

Package ‘skedastic’

January 10, 2020

Type Package

Title Heteroskedasticity Diagnostics for Linear Regression Models

Version 0.1.0

Description Implements numerous methods for detecting heteroskedasticity (sometimes called heteroscedasticity) in the classical linear regression model. These include the parametric and nonparametric tests of Goldfeld and Quandt (1965) <doi:10.1080/01621459.1965.10480811>, the test of Glejser (1969) <doi:10.1080/01621459.1969.10500976> as formulated by Mittelhammer, Judge and Miller (2000, ISBN: 0-521-62394-4), the BAMSET Test of Ramsey (1969) <doi:10.1111/j.2517-6161.1969.tb00796.x>, which uses the BLUS residuals derived by Theil (1965) <doi:10.1080/01621459.1965.10480851>, the test of Harvey (1976) <doi:10.2307/1913974>, the test of Breusch and Pagan (1979) <doi:10.2307/1911963> with and without the modification proposed by Koenker (1981) <doi:10.1016/0304-4076(81)90062-2>, the test of White (1980) <doi:10.2307/1912934>, the test and graphical Cook and Weisberg (1983) <doi:10.1093/biomet/70.1.1>, and the test of Li and Yao (2019) <doi:10.1016/j.ecosta.2018.01.001>. Homoskedasticity refers to the assumption of constant variance that is imposed on the model errors (disturbances); heteroskedasticity is the violation of this assumption.

RdMacros Rdpack

Depends R (>= 3.5.0)

Imports stats (>= 3.5.0), utils (>= 3.5.0), Rdpack (>= 0.11.0), tibble (>= 2.1.3), broom (>= 0.5.3), magrittr (>= 1.5.0), pracma (>= 2.2.5), gmp (>= 0.5.13), Rmpfr (>= 0.7.2), matrixcalc (>= 1.0.3), lubridate (>= 1.7.4), lmtest (>= 0.9.37), car (>= 3.0.5), het.test (>= 0.1), tseries (>= 0.10.47)

Suggests

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

URL <http://github.com/tjfarrar/skedastic>

BugReports <http://github.com/tjfarrar/skedastic/issues>

NeedsCompilation no

Author Thomas Farrar [aut, cre] (<<https://orcid.org/0000-0003-0744-6972>>),
University of the Western Cape [cph]

Maintainer Thomas Farrar <tjfarrar@alumni.uwaterloo.ca>

Repository CRAN

Date/Publication 2020-01-10 17:40:06 UTC

R topics documented:

bamset	2
blus	4
breusch_pagan	5
cook_weisberg	7
countpeaks	8
dpeak	9
dpeakdat	10
glejser	11
goldfeld_quandt	12
harvey	14
hetplot	15
li_yao	17
ppeak	18
white_lm	20

Index	22
--------------	-----------

bamset	<i>Ramsey's BAMSET Test for Heteroskedasticity in a Linear Regression Model</i>
--------	---

Description

This function implements the Bartlett's M Specification Error Test (BAMSET) method of Ramsey (1969) for testing for heteroskedasticity in a linear regression model.

Usage

```
bamset(  
  mainlm,  
  k = 3,  
  deflator = NULL,  
  correct = TRUE,  
  omitatmargins = TRUE,  
  omit = NULL  
)
```

Arguments

<code>mainlm</code>	Either an object of class "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
<code>k</code>	An integer. The number of subsets (≥ 2) into which the BLUS residuals are to be partitioned. Defaults to 3, the value suggested in Ramsey (1969).
<code>deflator</code>	Either a character specifying a column name from the design matrix of <code>mainlm</code> or an integer giving the index of a column of the design matrix. This variable is suspected to be related to the error variance under the alternative hypothesis. <code>deflator</code> may not correspond to a column of 1's (intercept). Default NULL means the data will be left in its current order (e.g. in case the existing index is believed to be associated with error variance).
<code>correct</code>	A logical. Should the test statistic be divided by a scaling constant to improve the chi-squared approximation? Defaults to TRUE.
<code>omitatmargins</code>	A logical. Should the indices of observations at the margins of the <code>k</code> subsets be passed to <code>blus</code> as the <code>omit</code> argument? If TRUE (the default), this overrides any <code>omit</code> argument passed directly. If FALSE, the <code>omit</code> argument must be specified and cannot be left as NULL.
<code>omit</code>	A numeric vector of length p (the number of columns in the linear model design matrix) giving the indices of p observations to omit in the BLUS residual vector; or a character partially matching "first" (for the first p observations), "last" (for the last p observations), or "random" (for a random sample of p indices between 1 and n). Defaults to "first".

Details

BAMSET is an analogue of Bartlett's M Test for heterogeneity of variances across independent samples from k populations. In this case the populations are k subsets of the residuals from a linear regression model. In order to meet the independence assumption, BLUS residuals are computed, meaning that only $n - p$ observations are used (where n is the number of rows and p the number of columns in the design matrix). Under the null hypothesis of homoskedasticity, the test statistic is asymptotically chi-squared distributed with $k - 1$ degrees of freedom. The test is right-tailed.

Value

An object of class "htest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Ramsey J (1969). "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **31**(2), 350–371.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
bamset(mtcars_lm, deflator = "wt", k = 3)

# BLUS residuals cannot be computed with given `omit` argument and so
# omitted indices are randomised:
bamset(mtcars_lm, deflator = "wt", k = 4, omitatargins = FALSE, omit = "last")
```

blus	<i>Compute Best Linear Unbiased Scalar-Covariance (BLUS) residuals from a linear model</i>
------	--

Description

This function computes the Best Linear Unbiased Scalar-Covariance (BLUS) residuals from a linear model, as defined in Theil (1965) and explained further in Theil (1968).

Usage

```
blus(
  mainlm,
  omit = c("first", "last", "random"),
  keepNA = TRUE,
  exhaust = NULL
)
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
omit	A numeric vector of length p (the number of columns in the linear model design matrix) giving the indices of p observations to omit in the BLUS residual vector; or a character partially matching "first" (for the first p) observations, "last" (for the last p observations), or "random" (for a random sample of p indices between 1 and n). Defaults to "first".
keepNA	A logical. Should BLUS residuals for omitted observations be returned as <code>NA_real_</code> to preserve the length of the residual vector?
exhaust	An integer. If singular matrices are encountered using the passed value of <code>omit</code> , how many random combinations of p indices should be attempted before an error is thrown? If <code>NULL</code> (the default), all possible combinations are attempted (which could result in very slow execution if n is large).

Details

Under the ideal linear model conditions, the BLUS residuals have a scalar covariance matrix $\sigma^2 I$ (meaning they have a constant variance and are mutually uncorrelated), unlike the OLS residuals, which have covariance matrix $\sigma^2 M$ where M is a function of the design matrix. Use of BLUS residuals could improve the performance of tests for heteroskedasticity and/or autocorrelation in the linear model. A linear model with n observations and an $n \times p$ design matrix yields only $n - p$ BLUS residuals. The choice of which p observations will not be represented in the BLUS residuals is specified within the algorithm.

Value

A double vector of length n containing the BLUS residuals (with NA_real_) for omitted observations), or a double vector of length $n - p$ containing the BLUS residuals only (if keepNA is set to FALSE)

References

Theil H (1965). "The Analysis of Disturbances in Regression Analysis." *Journal of the American Statistical Association*, **60**(312), 1067–1079.

Theil H (1968). "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances." *Journal of the American Statistical Association*, **63**(321), 242–251.

See Also

H. D. Vinod's online article, [Theil's BLUS Residuals and R Tools for Testing and Removing Auto-correlation and Heteroscedasticity](#), for an alternative function for extracting BLUS residuals.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
blus(mtcars_lm)
plot(mtcars_lm$residuals, blus(mtcars_lm))

# BLUS residuals cannot be computed with `omit = "last"` in this case so
# omitted indices are randomised:
blus(mtcars_lm, omit = "last")
```

breusch_pagan

Breusch-Pagan Test for Heteroskedasticity in a Linear Regression Model

Description

This function implements the popular method of Breusch and Pagan (1979) for testing for heteroskedasticity in a linear regression model, with or without the studentising modification of Koenker (1981).

Usage

```
breusch_pagan(mainlm, auxdesign = NULL, koenker = TRUE)
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
auxdesign	A <code>data.frame</code> or <code>matrix</code> representing an auxiliary design matrix of containing exogenous variables that (under alternative hypothesis) are related to error variance, or a character "fitted.values" indicating that the fitted \hat{y}_i values from OLS should be used. If set to NULL (the default), the design matrix of the original regression model is used. An intercept is included in the auxiliary regression even if the first column of auxdesign is not a vector of ones.
koenker	A logical. Should studentising modification of Koenker (1981) be implemented? Defaults to TRUE; if FALSE, the original form of the test proposed by Breusch and Pagan (1979) is used.

Details

The Breusch-Pagan Test entails fitting an auxiliary regression model in which the response variable is the vector of squared residuals from the original model and the design matrix Z consists of one or more exogenous variables that are suspected of being related to the error variance. In the absence of prior information on a possible choice of Z , one would typically use the explanatory variables from the original model. Under the null hypothesis of homoskedasticity, the distribution of the test statistic is asymptotically chi-squared with parameter degrees of freedom. The test is right-tailed.

Value

An object of `class` "htest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Breusch T, Pagan A (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica*, **47**(5), 1287–1294.

Koenker R (1981). "A Note on Studentizing a Test for Heteroscedasticity." *Journal of Econometrics*, **17**, 107–112.

See Also

`lmtest::bptest`, which performs exactly the same test as this function; `car::ncvTest`, which is not the same test but is implemented in `cook_weisberg`; `white_lm`, which is a special case of the Breusch-Pagan Test.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
breusch_pagan(mtcars_lm)
breusch_pagan(mtcars_lm, koenker = FALSE)
```

cook_weisberg	<i>Cook-Weisberg Score Test for Heteroskedasticity in a Linear Regression Model</i>
---------------	---

Description

This function implements the score test of Cook and Weisberg (1983) for testing for heteroskedasticity in a linear regression model.

Usage

```
cook_weisberg(mainlm, auxdesign = NULL, errorfun = "additive")
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
auxdesign	A <code>data.frame</code> or <code>matrix</code> representing an auxiliary design matrix of containing exogenous variables that (under alternative hypothesis) are related to error variance, or a character "fitted.values" indicating that the fitted \hat{y}_i values from OLS should be used. If set to NULL (the default), the design matrix of the original regression model is used. An intercept is included in the auxiliary regression even if the first column of auxdesign is not a vector of ones.
errorfun	A character describing the functional form of the error variance under the heteroskedastic alternative. Possible values are "additive" (the default) and "multiplicative", corresponding to the two cases considered in Cook and Weisberg (1983), or the name of a function in the environment (passed as a character). If the name of a function, it will be applied to auxiliary design element z_{ij} to obtain the corresponding element of D , according to the notation used in Cook and Weisberg (1983). The value "additive" corresponds to the function identity and "multiplicative" to the function log. Partial matching is NOT used.

Details

The Cook-Weisberg Score Test entails fitting an auxiliary regression model in which the response variable is the vector of standardised squared residuals $e_i^2/\hat{\sigma}^2$ from the original OLS model and the design matrix is some function of Z , an $n \times q$ matrix consisting of q exogenous variables. The test statistic is half the residual sum of squares from this auxiliary regression. Under the null hypothesis of homoskedasticity, the distribution of the test statistic is asymptotically chi-squared with q degrees of freedom. The test is right-tailed.

Value

An object of class "htest". If object is not assigned, its attributes are displayed in the console as a [tibble](#) using [tidy](#).

References

Cook RD, Weisberg S (1983). "Diagnostics for Heteroscedasticity in Regression." *Biometrika*, **70**(1), 1–10.

See Also

[car::ncvTest](#), which implements the same test. Calling `car::ncvTest` with `var.formula` argument omitted is equivalent to calling `skedastic::cook_weisberg` with `auxdesign = "fitted.values"`, `errorfun = "additive"`. Calling `car::ncvTest` with `var.formula = ~ X` (where X is the design matrix of the linear model with the intercept column omitted) is equivalent to calling `skedastic::cook_weisberg` with default `auxdesign` and `errorfun` values. The `errorfun = "multiplicative"` option has no equivalent in `car::ncvTest`.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
cook_weisberg(mtcars_lm)
cook_weisberg(mtcars_lm, auxdesign = "fitted.values", errorfun = "multiplicative")
```

countpeaks

Count peaks in a data series

Description

This function computes the number of peaks in a univariate data series, with peak defined as per Goldfeld and Quandt (1965). The function is used in the Goldfeld-Quandt nonparametric test for heteroskedasticity in a linear model.

Usage

```
countpeaks(x)
```


Arguments

x A double vector.

Value

An integer value between 0 and $\text{length}(x) - 1$ representing the number of peaks in the series.

References

Goldfeld SM, Quandt RE (1965). “Some Tests for Homoscedasticity.” *Journal of the American Statistical Association*, **60**(310), 539–547.

See Also

[goldfeld_quandt](#)

Examples

```
set.seed(9586)
countpeaks(rnorm(20))
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
countpeaks(mtcars_lm$residuals)
```

dpeak

Probability mass function of number of peaks in a random series

Description

This function computes $P(n, k)$ as defined by Goldfeld and Quandt (1965), i.e. the probability that a series of n independent and identically distributed random variables contains exactly k peaks, with peaks also as defined by Goldfeld and Quandt (1965). The function is used in [ppeak](#) to compute p -values for the Goldfeld-Quandt nonparametric test for heteroskedasticity in a linear model.

Usage

```
dpeak(n, k, usedata = FALSE)
```

Arguments

n A positive integer representing the number of observations in the series.

k An integer or a sequence of integers strictly incrementing by 1, with all values between 0 and $n - 1$ inclusive. Represents the number of peaks in the series.

usedata A logical. Should probability mass function values be read from [dpeakdat](#) rather than computing them? This option will save significantly on computation time if $n < 170$ but is currently only available for $n \leq 500$.

Value

A double between 0 and 1 representing the probability of exactly k peaks occurring in a series of n independent and identically distributed continuous random variables. Computation time is very slow for $n > 170$ (if `usedata` is `FALSE`) and for $n > 500$ regardless of `usedata` value.

References

Goldfeld SM, Quandt RE (1965). "Some Tests for Homoscedasticity." *Journal of the American Statistical Association*, **60**(310), 539–547.

See Also

[ppeak](#), [goldfeld_quandt](#)

Examples

```
dpeak(10, 0:9)
plot(0:9, dpeak(10, 0:9), type = "p", pch = 20, xlab = "Number of Peaks",
     ylab = "Probability")

# `dpeakdat` is a dataset containing probabilities generated from `dpeak`
utils::data(dpeakdat)
expval <- unlist(lapply(dpeakdat,
                      function(p) sum(p * 0:(length(p) - 1))))
plot(1:500, expval[1:500], type = "l", xlab = parse(text = "n"),
     ylab = "Expected Number of Peaks")
```

dpeakdat	<i>Probability distribution for number of peaks in a continuous, uncorrelated stochastic series</i>
----------	---

Description

A dataset containing the probability mass function for the distribution of the number of peaks in a continuous, uncorrelated stochastic series. These probabilities were generated from the [dpeak](#) function. This function is computationally very slow for $n > 170$; thus the functions of `skedastic` package that require peak probabilities ([ppeak](#) and [goldfeld_quandt](#)) by default obtain the probabilities from this data set rather than from [dpeak](#), provided that $n \leq 500$.

Usage

```
dpeakdat
```

Format

A list of 500 objects. The n th object is a double vector of length n , with elements representing the probability of k peaks for $k = 0, 1, \dots, n - 1$.

`glejser`*Glejser Test for Heteroskedasticity in a Linear Regression Model*

Description

This function implements the method of Glejser (1969) for testing for "multiplicative" heteroskedasticity in a linear regression model. Mittelhammer et al. (2000) gives the formulation of the test used here.

Usage

```
glejser(mainlm, auxdesign = NULL)
```

Arguments

<code>mainlm</code>	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
<code>auxdesign</code>	A <code>data.frame</code> or <code>matrix</code> representing an auxiliary design matrix of containing exogenous variables that (under alternative hypothesis) are related to error variance, or a character "fitted.values" indicating that the fitted \hat{y}_i values from OLS should be used. If set to NULL (the default), the design matrix of the original regression model is used. An intercept is included in the auxiliary regression even if the first column of <code>auxdesign</code> is not a vector of ones.

Details

Glejser's Test entails fitting an auxiliary regression model in which the response variable is the absolute residual from the original model and the design matrix Z consists of one or more exogenous variables that are suspected of being related to the error variance. In the absence of prior information on a possible choice of Z , one would typically use the explanatory variables from the original model. Under the null hypothesis of homoskedasticity, the distribution of the test statistic is asymptotically chi-squared with parameter degrees of freedom. The test is right-tailed.

Value

An object of `class` "hctest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Glejser H (1969). "A New Test for Heteroskedasticity." *Journal of the American Statistical Association*, **64**(325), 316–323.

Mittelhammer RC, Judge GG, Miller DJ (2000). *Econometric Foundations*. Cambridge, Cambridge University Press.

See Also

the description of the test in [SHAZAM](#) software (which produces identical results).

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
glejser(mtcars_lm)
```

goldfeld_quandt	<i>Goldfeld-Quandt Tests for Heteroskedasticity in a Linear Regression Model</i>
-----------------	--

Description

This function implements the two methods (parametric and nonparametric) of Goldfeld and Quandt (1965) for testing for heteroskedasticity in a linear regression model.

Usage

```
goldfeld_quandt(
  mainlm,
  method = "parametric",
  deflator = NULL,
  prop_central = 1/3,
  alternative = c("greater", "less", "two.sided"),
  pvals = NULL
)
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
method	A character indicating which of the two tests derived in Goldfeld and Quandt (1965) should be implemented. Possible values are "parametric" and "nonparametric". Default is "parametric". It is acceptable to specify only the first letter.

deflator	Either a character specifying a column name from the design matrix of <code>mainlm</code> or an integer giving the index of a column of the design matrix. This variable is suspected to be related to the error variance under the alternative hypothesis. <code>deflator</code> may not correspond to a column of 1's (intercept). Default NULL means the data will be left in its current order (e.g. in case the existing index is believed to be associated with error variance).
prop_central	A double specifying the proportion of central observations to exclude from the F test (when method is "parametric" only). <code>round</code> is used to ensure the number of central observations is an integer. The value must be small enough to allow the two auxiliary regressions to be fit; otherwise an error is thrown. Defaults to 1/3.
alternative	A character specifying the form of alternative hypothesis ((when method is "parametric" only; nonparametric test is strictly upper-tailed). If it is suspected that the error variance is positively associated with the deflator variable, "greater". If it is suspected that error variance is negatively associated with deflator variable, "less". If no prior information is available, "two.sided". Defaults to "greater".
pvals	A vector of p -values, i.e. upper cumulative probabilities corresponding to values of the test statistic from 0 to $n-1$ (used only when method is "nonparametric"). If NULL (the default), p -values are calculated within the function. To improve performance where the test is being used many times (e.g. in a Monte Carlo simulation), a vector of p -values for a particular n can be computed beforehand using <code>ppeak</code> and passed to the function, so that the p -values are not computed each time the function executes (see Examples). Computation of <code>ppeak</code> is extremely slow for $n > 170$.

Details

The parametric test entails putting the data rows in increasing order of some specified deflator (one of the explanatory variables). A specified proportion of the most central observations (under this ordering) is removed, leaving a subset of lower observations and a subset of upper observations. Separate OLS regressions are fit to these two subsets of observations (using all variables from the original model). The test statistic is the ratio of the sum of squared residuals from the 'upper' model to the sum of squared residuals from the 'lower' model. Under the null hypothesis, the test statistic is exactly F-distributed with numerator and denominator degrees of freedom equal to $(n - k)/2 - p$ where n is the number of observations in the original regression model, k is the number of central observations removed, and p is the number of columns in the design matrix (number of parameters to be estimated, including intercept).

The nonparametric test entails putting the residuals of the linear model in increasing order of some specified deflator (one of the explanatory variables). The test statistic is the number of peaks, with the j th absolute residual $|e_j|$ defined as a peak if $|e_j| \geq |e_i|$ for all $i < j$. The first observation does not constitute a peak. If the number of peaks is large relative to the distribution of peaks under the null hypothesis, this constitutes evidence for heteroskedasticity.

Value

An object of class "htest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Goldfeld SM, Quandt RE (1965). "Some Tests for Homoscedasticity." *Journal of the American Statistical Association*, **60**(310), 539–547.

See Also

`lmtest::gqtest`, which performs the parametric version of the Goldfeld-Quandt Test. The 'point' argument in that function allows the splitting point of data set into subsets to be other than the middle observation and thus allows the subsets to be of different sizes.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
goldfeld_quandt(mtcars_lm, deflator = "qsec", prop_central = 0.25)
goldfeld_quandt(mtcars_lm, deflator = "qsec", method = "nonparametric")
```

 harvey

Harvey Test for Heteroskedasticity in a Linear Regression Model

Description

This function implements the method of Harvey (1976) for testing for "multiplicative" heteroskedasticity in a linear regression model. Mittelhammer et al. (2000) gives the formulation of the test used here.

Usage

```
harvey(mainlm, auxdesign = NULL)
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
auxdesign	A <code>data.frame</code> or <code>matrix</code> representing an auxiliary design matrix of containing exogenous variables that (under alternative hypothesis) are related to error variance, or a character "fitted.values" indicating that the fitted \hat{y}_i values from OLS should be used. If set to NULL (the default), the design matrix of the original regression model is used. An intercept is included in the auxiliary regression even if the first column of auxdesign is not a vector of ones.

Details

Harvey's Test entails fitting an auxiliary regression model in which the response variable is the log of the vector of squared residuals from the original model and the design matrix Z consists of one or more exogenous variables that are suspected of being related to the error variance. In the absence of prior information on a possible choice of Z , one would typically use the explanatory variables from the original model. Under the null hypothesis of homoskedasticity, the distribution of the test statistic is asymptotically chi-squared with parameter degrees of freedom. The test is right-tailed.

Value

An object of class "htest". If object is not assigned, its attributes are displayed in the console as a [tibble](#) using [tidy](#).

References

Harvey AC (1976). "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica*, **44**(3), 461–465.

Mittelhammer RC, Judge GG, Miller DJ (2000). *Econometric Foundations*. Cambridge, Cambridge University Press.

See Also

the description of the test in [SHAZAM](#) software (which produces identical results).

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
harvey(mtcars_lm)
harvey(mtcars_lm, auxdesign = "fitted.values")
```

hetplot

Graphical Methods for Detecting Heteroskedasticity in a Linear Regression Model

Description

This function creates various two-dimensional scatter plots that can aid in detecting heteroskedasticity in a linear regression model.

Usage

```
hetplot(
  mainlm,
  horzvar = 1:n,
  vertvar = "res",
  vertfun = "identity",
```

```

    filetype = NULL,
    ...
)

```

Arguments

mainlm	Either an object of class "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
horzvar	A character vector describing the variable(s) to plot on horizontal axes ("index" for the data index i , "fitted.values" for the OLS predicted values \hat{y}_i , "fitted.values2" for transformed OLS predicted values $m_{ii}\hat{y}_i$, and/or names of explanatory variable columns). "explanatory" passes all explanatory variable columns. "log" concatenated with names of explanatory variable columns passes logs of those explanatory variables. "log_explanatory" passes logs of all explanatory variables. If more than one variable is specified, a separate plot is created for each.
vertvar	A character vector describing the variable to plot on the vertical axis ("res" for OLS residuals [the default], "res_blus" for BLUS residuals, "res_stand" for standardised OLS residuals: $e_i/\hat{\sigma}$, "res_constvar" for OLS residuals transformed to have constant variance: $e_i/\sqrt{m_{ii}}$, "res_stud" for studentised OLS residuals: $e_i/(s\sqrt{m_{ii}})$). If more than one value is specified, a separate plot is created for each.
vertfun	A character vector giving the names of functions to apply to the vertvar variable. Numerals such as "2" are taken to be powers to which vertvar should be set. If multiple values are specified, they are all applied to each element of vertvar.
filetype	A character giving the type of image file to which the plot(s) should be written. Values can be "png", "bmp", "jpeg", or "tiff". Image files are written to a subdirectory called "hetplot" within the R session's temporary directory, which can be located using <code>tempdir()</code> . The files should be moved or copied to another location if they are needed after the R session is ended. Default filenames contain timestamps for uniqueness. If NULL (the default), no image files are written, and in this case, if there are multiple plots, they are plotted on a single device using the "mfrow" graphical parameter. If many plots are requested at once, it is advisable to write them to image files.
...	Arguments to be passed to methods, such as graphical parameters (see <code>par</code>), parameters for <code>plot</code> , for <code>graphics devices</code> , and/or the <code>omit</code> argument for function <code>blus</code> , if BLUS residuals are being plotted. If it is desired to pass the <code>type</code> argument to a graphics device, use <code>gtype =</code> , since a <code>type</code> argument will be passed to <code>plot</code> .

Details

The variable plotted on the horizontal axis could be the original data indices, one of the explanatory variables, the OLS predicted (fitted) values, or any other numeric vector specified by the user.

The variable plotted on the vertical axis is some function of the OLS residuals or transformed version thereof such as the BLUS residuals Theil (1968) or standardised or studentised residuals as discussed in Cook and Weisberg (1983). A separate plot is created for each (horzvar, vertvar, vertfun) combination.

Value

A list containing two [data frames](#), one for vectors plotted on horizontal axes and one for vectors plotted on vertical axes.

References

Cook RD, Weisberg S (1983). "Diagnostics for Heteroscedasticity in Regression." *Biometrika*, **70**(1), 1–10.

Theil H (1968). "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances." *Journal of the American Statistical Association*, **63**(321), 242–251.

See Also

[plot.lm](#)

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
# Generates 2 x 2 matrix of plots in console
hetplot(mtcars_lm, horzvar = c("index", "fitted.values"),
vertvar = c("res_blus"), vertfun = c("2", "abs"), filetype = NULL)

# Generates 84 png files in tempdir() folder
## Not run
# hetplot(mainlm = mtcars_lm, horzvar = c("explanatory", "log_explanatory",
# "fitted.values2"), vertvar = c("res", "res_stand", "res_stud",
# "res_constvar"), vertfun = c("identity", "abs", "2"), filetype = "png")
```

li_yao

Li-Yao ALRT and CVT Tests for Heteroskedasticity in a Linear Regression Model

Description

This function implements the two methods of Li and Yao (2019) for testing for heteroskedasticity in a linear regression model.

Usage

```
li_yao(mainlm, method = "alrt")
```

Arguments

mainlm	Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important.
method	A character indicating which of the two tests derived in Li and Yao (2019) should be implemented. Possible values are "alrt" (approximate likelihood-ratio test) and "cvt" (coefficient-of-variation test). Default is "alrt". It is acceptable to specify only the first letter.

Details

These two tests are straightforward to implement; in both cases the test statistic is a function only of the residuals of the linear regression model. Furthermore, in both cases the test statistic is asymptotically normally distributed under the null hypothesis of homoskedasticity. Both tests are right-tailed. These tests are designed to be especially powerful in high-dimensional regressions, i.e. when the number of explanatory variables is large.

Value

An object of `class` "htest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Li Z, Yao J (2019). "Testing for heteroscedasticity in high-dimensional regressions." *Econometrics and Statistics*, **9**, 122–139.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
li_yao(mtcars_lm, method = "alrt")
li_yao(mtcars_lm, method = "cvt")
```

ppeak

Cumulative distribution function of number of peaks in a random series

Description

This function computes $\sum_k P(n, k)$, i.e. the probability that a series of n independent and identically distributed random variables contains $\geq k$ ($\leq k$) peaks, with peaks as defined in Goldfeld and Quandt (1965). The function may be used to compute p -values for the Goldfeld-Quandt nonparametric test for heteroskedasticity in a linear model. Computation time is very slow for $n > 170$ if `usedata` is set to `FALSE`.

Usage

```
ppeak(n, k, upper = TRUE, usedata = TRUE)
```

Arguments

n	A positive integer representing the number of observations in the series.
k	An integer or a sequence of integers strictly incrementing by 1, with all values between 0 and $n - 1$ inclusive. Represents the number of peaks in the series.
upper	A logical. Should upper tailed cumulative probability be calculated? Defaults to TRUE due to function being designed primarily for calculating p -values for the peaks test. Note that both upper and lower tailed cumulative probabilities are computed inclusive of k .
usedata	A logical. Should probability mass function values be read from dpeakdat rather than computing them from dpeak ? This option will save significantly on computation time if $n < 170$ but is currently only available for $n \leq 500$.

Value

A double between 0 and 1 representing the probability of at least (at most) k peaks occurring in a series of n independent and identically distributed continuous random variables.

References

Goldfeld SM, Quandt RE (1965). "Some Tests for Homoscedasticity." *Journal of the American Statistical Association*, **60**(310), 539–547.

See Also

[dpeak](#), [goldfeld_quandt](#)

Examples

```
# For an independent sample of size 250, the probability of at least 10
# peaks is 0.06186582
ppeak(250, 10, upper = TRUE, usedata = TRUE)
# For an independent sample of size 10, the probability of at most 2 peaks
# is 0.7060615
ppeak(10, 2, upper = FALSE, usedata = FALSE)
```

`white_lm`*White's Test for Heteroskedasticity in a Linear Regression Model*

Description

This function implements the popular method of White (1980) for testing for heteroskedasticity in a linear regression model.

Usage

```
white_lm(mainlm, interactions = FALSE)
```

Arguments

- | | |
|---------------------------|--|
| <code>mainlm</code> | Either an object of <code>class</code> "lm", or a list of two components: a response vector and a design matrix (in that order). If the latter, the design matrix must begin with a column of 1s if an intercept is to be included in the linear model. Passing an object of class "lm" is recommended in applications; passing a list containing the data is recommended where optimising computational performance is important. |
| <code>interactions</code> | A logical. Should two-way interactions between explanatory variables be included in the auxiliary regression? Defaults to FALSE, since when interaction terms are present the test is not a pure test of heteroskedasticity but also of model specification. |

Details

White's Test entails fitting an auxiliary regression model in which the response variable is the vector of squared residuals from the original model and the design matrix includes the original explanatory variables, their squares, and (optionally) their two-way interactions. The test statistic is the number of observations multiplied by the coefficient of determination from the auxiliary regression model:

$$T = nr_{\text{aux}}^2$$

White's Test is thus a special case of the method of Breusch and Pagan (1979). Under the null hypothesis of homoskedasticity, the distribution of the test statistic is asymptotically chi-squared with parameter degrees of freedom. The test is right-tailed.

Value

An object of `class` "htest". If object is not assigned, its attributes are displayed in the console as a `tibble` using `tidy`.

References

Breusch T, Pagan A (1979). “A Simple Test for Heteroscedasticity and Random Coefficient Variation.” *Econometrica*, **47**(5), 1287–1294.

White H (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, **48**(4), 817–838.

See Also

This function should not be confused with `het.test::whites.htest` and `tseries::white.test`, which do *not* implement the method of White (1980) for testing for heteroskedasticity in a linear model.

Examples

```
mtcars_lm <- lm(mpg ~ wt + qsec + am, data = mtcars)
white_lm(mtcars_lm)
white_lm(mtcars_lm, interactions = TRUE)
```

Index

*Topic **datasets**

dpeakdat, 10

bamset, 2

BLUS, 16

blus, 3, 4, 16

BLUS residuals, 3

breusch_pagan, 5

car::ncvTest, 6, 8

class, 3, 4, 6–8, 11–16, 18, 20

cook_weisberg, 6, 7

countpeaks, 8

data frames, 17

data.frame, 6, 7, 11, 14

dpeak, 9, 10, 19

dpeakdat, 9, 10, 19

glejser, 11

goldfeld_quandt, 9, 10, 12, 19

graphics devices, 16

harvey, 14

het.test::whites.htest, 21

hetplot, 15

li_yao, 17

lmtest::bptest, 6

lmtest::gqtest, 14

matrix, 6, 7, 11, 14

par, 16

plot, 16

plot.lm, 17

ppeak, 9, 10, 13, 18

round, 13

tibble, 3, 6, 8, 11, 13, 15, 18, 20

tidy, 3, 6, 8, 11, 13, 15, 18, 20

tseries::white.test, 21

white_lm, 6, 20