

Package ‘AutoSEARCH’

August 11, 2014

Type Package

Title General-to-Specific (GETS) Model Selection

Version 1.2

Date 2012-10-11

Author Genaro Sucarrat

Maintainer Genaro Sucarrat <genaro.sucarrat@bi.no>

Depends R (>= 2.4.1), zoo

Description Automated Multi-path General-to-Specific (GETS) model
selection of the mean and log-volatility specifications of a log-ARCH-X model

License GPL-2

LazyLoad yes

URL <http://www.sucarrat.net/>

Repository CRAN

Date/Publication 2012-10-11 18:58:30

NeedsCompilation no

R topics documented:

AutoSEARCH-package	2
eqwma	3
gedestp	4
gedlogl	5
gets.mean	6
gets.vol	10
gLag	14
gLog.ep	14
info.criterion	15

jb.test	16
leqhma	16
ols.fit1	18
ols.fit2	19
regs.mean.sm	20
regs.vol.sm	21
skewness.test	22
sm	22

Index	26
--------------	-----------

AutoSEARCH-package	<i>General-to-Specific (GETS) Model selection</i>
--------------------	---

Description

Automated multi-path General-to-Specific (GETS) model selection of the mean and log-volatility specification of a log-ARCH-X model.

The three main functions of the package are `sm`, `gets.mean` and `gets.vol`. `sm` estimates a model with an AR-X specification in the mean and a log-ARCH-X log-volatility specification. `gets.mean` undertakes General-to-Specific model selection of the mean specification, whereas `gets.vol` undertakes General-to-Specific model selection of the log-volatility specification.

The package also provides twelve auxiliary functions that are called by `sm`, `gets.mean` and `gets.vol`. They are documented very scarcely.

Details

Package:	AutoSEARCH
Type:	Package
Version:	1.2
Date:	2012-10-11
License:	GPL-2
LazyLoad:	yes

The code was developed in relation with the project G. Sucarrat and A. Escibano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', Oxford Bulletin of Economics and Statistics 74, Issue 5 (October), pp. 716-735. The model class in which model selection takes place is called SEARCH, which is an abbreviation of Stochastic Exponential ARCH. The acronym SEARCH also conveniently connotes the main motivation behind the model, namely that it is particularly suitable for specification SEARCH.

Author(s)

Genaro Sucarrat Maintainer: Genaro Sucarrat <genaro.sucarrat@bi.no>

References

G. Sucarrat and A. Escribano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', *Oxford Bulletin of Economics and Statistics* 74, Issue 5 (October), pp. 716-735

Examples

```
#Generate from AR(1) model:
set.seed(123)
y <- arima.sim(list(ar=0.4), 200)

#Estimate AR(2) with intercept as mean specification
#and log-ARCH(4) as log-volatility specification:
sm(y, mc=TRUE, ar=1:2, arch=1:4)

#General-to-Specific model selection of the mean:
mymodel <- gets.mean(y, mc=TRUE, ar=1:2, arch=1:4)

#General-to-Specific model selection of the
#simplified mean specification:
gets.vol(mymodel$resids, arch=1:4)
```

eqwma	<i>Equally Weighted Moving Average (EqWMA) of the pth. exponentiated values</i>
-------	---

Description

Returns an Equally Weighted Moving Average (EqWMA) of the pth. exponentiated values lagged. Optionally, the absolute values are computed before averaging, and the log of EqWMA can be returned.

Usage

```
eqwma(x, length=5, lag=1, start=1, p=1, log=FALSE, abs=FALSE)
```

Arguments

x	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the <code>na.trim</code> command from the <code>zoo</code> package
length	integer or vector of integers each equal to or greater than 1. The length or lengths of the moving window or windows of averages. Default: <code>length=20</code>
lag	integer equal to or greater than 0. If 0, then the moving averages are not lagged. Default: <code>lag=1</code>
start	integer equal to or greater than 1. Where to start the moving windows of averages. Default: <code>start=1</code> (the first observation)

p	numeric greater than zero. The exponent or p in x^p . Default: p=2
log	logical. If TRUE, then the logarithm of the moving average is returned. Default: log = FALSE
abs	logical. Whether x should be transformed to absolute values (TRUE) before averages are computed or not (FALSE). Default: abs = FALSE

Details

The function's intended primary use is to construct so-called 'mixed frequency regressors' for the mean. For the log-volatility specification the function `leqwma` is used.

Value

ewma	numeric vector, time series or zoo object (depending on the properties of x, the input) containing the values of the moving average
------	---

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

`leqwma`, `sm`, `gets.mean`, `gets.vol`

Examples

```
#Generate an iid normal series:
set.seed(123)
x <- rnorm(200)

#Compute lag of EqWMA(20) for x^2:
eqwma(x, p=2)

#Compute lag of log(EqWMA(30)) for abs(x)^1:
eqwma(x, length=30, p=1, log=TRUE, abs=TRUE)

#Compute lag of EqWMA(5) and lag of EqWMA(20) for x:
eqwma(x, length=c(5,20))
```

gedestp	<i>Estimates the shape parameter of a standardised Generalised Error Distribution (GED)</i>
---------	---

Description

The estimation method is based on an index of curtosis, and the code is based the `estimatep` function from the R package by Angelo M. Mineo called `normalp`

Usage

```
gedestp(x, method = c("inverse", "direct"))
```

Arguments

x	numeric vector
method	"inverse" or "direct"

Value

numeric, an estimate of the shape parameter

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

References

Mineo, A.M. (1994): 'Un nuovo metodo di stima di p per una corretta valutazione dei parametri di intensita e di scala di una curva normale di ordine p', Atti della XXXVII Riunione Scientifica della Societa Italiana di Statistica, San Remo, Vol. 2, pp. 147-154

See Also

gedlogl

Examples

```
#series of iid standard normal:
set.seed(123)
x <- rnorm(200)
gedestp(x)
```

gedlogl	<i>Log-likelihood of a standardised Generalised Error Distribution (GED)</i>
---------	--

Description

Compute the log-likelihood of a standardised (zero mean, unit variance) GED with shape parameter p

Usage

```
gedlogl(z, p = 2)
```

Arguments

z numeric vector
 p numeric value, the shape parameter

Value

numeric, the value of the log-likelihood

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

gedestp

Examples

```
#Log-likelihood of a Gaussian series:
set.seed(123)
x <- rnorm(200)
gedlogl(x, p=2)
```

gets.mean

General-to-Specific (GETS) model selection of the mean specification

Description

The starting model is referred to as the General Unrestricted Model (GUM), and may contain both a mean specification and a log-volatility specification. However, simplification is undertaken only with respect to the former. The diagnostic tests are optionally undertaken either on the residuals of the mean specification or on the standardised residuals. The latter might be desirable if a log-volatility specification is specified.

Usage

```
gets.mean(y, mc = NULL, ar = NULL, ewma = NULL, mx = NULL,
  arch = NULL, asym = NULL, log.ewma = NULL, vx = NULL,
  keep = NULL, p = 2, varcov.mat = c("ordinary", "white"),
  t.pval = 0.05, do.pet = TRUE, wald.pval = 0.05,
  ar.LjungB = c(2, 0.025), arch.LjungB = c(2, 0.025),
  tau = 2, info.method = c("sc", "aic", "hq"),
  info.resids = c("mean", "standardised"),
  include.empty = FALSE, zero.adj = 0.1, vc.adj = TRUE,
  tol = 1e-07, LAPACK = FALSE, max.regs = 1000,
  verbose = TRUE, smp1 = NULL, alarm = FALSE)
```

Arguments

y	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
mc	logical, TRUE or FALSE (default). TRUE includes intercept in the mean specification, FALSE does not
ar	integer vector, say, c(2,4) or 1:4. The AR-lags to include in the specification
ewma	either NULL (default) or a list with arguments sent to the eqwma function. In the latter case a lagged moving average of y is included as a regressor
mx	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
arch	integer vector, say, c(1,3) or 2:5. The log-ARCH terms to include in the log-volatility specification
asym	integer vector, say, c(1) or 1:3. The asymmetry or leverage terms to include in the log-volatility specification
log.ewma	NULL (default) or a list. If NULL then log(EWMA) is not included as volatility proxy. If a list, then log(EWMA) is included as a volatility proxy.
vx	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
keep	NULL (default) or an integer vector. If keep = NULL, then no regressors are excluded from removal. Otherwise, the regressors associated with the numbers in keep are excluded from the removal space. For example, keep=c(1) excludes the constant from removal. The regressor numbering is contained in the reg.no column of the gum.mean data frame (see below)
p	numeric value greater than zero. The power of the log-volatility specification.
varcov.mat	character vector, "ordinary" or "white". If "ordinary" then the ordinary variance-covariance matrix is used for inference. Otherwise the White (1980) heteroscedasticity robust matrix is used
t.pval	numeric value between 0 and 1. The significance level used for the two-sided regressor significance tests
do.pet	logical, TRUE (default) or FALSE. If TRUE then a Parsimonious Encompassing Test (PET) against the GUM is undertaken at each regressor removal for the joint significance of all the deleted regressors along the current path
wald.pval	numeric value between 0 and 1. The significance level used for the PETs
ar.LjungB	NULL or a two-element vector where the first element contains the order of a Ljung and Box (1979) test for serial correlation in the standardised residuals, and where the second element contains the significance level. If NULL, then the standardised residuals are not checked for serial correlation after each removal. The default is c(2, 0.025)
arch.LjungB	NULL or a two-element vector where the first element contains the order of a Ljung and Box (1979) test for ARCH (serial correlation in the squared standardised residuals), and where the second element contains the significance level. If

	NULL, then the standardised residuals are not checked for ARCH after each removal. The default is <code>c(2, 0.025)</code>
<code>tau</code>	NULL or a numeric value greater than 1. If NULL, then the shape parameter in a Generalised Error Distribution (GED) of the standardised residuals is estimated for the log-likelihood used in the calculation of the information criterion. If tau is equal to a numeric value, a GED(tau) is used. Default: <code>tau=2</code> (i.e. the standard normal density)
<code>info.method</code>	character string, "sc" (default), "aic" or "hq", which determines the information criterion used to select among terminal models. The abbreviations are short for the Schwarz or Bayesian information criterion (sc), the Akaike information criterion (aic) and the Hannan-Quinn (hq) information criterion
<code>info.resids</code>	character string, "mean" (default) or "standardised" which sets the residuals to be used in the computation of the information criterion
<code>include.empty</code>	logical, TRUE or FALSE (default). If TRUE then an empty model is included among the terminal models, if it passes the diagnostic tests, even if it is not equal to one of the terminals
<code>zero.adj</code>	numeric value between 0 and 1. The quantile adjustment for zero values. The default 0.1 means that the zero residuals are replaced by means of the 10 percent quantile of the absolute residuals before taking the logarithm
<code>vc.adj</code>	logical, TRUE (default) or FALSE. If true then the log-volatility constant is adjusted by means of the estimate of $E[\log(z^2)]$. This adjustment is needed for the standardised residuals to have unit variance. If FALSE then the log-volatility constant is not adjusted
<code>tol</code>	numeric value (default = $1e-07$). The tolerance for detecting linear dependencies in the columns of the regressors (see <code>qr()</code> function). Only used if LAPACK is FALSE
LAPACK	logical, TRUE or FALSE (default). If true use LAPACK otherwise use LINPACK (see <code>qr()</code> function)
<code>max.regs</code>	integer value, sets the maximum number of regressions along a deletion path. Default: <code>max.regs=1000</code>
<code>verbose</code>	logical, TRUE (default) or FALSE. FALSE returns less output and is therefore faster
<code>smpl</code>	Either NULL (default; the whole sample is used for estimation) or a two-element vector of dates with the start and end dates of the sample to be used in estimation. For example, <code>smpl=c("2001-01-01", "2009-12-31")</code>
<code>alarm</code>	Logical, either TRUE or FALSE (default). If TRUE, then a sound or beep is emitted when the specification search terminates in order to alert the user

Details

See Sucarrat and Escribano (2012)

Value

Returns a list with the following elements:

resids	zoo-object with the residuals of the final mean specification
resids.std	zoo-object with the standardised residuals
Elogzp	estimate of $E[\log(z^p)]$
call	the function call
gum.mean	a data frame with the estimation results of the GUM
gum.diagnostics	data frame with selected diagnostics of the GUM
keep	if any, the regressors that are excluded from deletion
insigs.in.gum	a numeric integer vector with the insignificant regressors of the GUM
paths	a list containing the simplification paths, that is, the sequences of deleted regressors
terminals	the distinct terminal models
terminals.results	the value and type of the information criterion (info) used in selecting among terminal specifications, and the number of observations (T) and parameters (k) used in the calculation of the information criterion
specific.mean	data frame with the estimation results of the final mean specification
specific.diagnostics	data frame with selected diagnostics of the standardised residuals

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

References

Genaro Sucarrat and Alvaro Escribano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', *Oxford Bulletin of Economics and Statistics* 74, Issue no. 5 (October), pp. 716-735

G. Ljung and G. Box (1979): 'On a Measure of Lack of Fit in Time Series Models'. *Biometrika* 66, pp. 265-270

See Also

sm, gets.vol, ewma

Examples

```
#Generate AR(1) model and four independent normal regressors:
set.seed(123)
y <- arima.sim(list(ar=0.4), 200)
xregs <- matrix(rnorm(4*200), 200, 4)

#General-to-Specific model selection of the mean:
mymodel <- gets.mean(y, mc=TRUE, ar=1:5, mx=xregs)

#General-to-Specific model selection of the mean
```

```

#with the intercept excluded from removal:
mymodel <- gets.mean(y, mc=TRUE, ar=1:5, mx=xregs, keep=1)

#General-to-Specific model selection of the mean
#with no intercept and with a log-ARCH(4) specification
#in the log-volatility using the standardised residuals
#when computing the log-likelihood for the information
#criterion:
mymodel <- gets.mean(y, mc=FALSE, ar=1:5, mx=xregs, arch=1:4,
  info.resids="standardised")

#General-to-Specific model selection of the mean with
#non-default serial-correlation diagnostics settings:
mymodel <- gets.mean(y, mc=TRUE, ar=1:5, mx=xregs,
  ar.LjungB=c(6, 0.05))

#General-to-Specific model selection of the mean with
#very liberal (i.e. 20 percent) significance levels (20 percent):
mymodel <- gets.mean(y, mc=TRUE, ar=1:5, mx=xregs, t.pval=0.2,
  wald.pval=0.2)

```

gets.vol

General-to-Specific (GETS) model selection of a log-volatility specification

Description

The starting model is referred to as the General Unrestricted Model (GUM) log-volatility specification, and takes the residuals from a mean specification as its modelling objective. But the modelling or model selection is undertaken in terms of the log-volatility specification. The diagnostic tests are undertaken on the standardised residuals (this is not optional)

Usage

```

gets.vol(e, arch=NULL, asym=NULL, log.ewma=NULL, vx=NULL,
  p=2, keep=c(1), t.pval=0.05, wald.pval=0.05, do.pet=TRUE,
  ar.LjungB=c(1, 0.025), arch.LjungB=c(1, 0.025), tau=2,
  info.method=c("sc", "aic", "hq"),
  info.resids=c("standardised", "log-sigma"), include.empty=FALSE,
  zero.adj=0.1, vc.adj=TRUE, tol=1e-07, LAPACK=FALSE, max.regs=1000,
  verbose=TRUE, alarm=FALSE, smp1=NULL)

```

Arguments

e	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
arch	integer vector, say, c(1,3) or 2:5. The ARCH-lags to include in the log-volatility specification

asym	integer vector, say, c(1) or 1:3. The asymmetry or leverage terms to include in the log-volatility specification
log.ewma	NULL (default) or a list. If NULL then log(EqWMA) is not included as volatility proxy. If a list, then log(EWMA) is included as a volatility proxy.
vx	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
p	numeric value greater than zero. The power of the log-volatility specification.
keep	NULL (default) or an integer vector. If keep = NULL, then no regressors are excluded from removal. Otherwise, the regressors associated with the numbers in keep are excluded from the removal space. For example, keep=c(1) excludes the constant from removal. The regressor numbering is contained in the reg.no column of the gum.mean data frame, see below
t.pval	numeric value between 0 and 1. The significance level used for the regressor tests
wald.pval	numeric value between 0 and 1. The significance level used for the PETs
do.pet	logical, TRUE (default) or FALSE. If TRUE then a Parsimonious Encompassing Test (PET) against the GUM is undertaken at each regressor removal for the joint significance of all the deleted regressors along the current path
ar.LjungB	NULL or a two-element vector where the first element contains the order of a Ljung and Box (1979) test for serial correlation in the standardised residuals, and where the second element contains the significance level. If NULL, then the standardised residuals are not checked for serial correlation after each removal. The default is c(2, 0.025)
arch.LjungB	NULL or a two-element vector where the first element contains the order of a Ljung and Box (1979) test for ARCH (serial correlation in the squared standardised residuals), and where the second element contains the significance level. If NULL, then the standardised residuals are not checked for ARCH after each removal. The default is c(2, 0.025)
tau	NULL or a numeric value greater than 1. If NULL, then the shape parameter in a Generalised Error Distribution (GED) of the standardised residuals is estimated for the log-likelihood used in the calculation of the information criterion. If tau is equal to a numeric value, a GED(tau) is used. Default: tau=2 (i.e. the standard normal density)
info.method	character string, "sc" (default), "aic" or "hq", which determines the information criterion used to select among terminal models. The abbreviations are short for the Schwarz or Bayesian information criterion (sc), the Akaike information criterion (aic) and the Hannan-Quinn (hq) information criterion
info.resids	character string, "mean" (default) or "standardised" which sets the residuals to be used in the computation of the information criterion
include.empty	logical, TRUE or FALSE (default). If TRUE then an empty model is included among the terminal models, if it passes the diagnostic tests, even if it is not equal to one of the terminals

zero.adj	numeric value between 0 and 1. The quantile adjustment for zero values. The default 0.1 means that the zero residuals are replaced by means of the 10 percent quantile of the absolute residuals before taking the logarithm
vc.adj	logical, TRUE (default) or FALSE. If true then the log-volatility constant is adjusted by means of the estimate of $E[\log(z^2)]$. This adjustment is needed for the standardised residuals to have unit variance. If FALSE then the log-volatility constant is not adjusted
tol	numeric value (default = $1e-07$). The tolerance for detecting linear dependencies in the columns of the regressors (see <code>qr()</code> function). Only used if LAPACK is FALSE
LAPACK	logical, TRUE or FALSE (default). If true use LAPACK otherwise use LINPACK (see <code>qr()</code> function)
max.regs	integer value, sets the maximum number of regressions along a deletion path. Default: max.regs=1000
verbose	logical, TRUE (default) or FALSE. FALSE returns less output and is therefore faster
alarm	Logical, either TRUE or FALSE (default). If TRUE, then a sound or beep is emitted when the specification search terminates in order to alert the user
smp1	Either NULL (default; the whole sample is used for estimation) or a two-element vector of dates with the start and end dates of the sample to be used in estimation. For example, <code>smp1=c("2001-01-01", "2009-12-31")</code>

Details

See Sucarrat and Escribano (2012)

Value

Returns a list with the following elements:

volatility.fit	zoo-object with the fitted values of the volatility (σ^p) of the final log-volatility specification
resids.ustar	zoo-object with the residuals of the AR-representation of the final log-volatility specification
resids.std	zoo-object with the standardised residuals of the final log-volatility specification
Elogzp	estimate of $E[\log(z^p)]$
call	the function call
gum.volatility	a data frame with the estimation results of the log-volatility GUM
gum.diagnostics	data frame with selected diagnostics of the log-volatility GUM
keep	if any, the regressors that are excluded from deletion
insigs.in.gum	a numeric integer vector with the insignificant regressors of the log-volatility GUM
paths	a list containing the simplification paths, that is, the sequences of deleted regressors

terminals the distinct terminal models
 terminals.results the value and type of the information criterion (info) used in selecting among terminal specifications, and the number of observations (T) and parameters (k) used in the calculation of the information criterion
 specific.volatility data frame with the estimation results of the final log-volatility specification
 specific.diagnostics data frame with selected diagnostics of the standardised residuals

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

References

Genaro Sucarrat and Alvaro Escribano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', *Oxford Bulletin of Economics and Statistics* 74, Issue no. 5 (October), pp. 716-735
 G. Ljung and G. Box (1979): 'On a Measure of Lack of Fit in Time Series Models'. *Biometrika* 66, pp. 265-270

See Also

sm, gets.mean, eqwma

Examples

```

#Generate iid normal residuals and a matrix of independent
#normals:
set.seed(123)
e <- rnorm(200)
xregs <- matrix(rnorm(4*200), 200, 4)

#General-to-Specific model selection of log-volatility:
mymodel <- gets.vol(e, arch=1:5, vx=log(xregs^2))

#General-to-Specific model selection of log-volatility
#with the log-ARCH(1) term excluded from removal:
mymodel <- gets.vol(e, arch=1:5, vx=log(xregs^2), keep=2)

#General-to-Specific model selection of log-volatility
#with all the log-ARCH terms excluded from removal:
mymodel <- gets.vol(e, arch=1:5, vx=log(xregs^2), asym=1:2,
  log.ewma=list(length=5), keep=2:6)

#If e is a daily (weekends excluded) financial return series,
#then the following specification includes a lagged volatility
#proxy both for the week (5-day average of squared return) and
#for the month (20-day average of squared returns), in addition
#to five log-ARCH terms:

```

```

mymodel <- gets.vol(e, arch=1:5, log.ewma=list(length=c(5,20)) )

#General-to-Specific model selection with very liberal
#(20 percent) significance levels:
mymodel <- gets.vol(e, arch=1:5, vx=log(xregs^2), t.pval=0.2,
  wald.pval=0.2)

```

gLag *Lag a series*

Description

Convenience function particularly suited for the lagging of zoo objects. By default the 'lost' values are padded with NAs so that the length of the series is retained. This is useful in creating the regressor matrices for the GETS model selection

Usage

```
gLag(y, k=1, na.value=NA)
```

Arguments

y	numeric vector, time-series or zoo object
k	integer equal to or greater than 1. Default: k=1
na.value	the value to replace the lost values with. Default: na.replace=NA

Value

the lagged vector, time series or zoo object

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

gLog.ep *Adjust for zero values and compute $\log(\text{abs}(e)^p)$*

Description

Adjusts a series called e - typically a series of residuals or financial returns - for zero values, so that the logarithm can be applied on the absolute pth. exponentiated values. Next, $\log(\text{abs}(e)^p)$ is computed

Usage

```
gLog.ep(e, zero.adj=0.1, p=2, na.replace=NA)
```

Arguments

e	numeric vector, time series or zoo object
zero.adj	numeric value between 0 and 1. The quantile adjustment for zero values. The default 0.1 means that the zero residuals are replaced by means of the 10 percent quantile of abs(e) before taking the logarithm
p	numeric value greater than zero. The power of the log-volatility specification.
na.replace	the value to replace NA values with. Default: na.replace=NA

Value

$\log(\text{abs}(e)^p)$, where the e values have been zero adjusted

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

References

Genaro Sucarrat and Alvaro Escribano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', Oxford Bulletin of Economics and Statistics 74, Issue no. 5 (October), pp. 716-735

info.criterion

Computes the the value of an information criterion

Description

Given a log-likelihood, the number of observations and the number of parameters, the value of a chosen information criterion is computed

Usage

```
info.criterion(logl, n=NULL, k=NULL, method=c("sc", "aic", "aicc", "hq"))
```

Arguments

logl	numeric, the value of the log-likelihood
n	integer, the number of observations
k	integer, the number of parameters
method	"sc", "aic", "aicc" or "hq"

Value

numeric, the value of the information criterion

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

jb.test	<i>Jarque-Bera test for normality</i>
---------	---------------------------------------

Description

Jarque-Bera test for normality

Usage

```
jb.test(x)
```

Arguments

x a numeric vector

Value

Returns a list with the following elements:

skewness	numeric, an estimate of the excess skewness relative to the normal
kurtosis	numeric, an estimate of the excess kurtosis relative to the normal
statistic	numeric, the Jarque-Bera test statistic
df	2, the degrees of freedom
p.value	numeric between 0 and 1, the p-value of the test under the null of normality

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

leqwma	<i>The logarithm of an Equally Weighted Moving Average (EqWMA) of the pth. exponentiated absolute values</i>
--------	--

Description

The default is that the logarithm of an Equally Weighted Moving Average (EqWMA) of the pth. exponentiated absolute values is returned. As a volatility model, this is also known as the 'historical' model or as an integrated ARCH model where the ARCH coefficients all have the same value with sum equal to one. In the log-volatility specification the lag of log(EqWMA) is thus a volatility proxy. It is an imperfect volatility proxy compared with high-frequency data (which can also be included as regressors anyway), but - in contrast to high-frequency data - is always available and easy to compute

Usage

```
leqwma(x, length = 5, lag = 1, start = 1, p = 2)
```

Arguments

x	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the <code>na.trim</code> command from the zoo package
length	integer or vector of integers each equal to or greater than 1. The length or lengths of the moving window or windows of averages. Default: length=20
lag	integer equal to or greater than 0. If 0, then the moving averages are not lagged. Default: lag=1
start	integer equal to or greater than 1. Where to start the moving windows of averages. Default: start=1 (the first observation)
p	numeric greater than zero. The exponent or p in $\text{abs}(x)^p$. Default: p=2

Details

The function's intended primary use is to construct volatility proxies in logarithms for inclusion in the log-volatility specification. The default is the lagged log of an equally weighted moving average of the squared residuals, where each average is made up of m observations. This is equivalent to an integrated ARCH(m) model where all the m ARCH-coefficients are equal. For further details on the use of log(EqWMA) as a volatility proxy, see Sucarrat and Escibano (2012)

Value

ewma	numeric vector, time series or zoo object (depending on the properties of x, the input) containing the values of the moving average
------	---

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

eqwma, sm, gets.mean, gets.vol

Examples

```
#Generate an iid normal series:
set.seed(123)
x <- rnorm(200)

#Compute lag of log(EqWMA(20)) for x^2:
leqwma(x, p=2)

#Compute lag of log(EqWMA(30)) for abs(x)^1:
eqwma(x, length=30, p=1)
```

```
#Compute lag of log(EqWMA(5)) and lag of log(EqWMA(20))
#for abs(x)^2:
eqwma(x, length=c(5,20))
```

ols.fit1

Fast and accurate OLS estimation by means of QR decomposition

Description

Fast and accurate OLS estimation by means of QR decomposition

Usage

```
ols.fit1(y, x, tol=1e-07, LAPACK=FALSE)
```

Arguments

y	numeric vector, the regressand
x	numeric matrix, the regressors
tol	numeric value (default = 1e-07). The tolerance for detecting linear dependencies in the columns of the regressors (see qr() function). Only used if LAPACK is FALSE
LAPACK	logical, TRUE or FALSE (default). If true use LAPACK otherwise use LINPACK (see qr() function)

Value

Returns a list with the following elements:

```
qr
rank
qraux
pivot
coefficients
```

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

ols.fit2, qr, solve

`ols.fit2`*Fast and accurate OLS estimation by means of QR decomposition*

Description

Returns slightly more information than `ols.fit1`, which makes the computation of variance-covariance matrices faster. However, if parameter estimation is the only objective, then `ols.fit1` is faster

Usage

```
ols.fit2(y, x, tol=1e-07, LAPACK=FALSE)
```

Arguments

<code>y</code>	numeric vector, the regressand
<code>x</code>	numeric matrix, the regressors
<code>tol</code>	numeric value (default = 1e-07). The tolerance for detecting linear dependencies in the columns of the regressors (see <code>qr()</code> function). Only used if <code>LAPACK</code> is <code>FALSE</code>
<code>LAPACK</code>	logical, <code>TRUE</code> or <code>FALSE</code> (default). If true use LAPACK otherwise use LINPACK (see <code>qr()</code> function)

Value

Returns a list with the following elements:

```
qr  
rank  
qraux  
pivot  
xtxinv  
xtx  
xty  
coefficients
```

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

`ols.fit1`, `qr`, `solve`

regs.mean.sm	<i>Creates the regressors of the mean specification of a SEARCH model</i>
--------------	---

Description

Creates the regressors of the mean specification of a SEARCH model

Usage

```
regs.mean.sm(y, mc=NULL, ar=NULL, ewma=NULL, mx=NULL)
```

Arguments

y	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
mc	logical, TRUE or FALSE (default). TRUE includes intercept in the specification, FALSE does not
ar	integer vector, say, c(2,4) or 1:4. The AR-lags to include in the specification
ewma	NULL or a list of arguments sent to the eqwma function
mx	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package

Value

Matrix with regressors

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

regs.vol.sm

regs.vol.sm	<i>Creates the regressors of the log-volatility specification</i>
-------------	---

Description

Creates the regressors of the log-volatility specification

Usage

```
regs.vol.sm(e, vc=TRUE, arch=NULL, asym=NULL, log.ewma=NULL, vx=NULL, p=2,  
zero.adj=0.1)
```

Arguments

e	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
vc	logical, TRUE (default) or FALSE. TRUE creates an intercept, FALSE does not
arch	integer vector, say, c(1,3) or 2:5. The ARCH-lags to include in the log-volatility specification
asym	integer vector, say, c(1) or 1:3. The asymmetry or leverage terms to include in the log-volatility specification
log.ewma	NULL (default) or a list. If NULL then log(EWMA) is not included as volatility proxy. If a list, then log(EWMA) is included as a volatility proxy.
vx	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the na.trim command from the zoo package
p	numeric value greater than zero. The power of the log-volatility specification.
zero.adj	numeric value between 0 and 1. The quantile adjustment for zero values. The default 0.1 means that the zero residuals are replaced by means of the 10 percent quantile of the absolute residuals before taking the logarithm

Value

Matrix with regressors

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

See Also

regs.mean.sm

skewness.test	<i>Chi-square test skewness in the standardised residuals</i>
---------------	---

Description

Chi-square test skewness in the standardised residuals

Usage

```
skewness.test(x)
```

Arguments

x	numeric vector
---	----------------

Value

Returns a list with the following elements:

statistic	the test statistic
p.value	numeric between 0 and 1, the p-value of the test under the null of no-skewness

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

sm	<i>Estimates a SEARCH model, that is, a model with an AR-X mean specification and a log-ARCH-X log-volatility specification</i>
----	---

Description

Estimates a SEARCH model by means of OLS. Estimation takes place in two steps. First the mean specification, then the log-volatility specification where the natural logarithm of the residuals constitutes the regressand. The mean specification can contain an intercept, AR-terms, lagged moving averages of the regressand and other conditioning covariates. The log-volatility specification can contain log-ARCH terms, asymmetry or 'leverage' terms, log(EqWMA) where EqWMA is a lagged equally weighted moving average of past p th. exponentiated residuals (a volatility proxy) and other conditioning covariates.

Usage

```
sm(y, mc=NULL, ar=NULL, ewma=NULL, mx=NULL, arch=NULL, asym=NULL,
  log.ewma=NULL, vx=NULL, p=2, zero.adj=0.1, vc.adj=TRUE,
  varcov.mat=c("ordinary", "white"), qstat.options=NULL,
  tol=1e-07, LAPACK=FALSE, verbose=TRUE, smpl=NULL)
```

Arguments

<code>y</code>	numeric vector, time-series or zoo object. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the <code>na.trim</code> command from the zoo package
<code>mc</code>	logical, TRUE or FALSE (default). TRUE includes intercept in the specification, FALSE does not
<code>ar</code>	integer vector, say, <code>c(2,4)</code> or <code>1:4</code> . The AR-lags to include in the specification
<code>ewma</code>	list of arguments sent to the <code>leqwma</code> function
<code>mx</code>	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the <code>na.trim</code> command from the zoo package
<code>arch</code>	integer vector, say, <code>c(1,3)</code> or <code>2:5</code> . The ARCH-lags to include in the log-volatility specification
<code>asym</code>	integer vector, say, <code>c(1)</code> or <code>1:3</code> . The asymmetry or leverage terms to include in the log-volatility specification
<code>log.ewma</code>	NULL (default) or a list. If NULL then <code>log(EWMA)</code> is not included as volatility proxy. If a list, then <code>log(EWMA)</code> is included as a volatility proxy.
<code>vx</code>	numeric matrix, time-series or zoo object of conditioning covariates. Note that missing values in the beginning or at the end of the series is allowed, as they are removed with the <code>na.trim</code> command from the zoo package
<code>p</code>	numeric value greater than zero. The power of the log-volatility specification.
<code>zero.adj</code>	numeric value between 0 and 1. The quantile adjustment for zero values. The default 0.1 means that the zero residuals are replaced by means of the 10 percent quantile of the absolute residuals before taking the logarithm
<code>vc.adj</code>	logical, TRUE (default) or FALSE. If true then the log-volatility constant is adjusted by means of the estimate of $E[\log(z^2)]$. This adjustment is needed for the standardised residuals to have unit variance. If FALSE then the log-volatility constant is not adjusted
<code>varcov.mat</code>	character vector, "ordinary" or "white". If "ordinary" then the ordinary variance-covariance matrix is used for inference. Otherwise the White (1980) heteroscedasticity robust matrix is used
<code>qstat.options</code>	NULL or an integer vector of length two, say, <code>c(2,5)</code> . The first value sets the order of the AR diagnostic test, whereas the second value sets the order of the ARCH diagnostic test. NULL (default) sets the vector to <code>c(1,1)</code>
<code>tol</code>	numeric value (default = $1e-07$). The tolerance for detecting linear dependencies in the columns of the regressors (see <code>qr()</code> function). Only used if LAPACK is FALSE
LAPACK	logical, TRUE or FALSE (default). If true use LAPACK otherwise use LINPACK (see <code>qr()</code> function)
<code>verbose</code>	logical, TRUE (default) or FALSE. FALSE returns less output and is therefore faster
<code>smp1</code>	Either NULL (default; the whole sample is used for estimation) or a two-element vector of dates with the start and end dates of the sample to be used in estimation. For example, <code>smp1=c("2001-01-01", "2009-12-31")</code>

Details

See Sucarrat and Escribano (2012)

Value

Returns a list with the following elements:

<code>call</code>	the function call
<code>mean.fit</code>	zoo-object with the fitted values of the mean specification
<code>resids</code>	zoo-object with the residuals of the mean specification
<code>volatility.fit</code>	zoo-object with the fitted values of the volatility (σ^p) specification
<code>resids.ustar</code>	zoo-object with the residuals of the AR-representation of the log-volatility specification
<code>resids.std</code>	zoo-object with the standardised residuals
<code>Elogzp</code>	estimate of $E(\log(z^p))$
<code>mean.results</code>	data frame with the estimation results of the mean specification
<code>volatility.results</code>	data frame with the estimation results of the log-volatility specification
<code>diagnostics</code>	data frame with selected diagnostics of the standardised residuals

Author(s)

Genaro Sucarrat (<http://www.sucarrat.net/>)

References

Genaro Sucarrat and Alvaro Escribano (2012): 'Automated Financial Model Selection: General-to-Specific Modelling of the Mean and Volatility Specifications', *Oxford Bulletin of Economics and Statistics* 74, Issue no. 5 (October), pp. 716-735

Halbert White (1980): 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica* 48, pp. 817-838

See Also

`gets.mean`, `gets.vol`, `ewma`

Examples

```
#Generate AR(1) model and independent normal regressors:
set.seed(123)
y <- arima.sim(list(ar=0.4), 200)
xregs <- matrix(rnorm(4*200), 200, 4)

#estimate AR(2) with intercept:
sm(y, mc=TRUE, ar=1:2)

#estimate AR(2) with intercept and four conditioning regressors
```



```
#in the mean:
sm(y, mc=TRUE, ar=1:2, mx=xregs)

#estimate a log-volatility specification with a log-ARCH(4)
#structure:
sm(y, arch=1:4)

#estimate a log-volatility specification with a log-ARCH(4)
#structure and an asymmetry or leverage term:
sm(y, arch=1:4, asym=1)

#estimate a log-volatility specification with a log-ARCH(4)
#structure, an asymmetry or leverage term, a 30-period log(EWMA) as
#volatility proxy, and the squareds of the conditioning regressors
#in the log-volatility specification:
sm(y, arch=1:4, asym=1, log.ewma=list(length=30), vx=log(xregs^2))

#estimate AR(2) with intercept and four conditioning regressors
#in the mean, and a log-volatility specification with a log-ARCH(4)
#structure, an asymmetry or leverage term, a 30-period log(EWMA) as
#volatility proxy, and the squareds of the conditioning regressors
#in the log-volatility specification:
sm(y, mc=TRUE, ar=1:2, mx=xregs, arch=1:4, asym=1,
    log.ewma=list(length=30), vx=log(xregs^2))
```

Index

AutoSEARCH (AutoSEARCH-package), [2](#)

AutoSEARCH-package, [2](#)

eqwma, [3](#)

gedestp, [4](#)

gedlogl, [5](#)

gets.mean, [6](#)

gets.vol, [10](#)

gLag, [14](#)

gLog.ep, [14](#)

info.criterion, [15](#)

jb.test, [16](#)

leqwma, [16](#)

ols.fit1, [18](#)

ols.fit2, [19](#)

regs.mean.sm, [20](#)

regs.vol.sm, [21](#)

skewness.test, [22](#)

sm, [22](#)