Package 'covmat'

September 28, 2015

Type Package

Title Covariance Matrix Estimation

Version 1.0

Date 2015-09-28

Author Rohit Arora

Maintainer Rohit Arora <emailrohitarora@gmail.com>

Description We implement a collection of techniques for estimating covariance matrices. Covariance matrices can be built using missing data. Stambaugh Estimation and FMMC methods can be used to construct such matrices. Covariance matrices can be built by denoising or shrinking the eigenvalues of a sample covariance matrix. Such techniques work by exploiting the tools in Random Matrix Theory to analyse the distribution of eigenvalues. Covariance matrices can also be built assuming that data has many underlying regimes. Each regime is allowed to follow a Dynamic Conditional Correlation model. Robust covariance matrices can be constructed by multivariate cleaning and smoothing of noisy data.

License Artistic-2.0

Suggests knitr, knitcitations, roxygen2, quantmod, PortfolioAnalytics, rmgarch

VignetteBuilder knitr

LazyLoad yes

Imports zoo, xts, robust, robustbase, VIM, ggplot2, reshape2, Matrix, parallel, doParallel, fGarch, lhs, scales, gridExtra, optimx, DEoptim, foreach

Depends mvtnorm, RMTstat, grid

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-09-28 18:46:22

R topics documented:

compareCov . . . 2
dow30data . . . 3
estRMT . . . 3
estSpikedCovariance . . . 4
etfdata . . . 5
factor.data . . . 6
isdccfit . . . 6
missingdata . . . 7
plot.isdcc . . . 8
plot.RMT . . . 8
plotmissing . . . 9
plotSpikedCovariance . . . 10
rmtdata . . . 10
robustMultExpSmoothing . . . 11
smoothing.matrix . . . 12

Index **13**

`compareCov` *This is a utility function to compare two covariance matrices*

Description

This is a utility function to compare two covariance matrices

Usage

```
compareCov(cov1, cov2, labels, corr = FALSE)
```

Arguments

`cov1`	covariance matrix using the first method
`cov2`	covariance matrix using the second method
`labels`	strings indicating the type of methods used for comparison
`corr`	flag indicating if the supplied matrices are of type covariance or correlation

Details

This method takes in two different covariance/correlation matrices computed using two different methods and visullay compares them using the ellipse plot. It produces a matrix with ellipses drawn in the upper triangle. The ellipse is drawn to be a contour of a standard bivariate normal with correlation given by the correlation of the two assets. One ellipse is drawn for each covariance matrix.

Author(s)

Rohit Arora

dow30data *Symbol Data*

Description

Contains returns for 30 symbols found in Dow Jones 30 starting from 2014-04-01 to 2015-07-23. Data maybe updated using the getSymbolData script in the inst folder

Usage

```
dow30data
```

Source

Yahoo finance

estRMT *Denoising of Covariance matrix using Random Matrix Theory*

Description

Denoising of Covariance matrix using Random Matrix Theory

Usage

```
estRMT(R, Q = NA, cutoff = c("max", "each"), eigenTreat = c("average",
  "delete"), numEig = 1, parallel = TRUE)
```

Arguments

R	xts or matrix of asset returns
Q	ratio of rows/size. Can be supplied externally or fit using data
cutoff	takes two values max/each. If cutoff is max, Q is fitted and cutoff for eigenvalues is calculated. If cutoff is each, Q is set to row/size. Individual cutoff for each eigenvalue is calculated and used for filteration.
eigenTreat	takes 2 values, average/delete. If average then the noisy eigenvalues are averged and each value is replaced by average. If delete then noisy eigenvalues are ignored and the diagonal entries of the correlation matrix are replaced with 1 to make the matrix psd.
numEig	number of eigenvalues that are known for variance calculation. Default is set to 1. If numEig = 0 then variance is assumed to be 1.
parallel	boolean to use all cores of a machine.

Details

This method takes in data as a matrix or an xts object. It then fits a marchenko pastur density to eigenvalues of the correlation matrix. All eigenvalues above the cutoff are retained and ones below the cutoff are replaced such that the trace of the correlation matrix is 1 or non-significant eigenvalues are deleted and diagonal of correlation matrix is changed to 1. Finally, correlation matrix is converted to covariance matrix.

Author(s)

Rohit Arora

Examples

```
## Not run:
 data("largereturn")
 model <- estRMT(largesymdata, numEig = 0)

## End(Not run)
```

---

estSpikedCovariance *(Donoho, Gavish, and Johnstone, 2013)*

Description

(Donoho, Gavish, and Johnstone, 2013)

Usage

```
estSpikedCovariance(R, gamma = NA, numOfSpikes = NA, method = c("KNTest",
  "median-fitting"), norm = c("Frobenius", "Operator", "Nuclear"),
  pivot = 1, statistical = NA, fit = NA)
```

Arguments

R	xts object of asset returns
gamma	ratio of varibales/observations. If NA it will be set to ratio of varibales/observations
numOfSpikes	number of spikes in the spike covariance model. If missing then it is estimated based on the number of eigenvalues above the cutoff.
method	KNTest/median-fitting. Default is KNTest
norm	Type of matrix norm that must be calculated. Defaults to Frobenius
pivot	takes values from 1...7. Details can be found in the paper
statistical	Stein/Entropy/Divergence/Affinity/Frechet. Default is set to NA. when a valid value is set norm and pivot values are ignored
fit	list with 5 elements, cutoff for the bulk of MP distribution, scaled lambdas, fitted gamma fitted scaling constant and numOfSpikes

Details

If the number of spikes are missing and the selected method is median-fitting then, firstly the scale factor is estimated. We guess the number of spikes by counting the number of breaks using the Freedman-Diaconis algorithm. The initial number of spikes are guessed by counting the number of elements after the first zero. We use this to lower bound the variance. Variance is computed by fitting the median to the spectrum of eigenvalues. If the method is KNTest then we follow the procedure in (Kritchman and Nadler, 2009)

Author(s)

Rohit Arora

Examples

```
## Not run:
 data("rmtdata")
 model <- estSpikedCovariance(rmtdata, numOfSpikes=15)

## End(Not run)
```

`etfdata` *Symbol Data*

Description

Contains returns for 9 etf symbols XLE, XLY, XLP, XLF, XLV,XLI, XLB, XLK,XLU starting from 2008-01-01 to 2010-12-31. Data maybe updated using the getSymbolData script in the inst folder

Usage

```
etfdata
```

Source

Yahoo finance

factor.data *Factor Data*

Description

Contains 8 factors. 5 Fama-French factors. 1 Momentum factor. 1 Liquidity factor, 1 Volatility factor.

Usage

```
factor.data
```

Source

Ken.French Website, Lubos Pastor website, Yahoo finance

isdccfit *Fit an Independent Regime Switching Model*

Description

Fit an Independent Regime Switching Model

Usage

```
isdccfit(R, numRegimes = NA, transMatbounds = c(2, 10), dccBounds = c(0,
  1), w = NA, ...)
```

Arguments

R	xts object of asset returns
numRegimes	number of regimes to fit to the data
transMatbounds	bounds on the parameter tau as described in (Lee, 2010). Each paramter is defaulted to lie in the range (2,10)
dccBounds	bounds on the paramter theta as described in (Lee, 2010). Each paramter is defaulted to lie in the range (0,1)
w	proportion of entries to consider in initializing correlation for for each regime. It is defualted to split data equally across all regimes
...	addition control paramters that can be passed to the control object in DEoptim

Details

This method takes in returns data and the number of regimes and fits sepearate covariances to each regime using the Expectation Maximization algorithm decribed in (Lee, 2010). IS-DCC model avoids the path dependency problem observed in other regime switching models and makes the solution more tractable by running a separate DCC process for each regime.

Fitting the IS-DCC model to data corresponds works in two steps. In the first step a time varying univariate volatility process, GARCH(1,1) is fitted to each time series. In the second step DCC parameters for each state are estimated along with the transition probabilities corresponding to the Hidden Markov model. This is done by maximising the log-likelihood of observing the residuals

Author(s)

Rohit Arora

References

Lee, H.-T. (2010). Regime switching correlation hedging. Journal of Banking &

Examples

```
## Not run:
 data("largereturn")
 model <- isdccfit(largesymdata, numRegimes=2, maxiter=50, paralletType=0)

## End(Not run)
```

`missingdata` *Symbol Data*

Description

Contains returns for 7 symbols ie BABA, TWTR, LNKD, YHOO, GE, LAZ, V starting from 2007-04-01 to 2015-07-02. Data maybe updated using the getSymbolData script in the inst folder

Usage

```
missingdata
```

Source

Yahoo finance

plot.isdcc *Implied State plot*

Description

Implied State plot

Usage

```
## S3 method for class 'isdcc'
plot(x, y = c(1, 2), ...)
```

Arguments

x	model of the type isdcc obtained by fitting an IS-DCC model to the data
y	type of plot. takes values 1/2. 1 = Implied States, 2 = Smoothed Proabability
...	additional arguments unused

Details

Plot implied states using the fitted Independent Switching DCC model

Author(s)

Rohit Arora

Examples

```
## Not run:
 data("largereturn")
 model <- isdccfit(largesymdata, numRegimes=2)
 plot(model)

## End(Not run)
```

plot.RMT *Eigenvalue plot*

Description

Eigenvalue plot

Usage

```
## S3 method for class 'RMT'
plot(x, y, ...)
```

Arguments

x	model of the type RMT obtained by fitting an RMT model to the data
y	unused
...	additional arguments unused

Details

Plots eigenvalues of the correlation matrix and overlays the Marchenko-Pastur density on top of it. There is a shap cutoff for the density. We are concerned with eigenvalues beyond this cutoff. Paramters used for plotting are added to the plot

Author(s)

Rohit Arora

Examples

```
## Not run:
 data("largereturn")
 model <- estRMT(largesymdata)
 plot(model)

## End(Not run)
```

---

plotmissing *Plot data to visualize missing values*

Description

Plot data to visualize missing values

Usage

```
plotmissing(data, which = c(3, 4))
```

Arguments

data	an xts/zoo object
which	takes values 3/4. 3 = Summary plot, 4 = Matrix plot

Details

This method takes in data as an xts object and plots the data. Missing values highlighted in red for matrix plot and time series of returns are shown in in Summary plot

Author(s)

Rohit Arora

plotSpikedCovariance *Eigenvalue plot. Similar to figure 1 in the paper*

Description

Eigenvalue plot. Similar to figure 1 in the paper

Usage

```
plotSpikedCovariance(R, gamma = NA, numOfSpikes = NA, method = "KNTest",
  ...)
```

Arguments

R	xts object of asset returns
gamma	ratio of varibales/observations. If NA it will be set to ratio of varibales/observations
numOfSpikes	model of the type spikedCovariance
method	KNTest/median-fitting. Default is KNTest
...	additional arguments unused

Details

Compares shrunk eigenvalues against sample eigenvalues for all norms and losses

Author(s)

Rohit Arora

Examples

```
## Not run:
 data("rmtdata")
 plot(rmtdata, numOfSpikes=10)

## End(Not run)
```

rmtdata *Simulated data for Spiked Covarianve Model*

Description

Contains large data for 500 observations and 100 variables based on standard normal independent random variables. It has 15 spikes. The number of spikes can be correctly detected using the fitting procedure as described in the code.

Usage

```
rmtdata
```

`robustMultExpSmoothing` *Robust Multivariate Exponential Smoothing*

Description

Robust Multivariate Exponential Smoothing

Usage

```
robustMultExpSmoothing(R, smoothMat = NA, startup_period = 10,
  training_period = 60, seed = 9999, trials = 50, method = "L-BFGS-B",
  lambda = 0.2)
```

Arguments

`R`	data
`smoothMat`	Optimal smoothing matrix. If missing it is estimated. The procedure maybe very slow for high-dimensional data. Also, the objective function being very noisy, optimization across multiple runs may lead to different smoothing matrices. #'
`startup_period`	length of samples required to calculate initial values
`training_period`	length of samples required to calculate forecast errors for evalualating the objective if smoothing matrix is estimated
`seed`	random seed to replicate the starting values for optimization
`trials`	number of strarting values to try for any optimization. Large number of trials for high dimensions can be time consuming
`method`	optimization method to use to evaluate an estimate of smoothing matrix. Default is L-BFGS-B
`lambda`	known constant as described in the paper. Defaults to 0.2

Details

Calculate Robust estimate of covariance matrix while also smoothing and cleaning the data using the procedure described in (Croux, Gelper, and Mahieu, 2010)

Author(s)

Rohit Arora

`smoothing.matrix` *Optimal Smoothing Matrix*

Description

Optimal Smoothing Matrix

Usage

```
smoothing.matrix(R, startup_period = 10, training_period = 60,
  seed = 9999, trials = 50, method = "L-BFGS-B", lambda = 0.2)
```

Arguments

`R`	data
`startup_period`	length of samples required to calculate initial values
`training_period`	length of samples required to calculate forecast errors for evalualating the objective
`seed`	random seed to replicate the starting values for optimization
`trials`	number of strarting values to try for any optimization. Large number of trials for high dimensions can be time consuming
`method`	optimization method to use to evaluate an estimate of smoothing matrix. Default is L-BFGS-B
`lambda`	known constant as described in the paper. Defaulted to 0.2

Details

Calcuation of smoothing matrix is done by assuming that the smoothing matrix is symmetrix and has a spectral decomposition. The orthogonal matrix in the decomposition is calculated using the product of givens rotation matrices and requires d(d-1)/2 angles for a d dimensional matrix. The eigenvalues are restricted to lie in [0,1].

Author(s)

Rohit Arora

Index

compareCov, 2

dow30data, 3

estRMT, 3
estSpikedCovariance, 4
etfdata, 5

factor.data, 6

isdccfit, 6

missingdata, 7

plot.isdcc, 8
plot.RMT, 8
plotmissing, 9
plotSpikedCovariance, 10

rmtdata, 10
robustMultExpSmoothing, 11

smoothing.matrix, 12