Package ‘ExtremeRisks’

August 27, 2020

Version 0.0.4
Date 2020-08-26
Title Extreme Risk Measures
Author Simone Padoan [cre, aut],
Gilles Stupfler [aut],
Bocconi Institute for Data Science and Analytics [fnd],
French National Research [fnd] (grant ANR-19-CE40-0013-01/ExtremReg)
Maintainer Simone Padoan <simone.padoan@unibocconi.it>
Imports evd, copula, mvtnorm, plot3D, tmvtnorm, pracma
Depends R (>= 3.5.0)
Description A set of procedures for estimating risks related to extreme events via risk measures such as Expectile, Value-at-Risk, etc. is provided. Estimation methods for univariate independent observations and temporal dependent observations are available. The methodology is extended to the case of independent multidimensional observations. The statistical inference is performed through parametric and non-parametric estimators. Inferential procedures such as confidence intervals, confidence regions and hypothesis testing are obtained by exploiting the asymptotic theory. Adapts the methodologies derived in Padoan and Stupfler (2020) <arxiv:2004.04078>, Padoan and Stupfler (2020) <arxiv:2007.08944>, Daouia et al. (2016) <doi:10.1007/978-3-319-48827-3>, de Haan et al. (2016) <doi:10.1007/s00780-015-0287-6>.
License GPL (>= 2)
URL mypage.unibocconi.it/simonepadoan/
NeedsCompilation yes
Repository CRAN
Repository/R-Forge/Project extremerisks
Repository/R-Forge/Revision 18
Repository/R-Forge/DateTimeStamp 2020-08-27 07:33:03
Date/Publication 2020-08-27 13:40:08 UTC
**R topics documented:**

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>dowjones</td>
<td>2</td>
</tr>
<tr>
<td>EBTailIndex</td>
<td>3</td>
</tr>
<tr>
<td>estExpectiles</td>
<td>4</td>
</tr>
<tr>
<td>estExtLevel</td>
<td>7</td>
</tr>
<tr>
<td>estMultiExpectiles</td>
<td>10</td>
</tr>
<tr>
<td>expectiles</td>
<td>13</td>
</tr>
<tr>
<td>ExpectMES</td>
<td>15</td>
</tr>
<tr>
<td>extMultiQuantile</td>
<td>18</td>
</tr>
<tr>
<td>extQuantile</td>
<td>21</td>
</tr>
<tr>
<td>HTailIndex</td>
<td>25</td>
</tr>
<tr>
<td>HypoTesting</td>
<td>27</td>
</tr>
<tr>
<td>MLTailIndex</td>
<td>31</td>
</tr>
<tr>
<td>MomTailIndex</td>
<td>33</td>
</tr>
<tr>
<td>MultiHTailIndex</td>
<td>35</td>
</tr>
<tr>
<td>predExpectiles</td>
<td>37</td>
</tr>
<tr>
<td>predMultiExpectiles</td>
<td>41</td>
</tr>
<tr>
<td>QuantMES</td>
<td>45</td>
</tr>
<tr>
<td>rbtimeseries</td>
<td>48</td>
</tr>
<tr>
<td>rmdata</td>
<td>50</td>
</tr>
<tr>
<td>rtimeseries</td>
<td>53</td>
</tr>
<tr>
<td>sp500</td>
<td>55</td>
</tr>
</tbody>
</table>

---

**Index**

<table>
<thead>
<tr>
<th>dowjones</th>
<th>Negative log-returns of DOW JONES.</th>
</tr>
</thead>
</table>

**Description**

Series of negative log-returns of the U.S. stock market index Dow Jones.

**Format**

A 8784 * 2 data frame.

**Details**

From the series of $n = 8785$ closing prices $S_t$, $t = 1, 2, ..., $ for the Dow Jones stock market index, recorded from January 29, 1985 to December 12, 2019, the series of negative log-returns:

$$X_{t+1} = -\log(S_{t+1}/S_t), \quad 1 \leq t \leq n - 1$$

is available. Hence the dataset (negative log-returns) contains 8784 observations.
Description

Computes a point estimate of the tail index based on the Expectile Based (EB) estimator.

Usage

EBTailIndex(data, tau, est=NULL)

Arguments

data A vector of \((1 \times n)\) observations.

tau A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.

est A real specifying the estimate of the expectile at the intermediate level tau.

Details

For a dataset \(\text{data}\) of sample size \(n\), the tail index \(\gamma\) of its (marginal) distribution is estimated using the EB estimator:

\[
\hat{\gamma}_{E} = \left(1 + \frac{\hat{F}_n(\tilde{\xi}_{\tau_n})}{1-\tau_n}\right)^{-1},
\]

where \(\hat{F}_n\) is the empirical survival function of the observations, \(\tilde{\xi}_{\tau_n}\) is an estimate of the \(\tau_n\)-th expectile. The observations can be either independent or temporal dependent. See Padoan and Stupfler (2020) and Daouia et al. (2018) for details.

- The so-called intermediate level \(\tau_n\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, \(\tau_n \in (0,1)\) is the ratio between the empirical mean distance of the \(\tau_n\)-th expectile from the smaller observations and the empirical mean distance of of the \(\tau_n\)-th expectile from all the observations. An estimate of \(\tau_n\)-th expectile is computed and used in turn to estimate \(\gamma\).

- The value \(\text{est}\), if provided, is meant to be an estimate of the \(\tau_n\)-th expectile which is used to estimate \(\gamma\). On the contrary, if \(\text{est}=NULL\), then the routine EBTailIndex estimate first the \(\tau_n\)-th expectile and then use it to estimate \(\gamma\).

Value

An estimate of the tail index \(\gamma\).

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;
References


See Also

HTailIndex, MomTailIndex, MLTailIndex,

Examples

# Tail index estimation based on the Expectile based estimator obtained with data
# simulated from an AR(1) with 1-dimensional Student-t distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <-  3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallblock <- 15

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.97

# sample size
ndata <- 2500

data <- rtimeseries(ndata, tsDist, tsType, par)

gammaHat <- EBTailIndex(data, tau)

gammaHat

---

estExpectiles

**High Expectile Estimation**

Description

Computes a point and interval estimate of the expectile at the intermediate level.
Usage

estExpectiles(data, tau, method="LAWS", tailest="Hill", var=FALSE, varType="asym-Dep-Adj", bigBlock=NULL, smallBlock=NULL, k=NULL, alpha=0.05)

Arguments

data A vector of \((1 \times n)\) observations.
tau A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.
method A string specifying the method used to estimate the expectile. By default est="LAWS" specifies the use of the direct LAWS estimator. See Details.
tailest A string specifying the type of tail index estimator. By default tailest="Hill" specifies the use of Hill estimator. See Details.
var If var=TRUE then an estimate of the variance of the expectile estimator is computed.
varType A string specifying the asymptotic variance to compute. By default varType="asym-Dep-Adj" specifies the variance estimator for serial dependent observations implemented with a suitable adjustment. See Details.
bigBlock An interger specifying the size of the big-block used to estiamte the asymptotic variance. See Details.
smallBlock An interger specifying the size of the small-block used to estiamte the asymptotic variance. See Details.
k An integer specifying the value of the intermediate sequence \(k_n\). See Details.
alpha A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence interval for the expectile at the intermediate level.

Details

For a dataset \(\text{data}\) of sample size \(n\), an estimate of the \(\tau_n\)-th expectile is computed. Two estimators are available: the so-called direct Least Asymmetrically Weighted Squares (LAWS) and indirect Quantile-Based (QB). The definition of the QB estimator depends on the estimation of the tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see HTailIndex) or in alternative using the the expectile based estimator (see EBTailIndex). The observations can be either independent or temporal dependent. See Section 3.1 in Padoan and Stupfler (2020) for details.

- The so-called intermediate level tau or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, \(\tau_n \in (0, 1)\) is the ratio between \(N\) (Numerator) and \(D\) (Denominator). Where \(N\) is the empirical mean distance of the \(\tau_n\)-th expectile from the observations smaller than it, and \(D\) is the empirical mean distance of \(\tau_n\)-th expectile from all the observations.
- If method='LAWS', then the expectile at the intermediate level \(\tau_n\) is estimated applying the direct LAWS estimator. Instead, If method='QB' the indirect QB estimator is used to estimate the expectile. See Section 3.1 in Padoan and Stupfler (2020) for details.
- When the expectile is estimated by the indirect QB estimator (method='QB'), an estimate of the tail index \(\gamma\) is needed. If tailest='Hill' then \(\gamma\) is estimated using the Hill estimator (see also HTailIndex). If tailest='ExpBased' then \(\gamma\) is estimated using the expectile based estimator (see EBTailIndex). See Section 3.1 in Padoan and Stupfler (2020) for details.
- $k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, when method='LAWs' and tau=NULL, $k_n$ specifies by $\tau_n = 1 - k_n/n$ the intermediate level of the expectile. Instead, when method='QB', if tailest="Hill" then the value $k_n$ specifies the number of $k+1$ larger order statistics to be used to estimate $\gamma$ by the Hill estimator and if tau=NULL then it also specifies by $\tau_n = 1 - k_n/n$ the confidence level $\tau_n$ of the quantile to estimate. Finally, if tailest="ExpBased" and tau=NULL then it also specifies by $\tau_n = 1 - k_n/n$ the intermediate level expectile based estimator of $\gamma$ (see EBTailIndex).

- If var=TRUE then the asymptotic variance of the expectile estimator is computed. With independent observations the asymptotic variance is computed by the formula Theorem 3.1 of Padoan and Stupfler (2020). This is achieved through varType="asym-Ind". With serial dependent observations the asymptotic variance is estimated by the formula in Theorem 3.1 of Padoan and Stupfler (2020). This is achieved through varType="asym-Dep". In this latter case the computation of the asymptotic variance is based on the "big blocks seperated by small blocks" technique which is a standard tools in time series, see Leadbetter et al. (1986). See also Section C.1 in Appendix of Padoan and Stupfler (2020). The size of the big and small blocks are specified by the parameters bigblock and smallblock, respectively. Still with serial dependent observations, If varType="asym-Dep-Adj", then the asymptotic variance is estimated using formula (C.79) in Padoan and Stupfler (2020), see Section C.1 of the Appendix for details.

- Given a small value $\alpha \in (0, 1)$ then an asymptotic confidence interval for the $\tau_n$-th expectile, with approximate nominal confidence level $(1 - \alpha)100\%$ is computed. See Sections 3.1 and C.1 in the Appendix of Padoan and Stupfler (2020).

**Value**

A list with elements:

- ExpctHat: a point estimate of the $\tau_n$-th expectile;
- VarExpHat: an estimate of the asymptotic variance of the expectile estimator;
- CIExpct: an estimate of the approximate $(1 - \alpha)100\%$ confidence interval for $\tau_n$-th expectile.

**Author(s)**

Simone Padoan, <simone.padoan@unibocconi.it>, [http://mypage.unibocconi.it/simonepadoan/](http://mypage.unibocconi.it/simonepadoan/);

**References**


**See Also**

HTailIndex, EBTailIndex, predExpectiles, extQuantile
Examples

# Extreme expectile estimation at the intermediate level tau obtained with
# 1-dimensional data simulated from an AR(1) with Student-t innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big-small-blocks setting
bigBlock <- 65
smallBlock <- 15

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.99

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# High expectile (intermediate level) estimation
expectHat <- estExpectiles(data, tau, var=TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
expectHat$ExpctHat
expectHat$CIExpct

---

estExtLevel

**Extreme Level Estimation**

**Description**

Estimates the expectile’s extreme level corresponding to a quantile’s extreme level.

**Usage**

```
estExtLevel(alpha_n, data=NULL, gammaHat=NULL, VarGamHat=NULL, tailest="Hill", k=NULL, var=FALSE, varType="asym-Dep", bigBlock=NULL, smallBlock=NULL, alpha=0.05)
```

**Arguments**

- `alpha_n`: A real in (0,1) specifying the extreme level $\alpha_n$ for the quantile. See Details.
- `data`: A vector of (1 x n) observations to be used to estimate the tail index in the case it is not provided. By default data=NULL specifies that no data are given.
- `gammaHat`: A real specifying an estimate of the tail index. By default gammaHat=NULL specifies that no estimate is given. See Details.
estExtLevel

VarGamHat  A real in specifying an estimate of the variance of the tail index estimator. By default VarGamHat=NULL specifies that no estimate is given. See Details.

tailest  A string specifying the type of tail index estimator to be used. By default tailest="Hill" specifies the use of Hill estimator. See Details.

k  An integer specifying the value of the intermediate sequence \( k_n \). See Details.

var  If var=TRUE then an estimate of the variance of the extreme level estimator is computed.

varType  A string specifying the asymptotic variance to compute. By default varType="asym-Dep" specifies the variance estimator for serial dependent observations. See Details.

bigBlock  An integer specifying the size of the big-block used to estimate the asymptotic variance. See Details.

smallBlock  An integer specifying the size of the small-block used to estimate the asymptotic variance. See Details.

alpha  A real in \((0,1)\) specifying the confidence level \((1-\alpha)100\%\) of the approximate confidence interval for the expecile at the intermediate level.

Details

For a given extreme level \( \alpha_n \) for the \( \alpha_n \)-th quantile, an estimate of the extreme level \( \tau'_n(\alpha_n) \) is computed such that \( \xi_{\tau'_n(\alpha_n)} = q_{\alpha_n} \). The estimator is defined by

\[
\hat{\tau}'_n(\alpha_n) = 1 - (1 - \alpha_n) \frac{\hat{\gamma}_n}{\hat{\gamma}_n}
\]

where \( \hat{\gamma}_n \) is a consistent estimator of the tail index \( \gamma \). If a value for the parameter gammaHat is given, then such a value is used to compute \( \hat{\tau}'_n \). If gammaHat is NULL and a dataset is provided through the parameter data, then the index \( \gamma \) is estimated by a suitable estimator \( \hat{\gamma}_n \). See Section 6 in Padoan and Stupfler (2020) for more details.

- If VarGamHat is specified, i.e. the variance of the tail index estimator, then the variance of the extreme level estimator \( \hat{\tau}'_n \) is computed by using such value.
- When estimating the tail index, if tailest='Hill' then \( \gamma \) is estimated using the Hill estimator (see also HTailIndex). If tailest='ML' then \( \gamma \) is estimated using the Maximum Likelihood estimator (see MLTailIndex). If tailest='ExpBased' then \( \gamma \) is estimated using the expectation based estimator (see EBTailIndex). If tailest='Moment' then \( \gamma \) is estimated using the moment based estimator (see MomentTailIndex). See Padoan and Stupfler (2020) for details.
- \( k \) or \( k_n \) is the value of the so-called intermediate sequence \( k_n, n = 1, 2, \ldots \). Its represents a sequence of positive integers such that \( k_n \to \infty \) and \( k_n/n \to 0 \) as \( n \to \infty \). Practically, when tailest='Hill' then the value \( k_n \) specifies the number of \( k+1 \) larger order statistics to be used to estimate \( \gamma \) by the Hill estimator. See MLTailIndex, EBTailIndex and MomentTailIndex for the other estimators.
- If var=TRUE then the asymptotic variance of the extreme level estimator is computed by applying the delta method, i.e.

\[
\text{Var}(\hat{\tau}'_n) = \text{Var}(\hat{\gamma}_n) + (\alpha_n - 1)^2/(1 - \hat{\gamma}_n)^4
\]

where \( \text{Var}(\hat{\gamma}_n) \) is provided by VarGamHat or is estimated when estimating the tail index through tailest='Hill' and tailest='ML'. See HTailIndex and MLTailIndex for details on how the variance is computed.
- Given a small value \( \alpha \in (0,1) \) then an asymptotic confidence interval for the extreme level, \( \tau'_n(\alpha_n) \), with approximate nominal confidence level \((1-\alpha)100\%\) is computed.
Value

A list with elements:

- \( \text{tauHat} \): an estimate of the extreme level \( \tau_n' \);
- \( \text{tauVar} \): an estimate of the asymptotic variance of the extreme level estimator \( \hat{\tau}_n'(\alpha_n) \);
- \( \text{tauCI} \): an estimate of the approximate \((1 - \alpha)100\%\) confidence interval for the extreme level \( \tau_n'(\alpha_n) \).

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, [http://mypage.unibocconi.it/simonepadoan/](http://mypage.unibocconi.it/simonepadoan/);

References


See Also

estExpectiles, predExpectiles, extQuantile

Examples

```r
# Extreme level estimation for a given quantile's extreme level alpha_n
# obtained with 1-dimensional data simulated from an AR(1) with Student-t innovations

tsDist <- "studentT"
tstype <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# quantile's extreme level
alpha_n <- 0.999

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)
```
# expectile's extreme level estimation
tau1Hat <- estExtLevel(alpha_n, data, var=TRUE, k=150, bigBlock=bigBlock, 
smallBlock=smallBlock)
tau1Hat

---

**estMultiExpectiles**  
**Multidimensional High Expectile Estimation**

**Description**

Computes point estimates and \((1 - \alpha)100\%\) confidence regions for \(d\)-dimensional expectiles at the intermediate level.

**Usage**

```r
estMultiExpectiles(data, tau, method="LAWS", tailest="Hill", var=FALSE, 
varType="asym-Ind-Adj", k=NULL, alpha=0.05, plot=FALSE)
```

**Arguments**

- `data`  
  A matrix of \((n \times d)\) observations.

- `tau`  
  A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See **Details**.

- `method`  
  A string specifying the method used to estimate the expectile. By default `est="LAWS"` specifies the use of the direct LAWS estimator. See **Details**.

- `tailest`  
  A string specifying the type of tail index estimator. By default `tailest="Hill"` specifies the use of Hill estimator. See **Details**.

- `var`  
  If `var=TRUE` then an estimate of the variance of the expectile estimator is computed.

- `varType`  
  A string specifying the asymptotic variance-covariance matrix to compute. By default `varType="asym-Ind-Adj"` specifies that the variance-covariance matrix is computed assuming dependent variables and exploiting a suitable adjustment. See **Details**.

- `k`  
  An integer specifying the value of the intermediate sequence \(k_n\). See **Details**.

- `alpha`  
  A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence region for the \(d\)-dimensional expectile at the intermediate level.

- `plot`  
  A logical value. By default `plot=FALSE` specifies that no graphical representation of the estimates is not provided. See **Details**.

**Details**

For a dataset `data` of \(d\)-dimensional observations and sample size \(n\), an estimate of the \(\tau_n\)-th \(d\)-dimensional is computed. Two estimators are available: the so-called direct Least Asymmetrically Weighted Squares (LAWS) and indirect Quantile-Based (QB). The QB estimator depends on the estimation of the \(d\)-dimensional tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimator (see `MultiHTailIndex`). The data are regarded as \(d\)-dimensional temporal independent observations coming from dependent variables. See Padoan and Stupfler (2020) for details.
• The so-called intermediate level \( \tau_n \) or \( \tau_n \) is a sequence of positive reals such that \( \tau_n \to 1 \) as \( n \to \infty \). Practically, for each individual marginal distribution \( \tau_n \in (0, 1) \) is the ratio between \( N \) (Numerator) and \( D \) (Denominator). Where \( N \) is the empirical mean distance of the \( \tau_n \)-th expectile from the observations smaller than it, and \( D \) is the empirical mean distance of \( \tau_n \)-th expectile from all the observations.

• If \texttt{method=’LAWS’}, then the expectile at the intermediate level \( \tau_n \) is estimated applying the direct LAWS estimator. Instead, If \texttt{method=’QB’} the indirect QB esimtator is used to estimate the expectile. See Section 2.1 in Padoan and Stupfler (2020) for details.

• When the expectile is estimated by the indirect QB esimtator (\texttt{method=’QB’}), an estimate of the \( d \)-dimensional tail index \( \gamma \) is needed. Here the \( d \)-dimensional tail index \( \gamma \) is estimated using the \( d \)-dimensional Hill estimator (\texttt{tau=’Hill’}, see \texttt{MultiHTailIndex}). This is the only available option so far (soon more results will be available).

• \( k \) or \( k_n \) is the value of the so-called intermediate sequence \( k_n, n = 1, 2, \ldots \). Its represents a sequence of positive integers such that \( k_n \to \infty \) and \( k_n/n \to 0 \) as \( n \to \infty \). Practically, for each marginal distribution, when \texttt{method=’LAWS’} and \texttt{tau=NULL}, \( k_n \) specifies by \( \tau_n = 1 - k_n/n \) the intermediate level of the expectile. Instead, for each marginal distribution, when \texttt{method=’QB’}, then the value \( k_n \) specifies the number of \( k+1 \) larger order statistics to be used to estimate \( \gamma \) by the Hill estimator and if \texttt{tau=NULL} then it also specifies by \( \tau_n = 1 - k_n/n \) the confidence level \( \tau_n \) of the quantile to estimate.

• If \texttt{var=TRUE} then an estimate of the asymptotic variance-covariance matrix of the \( d \)-dimensional expectile estimator is computed. If the data are regarded as \( d \)-dimensional temporal independent observations coming from dependent variables. Then, the asymptotic variance-covariance matrix is estimated by the formulas in section 3.1 of Padoan and Stupfler (2020). In particular, the variance-covariance matrix is computed exploiting the asymptotic behaviour of the relative expectile estimator appropriately normalized and using a suitable adjustment. This is achieved through \texttt{varType=’asym-Ind-Adj’}. The data can also be regarded as coded-dimensional temporal independent observations coming from independent variables. In this case the asymptotic variance-covariance matrix is diagonal and is also computed exploiting the formulas in section 3.1 of Padoan and Stupfler (2020). This is achieved through \texttt{varType=’asym-Ind’}.

• Given a small value \( \alpha \in (0, 1) \) then an asymptotic confidence region for the \( \tau_n \)-th expectile, with approximate nominal confidence level \((1 - \alpha)100\%\) is computed. In particular, a "symmetric" confidence regions is computed exploiting the asymptotic behaviour of the relative expectile estimator appropriately normalized. See Sections 3.1 of Padoan and Stupfler (2020) for detailed.

• If \texttt{plot=TRUE} then a graphical representation of the estimates is not provided.

Value

A list with elements:

• \texttt{ExpctHat}: an point estimate of the \( \tau_n \)-th \( d \)-dimensional expecile;
• \texttt{biasTerm}: an point estimate of the bias term of the estimated expecile;
• \texttt{VarCovEHat}: an estimate of the asymptotic variance of the expectile estimator;
• \texttt{EstConReg}: an estimate of the approximate \((1 - \alpha)100\%\) confidence region for \( \tau_n \)-th \( d \)-dimensional expecile.
Author(s)
Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/

References

See Also
MultiHTailIndex, predMultiExpectiles, extMultiQuantile

Examples
# Extreme expectile estimation at the intermediate level tau obtained with
# d-dimensional observations simulated from a joint distribution with
# a Gumbel copula and equal Frechet marginal distributions.
library(plot3D)
library(copula)
library(evd)

# distributional setting
copula <- "Gumbel"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Intermediate level (or sample tail probability 1-tau)
tau <- .95

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet
# marginals distributions and a Gumbel copula
data <- rmdata(ndata, dist, copula, par)
scatter3D(data[,1], data[,2], data[,3])

# High d-dimensional expectile (intermediate level) estimation
expectHat <- estMultiExpectiles(data, tau, var=TRUE)
expectHat$ExpctHat
expectHat$VarCovEHat
# run the following command to see the graphical representation
expectHat <- estMultiExpectiles(data, tau, var=TRUE, plot=TRUE)
expectiles

Expectile Computation

Description
Computes the true expectile for some families of parametric models.

Usage

```r
expectiles(par, tau, tsDist="gPareto", tsType="IID", trueMethod="true",
estMethod="LAWS", nrep=1e+05, ndata=1e+06, burnin=1e+03)
```

Arguments

- `par`: A vector of \((1 \times p)\) parameters of the time series parametric family. See Details.
- `tau`: A real in \((0, 1)\) specifying the level \(\tau\) of the expectile to be computed. See Details.
- `tsDist`: A string specifying the parametric family of the innovations distribution. By default `tsDist="gPareto"` specifies a Pareto family of distributions. See Details.
- `tsType`: A string specifying the type of time series. By default `tsType="IID"` specifies a sequence of independent and identically distributed random variables. See Details.
- `trueMethod`: A string specifying the method used to computed the expectile. By default `trueMethod="true"` specifies that the true analytical expression to computed the expectile is used. See Details.
- `estMethod`: A string specifying the method used to estimate the expectile. By default `estMethod="LAWS"` specifies the use of the direct LAWS estimator. See Details.
- `nrep`: A positive integer specifying the number of simulations to use for computing an approximation of the expectile. See Details.
- `ndata`: A positive integer specifying the number of observations to generated for each simulation. See Details.
- `burnin`: A positive integer specifying the number of initial observations to discard from the simulated sample.

Details
For a parametric family of time series models or a parametric family of distributions (for the case of independent observations) the \(\tau\)-th expectile (or expectile of level \(\tau\)) is computed.

- There are two methods to compute the \(\tau\)-th expectile. For the Generalised Pareto and Student-
\(t\) parametric families of distributions, the analytical expression of the expectile is available. This is used to compute the \(\tau\)-th expectile if the parameter `trueMethod="true"` is specified. For most of parametric family of distributions or parametric families of time series models the analytical expression of the expectile is not available. In this case an approximate value of the \(\tau\)-th expectile is computed via a Monte Carlo method if the parameter
trueMethod="approx" is specified. In particular, ndata observations from a family of
time series models (e.g. tsType="AR" and tsDist="studentT") or a sequence of inde-
pendent and identically distributed random variables with common family of distributions
(e.g. tsType="IID" and tsDist="gPareto") are simulated nrep times. For each simula-
tion the \( \tau \)-th expectile is estimate by the estimation method specified by estMethod. The
mean of such estimate provides an approximate value of the \( \tau \)-th expectile. The available
estimator to estimate the expcile are the direct LAWS (estMethod="LAWS") and the indirect
QB (estMethod="QB"), see estExpectiles for details. The available families of distributions
are: Generalised Pareto (tsDist="gPareto"), Student-t (tsDist="studentT") and Frechet
(tsDist="Frechet"). The available classes of time series with parametric innovations fami-
lies of distributions are specified in rtimeseries.

Value

The \( \tau \)-th expectile.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

rtimeseries

Examples

# Derivation of the true tau-th expectile for the Pareto distribution
# via accurate simulation

# parameter value
par <- c(1, 0.3)

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.99

trueExp <- expectiles(par, tau)
trueExp

# tau-th expectile of the AR(1) with Student-t innovations
tsDist <- "studentT"
tsType <- "AR"

# Approximation via Monte Carlo methods
trueMethod <- "approx"
# ExpectMES

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.99

trueExp <- expectiles(par, tau, tsDist, tsType, trueMethod)
trueExp

---

## ExpectMES

*Marginal Expected Shortfall Expectile Based Estimation*

### Description

Computes a point and interval estimate of the Marginal Expected Shortfall (MES) using an expectile based approach.

### Usage

```
ExpectMES(data, tau, tau1, method="LAWS", var=FALSE, varType="asym-Dep", bias=FALSE, 
bigBlock=NULL, smallBlock=NULL, k=NULL, alpha_n=NULL, alpha=0.05)
```

### Arguments

- **data**: A vector of $(1 \times n)$ observations.
- **tau**: A real in $(0, 1)$ specifying the intermediate level $\tau_n$. See **Details**.
- **tau1**: A real in $(0, 1)$ specifying the extreme level $\tau'_n$. See **Details**.
- **method**: A string specifying the method used to estimate the expecile. By default est="LAWS" specifies the use of the LAWS based estimator. See **Details**.
- **var**: If var=TRUE then an estimate of the asymptotic variance of the MES estimator is computed.
- **varType**: A string specifying the type of asymptotic variance to compute. By default varType="asym-Dep" specifies the variance estimator for serial dependent observations. See **Details**.
- **bias**: A logical value. By default bias=FALSE specifies that no bias correction is computed. See **Details**.
- **bigBlock**: An integer specifying the size of the big-block used to estimate the asymptotic variance. See **Details**.
- **smallBlock**: An integer specifying the size of the small-block used to estimate the asymptotic variance. See **Details**.
- **k**: An integer specifying the value of the intermediate sequence $k_n$. See **Details**.
alpha_n  A real in (0, 1) specifying the quantile’s extreme level to be use in order to estimate the expectile’s extreme level.

alpha  A real in (0, 1) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence interval for the expecile at the intermediate level.

Details

For a dataset data of sample size \(n\), an estimate of the \(\tau_n\)-th MES is computed. The estimation of the MES at the extreme level \(\tau_1\) \((\tau_n')\) is indeed meant to be a prediction. Two estimators are available: the so-called Least Asymmetrically Weighted Squares (LAWS) based estimator and the Quantile-Based (QB) estimator. The definition of both estimators depends on the estimation of the tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see HTailIndex for details). The observations can be either independent or temporal dependent. See Section 4 in Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(\tau\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). See predExpectiles for details.
- The so-called extreme level \(\tau_1\) or \(\tau_n'\) is a sequence of positive reals such that \(\tau_n' \to 1\) as \(n \to \infty\). See predExpectiles for details.
- When \(method='\text{LAWS}'\), then the \(\tau_n'\)-th MES is estimated using the LAWS based estimator. When \(method='\text{QB}'\), the expectile is instead estimated using the QB estimator. See Sectio 4 in Padoan and Stupfler (2020) and in particular Corollary 4.3 and 4.4 for details. The definition of both estimators depend on the estimation of the tail index \(\gamma\). In particular, the tail index \(\gamma\) is estimated using the Hill estimator (see HTailIndex).
- If \(var=\text{TRUE}\) then an estimate of the asymptotic variance of the \(\tau_n'\)-th MES is computed. Notice that the estimation of the asymptotic variance is only available when \(\gamma\) is estimated using the Hill estimator (see HTailIndex). With independent observations the asymptotic variance is estimated by \(\hat{\gamma}^2\), see Corollary 4.3 in Padoan and Stupfler (2020). This is achieved through \(varType='\text{asym-Ind}'\). With serial dependent observations the asymptotic variance is estimated by the formula in Corollary 4.3 of Padoan and Stupfler (2020). This is achieved through \(varType='\text{asym-Dep}'\). See Section 4 adn 5 in Padoan and Stupfler (2020) for details. In this latter case the computation of the serial dependence is based on the "big blocks seperated by small blocks" technique which is a standard tools in time series, see e.g. Leadbetter et al. (1986). The size of the big and small blocks are specified by the parameters bigBlock and smallBlock, respectively.
- If \(bias=\text{TRUE}\) then \(\gamma\) is estimated using formula (4.2) of Haan et al. (2016). This is used by the LAWS and QB estimators. Furthermore, the \(\tau_n'-th\) quantile is estimated using the formula in page 330 of de Haan et al. (2016). This provides a bias corrected version of the Weissman estimator. This is used by the QB estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity the asymptotic variance is estimated by the formula in Corollary 3.8, with serial dependent observations, and \(\hat{\gamma}^2\) with independent observation (see e.g. de Drees 2000, for the details).
- \(k\) or \(k_n\) is the value of the so-called intermediate sequence \(k_n, n = 1, 2, \ldots\). It represents a sequence of positive integers such that \(k_n \to \infty\) and \(k_n/n \to 0\) as \(n \to \infty\). Practically, when \(tau=NULL\) and \(method='\text{LAWS}'\), then \(\tau_n = 1 - k_n/n\) is the intermediate level of the expectile to be stimated. \(k_n\) also specifies the number of \(k+1\) larger order statistics used in the definition of the Hill estimator (see HTailIndex for detail). Differently, When \(tau=NULL\) and \(method='\text{QB}'\), then \(\tau_n = 1 - k_n/n\) is the intermediate level of the quantile to be stimated.
• If the quantile’s extreme level is provided by \( \alpha_n \), then expectile’s extreme level \( \tau'_n \) is replaced by \( \tau'_n(\alpha_n) \) which is estimated by the method described in Section 6 of Padoan and Stupfler (2020). See \texttt{estExtLevel} for details.

• Given a small value \( \alpha \in (0, 1) \) then an estimate of an asymptotic confidence interval for \( \tau'_n \)-th expectile, with approximate nominal confidence level \( (1 - \alpha)100\% \), is computed. The confidence intervals are computed exploiting formula in Corollary 4.3, 4.4 and Theorem 6.2 of Padoan and Stupfler (2020) and (46) in Drees (2003). See Sections 4-6 in Padoan and Stupfler (2020) for details. When \( \texttt{bias=TRUE} \) confidence intervals are computed in the same way but after correcting the tail index estimate by an estimate of the bias term, see formula (4.2) in de Haan et al. (2016) for details.

Value

A list with elements:

- \( \texttt{HatXMES} \): an estimate of the \( \tau'_n \)-th expectile based MES;

- \( \texttt{VarHatXMES} \): an estimate of the asymptotic variance of the expectile based MES estimator;

- \( \texttt{CIHatXMES} \): an estimate of the approximate \( (1 - \alpha)100\% \) confidence interval for \( \tau'_n \)-th MES.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

\texttt{QuantMES}, \texttt{HTailIndex}, \texttt{predExpectiles}, \texttt{extQuantile}
Examples

# Marginl Expected Shortfall expectile based estimation at the extreme level
# obtained with 2-dimensional data simulated from an AR(1) with bivariate
# Student-t distributed innovations

tsDist <- "AStudentT"
tsType <- "AR"
tsCopula <- "studentT"

# parameter setting
corr <- 0.8
dep <- 0.8
df <- 3
par <- list(corr=corr, dep=dep, df=df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# quantile's extreme level
alpha_n <- 0.999

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rbtimeseries(ndata, tsDist, tsType, tsCopula, par)

# Extreme MES expectile based estimation
MESHat <- ExpectMES(data, NULL, NULL, var=TRUE, k=150, bigBlock=bigBlock,
        smallBlock=smallBlock, alpha_n=alpha_n)

MESHat

---

extMultiQuantile  Multidimensional Value-at-Risk (VaR) or Extreme Quantile (EQ) Estimation

Description

Computes point estimates and $(1 - \alpha)100\%$ confidence regions for d-dimensional VaR based on the Weissman estimator.

Usage

extMultiQuantile(data, tau, tau1, var=FALSE, varType="asym-Ind-Log", bias=FALSE, k=NULL, alpha=0.05, plot=FALSE)
Arguments

- **data** A matrix of \((n \times d)\) observations.
- **tau** A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.
- **tau1** A real in \((0, 1)\) specifying the extreme level \(\tau'_n\). See Details.
- **var** If \(\text{var}=\text{TRUE}\) then an estimate of the asymptotic variance-covariance matrix of the \(d\)-dimensional VaR estimator is computed.
- **varType** A string specifying the type of asymptotic variance-covariance matrix to compute. By default \(\text{varType}=\text{"asym-Ind-Log"}\) specifies that the variance-covariance matrix is obtained assuming dependent variables and exploiting the logarithmic scale. See Details.
- **bias** A logical value. By default \(\text{bias}=\text{FALSE}\) specifies that no bias correction is computed. See Details.
- **k** An integer specifying the value of the intermediate sequence \(k_n\). See Details.
- **alpha** A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence region for the \(d\)-dimensional VaR.
- **plot** A logical value. By default \(\text{plot}=\text{FALSE}\) specifies that no graphical representation of the estimates is not provided. See Details.

Details

For a dataset \(\text{data}\) of \(d\)-dimensional observations and sample size \(n\), the VaR or EQ, corresponding to the extreme level \(\text{tau1}\), is computed by applying the \(d\)-dimensional Weissman estimator. The definition of the Weissman estimator depends on the estimation of the \(d\)-dimensional tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see MultiHTailIndex). The data are regarded as \(d\)-dimensional temporal independent observations coming from dependent variables. See Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(\text{tau}\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, for each variable, \((1 - \tau_n) \in (0, 1)\) is a small proportion of observations in the observed data sample that exceed the \(\tau_n\)-th empirical quantile. Such proportion of observations is used to estimate the individual \(\tau_n\)-th quantile and tail index \(\gamma\).
- The so-called extreme level \(\text{tau1}\) or \(\tau'_n\) is a sequence of positive reals such that \(\tau'_n \to 1\) as \(n \to \infty\). For each variable, the value \((1 - \tau'_n) \in (0, 1)\) is meant to be a small tail probability such that \((1 - \tau'_n) = 1/n\) or \((1 - \tau'_n) < 1/n\). It is also assumed that \(n(1 - \tau'_n) \to C\) as \(n \to \infty\), where \(C\) is a positive finite constant. The value \(C\) is the expected number of exceedances of the individual \(\tau'_n\)-th quantile. Typically, \(C \in (0, 1)\) which means that it is expected that there are no observations in a data sample exceeding the individual quantile of level \((1 - \tau'_n)\).
- If \(\text{var}=\text{TRUE}\) then an estimate of the asymptotic variance-covariance matrix of the \(\tau'_n\)-th \(d\)-dimensional quantile is computed. The data are regarded as temporal independent observations coming from dependent variables. The asymptotic variance-covariance matrix is estimated exploiting the formula in Section 5 of Padoan and Stupfler (2020). In particular, the variance-covariance matrix is computed exploiting the asymptotic behaviour of the normalized quantile estimator which is expressed in logarithmic scale. This is achieved through \(\text{varType}=\text{"asym-Ind-Log"}\). If \(\text{varType}=\text{"asym-Ind"}\) then the variance-covariance matrix is computed exploiting the asymptotic behaviour of the \(d\)-dimensional relative quantile estimator appropriately normalized (see formula in Section 5 of Padoan and Stupfler (2020)).
If bias=TRUE then an estimate of each individual $\tau_n^{(\prime)}$-th quantile is estimated using the formula in page 330 of de Haan et al. (2016), which provides a bias corrected version of the Weissman estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. For simplicity standard the variance-covariance matrix is still computed using formula in Section 5 of Padoan and Stupfler (2020).

$k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, for each marginal distribution, the value $k_n$ specifies the number of $k+1$ larger order statistics to be used to estimate the individual $\tau_n$-th empirical quantile and individual tail index $\gamma_j$ for $j = 1, \ldots, d$. The intermediate level $\tau_n$ can be seen defined as $\tau_n = 1 - k_n/n$.

Given a small value $\alpha \in (0, 1)$ then an estimate of an asymptotic confidence region for $\tau_n^{(\prime)}$-th d-dimensional quantile, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence regions are computed exploiting the asymptotic behaviour of the normalized quantile estimator in logarithmic scale. This is an "asymmetric" region and it is achieved through varType="asym-Ind-Log". A "symmetric" region is obtained exploiting the asymptotic behaviour of the relative quantile estimator appropriately normalized, see formula in Section 5 of Padoan and Stupfler (2020). This is achieved through varType="asym-Ind".

If plot=TRUE then a graphical representation of the estimates is not provided.

**Value**

A list with elements:

- **ExtQHat**: an estimate of the d-dimensional VaR or $\tau_n^{(\prime)}$-th d-dimensional quantile;
- **VarCovExQHat**: an estimate of the asymptotic variance-covariance of the d-dimensional VaR estimator;
- **EstConReg**: an estimate of the approximate $(1 - \alpha)100\%$ confidence regions for the d-dimensional VaR.

**Author(s)**

Simone Padoan, <simone.padoan@unibocconi.it>, [http://mypage.unibocconi.it/simonepadoan/](http://mypage.unibocconi.it/simonepadoan/);

**References**


**See Also**

*MultiHTailIndex, estMultiExpectiles, predMultiExpectiles*
Examples

# Extreme quantile estimation at the extreme level tau1 obtained with
# d-dimensional observations simulated from a joint distribution with
# a Gumbel copula and equal Frechet marginal distributions.
library(plot3D)
library(copula)
library(evd)

# distributional setting
copula <- "Gumbel"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.95
# Extreme level (or tail probability 1-tau1 of unobserved quantile)
tau1 <- 0.9995

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet
# marginals distributions and a Gumbel copula
data <- rmdata(ndata, dist, copula, par)
scatter3D(data[,1], data[,2], data[,3])

# High d-dimensional expectile (intermediate level) estimation
extQHat <- extMultiQuantile(data, tau, tau1, TRUE)

extQHat$ExtQHat
extQHat$VarCovExQHat
# run the following command to see the graphical representation
extQHat <- extMultiQuantile(data, tau, tau1, TRUE, plot=TRUE)

---

extQuantile

Value-at-Risk (VaR) or Extreme Quantile (EQ) Estimation

Description

Computes a point and interval estimate of the VaR based on the Weissman estimator.
extQuantile

Usage

extQuantile(data, tau, tau1, var=FALSE, varType="asym-Dep", bias=FALSE, bigBlock=NULL, smallBlock=NULL, k=NULL, alpha=0.05)

Arguments

data  A vector of $(1 \times n)$ observations.
tau  A real in $(0, 1)$ specifying the intermediate level $\tau_n$. See Details.
tau1  A real in $(0, 1)$ specifying the extreme level $\tau'_n$. See Details.
var  If var=TRUE then an estimate of the asymptotic variance of the VaR estimator is computed.
varType  A string specifying the type of asymptotic variance to compute. By default varType="asym-Dep" specifies the variance estimator for serial dependent observations. See Details.
bias  A logical value. By default bias=FALSE specifies that no bias correction is computed. See Details.
bigBlock  An integer specifying the size of the big-block used to estimate the asymptotic variance. See Details.
smallBlock  An integer specifying the size of the small-block used to estimate the asymptotic variance. See Details.
k  An integer specifying the value of the intermediate sequence $k_n$. See Details.
alpha  A real in $(0, 1)$ specifying the confidence level $(1 - \alpha)100\%$ of the approximate confidence interval for the VaR.

Details

For a dataset data of sample size $n$, the VaR or EQ, corresponding to the extreme level tau1, is computed by applying the Weissman estimator. The definition of the Weissman estimator depends on the estimation of the tail index $\gamma$. Here, $\gamma$ is estimated using the Hill estimation (see HTailIndex). The observations can be either independent or temporal dependent (see e.g. de Haan and Ferreira 2006; Drees 2003; de Haan et al. 2016 for details).

- The so-called intermediate level tau or $\tau_n$ is a sequence of positive reals such that $\tau_n \to 1$ as $n \to \infty$. Practically, $(1 - \tau_n) \in (0, 1)$ is a small proportion of observations in the observed data sample that exceed the $\tau_n$-th empirical quantile. Such proportion of observations is used to estimate the $\tau_n$-th quantile and $\gamma$.

- The so-called extreme level tau1 or $\tau_n'$ is a sequence of positive reals such that $\tau_n' \to 1$ as $n \to \infty$. The value $(1 - \tau_n') \in (0, 1)$ is meant to be a small tail probability such that $(1 - \tau_n') = 1/n$ or $(1 - \tau_n') < 1/n$. It is also assumed that $n(1 - \tau_n') \to C$ as $n \to \infty$, where $C$ is a positive finite constant. The value $C$ is the expected number of exceedances of the $\tau_n'$-th quantile. Typically, $C \in (0, 1)$ which means that it is expected that there are no observations in a data sample exceeding the quantile of level $(1 - \tau_n')$.

- If var=TRUE then an estimate of the asymptotic variance of the $\tau_n'$-th quantile is computed. With independent observations the asymptotic variance is estimated by the formula $\hat{\gamma}^2$ (see e.g. de Drees 2000, 2003, for details). This is achieved through varType="asym-Ind".
serial dependent data the asymptotic variance is estimated by the formula in 1288 in Drees (2000). This is achieved through varType="asym-Dep". In this latter case the computation of the serial dependence is based on the "big blocks seperated by small blocks" technique which is a standard tools in time series, see e.g. Leadbetter et al. (1986). The size of the big and small blocks are specified by the parameters bigBlock and smallBlock, respectively. With serial dependent data the asymptotic variance can also be estimated by formula (32) of Drees (2003). This is achieved through varType="asym-Alt-Dep".

- If bias=TRUE then an estimate of the $\tau'_n$-th quantile is computed using the formula in page 330 of de Haan et al. (2016), which provides a bias corrected version of the Weissman estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity standard formula in Drees (2000) page 1288 is used.

- $k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, the value $k_n$ specifies the number of $k+1$ larger order statistics to be used to estimate the $\tau_n$-th empirical quantile and $\gamma$. The intermediate level $\tau_n$ can be seen defined as $\tau_n = 1 - k_n/n$.

- Given a small value $\alpha \in (0,1)$ then an estimate of an asymptotic confidence interval for the $\tau'_n$-th quantile, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence intervals are computed exploiting the formulas (33) and (46) of Drees (2003). When bias=TRUE confidence intervals are computed in the same way but after correcting the tail index estimate by an estimate of the bias term, see formula (4.2) in de Haan et al. (2016) for details. Furthermore, in this case with serial dependent data the asymptotic variance is estimated using the formula in Drees (2000) page 1288.

### Value

A list with elements:

- ExtQHat: an estimate of the VaR or $\tau'_n$-th quantile;
- VarExQHat: an estimate of the asymptotic variance of the VaR estimator;
- CIExtQ: an estimate of the approximate $(1 - \alpha)100\%$ confidence interval for the VaR.

### Author(s)


### References


See Also

HTailIndex, EBTailIndex, estExpectiles

Examples

# Extreme quantile estimation at the level tau1 obtained with 1-dimensional data
# simulated from an AR(1) with univariate Student-t distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.97
# Extreme level (or tail probability 1-tau1 of unobserved quantile)
tau1 <- 0.9995

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# VaR (extreme quantile) estimation
extQHat1 <- extQuantile(data, tau, tau1, TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
extQHat1$ExtQHat
nextQHat1$CIExtQ

# VaR (extreme quantile) estimation with bias correction
extQHat2 <- extQuantile(data, tau, tau1, TRUE, bias=TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
extQHat2$ExtQHat
nextQHat2$CIExtQ
HTailIndex

Hill Tail Index Estimation

Description
Computes a point and interval estimate of the tail index based on the Hill’s estimator.

Usage
HTailIndex(data, k, var=FALSE, varType="asym-Dep", bias=FALSE, bigBlock=NULL, smallBlock=NULL, alpha=0.05)

Arguments
data A vector of \((1 \times n)\) observations.
k An integer specifying the value of the intermediate sequence \(k_n\). See Details.
var If var=TRUE then an estimate of the variance of the tail index estimator is computed.
varType A string specifying the asymptotic variance to compute. By default varType="asym-Dep" specifies the variance estimator for serial dependent observations. See Details.
bias A logical value. By default bias=FALSE specifies that no bias correction is computed. See Details.
bigBlock An integer specifying the size of the big-block used to estimate the asymptotic variance. See Details.
smallBlock An integer specifying the size of the small-block used to estimate the asymptotic variance. See Details.
alpha A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence interval for the tail index.

Details
For a dataset \(\text{data}\) of sample size \(n\), the tail index \(\gamma\) of its (marginal) distribution is computed by applying the Hill estimator. The observations can be either independent or temporal dependent.

- k or \(k_n\) is the value of the so-called intermediate sequence \(k_n, n = 1, 2, \ldots\). Its represents a sequence of positive integers such that \(k_n \to \infty\) and \(k_n/n \to 0\) as \(n \to \infty\). Practically, the value \(k_n\) specifies the number of \(k+1\) larger order statistics to be used to estimate \(\gamma\).

- If var=TRUE then an estimate of the asymptotic variance of the Hill estimator is computed. With independent observations the asymptotic variance is estimated by the formula \(\hat{\gamma}^2\), see Theorem 3.2.5 of de Haan and Ferreira (2006). This is achieved through varType="asym-Ind". With serial dependent observations the asymptotic variance is estimated by the formula in 1288 in Drees (2000). This is achieved through varType="asym-Dep". In this latter case the serial dependence is estimated by exploiting the "big blocks separated by small blocks" technique which is a standard tool in time series, see Leadbetter et al. (1986). See also formula (11) in Drees (2003). The size of the big and small blocks are specified by the parameters bigBlock and smallBlock, respectively.
• If bias=TRUE then an estimate of the bias term of the Hill estimator is computed implementing using formula (4.2) in de Haan et al. (2016). However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.1. Instead for simplicity standard formulas have been used (see de Haan and Ferreira 2006 Theorem 3.2.5 and Drees 2000 page 1288).

• Given a small value $\alpha \in (0,1)$ then an estimate of an asymptotic confidence interval for $\gamma$, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence intervals are computed exploiting the formulas in de Haan and Ferreira (2006) Theorem 3.2.5 and Drees (2000) page 1288. When bias=TRUE the confidence intervals are computed in the same way but after correcting the tail index estimate by an estimate of the bias term, see formula (4.2) in de Haan et al. (2016) for details.

Value

A list with elements:

- gammaHat: an estimate of tail index $\gamma$;
- VarGamHat: an estimate of the asymptotic variance of the Hill estimator;
- BiasGamHat: an estimate of bias term of the Hill estimator;
- AdjExtQHat: the adjustment to correct the Weissman estimator of an extreme quantile.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

MLTailIndex, MomTailIndex, EBTailIndex
Examples

# Tail index estimation based on the Hill estimator obtained with
# 1-dimensional data simulated from an AR(1) with univariate Student-t
# distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# Number of larger order statistics
k <- 150

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# tail index estimation
gammaHat1 <- HTailIndex(data, k, TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
gammaHat1$gammaHat
gammaHat1$CIgamHat

# tail index estimation with bias correction
gammaHat2 <- HTailIndex(data, 2*k, TRUE, bias=TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
gammaHat2$gammaHat-gammaHat2$BiasGamHat
gammaHat2$CIgamHat

Wald-Type Hypothesis Testing

Description

Wald-type hypothesis tes for testing equality of high or extreme expectiles and quantiles

Usage

HypoTesting(data, tau, tau1=NULL, type="ExpectRisks", level="extreme",
method="LAWS", bias=FALSE, k=NULL, alpha=0.05)
HypoTesting

Arguments

- **data**: A matrix of \((n \times d)\) observations.
- **tau**: A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.
- **tau1**: A real in \((0, 1)\) specifying the extreme level \(\tau'_n\). See Details.
- **type**: A string specifying the type of test. By default type="ExpectRisks" specifies the test for testing the equality of expectiles. See Details.
- **level**: A string specifying the level of the expectile. This makes sense when type="ExpectRisks". By default level="extreme" specifies that the test concerns expectiles at the extreme level. See Details.
- **method**: A string specifying the method used to estimate the expectile. By default est="LAWs" specifies the use of the LAWS based estimator. See Details.
- **bias**: A logical value. By default bias=FALSE specifies no bias correction is computed. See Details.
- **k**: An integer specifying the value of the intermediate sequence \(k_n\). See Details.
- **alpha**: A real in \((0, 1)\) specifying the significance level of the test.

Details

With a dataset \(data\) of \(d\)-dimensional observations and sample size \(n\), a Wald-type hypothesis testing is performed in order to check whether the is empirical evidence against the null hypothesis. The null hypothesis concerns the equality among the expectiles or quantiles or tail indices of the marginal distributions. The three tests depend on the depends on the estimation of the \(d\)-dimensional tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see MultiHTailIndex for details). The data are regarded as \(d\)-dimensional temporal independent observations coming from dependent variables. See Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(tau\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, for each marginal distribution, \(\tau_n \in (0, 1)\) is the ratio between \(N\) (Numerator) and \(D\) (Denominator). Where \(N\) is the empirical mean distance of the \(\tau_n\)-th expectile from the observations smaller than it, and \(D\) is the empirical mean distance of \(\tau_n\)-th expectile from all the observations.

- The so-called extreme level \(tau1\) or \(\tau'_n\) is a sequence of positive reals such that \(\tau'_n \to 1\) as \(n \to \infty\). For each marginal distribution, the value \((1 - tau1) \in (0, 1)\) is meant to be a small tail probability such that \((1 - \tau'_n) = 1/n\) or \((1 - \tau'_n) < 1/n\). It is also assumed that \(n(1 - \tau'_n) \to C\) as \(n \to \infty\), where \(C\) is a positive finite constant. Typically, \(C \in (0, 1)\) so it is expected that there are no observations in a data sample that are greater than the expectile at the extreme level \(\tau'_n\).

- When type="ExpectRisks", the null hypothesis of the hypothesis testing concerns the equality among the expectiles of the marginal distributions. See Section 3.3 of Padoan and Stupfler (2020) for details. When type="QuantRisks", the null hypothesis of the hypothesis testing concerns the equality among the quantiles of the marginal distributions. See Section 5 of Padoan and Stupfler (2020) for details. Note that in this case the test is based on the asymptotic distribution of normalized quantile estimator in the logarithmic scale. When type="tails", the null hypothesis of the hypothesis testing concerns the equality among the tail indices of the marginal distributions. See Sections 3.2 and 3.3 of Padoan and Stupfler (2020) for details.
• When `type='ExpectRisks'`, the null hypothesis concerns the equality among the expectiles of the marginal distributions at the intermediate level and this is achieved through `level='inter'`. In this case the test is obtained exploiting the asymptotic distribution of relative expectile appropriately normalised. See Section 2.1, 3.1 and 3.3 of Padoan and Stupfler (2020) for details. Instead, if `level='extreme'` the null hypothesis concerns the equality among the expectiles of the marginal distributions at the extreme level.

• When `method='VarLAWS'`, then the $\tau_n^{'th}$ $d$-dimensional expectile is estimated using the LAWS based estimator. When `method='QB'`, the expectile is instead estimated using the QB estimator. The definition of both estimators depend on the estimation of the $d$-dimensional tail index $\gamma$. The $d$-dimensional tail index $\gamma$ is estimated using the $d$-dimensional Hill estimator (tailest='Hill'), see `MultiHTailIndex`). See Section 2.2 in Padoan and Stupfler (2020) for details.

• If `bias=TRUE` then $d$-dimensional $\gamma$ is estimated using formula (4.2) of Haan et al. (2016). This is used by the LAWS and QB estimators. Furthermore, the $\tau_n^{'th}$ quantile is estimated using the formula in page 330 of de Haan et al. (2016). This provides a bias corrected version of the Weissman estimator. This is used by the QB estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity the asymptotic variance-covariance matrix is estimated by the formulas Section 3.2 of Padoan and Stupfler (2020).

• $k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$ as $n \rightarrow \infty$. Practically, for each marginal distribution when `tau=NULL` and `method='VarLAWS'` or `method='QB'`, then $\tau_n = 1 - k_n/n$ is the intermediate level of the expectile to be estimated. When `tailest='Hill'`, for each marginal distributions, then $k_n$ specifies the number of $k+1$ larger order statistics used in the definition of the Hill estimator.

• A small value $\alpha \in (0, 1)$ specifies the significance level of Wald-type hypothesis testing.

Value

A list with elements:

• `logLikR`: the observed value of log-likelihood ratio statistic test;
• `critVal`: the quantile (critical level) of a chi-square distribution with $d$ degrees of freedom and confidence level $\alpha$.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

`MultiHTailIndex`, `predMultiExpectiles`, `extMultiQuantile`
**Examples**

# Hypothesis testing on the equality extreme expectiles based on a sample of d-dimensional observations simulated from a joint distribution with a Gumbel copula and equal Frechet marginal distributions.

```r
library(plot3D)
library(copula)
library(evd)

# distributional setting
copula <- "Gumbel"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.95
# Extreme level (or tail probability 1-tau1 of unobserved expectile)
tau1 <- 0.9995

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet marginals distributions and a Gumbel copula
data <- rmdata(ndata, dist, copula, par)
scatter3D(data[,1], data[,2], data[,3])

# Performs Wald-type hypothesis testing
HypoTesting(data, tau, tau1)
```

# Hypothesis testing on the equality extreme expectiles based on a sample of d-dimensional observations simulated from a joint distribution with a Clayton copula and different Frechet marginal distributions.

```r
# distributional setting
copula <- "Clayton"
dist <- "Frechet"

# parameter setting
dim <- 3
dep <- 2
scale <- rep(1, dim)
shape <- c(2.1, 3, 4.5)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Intermediate level (or sample tail probability 1-tau)
tau <- 0.95
```
Extreme level (or tail probability 1-\tau_1 of unobserved expectile)
\tau_1 \leftarrow 0.9995

Sample size
ndata \leftarrow 1000

Simulates a sample from a multivariate distribution with equal Frechet
marginals distributions and a Gumbel copula
data \leftarrow \text{rmdata}(ndata, \text{dist}, \text{copula}, \text{par})
scatter3D(data[,1], data[,2], data[,3])

Performs Wald-type hypothesis testing
\text{HypoTesting}(data, \tau, \tau_1)

MLTailIndex

Maximum Likelihood Tail Index Estimation

Description
Computes a point and interval estimate of the tail index based on the Maximum Likelihood (ML) estimator.

Usage
MLTailIndex(data, k, var=FALSE, varType=\"asym-Dep\", bigBlock=NULL, smallBlock=NULL, alpha=0.05)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>A vector of ((1 \times n)) observations.</td>
</tr>
<tr>
<td>k</td>
<td>An integer specifying the value of the intermediate sequence (k_n). See Details.</td>
</tr>
<tr>
<td>var</td>
<td>If (\text{var=TRUE}) then an estimate of the asymptotic variance of the tail index estimator is computed.</td>
</tr>
<tr>
<td>varType</td>
<td>A string specifying the asymptotic variance to compute. By default (\text{varType=&quot;asym-Dep&quot;}) specifies the variance estimator for serial dependent observations. See Details.</td>
</tr>
<tr>
<td>bigBlock</td>
<td>An integer specifying the size of the big-block used to estimate the asymptotic variance. See Details.</td>
</tr>
<tr>
<td>smallBlock</td>
<td>An integer specifying the size of the small-block used to estimate the asymptotic variance. See Details.</td>
</tr>
<tr>
<td>alpha</td>
<td>A real in ((0,1)) specifying the confidence level ((1-\alpha)100%) of the approximate confidence interval for the tail index.</td>
</tr>
</tbody>
</table>
Details

For a dataset data of sample size $n$, the tail index $\gamma$ of its (marginal) distribution is computed by applying the ML estimator. The observations can be either independent or temporal dependent.

- $k$ or $k_n$ is the value of the so-called intermediate sequence $k_n, n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, the value $k_n$ specifies the number of $k+1$ larger order statistics to be used to estimate $\gamma$.

- If var=TRUE then the asymptotic variance of the Hill estimator is computed. With independent observations the asymptotic variance is estimated by the formula in Theorem 3.4.2 of de Haan and Ferreira (2006). This is achieved through varType="asym-Ind". With serial dependent observations the asymptotic variance is estimated by the formula in 1288 in Drees (2000). This is achieved through varType="asym-Dep". In this latter case the serial dependence is estimated by exploiting the "big blocks separated by small blocks" technique which is a standard tool in time series, see Leadbetter et al. (1986). See also formula (11) in Drees (2003). The size of the big and small blocks are specified by the parameters bigBlock and smallBlock, respectively.

- Given a small value $\alpha \in (0, 1)$ then an asymptotic confidence interval for the tail index, with approximate nominal confidence level $(1 - \alpha)100\%$ is computed.

Value

A list with elements:

- gammaHat: an estimate of tail index $\gamma$;
- VarGamHat: an estimate of the variance of the ML estimator;
- CIgamHat: an estimate of the approximate $(1 - \alpha)100\%$ confidence interval for $\gamma$.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

HTailIndex, MomTailIndex, EBTailIndex
Examples

# Tail index estimation based on the Maximum Likelihood estimator obtained with # 1-dimensional data simulated from an AR(1) with univariate Student-t # distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# Number of larger order statistics
k <- 150

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# tail index estimation
gammaHat <- MLTailIndex(data, k, TRUE, bigBlock=bigBlock, smallBlock=smallBlock)
gammaHat$gammaHat
gammaHat$CIgamHat

MomTailIndex

Moment based Tail Index Estimation

Description

Computes a point estimate of the tail index based on the Moment Based (MB) estimator.

Usage

MomTailIndex(data, k)

Arguments

data A vector of \((1 \times n)\) observations.

k An integer specifying the value of the intermediate sequence \(k_n\). See Details.
Details

For a dataset data of sample size $n$, the tail index $\gamma$ of its (marginal) distribution is computed by applying the MB estimator. The observations can be either independent or temporal dependent. For details see de Haan and Ferreira (2006).

- $k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. It represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, the value $k_n$ specifies the number of $k+1$ larger order statistics to be used to estimate $\gamma$.

Value

An estimate of the tail index $\gamma$.

Author(s)


References


See Also

HTailIndex, MLTailIndex, EBTailIndex

Examples

# Tail index estimation based on the Moment estimator obtained with
# 1-dimensional data simulated from an AR(1) with univariate Student-t
# distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallblock <- 15

# Number of larger order statistics
k <- 150

# sample size
ndata <- 2500
MultiHTailIndex

Description

Computes point estimates and \((1 - \alpha)100\%\) confidence regions estimate of \(d\)-dimensional tail indices based on the Hill’s estimator.

Usage

```r
MultiHTailIndex(data, k, var=FALSE, varType="asym-Dep", bias=FALSE, alpha=0.05, plot=FALSE)
```

Arguments

- `data`: A matrix of \((n \times d)\) observations.
- `k`: An integer specifying the value of the intermediate sequence \(k_n\). See Details.
- `var`: If `var=TRUE` then an estimate of the variance-covariance matrix of the tail indices estimators is computed.
- `varType`: A string specifying the asymptotic variance to compute. By default `varType="asym-Dep"` specifies the variance estimator for \(d\) dependent marginal variables. See Details.
- `bias`: A logical value. By default `bias=FALSE` specifies that no bias correction is computed. See Details.
- `alpha`: A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence interval for the tail index.
- `plot`: A logical value. By default `plot=FALSE` specifies that no graphical representation of the estimates is provided. See Details.

Details

For a dataset \(data\) of \((n \times d)\) observations, where \(d\) is the number of variables and \(n\) is the sample size, the tail index \(\gamma\) of the \(d\) marginal distributions is estimated by applying the Hill estimator. Together with a point estimate a \((1 - \alpha)100\%\) confidence region is computed. The data are regarded as \(d\)-dimensional temporal independent observations coming from dependent variables.

- \(k\) or \(k_n\) is the value of the so-called intermediate sequence \(k_n\), \(n = 1, 2, \ldots\). Its represents a sequence of positive integers such that \(k_n \to \infty\) and \(k_n/n \to 0\) as \(n \to \infty\). Practically, the value \(k_n\) specifies the number of \(k+1\) larger order statistics to be used to estimate each marginal tail index \(\gamma_j\) for \(j = 1, \ldots, d\).
• If var=TRUE then an estimate of the asymptotic variance-covariance matrix of the multivariate Hill estimator is computed. With independent observations the asymptotic variance-covariance matrix is estimated by the matrix $\hat{\Sigma}_{\text{LAWS}}(\gamma, R)(1, 1)$, see bottom formula in page 14 of Padoan and Stupfler (2020). This is achieved through varType="asym-Dep" which means $d$ dependent marginal variables. When varType="asym-Ind" $d$ marginal variables are regarded as independent and the returned variance-covariance matrix $\hat{\Sigma}_{\text{LAWS}}(\gamma, R)(1, 1)$ is a diagonal matrix with only variance terms.

• If bias=TRUE then an estimate of the bias term of the Hill estimator is computed implementing using formula (4.2) in de Haan et al. (2016). In this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.1 but instead for simplicity the formula at the bottom of page 14 in Padoan and Stupfler (2020) is still used.

• Given a small value $\alpha \in (0, 1)$ then an estimate of an asymptotic confidence region for $\gamma_j$, for $j = 1, \ldots, d$, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence intervals are computed exploiting the asymptotic normality of multivariate Hill estimator appropriately normalized (the logarithmic scale is not used), see Padoan and Stupfler (2020) for details.

• If plot=TRUE then a graphical representation of the estimates is not provided.

Value

A list with elements:

• gammaHat: an estimate of the $d$ tail indices $\gamma_j$, for $j = 1, \ldots, d$;

• VarCovGHat: an estimate of the asymptotic variance-covariance matrix of the multivariate Hill estimator;

• biasTerm: an estimate of bias term of the multivariate Hill estimator;

• EstConReg: an estimate of the $(1 - \alpha)100\%$ confidence region.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

HTailIndex, rmdata
Examples

# Tail index estimation based on the multivariate Hill estimator obtained with
# n observations simulated from a d-dimensional random vector with a multivariate
# distribution with equal Frechet margins and a Clayton copula.
library(plot3D)
library(copula)
library(evd)

# distributional setting
copula <- "Clayton"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Number of larger order statistics
k <- 150

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet
# marginals distributions and a Clayton copula
data <- rmdata(ndata, dist, copula, par)
scatter3D(data[,1], data[,2], data[,3])

# tail indices estimation
est <- MultiHTailIndex(data, k, TRUE)
est$gammaHat
est$VarCovGHat

# run the following command to see the graphical representation
est <- MultiHTailIndex(data, k, TRUE, plot=TRUE)

predExpectiles

Extreme Expectile Estimation

Description

Computes a point and interval estimate of the expectile at the extreme level (Expectile Prediction).

Usage

predExpectiles(data, tau, tau1, method="LAWS", tailest="Hill", var=FALSE, varType="asym-Dep", bias=FALSE, bigBlock=NULL, smallBlock=NULL, k=NULL, alpha_n=NULL, alpha=0.05)
Arguments

- **data**: A vector of \((1 \times n)\) observations.
- **tau**: A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.
- **tau1**: A real in \((0, 1)\) specifying the extreme level \(\tau'_n\). See Details.
- **method**: A string specifying the method used to estimate the expecile. By default est="LAWS" specifies the use of the LAWS based estimator. See Details.
- **tailest**: A string specifying the tail index estimator. By default tailest="Hill" specifies the use of Hill estimator. See Details.
- **var**: If var=TRUE then an estimate of the asymptotic variance of the expectile estimator is computed.
- **varType**: A string specifying the type of asymptotic variance to compute. By default varType="asym-Dep" specifies the variance estimator for serial dependent observations. See Details.
- **bias**: A logical value. By default bias=FALSE specifies that no bias correction is computed. See Details.
- **bigBlock**: An integer specifying the size of the big-block used to estimate the asymptotic variance. See Details.
- **smallBlock**: An integer specifying the size of the small-block used to estimate the asymptotic variance. See Details.
- **k**: An integer specifying the value of the intermediate sequence \(k_n\). See Details.
- **alpha_n**: A real in \((0, 1)\) specifying the quantile’s extreme level to be use in order to estimate the expectile’s extreme level.
- **alpha**: A real in \((0, 1)\) specifying the confidence level \((1-\alpha)100\%\) of the approximate confidence interval for the expectile at the intermediate level.

Details

For a dataset \( \text{data} \) of sample size \(n\), an estimate of the \(\tau'_n\)-th expectile is computed. The estimation of the expectile at the extreme level \(\tau_1\) \((\tau'_n)\) is meant to be a prediction beyond the observed sample. Two estimators are available: the so-called Least Asymmetrically Weighted Squares (LAWS) based estimator and the Quantile-Based (QB) estimator. The definition of both estimators depends on the estimation of the tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see HTailIndex for details) or in alternative using the the expectile based estimator (see EBTailIndex). The observations can be either independent or temporal dependent. See Section 3.2 in Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(\tau\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, \(\tau_n \in (0, 1)\) is the ratio between \(N\) (Numerator) and \(D\) (Denominator). Where \(N\) is the empirical mean distance of the \(\tau_n\)-th expectile from the observations smaller than it, and \(D\) is the empirical mean distance of the \(\tau_n\)-th expectile from all the observations.
- The so-called extreme level \(\tau_1\) or \(\tau'_n\) is a sequence of positive reals such that \(\tau'_n \to 1\) as \(n \to \infty\). The value \(1 - \tau'_n\) \((\tau_n)\) is meant to be a small tail probability such that \((1 - \tau'_n) = 1/n\) or \((1 - \tau'_n) < 1/n\). It is also assumed that \(n(1 - \tau_n) \to C\) as \(n \to \infty\), where \(C\) is a positive finite constant. Typically, \(C \in (0, 1)\) so it is expected that there are no observations in a data sample that are greater than the expectile at the extreme level \(\tau'_n\).
• When `method='LAWS'`, then the $\tau'_{n_k}$-th expectile is estimated using the LAWS based estimator. When `method='QB'`, the expectile is instead estimated using the QB esimtator. The definition of both estimators depend on the estimation of the tail index $\gamma$. When $\text{tailest}='\text{Hill}'$ then $\gamma$ is estimated using the Hill estimator (see `HTailIndex`). When $\text{tailest}='\text{ExpBased}'$, then $\gamma$ is estimated using the expectile based estimator (see `EBTailIndex`). See Section 3.2 in Padoan and Stupfler (2020) for details.

• If `var=TRUE` then an estimate of the asymptotic variance of the $\tau'_{n_k}$-th expectile is computed. Notice that the estimation of the asymptotic variance is only available when $\gamma$ is estimated using the Hill estimator (see `HTailIndex`). With independent observations the asymptotic variance is estimated by $\gamma^2$, see the remark below Theorem 3.5 in Padoan and Stupfler (2020). This is achieved through `varType='asym-Ind'`. With serial dependent observations the asymptotic variance is estimated by the formula in Throrem 3.5 of Padoan and Stupfler (2020). This is achieved through `varType='asym-Dep'`. See Section 3.2 in Padoan and Stupfler (2020) for details. In this latter case the computation of the serial dependence is based on the "big blocks seperated by small blocks" techinque which is a standard tools in time series, see e.g. Leadbetter et al. (1986). The size of the big and small blocks are specified by the parameters `bigBlock` and `smallBlock`, respectively.

• If `bias=TRUE` then $\gamma$ is estimated using formula (4.2) of Haan et al. (2016). This is used by the LAWS and QB estimators. Furthermore, the $\tau'_{n_k}$-th quantile is estimated using the formula in page 330 of de Haan et al. (2016). This provides a bias corrected version of the Weissman estimator. This is used by the QB estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity the asymptotic variance is estimated by the formula in Corollary 3.8, with serial dependent observations, and $\gamma^2$ with independent observation (see e.g. de Drees 2000, for the details).

• $k$ or $n_k$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, when $\text{tau=NULL}$ and method='LAWS', then $\tau_n = 1 - k_n/n$ is the intermediate level of the expectile to be estimated. The latter is also used to estimate the tail index when $\text{tailest}='\text{ExpBased}'$. Instead, if $\text{tailest}='\text{Hill}'$, then $n_k$ specifies the number of $k+1$ larger order statistics used in the definition of the Hill estimator. Differently, When $\text{tau=NULL}$ and method='QB', then $\tau_n = 1 - k_n/n$ is the intermediate level of the quantile to be estimated and of the expectile to be estimated when $\text{tailest}='\text{ExpBased}'$. Instead, when $\text{tailest}='\text{Hill}'$ it is the number of $k+1$ larger order statistics used in the definition of the Hill estimator.

• If quantile’s extreme level is provided by $\alpha_n$, then expectile’s extreme level $\tau'_{n_k}(\alpha_n)$ is replaced by $\tau'_{n_k}(\alpha_n)$ which is estimated using the method described in Section 6 of Padoan and Stupfler (2020). See `estExtLevel` for details.

• Given a small value $\alpha \in (0, 1)$ then an estimate of an asymptotic confidence interval for $\tau'_{n_k}$-th expectile, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence intervals are computed exploiting formula (10) and (11) in Padoan and Stupfler (2020) and (46) in Drees (2003). See Section 5 in Padoan and Stupfler (2020) for details. When `biast=TRUE` confidence intervals are computed in the same way but after correcting the tail index estimate by an estimate of the bias term, see formula (4.2) in de Haan et al. (2016) for details.

**Value**

A list with elements:
• EExpChat: an estimate of the \( \tau_{n}^{\prime} \)-th expecile;
• VarExtHat: an estimate of the asymptotic variance of the expecile estimator;
• CIExpct: an estimate of the approximate \((1 - \alpha)100\%\) confidence interval for \( \tau_{n}^{\prime} \)-th expecile.

**Author(s)**
Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

**References**

**See Also**
HTailIndex, EBTailIndex, estExpectiles, extQuantile

**Examples**

```r
# Extreme expecile estimation at the extreme level tau1 obtained with
# 1-dimensional data simulated from an AR(1) with univariate
# Student-t distributed innovations

tsDist <- "studentT"
tsType <- "AR"

corr <- 0.8
df <- 3
par <- c(corr, df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

tau <- 0.95
```

```r
# Intermediate level (or sample tail probability 1-tau)
```
predMultiExpectiles

# Extreme level (or tail probability 1-tau1 of unobserved expectile)
tau1 <- 0.9995

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# Extreme expectile estimation
expectHat1 <- predExpectiles(data, tau, tau1, var=TRUE, bigBlock=bigBlock,
                             smallBlock=smallBlock)
expectHat1$EExpcHat
expectHat1$CIExpct

# Extreme expectile estimation with bias correction
tau <- 0.80
expectHat2 <- predExpectiles(data, tau, tau1, "QB", var=TRUE, bias=TRUE, bigBlock=bigBlock,
                             smallBlock=smallBlock)
expectHat2$EExpcHat
expectHat2$CIExpct

predMultiExpectiles  Multidimensional Extreme Expectile Estimation

Description
Computes point estimates and \((1 - \alpha)\)100% confidence regions for d-dimensional expectile at the extreme level (Expectile Prediction).

Usage
predMultiExpectiles(data, tau, tau1, method="LAWS", tailest="Hill", var=FALSE, varType="asym-Ind-Adj-Log", bias=FALSE, k=NULL, alpha=0.05, plot=FALSE)

Arguments

data A matrix of \((n \times d)\) observations.
tau A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.
tau1 A real in \((0, 1)\) specifying the extreme level \(\tau'_n\). See Details.
method A string specifying the method used to estimate the expectile. By default est="LAWS" specifies the use of the LAWS based estimator. See Details.
tailest A string specifying the tail index estimator. By default tailest="Hill" specifies the use of Hill estimator. See Details.
var If var=TRUE then an estimate of the asymptotic variance of the expectile estimator is computed.
predMultiExpectiles

**varType**
A string specifying the type of asymptotic variance-covariance matrix to compute. By default `varType="asym-Ind-Adj-Log"` specifies that the variance-covariance matrix is computed assuming dependent variables and exploiting the log scale and a suitable adjustment. See Details.

**bias**
A logical value. By default `bias=FALSE` specifies that no bias correction is computed. See Details.

**k**
An integer specifying the value of the intermediate sequence \( k_\tau \). See Details.

**alpha**
A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence region for the \(d\)-dimensional expectile at the extreme level.

**plot**
A logical value. By default `plot=FALSE` specifies that no graphical representation of the estimates is provided. See Details.

### Details

For a dataset `data` of \(d\)-dimensional observations and sample size \(n\), an estimate of the \(\tau_n^{(d)}\)th \(d\)-dimensional expectile is computed. The estimation of the \(d\)-dimensional expectile at the extreme level \(\tau_1\) \((\tau_n^{(d)}\) is meant to be a prediction beyond the observed sample. Two estimators are available: the so-called Least Asymmetrically Weighted Squares (LAWS) based estimator and the Quantile-Based (QB) estimator. The definition of both estimators depends on the estimation of the \(d\)-dimensional tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see `MultiHTailIndex` for details). The data are regarded as \(d\)-dimensional temporal independent observations coming from dependent variables. See Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(\tau_1\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). Practically, for each marginal distribution, \(\tau_n \in (0, 1)\) is the ratio between \(N\) (Numerator) and \(D\) (Denominator). Where \(N\) is the empirical mean distance of the \(\tau_n\)th expectile from the observations smaller than it, and \(D\) is the empirical mean distance of the \(\tau_n\)th expectile from all the observations.

- The so-called extreme level \(\tau_1\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \to 1\) as \(n \to \infty\). For each marginal distribution, the value \((1 - \tau_n) \in (0, 1)\) is meant to be a small tail probability such that \((1 - \tau_n) = 1/n\) or \((1 - \tau_n) < 1/n\). It is also assumed that \(n(1 - \tau_n) \to C\) as \(n \to \infty\), where \(C\) is a positive finite constant. Typically, \(C \in (0, 1)\) so it is expected that there are no observations in a data sample that are greater than the expectile at the extreme level \(\tau_n\).

- When `method="LAWS"`, then the \(\tau_n\)th \(d\)-dimensional expectile is estimated using the LAWS based estimator. When `method="QB"`, the expectile is instead estimated using the QB estimator. The definition of both estimators depend on the estimation of the \(d\)-dimensional tail index \(\gamma\). The \(d\)-dimensional tail index \(\gamma\) is estimated using the \(d\)-dimensional Hill estimator (tailest="Hill"), see `MultiHTailIndex`). This is the only available option so far (soon more results will be available). See Section 2.2 in Padoan and Stupfler (2020) for details.

- If `var=TRUE` then an estimate of the asymptotic variance-covariance matrix of the \(\tau_n\)th \(d\)-dimensional expectile is computed. Notice that the estimation of the asymptotic variance-covariance matrix is only available when \(\gamma\) is estimated using the Hill estimator (see `MultiHTailIndex`). The data are regarded as temporal independent observations coming from dependent variables. The asymptotic variance-covariance matrix is estimated exploiting the formulas in Section 3.2 of Padoan and Stupfler (2020). The variance-covariance matrix is computed exploiting the asymptotic behaviour of the normalized expectile estimator which is expressed...
predMultiExpectiles

in logarithmic scale. In addition, a suitable adjustment is considered. This is achieved through varType="asym-Ind-Adj-Log". The data can also be regarded as coded-dimensional temporal independent observations coming from independent variables. In this case the asymptotic variance-covariance matrix is diagonal and is also computed exploiting the formulas in Section 3.2 of Padoan and Stupfler (2020). This is achieved through varType="asym-Ind-Log". If varType="asym-Ind-Adj", then the variance-covariance matrix is computed exploiting the asymptotic behaviour of the relative expectile estimator appropriately normalized and exploiting a suitable adjustment. This concerns the case of dependent variables. The case of independent variables is achieved through varType="asym-Ind".

• If bias=TRUE then d-dimensional $\gamma$ is estimated using formula (4.2) of Haan et al. (2016). This is used by the LAWS and QB estimators. Furthermore, the $\tau_n'$-th quantile is estimated using the formula in page 330 of de Haan et al. (2016). This provides a bias corrected version of the Weissman estimator. This is used by the QB estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity the asymptotic variance-covariance matrix is estimated by the formulas Section 3.2 of Padoan and Stupfler (2020).

• k or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. Practically, for each marginal distribution when tau=NULL and method='LAWS' or method='QB', then $\tau_n = 1 - k_n/n$ is the intermediate level of the expectile to be estimated. When tail.est='Hill', for each marginal distributions, then $k_n$ specifies the number of k+1 larger order statistics used in the definition of the Hill estimator.

• Given a small value $\alpha \in (0, 1)$ then an estimate of an asymptotic confidence region for $\tau_n'$-$th$ d-dimensional expectile, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence regions are computed exploiting the formulas in Section 3.2 of Padoan and Stupfler (2020). If varType="asym-Ind-Adj-Log", then an "asymmetric" confidence regions is computed exploiting the asymptotic behaviour of the normalized expectile estimator in logarithmic scale and using a suitable adjustment. This choice is recommended. If varType="asym-Ind-Adj", then the a "symmetric" confidence regions is computed exploiting the asymptotic behaviour of the relative expectile estimator appropriately normalized.

• If plot=TRUE then a graphical representation of the estimates is not provided.

Value

A list with elements:

• ExpctHat: an estimate of the $\tau_n'$-$th$ d-dimensional expectile;
• biasTerm: an estimate of the bias term of yje $\tau_n'$-$th$ d-dimensional expectile;
• VarCovEHat: an estimate of the asymptotic variance-covariance of the d-dimensional expectile estimator;
• EstConReg: an estimate of the approximate $(1 - \alpha)100\%$ confidence regions for $\tau_n'$-$th$ d-dimensional expectile.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;
References


See Also

MultiHTailIndex, estMultiExpectiles, extMultiQuantile

Examples

```r
# Extreme expectile estimation at the extreme level tau1 obtained with
d-dimensional observations simulated from a joint distribution with
# a Gumbel copula and equal Frechet marginal distributions.
library(plot3D)
library(copula)
library(evd)

# distributional setting
copula <- "Gumbel"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# Intermediate level (or sample tail probability 1-\(\tau\))
\(\tau\) <- 0.95
# Extreme level (or tail probability 1-\(\tau_1\) of unobserved expectile)
\(\tau_1\) <- 0.9995

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet
# marginals distributions and a Gumbel copula
data <- rmdata(ndata, dist, copula, par)
scatter3D(data[,1], data[,2], data[,3])

# High d-dimensional expectile (intermediate level) estimation
expectHat <- predMultiExpectiles(data, \(\tau\), \(\tau_1\), var=TRUE)
expectHat$ExpctHat
expectHat$VarCovEHat
# run the following command to see the graphical representation
expectHat <- predMultiExpectiles(data, \(\tau\), \(\tau_1\), var=TRUE, plot=TRUE)
```
QuantMES

Marginal Expected Shortfall Quantile Based Estimation

Description

Computes a point and interval estimate of the Marginal Expected Shortfall (MES) using a quantile based approach.

Usage

QuantMES(data, tau, tau1, var=FALSE, varType="asym-Dep", bias=FALSE, bigBlock=NULL, smallBlock=NULL, k=NULL, alpha=0.05)

Arguments

data: A vector of \((1 \times n)\) observations.

tau: A real in \((0, 1)\) specifying the intermediate level \(\tau_n\). See Details.

tau1: A real in \((0, 1)\) specifying the extreme level \(\tau'_n\). See Details.

var: If \(\text{var}=\text{TRUE}\) then an estimate of the asymptotic variance of the MES estimator is computed.

varType: A string specifying the type of asymptotic variance to compute. By default \(\text{varType}="\text{asym-Dep}"\) specifies the variance estimator for serial dependent observations. See Details.

bias: A logical value. By default \(\text{bias}=\text{FALSE}\) specifies that no bias correction is computed. See Details.

bigBlock: An interger specifying the size of the big-block used to estimate the asymptotic variance. See Details.

smallBlock: An interger specifying the size of the small-block used to estimate the asymptotic variance. See Details.

k: An integer specifying the value of the intermediate sequence \(k_n\). See Details.

alpha: A real in \((0, 1)\) specifying the confidence level \((1 - \alpha)100\%\) of the approximate confidence interval for the expecile at the intermediate level.

Details

For a dataset \(\text{data}\) of sample size \(n\), an estimate of the \(\tau'_n\)-th MES is computed. The estimation of the MES at the extreme level \(\tau_1\) (\(\tau'_n\)) is indeed meant to be a prediction. Estimates are obtained through the quantile based estimator defined in page 12 of Padoan and Stupfler (2020). Such an estimator depends on the estimation of the tail index \(\gamma\). Here, \(\gamma\) is estimated using the Hill estimation (see HTailIndex for details). The observations can be either independent or temporal dependent. See Section 4 in Padoan and Stupfler (2020) for details.

- The so-called intermediate level \(\tau_0\) or \(\tau_n\) is a sequence of positive reals such that \(\tau_n \rightarrow 1\) as \(n \rightarrow \infty\). See predExpectiles for details.
The so-called extreme level $\tau_n'$ or $\tau'_n$ is a sequence of positive reals such that $\tau'_n \to 1$ as $n \to \infty$. See `predExpectiles` for details.

If `var=TRUE` then an estimate of the asymptotic variance of the $\tau'_n$-th MES is computed. Notice that the estimation of the asymptotic variance is only available when $\gamma$ is estimated using the Hill estimator (see `HTailIndex`). With independent observations the asymptotic variance is estimated by $\hat{\gamma}^2$, see Corollary 4.3 in Padoan and Stupfler (2020). This is achieved through `varType="asym-Ind"`. With serial dependent observations the asymptotic variance is estimated by the formula in Corollary 4.2 of Padoan and Stupfler (2020). This is achieved through `varType="asym-Dep"`. See Section 4 and 5 in Padoan and Stupfler (2020) for details. In this latter case the computation of the serial dependence is based on the "big blocks separated by small blocks" technique which is a standard tool in time series, see e.g. Leadbetter et al. (1986). The size of the big and small blocks are specified by the parameters `bigBlock` and `smallBlock`, respectively.

If `bias=TRUE` then $\gamma$ is estimated using formula (4.2) of Haan et al. (2016). This is used by the LAWS and QB estimators. Furthermore, the $\tau'_n$-th quantile is estimated using the formula in page 330 of de Haan et al. (2016). This provides a bias corrected version of the Weisman estimator. This is used by the QB estimator. However, in this case the asymptotic variance is not estimated using the formula in Haan et al. (2016) Theorem 4.2. Instead, for simplicity the asymptotic variance is estimated by the formula in Corollary 3.8, with serial dependent observations, and $\hat{\gamma}^2$ with independent observation (see e.g. de Drees 2000, for the details).

$k$ or $k_n$ is the value of the so-called intermediate sequence $k_n$, $n = 1, 2, \ldots$. Its represents a sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. $k_n$ specifies the number of $k+1$ larger order statistics used in the definition of the Hill estimator (see `HTailIndex` for details).

If the quantile’s extreme level is provided by $\alpha_n$, then expectile’s extreme level $\tau'_n$ is replaced by $\tau'_n(\alpha_n)$ which is estimated by the method described in Section 6 of Padoan and Stupfler (2020). See `estExtLevel` for details.

Given a small value $\alpha \in (0, 1)$ then an estimate of an asymptotic confidence interval for $\tau'_n$-th expectile, with approximate nominal confidence level $(1 - \alpha)100\%$, is computed. The confidence intervals are computed exploiting formula in Corollary 4.2, Theorem 6.2 of Padoan and Stupfler (2020) and (46) in Drees (2003). See Sections 4-6 in Padoan and Stupfler (2020) for details. When `biast=TRUE` confidence intervals are computed in the same way but after correcting the tail index estimate by an estimate of the bias term, see formula (4.2) in de Haan et al. (2016) for details.

### Value

A list with elements:

- `HatQMES`: an estimate of the $\tau'_n$-th quantile based MES;
- `VarHatQMES`: an estimate of the asymptotic variance of the quantile based MES estimator;
- `CIHatQMES`: an estimate of the approximate $(1 - \alpha)100\%$ confidence interval for $\tau'_n$-th MES.

### Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, [http://mypage.unibocconi.it/simonepadoan/](http://mypage.unibocconi.it/simonepadoan/);
References


See Also

ExpectMES, HTailIndex, predExpectiles, extQuantile

Examples

# Marginl Expected Shortfall quantile based estimation at the extreme level
# obtained with 2-dimensional data simulated from an AR(1) with bivariate
# Student-t distributed innovations

tsDist <- "AStudentT"
tsType <- "AR"
tsCopula <- "studentT"

# parameter setting
corr <- 0.8
dep <- 0.8
df <- 3
par <- list(corr=corr, dep=dep, df=df)

# Big- small-blocks setting
bigBlock <- 65
smallBlock <- 15

# quantile's extreme level
tau1 <- 0.9995

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rbtimeseries(ndata, tsDist, tsType, tsCopula, par)
# Extreme MES expectile based estimation
MESHat <- QuantMES(data, NULL, tau1, var=TRUE, k=150, bigBlock=bigBlock, smallBlock=smallBlock)

## rbtimeseries

**Simulation of Two-Dimensional Temporally Dependent Observations**

**Description**

Simulates samples from parametric families of bivariate time series models.

**Usage**

```r
rbtimeseries(n, dist="studentT", type="AR", copula="Gumbel", par, burnin=1e+03)
```

**Arguments**

- `ndata`: A positive integer specifying the number of observations to simulate.
- `dist`: A string specifying the parametric family of the innovations distribution. By default `dist="studentT"` specifies a Student-t family of distributions. See Details.
- `type`: A string specifying the type of time series. By default `type="AR"` specifies a linear Auto-Regressive time series. See Details.
- `copula`: A string specifying the type copula to be used. By default `copula="Gumbel"` specifies the Gumbel copula. See Details.
- `par`: A list of parameters to be specified for the bivariate time series parametric family. See Details.
- `burnin`: A positive integer specifying the number of initial observations to discard from the simulated sample.

**Details**

For a time series class (type), with a parametric family (dist) for the innovations, a sample of size ndata is simulated. See for example Brockwell and Davis (2016).

- The available categories of bivariate time series models are: Auto-Regressive (type="AR"), Auto-Regressive and Moving-Average (type="ARMA"), Generalized-Autoregressive-Conditional-Heteroskedasticity (type="GARCH") and Auto-Regressive.

- With AR(1) times series the available families of distributions for the innovations and the dependence structure (copula) are:
  - Student-t (dist="studentT" and copula="studentT") with marginal parameters (equal for both distributions): \( \phi \in (-1,1) \) (autoregressive coefficient), \( \nu > 0 \) (degrees of freedom) and dependence parameter \( dep \in (-1,1) \). The parameters are specified as `par <- list(corr, df, dep);`
- Asymmetric Student-\(t\) (\texttt{dist=\"AStudentT\" and \texttt{copula=\"studentT\"}) with marginal parameters (equal for both distributions): \(\phi \in (-1, 1)\) (autoregressive coefficient), \(\nu > 0\) (degrees of freedom) and dependence parameter \(\text{dep} \in (-1, 1)\). The parameters are specified as \(\text{par} \leftarrow \text{list}(\text{corr}, \text{df}, \text{dep})\). Note that in this case the tail index of the lower and upper tail of the first marginal are different, see Padoan and Stupfler (2020) for details;

- With ARMA(1,1) times series the available families of distributions for the innovations and the dependence structure (copula) are:
  - symmetric Pareto (\texttt{dist=\"double-Pareto\" and \texttt{copula=\"Gumbel\") with marginal parameters (equal for both distributions): \(\phi \in (-1, 1)\) (autoregressive coefficient), \(\sigma > 0\) (scale), \(\alpha > 0\) (shape), \(\theta\) (movingaverage coefficient), and dependence parameter \(\text{dep}\) (\(\text{dep} > 0\) if \texttt{copula=\"Gumbel\") or \(\text{dep} \in (-1, 1)\) if \texttt{copula=\"Gaussian\")}. The parameters are specified as \(\text{par} \leftarrow \text{list}(\text{corr}, \text{scale}, \text{shape}, \text{smooth}, \text{dep})\).
  - symmetric Pareto (\texttt{dist=\"double-Pareto\" and \texttt{copula=\"Gumbel\") with marginal parameters (equal for both distributions): \(\phi \in (-1, 1)\) (autoregressive coefficient), \(\sigma > 0\) (scale), \(\alpha > 0\) (shape), \(\theta\) (movingaverage coefficient), and dependence parameter \(\text{dep}\) (\(\text{dep} > 0\) if \texttt{copula=\"Gumbel\") or \(\text{dep} \in (-1, 1)\) if \texttt{copula=\"Gaussian\")}. The parameters are specified as \(\text{par} \leftarrow \text{list}(\text{corr}, \text{scale}, \text{shape}, \text{smooth}, \text{dep})\). Note that in this case the tail index of the lower and upper tail of the first marginal are different, see Padoan and Stupfler (2020) for details;

- With ARCH(1)/GARCH(1,1) time series the distribution of the innovations are symmetric Gaussian (\texttt{dist=\"Gaussian\") or asymmetric Gaussian \texttt{dist=\"AGaussian\}). In both cases the marginal parameters (equal for both distributions) are: \(\alpha_0, \alpha_1, \beta\). In the asymmetric Gaussian case the tail index of the lower and upper tail of the first marginal are different, see Padoan and Stupfler (2020) for details. The available copulas are: Gaussian (\texttt{copula=\"Gaussian\") with dependence parameter \(\text{dep} \in (-1, 1)\), Student-\(t\) (\texttt{copula=\"studentT\") with dependence parameters \(\text{dep} \in (-1, 1)\) and \(\nu > 0\) (degrees of freedom), Gumbel (\texttt{copula=\"Gumbel\") with dependence parameter \(\text{dep} > 0\). The parameters are specified as \(\text{par} \leftarrow \text{list}(\alpha_0, \alpha_1, \beta, \text{dep})\) or \(\text{par} \leftarrow \text{list}(\alpha_0, \alpha_1, \beta, \text{dep}, \text{df})\).

**Value**

A vector of \((2 \times n)\) observations simulated from a specified bivariate time series model.

**Author(s)**

Simone Padoan, <simone.padoan@unibocconi.it>, \url{http://mypage.unibocconi.it/simonepadoan/};
Gilles Stupfler, <gilles.stupfler@ensai.fr>, \url{http://ensai.fr/en/equipe/stupfler-gilles/}

**References**


**See Also**

\texttt{rtimeseries}, \texttt{expectiles}
Examples

```r
# Data simulation from a 2-dimensional AR(1) with bivariate Student-t distributed
# innovations, with one marginal distribution whose lower and upper tail indices
# that are different

tsDist <- "AStudentT"
tsType <- "AR"
tsCopula <- "studentT"

data <- rbtimeseries(ndata, tsDist, tsType, tsCopula, par)

# Extreme expectile estimation
plot(data, pch=21)
plot(data[,1], type="l")
plot(data[,2], type="l")
```

Description

Simulates samples of independent $d$-dimensional observations from parametric families of joint
distributions with a given copula and equal marginal distributions.

Usage

```r
rmdta (ndata, dist="studentT", copula="studentT", par)
```

Arguments

- `ndata` A positive integer specifying the number of observations to simulate.
- `dist` A string specifying the parametric family of equal marginal distributions. By
default `dist="studentT"` specifies a Student-$t$ family of distributions. See Details.
- `copula` A string specifying the type copula to be used. By default `copula="studentT"`
specifies the Student-$t$ copula. See Details.
- `par` A list of $p$ parameters to be specified for the multivariate parametric family of
distributions. See Details.
Details

For a joint multivariate distribution with a given parametric copula class (copula) and a given parametric family of equal marginal distributions (dist), a sample of size ndata is simulated.

- The available copula classes are: Student-t (copula="studentT") with $\nu > 0$ degrees of freedom (df) and scale parameters $\rho_{i,j} \in (-1,1)$ for $i \neq j = 1,\ldots,d$ (sigma), Gaussian (copula="Gaussian") with correlation parameters $\rho_{i,j} \in (-1,1)$ for $i \neq j = 1,\ldots,d$ (sigma), Clayton (copula="Clayton") with dependence parameter $\theta > 0$ (dep), Gumbel (copula="Gumbel") with dependence parameter $\theta \geq 1$ (dep) and Frank (copula="Frank") with dependence parameter $\theta > 0$ (dep).

- The available families of marginal distributions are:
  - Student-t (dist="studentT") with $\nu > 0$ degrees of freedom (df);
  - Asymmetric Student-t (dist="AstudentT") with $\nu > 0$ degrees of freedom (df). In this case all the observations are only positive;
  - Frechet (dist="Frechet") with scale $\sigma > 0$ (scale) and shape $\alpha > 0$ (shape) parameters.
  - Frechet (dist="double-Frechet") with scale $\sigma > 0$ (scale) and shape $\alpha > 0$ (shape) parameters. In this case positive and negative observations are allowed;
  - symmetric Pareto (dist="double-Pareto") with scale $\sigma > 0$ (scale) and shape $\alpha > 0$ (shape) parameters. In this case positive and negative observations are allowed.

- The available classes of multivariate joint distributions are:
  - studentT-studentT (dist="studentT" and copula="studentT") with parameters par <-list(df,sigma);
  - studentT (dist="studentT" and copula="None" with parameters par <-list(df,dim). In this case the d variables are regarded as independent;
  - studentT-AstudentT (dist="AstudentT" and copula="studentT") with parameters par <-list(df,sigma,shape);
  - Gaussian-studentT (dist="studentT" and copula="Gaussian") with parameters par <-list(df,sigma);
  - Gaussian-AstudentT (dist="AstudentT" and copula="Gaussian") with parameters par <-list(df,sigma,shape);
  - Frechet (dist="Frechet" and copula="None") with parameters par <-list(shape,dim). In this case the d variables are regarded as independent;
  - Clayton-Frechet (dist="Frechet" and copula="Clayton") with parameters par <-list(dep,dim,scale,shape);
  - Gumbel-Frechet (dist="Frechet" and copula="Gumbel") with parameters par <-list(dep,dim,scale,shape);
  - Frank-Frechet (dist="Frechet" and copula="Frank") with parameters par <-list(dep,dim,scale,shape);
  - Clayton-double-Frechet (dist="double-Frechet" and copula="Clayton") with parameters par <-list(dep,dim,scale,shape);
  - Gumbel-double-Frechet (dist="double-Frechet" and copula="Gumbel") with parameters par <-list(dep,dim,scale,shape);
  - Frank-double-Frechet (dist="double-Frechet" and copula="Frank") with parameters par <-list(dep,dim,scale,shape);
  - Clayton-double-Pareto (dist="double-Pareto" and copula="Clayton") with parameters par <-list(dep,dim,scale,shape);
- Gumbel-double-Pareto (dist="double-Pareto" and copula="Gumbel") with parameters `par <- list(dep, dim, scale, shape);`
- Frank-double-Pareto (dist="double-Pareto" and copula="Frank") with parameters `par <- list(dep, dim, scale, shape).

Note that above `dim` indicates the number of d marginal variables.

**Value**

A matrix of \((n \times d)\) observations simulated from a specified multivariate parametric joint distribution.

**Author(s)**


**References**


**See Also**

rtimeseries, rbtimeseries

**Examples**

```r
library(plot3D)
library(copula)
library(evd)

# Data simulation from a 3-dimensional random vector a with multivariate distribution # given by a Gumbel copula and three equal Frechet marginal distributions

# distributional setting
copula <- "Gumbel"
dist <- "Frechet"

# parameter setting
dep <- 3
dim <- 3
scale <- rep(1, dim)
shape <- rep(3, dim)
par <- list(dep=dep, scale=scale, shape=shape, dim=dim)

# sample size
ndata <- 1000

# Simulates a sample from a multivariate distribution with equal Frechet
```
rtimeseries

Simulation of One-Dimensional Temporally Dependent Observations

Description

Simulates samples from parametric families of time series models.

Usage

rtimeseries(ndata, dist="studentT", type="AR", par, burnin=1e+03)

Arguments

ndata A positive integer specifying the number of observations to simulate.
dist A string specifying the parametric family of the innovations distribution. By default dist="studentT" specifies a Student-t family of distributions. See Details.
type A string specifying the type of time series. By default type="AR" specifies a linear Auto-Regressive time series. See Details.

par A vector of \((1 \times p)\) parameters to be specified for the univariate time series parametric family. See Details.

burnin A positive integer specifying the number of initial observations to discard from the simulated sample.

Details

For a time series class (type) with a parametric family (dist) for the innovations, a sample of size ndata is simulated. See for example Brockwell and Davis (2016).

• The available categories of time series models are: Auto-Regressive (type="AR"), Auto-Regressive and Moving-Average (type="ARMA"), Generalized-Autoregressive-Conditional-Heteroskedasticity (type="GARCH") and Auto-Regressive and Moving-Maxima (type="ARMAX").

• With AR(1) and ARMA(1,1) times series the available families of distributions for the innovations are:
  
  – Student-t (dist="studentT") with parameters: \(\phi \in (-1,1)\) (autoregressive coefficient), \(\nu > 0\) (degrees of freedom) specified by par=c(corr,df);
  
  – symmetric Frechet (dist="double-Frechet") with parameters \(\phi \in (-1,1)\) (autoregressive coefficient), \(\sigma > 0\) (scale), \(\alpha > 0\) (shape), \(\theta\) (movingaverage coefficient), specified by par=c(corr,scale,shape,smooth);
  
  – symmetric Pareto (dist="double-Pareto") with parameters \(\phi \in (-1,1)\) (autoregressive coefficient), \(\sigma > 0\) (scale), \(\alpha > 0\) (shape), \(\theta\) (movingaverage coefficient), specified by par=c(corr,scale,shape,smooth).

With ARCH(1)/GARCH(1,1) time series the Gaussian family of distributions is available for the innovations (dist="Gaussian") with parameters, \(\alpha_0, \alpha_1, \beta\) specified by par=c(alpha0,alpha1,beta). Finally, with ARMAX(1) times series the Frechet families of distributions is available for the innovations (dist="Frechet") with parameters, \(\phi \in (-1,1)\) (autoregressive coefficient), \(\sigma > 0\) (scale), \(\alpha > 0\) (shape) specified by par=c(corr,scale,shape).

Value

A vector of \((1 \times n)\) observations simulated from a specified time series model.

Author(s)

Simone Padoan, <simone.padoan@unibocconi.it>, http://mypage.unibocconi.it/simonepadoan/;

References


See Also

expectiles

Examples

# Data simulation from a 1-dimensional AR(1) with univariate Student-t
# distributed innovations

tsDist <- "studentT"
tsType <- "AR"

# parameter setting
corr <- 0.8
df <- 3
par <- c(corr, df)

# sample size
ndata <- 2500

# Simulates a sample from an AR(1) model with Student-t innovations
data <- rtimeseries(ndata, tsDist, tsType, par)

# Graphic representation
plot(data, type="l")
acf(data)

sp500

Negative log-returns of S&P 500.

Description

Series of negative log-returns of the U.S. stock market index Standard and Poor 500.

Format

A 8784 * 2 data frame.

Details

From the series of n = 8785 closing prices $S_t, t = 1, 2, ..., for the Standard and Poor 500 stock market index, recorded from January 29, 1985 to December 12, 2019, the series of negative log-returns.

$X_{t+1} = -\log(S_{t+1}/S_t), \quad 1 \leq t \leq n - 1$

is available. Hence the dataset (negative log-returns) contains 8784 observations.
Index

* datasets
  dowjones, 2
  sp500, 55
dowjones, 2
EBTailIndex, 3, 5, 6, 8, 24, 26, 32, 34, 38–40
estExpectiles, 4, 9, 14, 24, 40
estExtLevel, 7, 17, 39, 46
estMultiExpectiles, 10, 20, 44
expectiles, 13, 49, 55
ExpectMES, 15, 47
extMultiQuantile, 12, 18, 29, 44
extQuantile, 6, 9, 17, 21, 40, 47
HTailIndex, 4–6, 8, 16, 17, 22, 24, 25, 32, 34, 36, 38–40, 45–47
HypoTesting, 27
MLTailIndex, 4, 8, 26, 31, 34
MomTailIndex, 4, 8, 26, 32, 33
MultiHTailIndex, 10–12, 19, 20, 28, 29, 35, 42, 44
predExpectiles, 6, 9, 16, 17, 37, 45–47
predMultiExpectiles, 12, 20, 29, 41
QuantMES, 17, 45
rbtimeseries, 48, 52
rmdata, 36, 50
rtimeseries, 14, 49, 52, 53
sp500, 55