Hoehle, D. (2007) Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal*, **7(3)**, pp. 281–312.
- MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* **29(3)**, pp. 305–325.
- Zeileis, A. (2004) Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, **11(10)**, pp. 1–17. URL <http://www.jstatsoft.org/v11/i10/>.

### See Also

`vcovHC` from the `sandwich` package for weighting schemes (`type` argument).

### Examples

```
library(lmtest)
library(car)
data("Produc", package="plm")
zz <- plm(log(gsp)~log(pcap)+log(pc)+log(emp)+unemp, data=Produc, model="pooling")
## standard coefficient significance test
coefstest(zz)
## SCC robust significance test, default
coefstest(zz, vcov.=vcovSCC)
## idem with parameters, pass vcov as a function argument
coefstest(zz, vcov.=function(x) vcovSCC(x, type="HC1", maxlag=4))
## joint restriction test
waldtest(zz, update(zz, .~-log(emp)-unemp), vcov=vcovSCC)
## test of hyp.: 2*log(pc)=log(emp)
linearHypothesis(zz, "2*log(pc)=log(emp)", vcov.=vcovSCC)
```

---

Wages

*Panel Data of Individual Wages*

---

### Description

A panel of 595 individuals from 1976 to 1982, taken from the Panel Study of Income Dynamics (PSID).

The data are organized as a stacked time series/balanced panel, see **Examples** on how to convert to a `pdata.frame`.

*total number of observations* : 4165

*observation* : individuals

*country* : United States

### Usage

```
data(Wages)
```

### Format

A data frame containing:

**exp** years of full-time work experience.

**wks** weeks worked.

**bluecol** blue collar?

**ind** works in a manufacturing industry?

**south** resides in the south?

**smsa** resides in a standard metropolitan statistical area?

**married** married?

**sex** a factor with levels "male" and "female"

**union** individual's wage set by a union contract?

**ed** years of education.

**black** is the individual black?

**lwage** logarithm of wage.

### Source

Online complements to Baltagi (2001):

<http://www.wiley.com/legacy/wileychi/baltagi/>

Online complements to Baltagi (2013):

<http://bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4338&itemId=1118672321&resourceId=13452>

## References

- Baltagi, Badi H. (2001) *Econometric Analysis of Panel Data*, 2nd ed., John Wiley and Sons.
- Baltagi, Badi H. (2013) *Econometric Analysis of Panel Data*, 5th ed., John Wiley and Sons.
- Cornwell, C. and P. Rupert (1988) “Efficient estimation with panel data: an empirical comparison of instrumental variables estimators”, *Journal of Applied Econometrics*, **3**(2), pp. 149–155.

## Examples

```
# data set 'Wages' is organized as a stacked time series/balanced panel
data("Wages", package = "plm")
Wag <- pdata.frame(Wages, index=595)
```

---

within_intercept	<i>Overall Intercept for Within Models Along its Standard Error</i>
------------------	---

---

## Description

This function gives an overall intercept for within models and its accompanying standard error

## Usage

```
within_intercept(object, ...)
## S3 method for class 'plm'
within_intercept(object, vcov = NULL, ...)
```

## Arguments

object	object of class plm which must be a within model (fixed effects model),
vcov	if not NULL (default), a function to calculate a user defined variance–covariance matrix (function for robust vcov),
...	further arguments (currently none).

## Details

The (somewhat artificial) intercept for within models (fixed effects models) was made popular by Stata of StataCorp (see Gould (2013)), EViews of IHS, and gretl (see Cottrell/Lucchetti (2016), p. 160-161 (example 18.1)), see for treatment in the literature, e.g. Greene (2012), Ch. 11.4.4, p. 364. It can be considered an overall intercept in the within model framework and is the weighted mean of fixed effects (see **Examples** for the relationship).

`within_intercept` estimates a new model which is computationally more demanding than just taking the weighted mean. However, with `within_intercept` one also gets the associated standard error and it is possible to get an overall intercept for twoway fixed effect models.

Users can set argument `vcov` to a function to calculate a specific (robust) variance–covariance matrix and get the respective (robust) standard error for the overall intercept, e.g. the function `vcovHC`, see examples for usage. Note: The argument `vcov` must be a function, not a matrix, because the model to calculate the overall intercept for the within model is different from the within model itself.

**Value**

A named numeric of length one: The overall intercept for the estimated within model along attribute "se" which contains the standard error for the intercept.

**Author(s)**

Kevin Tappe

**References**

Cottrell, A./Lucchetti, R., Gretl User's Guide, <http://gretl.sourceforge.net/gretl-help/gretl-guide.pdf>.

Gould, W. (2013), "How can there be an intercept in the fixed-effects model estimated by xtreg, fe?", <http://www.stata.com/support/faqs/statistics/intercept-in-fixed-effects-model/>.

Greene, W. H. (2012), Econometric Analysis, 7th ed., Prentice Hall, Ch. 11.4.4.

**See Also**

[fixef](#) to extract the fixed effects of a within model.

**Examples**

```
# estimate within model (unbalanced data)
data("Hedonic", package = "plm")
mod_fe <- plm(mv ~ age + crim, data = Hedonic, index = "townid")
overallint <- within_intercept(mod_fe)
attr(overallint, "se") # standard error

# overall intercept is the weighted mean of fixed effects in the one-way case
weighted.mean(fixef(mod_fe), as.numeric(table(index(mod_fe)[[1]])))

# relationship of type="dmean", "level" and within_intercept in the one-way case
data("Grunfeld", package = "plm")
gi <- plm(inv ~ value + capital, data = Grunfeld, model = "within")
fx_level <- fixef(gi, type = "level")
fx_dmean <- fixef(gi, type = "dmean")
overallint <- within_intercept(gi)
all.equal(overallint + fx_dmean, fx_level, check.attributes = FALSE) # TRUE

# overall intercept with robust standard error
within_intercept(gi, vcov = function(x) vcovHC(x, method="arellano", type="HC0"))
```

# Index

- \*Topic **aneweytest**
  - [aneweytest](#), 4
- \*Topic **array**
  - [detect\\_lin\\_dep](#), 10
- \*Topic **attribute**
  - [index](#), 23
  - [is.pbalanced](#), 24
  - [is.pconsecutive](#), 25
  - [make.pbalanced](#), 31
  - [make.pconsecutive](#), 33
  - [nobs](#), 40
  - [pdim](#), 55
  - [plm.data](#), 76
  - [punbalancedness](#), 87
  - [pvar](#), 93
  - [within\\_intercept](#), 124
- \*Topic **classes**
  - [dynformula](#), 13
  - [lag](#), 29
  - [model.frame.pFormula](#), 37
  - [pdata.frame](#), 52
  - [pFormula](#), 58
  - [pseries](#), 85
- \*Topic **datasets**
  - [Cigar](#), 5
  - [Crime](#), 8
  - [EmplUK](#), 14
  - [Gasoline](#), 19
  - [Grunfeld](#), 20
  - [Hedonic](#), 21
  - [LaborSupply](#), 28
  - [Males](#), 36
  - [Parity](#), 41
  - [Produc](#), 84
  - [RiceFarms](#), 105
  - [Snmesp](#), 107
  - [SumHes](#), 108
  - [Wages](#), 123
- \*Topic **htest**
  - [cipstest](#), 6
  - [cortab](#), 8
  - [mtest](#), 39
  - [pbgtest](#), 42
  - [pbltest](#), 44
  - [pbsytest](#), 45
  - [pcdtest](#), 49
  - [pdwtest](#), 56
  - [pFtest](#), 59
  - [pgrangertest](#), 65
  - [phtest](#), 68
  - [plmtest](#), 78
  - [pooltest](#), 83
  - [purtest](#), 90
  - [pwaldtest](#), 96
  - [pwartest](#), 98
  - [pwfdtest](#), 100
  - [pwtest](#), 101
  - [r.squared](#), 103
  - [sargan](#), 106
- \*Topic **manip**
  - [detect\\_lin\\_dep](#), 10
  - [pmodel.response](#), 82
- \*Topic **model**
  - [plm-package](#), 3
- \*Topic **package**
  - [plm-package](#), 3
- \*Topic **piest**
  - [piest](#), 70
- \*Topic **regression**
  - [ercomp](#), 15
  - [fixef.plm](#), 17
  - [pcce](#), 48
  - [pggls](#), 60
  - [pgmm](#), 62
  - [pht](#), 67
  - [pldv](#), 71
  - [plm](#), 72
  - [plm-package](#), 3

- pmg, 80
- pvcn, 94
- ranef.plm, 104
- summary.plm, 109
- vcovBK, 111
- vcovDC, 113
- vcovG, 115
- vcovHC, 116
- vcovNW, 119
- vcovSCC, 121
- [.pdata.frame (pdata.frame), 52
- [[.pdata.frame (pdata.frame), 52
- \$.pdata.frame (pdata.frame), 52
- alias, 11
- alias (detect\_lin\_dep), 10
- alias.lm, 11
- aneweytest, 4
- as.data.frame, 53
- as.data.frame.pdata.frame  
(pdata.frame), 52
- as.Formula.pFormula (pFormula), 58
- as.list.data.frame, 54
- as.list.pdata.frame (pdata.frame), 52
- as.matrix.pseries, 30
- as.matrix.pseries (pseries), 85
- Between, 30
- Between (pseries), 85
- between, 30
- between (pseries), 85
- Between.pseries (pseries), 85
- between.pseries (pseries), 85
- bgtest, 42, 43, 57
- Cigar, 5
- cipstest, 6, 92
- cortab, 8
- Crime, 8
- data.frame, 25, 27, 32, 35
- detect\_lin\_dep, 10, 75
- diff, 86
- diff (lag), 29
- dwtest, 57
- dynformula, 13, 64
- EmplUK, 14
- ercomp, 15
- Extract, 52, 53
- fixef, 75, 104, 111, 125
- fixef (fixef.plm), 17
- fixef.plm, 17
- Formula, 38, 58
- formula.dynformula (dynformula), 13
- Gasoline, 19
- grangertest, 66
- Grunfeld, 20, 21
- has.intercept (plm), 72
- Hedonic, 21
- index, 23, 54
- is.pbalanced, 24, 32, 56
- is.pconsecutive, 25, 25, 30, 32, 35, 54
- LaborSupply, 28
- lag, 27, 29, 30, 32, 35, 86
- lag.zoo, 30
- lead, 86
- lead (lag), 29
- lm, 5, 60, 62, 67, 70, 73, 80
- make.pbalanced, 25, 27, 31, 34, 35
- make.pconsecutive, 25, 27, 32, 33
- Males, 36
- model.frame, 11, 58, 82
- model.frame (model.frame.pFormula), 37
- model.frame.pFormula, 37
- model.matrix, 11, 58, 82
- model.matrix (model.frame.pFormula), 37
- mtest, 39, 64
- nobs, 40, 56, 89
- Parity, 41
- pbgtest, 42, 47, 57, 99, 102
- pbltest, 43, 44, 47, 57, 99, 102
- pbsytest, 43, 45, 45, 57, 99, 102
- pcce, 48
- pcdtest, 49
- pdata.frame, 24–27, 31, 32, 34, 35, 48, 50,  
52, 55, 56, 60, 65, 71, 76, 80, 88, 89,  
93, 95
- pdim, 25, 27, 32, 35, 41, 54, 55, 88, 89, 94
- pdwtest, 43, 45, 47, 56, 99, 102
- pFormula, 38, 58

- pFtest, [59](#), [79](#)
- pggls, [60](#)
- pgmm, [27](#), [40](#), [62](#), [107](#)
- pgrangertest, [65](#)
- pht, [67](#)
- phtest, [68](#)
- piest, [70](#)
- pldv, [71](#)
- plm, [10](#), [15](#), [16](#), [18](#), [24](#), [27](#), [72](#), [103](#), [110](#), [111](#)
- plm-package, [3](#)
- plm.data, [76](#)
- plmtest, [47](#), [59](#), [78](#), [102](#)
- plot.plm (plm), [72](#)
- pmg, [80](#)
- pmodel.response, [38](#), [58](#), [82](#)
- pooltest, [83](#)
- print.dynformula (dynformula), [13](#)
- print.ercomp (ercomp), [15](#)
- print.fixef (fixef.plm), [17](#)
- print.htest, [111](#)
- print.panelmodel (plm), [72](#)
- print.pdata.frame (pdata.frame), [52](#)
- print.pdim (pdim), [55](#)
- print.purtest (purtest), [90](#)
- print.pvar (pvar), [93](#)
- print.summary.aneweytest (aneweytest), [4](#)
- print.summary.fixef (fixef.plm), [17](#)
- print.summary.pcce (pcce), [48](#)
- print.summary.pggls (pggls), [60](#)
- print.summary.pgmm (pgmm), [62](#)
- print.summary.pht (pht), [67](#)
- print.summary.piest (piest), [70](#)
- print.summary.plm (summary.plm), [109](#)
- print.summary.pmg (pmg), [80](#)
- print.summary.pseries, [30](#)
- print.summary.pseries (pseries), [85](#)
- print.summary.purtest (purtest), [90](#)
- print.summary.pvcm (pvcm), [94](#)
- Produc, [84](#)
- pseries, [25](#), [27](#), [32](#), [35](#), [85](#)
- punbalancedness, [25](#), [32](#), [35](#), [56](#), [87](#)
- purtest, [7](#), [90](#)
- pvar, [25](#), [27](#), [32](#), [35](#), [54](#), [56](#), [93](#)
- pvcm, [94](#)
- pvcovHC (vcovHC), [116](#)
- pwaldtest, [96](#), [110](#), [111](#)
- pwartest, [43](#), [45](#), [57](#), [98](#), [102](#)
- pwfdtest, [43](#), [45](#), [57](#), [99](#), [100](#), [102](#)
- pwtest, [101](#)
- r.squared, [103](#), [111](#)
- ranef, [18](#)
- ranef (ranef.plm), [104](#)
- ranef.plm, [104](#)
- RiceFarms, [105](#)
- sargan, [64](#), [106](#)
- Snmesp, [107](#)
- SumHes, [108](#)
- summary.aneweytest (aneweytest), [4](#)
- summary.fixef (fixef.plm), [17](#)
- summary.pcce (pcce), [48](#)
- summary.pggls (pggls), [60](#)
- summary.pgmm (pgmm), [62](#)
- summary.pht (pht), [67](#)
- summary.piest (piest), [70](#)
- summary.plm, [75](#), [97](#), [98](#), [103](#), [109](#)
- summary.pmg (pmg), [80](#)
- summary.pseries, [30](#)
- summary.pseries (pseries), [85](#)
- summary.purtest (purtest), [90](#)
- summary.pvcm (pvcm), [94](#)
- vcovBK, [111](#), [116](#)
- vcovDC, [113](#), [116](#)
- vcovG, [115](#)
- vcovHC, [98](#), [111–114](#), [116](#), [116](#), [117](#), [118](#), [120](#), [122](#), [124](#)
- vcovNW, [116](#), [119](#)
- vcovSCC, [116](#), [120](#), [121](#)
- Wages, [123](#)
- waldtest, [97](#)
- Within, [30](#)
- Within (pseries), [85](#)
- within\_intercept, [18](#), [111](#), [124](#)