Package 'factorcpt'

December 26, 2016

Type Package

Title Simultaneous Change-Point and Factor Analysis

Version 0.1.2

Author Haeran Cho [aut, cre],
Matteo Barigozzi [aut],
Piotr Fryzlewicz [aut]

Maintainer Haeran Cho <haeran.cho@bristol.ac.uk>

Description
Identifies change-points in the common and the idiosyncratic components via factor modelling.

License GPL-2

LazyData TRUE

Depends R (>= 2.14.0)

Suggests RcppArmadillo

Imports Rcpp (>= 0.12.4), foreach, parallel, iterators, fields,
doParallel

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-12-26 12:25:11

R topics documented:

`factorcpt-package` *Simultaneous multiple change-point and factor analysis for high-dimensional time series*

Description

The package implements a two-stage methodology for consistent multiple change-point detection under factor modelling. It performs multiple change-point analysis on the common and idiosyncratic components separately, and thus automatically identifies their origins. The package also implements the Double CUSUM Binary Segmentation algorithm, which is proposed for multiple change-point detection in high-dimensional panel data with breaks in its mean.

Details

Package:	factorcpt
Type:	Package
Version:	0.1.2
Date:	2016-12-22
License:	GPL (>= 2)

The main routines of the package are `factor.seg.alg` and `func_dc`.

Author(s)

Haeran Cho, Matteo Barigozzi, Piotr Fryzlewicz

Maintainer: Haeran Cho <haeran.cho@bristol.ac.uk>

References

M. Barigozzi, H. Cho and P. Fryzlewicz (2016) Simultaneous multiple change-point and factor analysis for high-dimensional time series, Preprint.

H. Cho (2016) Change-point detection in panel data via double CUSUM statistic. Electronic Journal of Statistics. 10: 2000-2038.

`factor.seg.alg` *Simultaneous multiple change-point and factor analysis for high-dimensional time series*

Description

The function applies the two-stage methodology for multiple change-point detection under factor modelling. It first transforms the input time series into panels of statistics that contain the change-points in the second-order structure of the common and idiosyncratic components, as change-points in their 'means', to which the Double CUSUM Binary Segmentation algorithm is applied in the second stage. The function returns change-point estimates from the common and idiosyncratic components separately.

Usage

```
factor.seg.alg(x, r=NULL, bn.op=2, sig.lev=.05, max.q=NULL, q.seq=NULL,
               qlen=5, qby=0, dw=NULL, p=NULL, B=200, scales=NULL,
               rule=NULL, mby=NULL, tby=NULL, idio.diag=FALSE,
               do.parallel=TRUE, no.proc=2)
```

Arguments

x	input time series matrix, with each row representing a time series
r	the number of factors, if r=NULL, screening over a range of factor number candidates is performed as described in the paper
bn.op	an index number for the information criterion-based estimator of Bai and Ng (2002) for the number of factors, the default value bn.op=2 is used in Barigozzi, Cho & Fryzlewicz (2016)
sig.lev	sets the level of significance for drawing the bootstrap-based threshold
max.q	the maximum number of factors, if max.q=NULL, a default value is chosen as described in the paper
q.seq	a vector of factor number candidates; if q.seq=NULL, it is chosen as described in Barigozzi, Cho & Fryzlewicz (2016) using qlen or qby
qlen	specifies the length of the sequence of factor number candidates
qby	specifies the increment of the sequence of factor number candidates when qlen=0; if both qlen=0 and qby=0, the sequence of factor number candidates is generated as if qby=1
dw	trims off the interval of consideration in the binary segmentation algorithm and determines the minimum length of a stationary segment; if dw=NULL, a default value is chosen as described in the Appendix of Barigozzi, Cho & Fryzlewicz (2016)
p	inverse of the average length of blocks in the stationary bootstrap procedure, if p=NULL, a data-driven value is chosen as described in the Appendix of Barigozzi, Cho & Fryzlewicz (2016)
B	the number of bootstrap samples for threshold selection
scales	the number of wavelet scales used for wavelet filtering of the common and idiosyncratic components estimated via PCA
rule	the depth of a binary tree for change-point analysis, see the Appendix of Barigozzi, Cho & Fryzlewicz (2016)

mby	see dmby in func_dc_by
tby	see dtby in func_dc_by
idio.diag	if idio.diag=TRUE, only the diagonal wavelet-transform is employed in order to generate the panel of statistics from the idiosyncratic components
do.parallel	if do.parallel=TRUE, a set of copies of R running in parallel are created and used for bootstrap procedure
no.proc	sets the number of processes to be spawned when do.parallel=TRUE,

Value

cs.list	a list of objects returned by an internal function common.seg, which contain the results from performing change-point analysis on the common components estimated with a range of factor number candidates
r	the factor number selected from performing change-point analysis on the common component
common.est.cps	change-points detected from the common component estimated with r as the factor number
idio.seg.res	an object returned by an internal function idio.seg, which contains the results from performing change-point analysis on the idiosyncratic component estimated with r as the factor number
idio.est.cps	change-points detected from the idiosyncratic component
gfm	factor structure of x estimated from get.factor.model
q.seq	a vector containing the range of factor number candidates

Author(s)

Haeran Cho

References

J. Bai and S. Ng (2002) Determining the number of factors in approximate factor models. Econometrica. 70: 191-221.

M. Barigozzi, H. Cho and P. Fryzlewicz (2016) Simultaneous multiple change-point and factor analysis for high-dimensional time series, Preprint.

H. Cho (2016) Change-point detection in panel data via double CUSUM statistic. Electronic Journal of Statistics. 10: 2000-2038.

Examples

```
n <- 50; T <- 200
e <- matrix(rnorm(n*T), nrow=n) # idiosyncratic components
r <- 3 # factor number
Lam <- matrix(rnorm(n*r, 1, 1), nrow=n) # loadings
f <- matrix(rnorm(r*T), nrow=r) # factors
```

```
chi <- e*0 # common component
chp <- T/2 # change-point
chi[, 1:chp] <- Lam%*%f[, 1:chp]
Lam <- Lam + matrix(rnorm(n*r, 0, sqrt(2)), nrow=n) # new loadings
chi[, (chp+1):T] <- Lam%*%f[, (chp+1):T]
x <- chi + sqrt(r)*e

fsa <- factor.seg.alg(x, idio.diag=TRUE)
fsa$common.est.cps
fsa$idio.est.cps
fsa$q.seq
```

---

func_dc *Double CUSUM statistics*

Description

Aggregates high-dimensional CUSUM statistic series via Double CUSUM approach.

Usage

```
func_dc(z)
```

Arguments

z	input data matrix

Value

cs	a matrix of CUSUM series
acs	a matrix of absolute values of CUSUM series
res	pointwise maximum of Double CUSUM statistics
mat	a matrix of Double CUSUM statistics

Author(s)

Haeran Cho

References

H. Cho (2016) Change-point detection in panel data via double CUSUM statistic. Electronic Journal of Statistics. 10: 2000-2038.

Examples

```
e <- matrix(rnorm(100*100), nrow=100)
f <- matrix(0, nrow=100, ncol=100)
f[, 51:100] <- 1 # t = 50 represents the change-point in the means
x <- f + e
fd <- func_dc(x)
plot(fd$res, type='l'); abline(v=50, col=2) # pointwise maximum of DC statistics
image(t(fd$mat)) # heatmap of DC statistics with x-axis representing the time
```

`get.factor.model` *Factor model estimation via Principal Component Analysis*

Description

Estimates the components of the factor structure for an input time series, such as loadings and factors, as well as estimating the number of factors.

Usage

```
get.factor.model(x, max.q = NULL, q = NULL, bn = TRUE,
                 bn.op = 2, normalisation = TRUE)
```

Arguments

`x`	input time series matrix, with each row representing a time series
`max.q`	see `max.q` in `factor.seg.alg`
`q`	the number of factors; if `bn=FALSE`, `q` needs to be given
`bn`	if `bn=TRUE`, the factor number is estimated using the information criterion-based estimator of Bai and Ng (2002)
`bn.op`	an index number for the information criterion-based estimator of Bai and Ng (2002); the default value `bn.op=2` is used in Barigozzi, Cho & Fryzlewicz (2016)
`normalisation`	if `normalisation=TRUE`, each row `x` is standardised prior to PCA

Value

`lam`	an `n`-by-`(min(n, T)-1)` matrix containing the estimated loadings
`f`	a `(min(n, T)-1)`-by-`T` matrix containing the estimated factors
`norm.x`	if `normalisation=TRUE`, row-wise standardised `x`; else `norm.x=x`
`q.hat`	estimated number of factors
`max.q`	the maximum factor number used for factor number estimation
`ic`	information criterion values computed at a range of factor numbers from `0` to `max.q`

Author(s)

Haeran Cho

References

J. Bai and S. Ng (2002) Determining the number of factors in approximate factor models. Econometrica. 70: 191-221.

M. Barigozzi, H. Cho and P. Fryzlewicz (2016) Simultaneous multiple change-point and factor analysis for high-dimensional time series, Preprint.

Examples

```
n <- 50; T <- 200
e <- matrix(rnorm(n*T), nrow=n) # idiosyncratic components
r <- 3 # factor number
Lam <- matrix(rnorm(n*r, 1, 1), nrow=n) # loadings
f <- matrix(rnorm(r*T), nrow=r) # factors
chi <- e*0 # common component
chp <- T/2 # change-point
chi <- Lam%*%f
x <- chi + sqrt(r)*e

gfm <- get.factor.model(x)
gfm$q.hat
```

---

post.cpts.analysis *Display the results from change-point analysis and factor modelling*

Description

Computes the minimum number of eigenvalues required so that the given proportion of the variance of x over each segment, defined by the change-points detected from the common components, by varying the proportion and plots the number of eigenvalues.

Usage

```
post.cpts.analysis(x, est.cps, cutoff.seq = seq(0.5, 0.95, by = 0.05), do.plot = TRUE)
```

Arguments

x	input time series
est.cps	a vector containing the change-points estimated for the common components
cutoff.seq	a sequence containing the proportions between zero and one
do.plot	if do.plot=TRUE, produces a plot of minimum numbers of eigenvalues required for accounting for the given proportion of variance, over time (x-axis) and the level of proportion (y-axis).

Value

a matrix containing the minimum numbers of eigenvalues for varying levels of proportions over the segments defined by two neighbouring change-points in the common components

Author(s)

Haeran Cho

References

M. Barigozzi, H. Cho and P. Fryzlewicz (2016) Simultaneous multiple change-point and factor analysis for high-dimensional time series, Preprint.

Examples

```
n <- 50; T <- 200
e <- matrix(rnorm(n*T), nrow=n) # idiosyncratic components
r <- 3 # factor number
Lam <- matrix(rnorm(n*r, 1, 1), nrow=n) # loadings
f <- matrix(rnorm(r*T), nrow=r) # factors
chi <- e*0 # common component
chp <- T/2 # change-point
chi[, 1:chp] <- Lam%*%f[, 1:chp]
Lam <- Lam + matrix(rnorm(n*r, 0, sqrt(2)), nrow=n) # new loadings
chi[, (chp+1):T] <- Lam%*%f[, (chp+1):T]
x <- chi + sqrt(r)*e

post.cpts.analysis(x, 100)
```

Index