

Package ‘KOTORY’

October 8, 2026

Type Package

Title Robust Three-Group Tests for Heteroscedasticity in Linear Regression

Version 0.1.0

Date 2026-09-28

Maintainer Ebrahim Khaled Ebrahim <ebrahimkhaled@alexu.edu.eg>

Description Tests for heteroscedasticity in the linear regression model that sort the data by a regressor, split them into three equal parts and compare the error scale of the parts. The ordinary least squares version 'kah3.test()' refers the ratio of the largest to the smallest residual mean square to its exact null distribution, Hartley's maximum F-ratio with three groups; the robust version 'kah.robust.test()' replaces the mean squares by least trimmed squares scales, so that outliers neither create nor hide heteroscedasticity, and refers the ratio to a maximum F-ratio with simulated effective degrees of freedom, to a Monte Carlo reference or to a residual bootstrap. The distribution, density, quantile and random generation functions of the maximum F-ratio are provided, together with 'run.all.het()', which runs the proposed tests next to the Goldfeld-Quandt, Breusch-Pagan, White and robust modified Goldfeld-Quandt tests in one call. For more details see Hartley (1950) <doi:10.1093/biomet/37.3-4.308>, Goldfeld and Quandt (1965) <doi:10.1080/01621459.1965.10480811> and Rousseeuw (1984) <doi:10.1080/01621459.1984.10477105>.

License GPL-3

URL <https://github.com/ebrahimkhaled/KOTORY>

BugReports <https://github.com/ebrahimkhaled/KOTORY/issues>

Depends R (>= 3.5.0)

Imports robustbase, stats

Suggests lmtest, skedastic, testthat (>= 3.0.0)

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no
Author Ahmed El-Kotory [aut] (ORCID: <<https://orcid.org/0009-0000-1485-0768>>),
Ebrahim Khaled Ebrahim [aut, cre] (ORCID:
<<https://orcid.org/0009-0006-7839-8778>>)
Repository CRAN
Date/Publication 2026-10-08 18:00:02 UTC

Contents

fmax	2
kah.nu.star	3
kah.robust.test	4
kah3.test	6
run.all.het	8
Index	10

fmax	<i>Hartley’s Maximum F-Ratio Distribution</i>
------	---

Description

Distribution function, density, quantile function and random generation for Hartley’s maximum F-ratio: the ratio of the largest to the smallest of k independent mean squares, each with df degrees of freedom, under equal population variances and normal data.

Usage

```
pfmax(q, df, k = 3, lower.tail = TRUE)

dfmax(x, df, k = 3)

qfmax(p, df, k = 3, lower.tail = TRUE)

rfmax(n, df, k = 3)
```

Arguments

q, x	Vector of quantiles (values of the ratio, ≥ 1).
df	Degrees of freedom of each mean square (positive, may be non-integer).
k	Number of groups (integer ≥ 2). Default 3.
lower.tail	Logical; if TRUE (default) probabilities are $P(F_{max} \leq q)$, otherwise $P(F_{max} > q)$.
p	Vector of probabilities.
n	Number of observations to generate.

Details

Let $S_1, \dots, S_k$ be independent χ_ν^2 variables. The statistic $F_{max} = \max_j S_j / \min_j S_j$ has distribution function

$$P(F_{max} \leq c) = k \int_0^1 [G\{cG^{-1}(u)\} - u]^{k-1} du, \quad c \geq 1,$$

where G is the χ_ν^2 distribution function. The integral is taken on the probability scale, where the integrand is bounded for every $\nu > 0$, so non-integer degrees of freedom are allowed. The density is obtained by differentiating under the integral sign.

For $k = 2$, $P(F_{max} \leq c) = P(1/c \leq F_{\nu,\nu} \leq c)$, the two-sided variance-ratio test.

The quantile function in **SuppDists** (`qmaxFratio`) is inaccurate in the far upper tail; these functions integrate numerically instead.

Value

`pfmax` gives the distribution function, `dfmax` the density, `qfmax` the quantile function and `rfmax` random deviates. Arguments are recycled to a common length.

References

Hartley HO (1950). "The maximum F-ratio as a short-cut test for heterogeneity of variance." *Biometrika*, **37**(3/4), 308-312. doi:[10.1093/biomet/37.34.308](https://doi.org/10.1093/biomet/37.34.308)

Examples

```
# Hartley's tabulated 5% and 1% points for k = 3 groups, 4 d.f.: 15.5 and 37
qfmax(c(0.95, 0.99), df = 4, k = 3)

# p-value of an observed ratio of 6.2 with three groups of 10 d.f.
pfmax(6.2, df = 10, lower.tail = FALSE)

# k = 2 reduces to the two-sided F test
pfmax(3, df = 8, k = 2)
pf(3, 8, 8) - pf(1 / 3, 8, 8)
```

Description

The effective degrees of freedom ν^* used by `kah.robust.test` with `method = "fmax"`: the raw least trimmed squares scale of a part with m observations and p regressors behaves approximately like $\sigma^2 \chi_{\nu^*}^2 / \nu^*$ under normal errors.

Usage

```
kah.nu.star(m, p, alpha = 0.75)
```

Arguments

m	Number of observations in each part, $\lfloor n/3 \rfloor$.
p	Number of regressors (excluding the intercept).
alpha	LTS coverage: 0.5, 0.75 or 0.9.

Details

For a scaled chi-square variable, $Var\{\log(\chi_\nu^2/\nu)\} = \psi'(\nu/2)$, where ψ' is the trigamma function. For every grid point, the LTS scale of one part was simulated under normal errors and ν^* was obtained by solving $\psi'(\nu^*/2) = \widehat{Var}(\log \hat{\sigma}_{LTS}^2)$; the same step applied to least squares recovers $\nu = m - p - 1$. Values between grid points are interpolated on $\log m$; beyond the largest part size the stable ratio ν^*/ν is used. The table is available for $\alpha = 0.5, 0.75$ and 0.9 and $p = 1, \dots, 5$; for other settings use `kah.robust.test(method = "mc")`.

Value

The effective degrees of freedom ν^* (a single number).

Examples

```
kah.nu.star(m = 20, p = 1, alpha = 0.75)
# ratio to the least squares degrees of freedom
kah.nu.star(100, 2, 0.9) / (100 - 2 - 1)
```

kah.robust.test	<i>Robust KaH Test for Heteroscedasticity (Least Trimmed Squares)</i>
-----------------	---

Description

The outlier-resistant version of [kah3.test](#). The data are sorted by a regressor and split into three equal parts; a least trimmed squares (LTS) regression is fitted in each part, and the ratio of the largest to the smallest raw LTS scale is the test statistic. Outliers therefore neither create nor hide heteroscedasticity as easily as with least squares.

Usage

```
kah.robust.test(
  x,
  data = NULL,
  order.by = NULL,
  alpha = 0.75,
  method = c("fmax", "mc", "bootstrap"),
  B = 499,
  seed = 1
)
```

Arguments

x	A fitted lm object, or a model formula.
data	A data frame, used when x is a formula.
order.by	The variable the data are sorted by before splitting: the name of a regressor or of a column of data, or a numeric vector of length n . Default: the first regressor.
alpha	LTS coverage: the fraction of each part that is kept (between 0.5 and 1). Lower values resist more outliers but cost power.
method	Reference distribution: "fmax", "mc" or "bootstrap". See Details.
B	Number of Monte Carlo or bootstrap samples.
seed	Integer seed for the LTS subsampling and the resampling, or NULL to use the current random-number stream.

Details

The statistic is $T = \max_j \hat{\sigma}_j / \min_j \hat{\sigma}_j$, a ratio of *scales* (standard deviations), computed with `robustbase::ltsReg(alpha = alpha)`. Three references are available:

"fmax" (default) T^2 is referred to Hartley's maximum F-ratio with three groups and ν^* effective degrees of freedom. The LTS scale behaves like a scaled $\chi^2_{\nu^*}$ variable with $\nu^* < \nu = m - p - 1$, because trimming loses information; ν^* was estimated by simulation for $\alpha = 0.5, 0.75$ and 0.9 , $p = 1, \dots, 5$ and part sizes up to 200 (interpolated between grid points; the large- m ratio ν^*/ν is used beyond). See kah.nu.star.

"mc" A Monte Carlo reference under homoscedastic normal errors for the observed regressors: B data sets are simulated and the statistic recomputed. Works for any alpha and any number of regressors.

"bootstrap" A residual bootstrap: the standardized LTS residuals of the three parts are pooled and resampled, which keeps the error distribution of the data (heavy tails, outliers) under the null hypothesis. This is the reference to use when the errors may be non-normal.

The statistic is invariant to the regression coefficients and to the error scale, so the Monte Carlo and bootstrap samples are generated with zero coefficients.

LTS uses random subsampling. For reproducible results the fits are run under seed; the caller's random-number stream is left unchanged.

Value

An object of class "htest" with components statistic, parameter (k, df and alpha), p.value, estimate (the three LTS scales), method, alternative, data.name, critical (critical values of T at the 0.5\ 1\

References

- Rousseeuw PJ (1984). "Least median of squares regression." *Journal of the American Statistical Association*, **79**(388), 871-880. doi:10.1080/01621459.1984.10477105
- Rousseeuw PJ, Van Driessen K (2006). "Computing LTS regression for large data sets." *Data Mining and Knowledge Discovery*, **12**(1), 29-45. doi:10.1007/s1061800500244
- Hartley HO (1950). "The maximum F-ratio as a short-cut test for heterogeneity of variance." *Biometrika*, **37**(3/4), 308-312. doi:10.1093/biomet/37.34.308

See Also

[kah3.test](#), [run.all.het](#), [kah.nu.star](#).

Examples

```
kah.robust.test(dist ~ speed, data = cars)

# one outlier in a homoscedastic sample
set.seed(2)
x <- runif(60)
y0 <- 1 + x + rnorm(60)
y <- y0
y[which.min(x)] <- 12
c(clean = kah3.test(y0 ~ x)$p.value, outlier = kah3.test(y ~ x)$p.value)
# least squares is fooled; the robust statistic does not move
c(clean = kah.robust.test(y0 ~ x)$p.value, outlier = kah.robust.test(y ~ x)$p.value)

# bootstrap reference, safest when the errors may be non-normal
kah.robust.test(y ~ x, method = "bootstrap", B = 199)
```

kah3.test

KaH-III Test for Heteroscedasticity (Exact Null Distribution)

Description

The three-group variance-ratio test for heteroscedasticity in the linear regression model. The data are sorted by a regressor and split into three equal parts; the regression is fitted by ordinary least squares in each part, and the ratio of the largest to the smallest residual mean square is referred to its exact null distribution.

Usage

```
kah3.test(x, data = NULL, order.by = NULL)
```

Arguments

<code>x</code>	A fitted <code>lm</code> object, or a model formula.
<code>data</code>	A data frame, used when <code>x</code> is a formula.
<code>order.by</code>	The variable the data are sorted by before splitting: the name of a regressor or of a column of data, or a numeric vector of length n . Default: the first regressor.

Details

With $m = \lfloor n/3 \rfloor$ observations per part and p regressors, each residual mean square is $s_j^2 = \sigma^2 \chi_\nu^2 / \nu$ with $\nu = m - p - 1$ under homoscedastic normal errors, and the three are independent because the parts share no observations. The statistic

$$T = \max_j s_j^2 / \min_j s_j^2$$

therefore follows Hartley's maximum F-ratio with $k = 3$ groups and ν degrees of freedom *exactly*, at every sample size (see [pfmax](#)). This is the three-group extension of the Goldfeld-Quandt test; large values indicate that the error variance changes along the sorting variable.

When n is not a multiple of 3, the one or two observations at the part boundaries are left out, exactly as in the original implementation.

The test assumes normal errors. With heavy-tailed errors or outliers it rejects too often; use [kah.robust.test](#) in that case.

Value

An object of class "htest" with components statistic (the ratio T), parameter (k and df), p.value, estimate (the three residual mean squares), method, alternative and data.name; plus critical, the critical values at the 0.5\ and n.used, the number of observations in the three parts.

References

Hartley HO (1950). "The maximum F-ratio as a short-cut test for heterogeneity of variance." *Biometrika*, **37**(3/4), 308-312. doi:[10.1093/biomet/37.34.308](https://doi.org/10.1093/biomet/37.34.308)

Goldfeld SM, Quandt RE (1965). "Some tests for homoscedasticity." *Journal of the American Statistical Association*, **60**(310), 539-547. doi:[10.1080/01621459.1965.10480811](https://doi.org/10.1080/01621459.1965.10480811)

See Also

[kah.robust.test](#) for the outlier-resistant version, [run.all.het](#) to run it next to other tests.

Examples

```
# stopping distance of cars: the spread grows with speed
kah3.test(dist ~ speed, data = cars)

# the same from a fitted model
fit <- lm(dist ~ speed, data = cars)
kah3.test(fit)$critical
```

run.all.het

*Run the KaH Tests Next to Other Heteroscedasticity Tests***Description**

Runs the proposed tests and a set of established heteroscedasticity tests on the same model in one call, and returns one row per test. The established tests are provided for comparison and are attributed to their authors; they are not claimed as original to this package.

Usage

```
run.all.het(
  x,
  data = NULL,
  order.by = NULL,
  alpha = 0.75,
  include.optional = TRUE,
  bootstrap = FALSE,
  B = 499,
  seed = 1
)
```

Arguments

<code>x</code>	A fitted <code>lm</code> object, or a model formula.
<code>data</code>	A data frame, used when <code>x</code> is a formula.
<code>order.by</code>	The variable the data are sorted by before splitting: the name of a regressor or of a column of data, or a numeric vector of length n . Default: the first regressor.
<code>alpha</code>	LTS coverage: the fraction of each part that is kept (between 0.5 and 1). Lower values resist more outliers but cost power.
<code>include.optional</code>	Run the tests from skedastic if it is installed.
<code>bootstrap</code>	Also run the robust KaH test with the bootstrap reference (slower).
<code>B</code>	Number of Monte Carlo or bootstrap samples.
<code>seed</code>	Integer seed for the LTS subsampling and the resampling, or <code>NULL</code> to use the current random-number stream.

Details

Tests always run:

KaH-III [kah3.test](#), exact Hartley reference.

KaH robust [kah.robust.test](#) with `alpha` and the "fmax" reference.

Goldfeld-Quandt Goldfeld and Quandt (1965): data sorted by `order.by`, the middle third omitted, ratio of the residual mean squares of the outer thirds, two-sided F reference.

Breusch-Pagan (Koenker) Koenker's (1981) studentized version of the Breusch and Pagan (1979) test: nR^2 of the squared residuals on the regressors, χ_p^2 reference.

White White (1980): nR^2 of the squared residuals on the regressors, their squares and cross products.

MGQ (Rana et al.) The robust modified Goldfeld-Quandt test of Rana, Midi and Imon (2008): median squared deletion residuals of the outer thirds after an LTS outlier screen, F reference as proposed by the authors (two-sided).

Optional tests (include.optional = TRUE, need **skedastic**): Wilcox and Keselman (2006), Zhou, Song and Thompson (2015) and Li and Yao (2019). Optional: bootstrap = TRUE adds the robust KaH test with the residual-bootstrap reference.

One failing test never stops the run; its row gets NA and a note.

Value

A data frame of class "het_battery" with columns Test, Group, Statistic, df, p_value and Note. It prints grouped, with the p-values formatted.

References

- Breusch TS, Pagan AR (1979). "A simple test for heteroscedasticity and random coefficient variation." *Econometrica*, **47**(5), 1287-1294. doi:[10.2307/1911963](https://doi.org/10.2307/1911963)
- Goldfeld SM, Quandt RE (1965). "Some tests for homoscedasticity." *Journal of the American Statistical Association*, **60**(310), 539-547. doi:[10.1080/01621459.1965.10480811](https://doi.org/10.1080/01621459.1965.10480811)
- Koenker R (1981). "A note on studentizing a test for heteroscedasticity." *Journal of Econometrics*, **17**(1), 107-112. doi:[10.1016/03044076\(81\)900622](https://doi.org/10.1016/03044076(81)900622)
- Li Z, Yao J (2019). "Testing for heteroscedasticity in high-dimensional regressions." *Econometrics and Statistics*, **9**, 122-139. doi:[10.1016/j.ecosta.2018.01.001](https://doi.org/10.1016/j.ecosta.2018.01.001)
- Rana MS, Midi H, Imon AHMR (2008). "A robust modification of the Goldfeld-Quandt test for the detection of heteroscedasticity in the presence of outliers." *Journal of Mathematics and Statistics*, **4**(4), 277-283. doi:[10.3844/jmssp.2008.277.283](https://doi.org/10.3844/jmssp.2008.277.283)
- White H (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, **48**(4), 817-838. doi:[10.2307/1912934](https://doi.org/10.2307/1912934)
- Wilcox RR, Keselman HJ (2006). "Detecting heteroscedasticity in a simple regression model via quantile regression slopes." *Journal of Statistical Computation and Simulation*, **76**(8), 705-712. doi:[10.1080/10629360500107923](https://doi.org/10.1080/10629360500107923)
- Zhou QM, Song PX-K, Thompson ME (2015). "Profiling heteroscedasticity in linear regression models." *Canadian Journal of Statistics*, **43**(3), 358-377. doi:[10.1002/cjs.11252](https://doi.org/10.1002/cjs.11252)

See Also

[kah3.test](#), [kah.robust.test](#).

Examples

```
run.all.het(dist ~ speed, data = cars, include.optional = FALSE)
```

Index

`dfmax (fmax)`, [2](#)

`fmax`, [2](#)

`kah.nu.star`, [3](#), [5](#), [6](#)

`kah.robust.test`, [3](#), [4](#), [7–9](#)

`kah3.test`, [4](#), [6](#), [6](#), [8](#), [9](#)

`pfmax`, [7](#)

`pfmax (fmax)`, [2](#)

`qfmax (fmax)`, [2](#)

`rfmax (fmax)`, [2](#)

`run.all.het`, [6](#), [7](#), [8](#)