Package 'fvarseg'

February 23, 2022

Type Package

Title High-dimensional Time Series Segmentation via Factor-adjusted Vector Autoregressive Modelling

Version 0.1.0

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

Description Implements a two-stage time series segmentation methodology that first detects change points in the factor-driven common component, and then detects change points in the idiosyncratic vector autoregressive process.

Depends R (>= 4.1.0)

Imports lpSolve, parallel, doParallel, foreach

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

R topics documented:

`common.seg` *Segment factor-driven common component*

Description

Segment factor-driven common component

Usage

```
common.seg(
  x,
  center = TRUE,
  G.seq = NULL,
  thr = NULL,
  tt.by = floor(2 * log(dim(x)[2])),
  eta = 0.5
)
```

Arguments

x	input time series matrix, with each row representing a variable
center	whether to de-mean the input x row-wise
G.seq	an integer vector of bandwidth; if G.seq = NULL, a default choice G.seq = round(n * 1/c(10,8,6,4)) is used
thr	a vector of thresholds which is of the same length as G.seq; if thr = NULL, a default choice based on simulations is used
tt.by	an integer specifying the grid over which the test statistic is computed, which is round(seq(G,dim(x)[2] -G,by = tt.by)) for each bandwidth G
eta	a constant between 0 and 1; each local maximiser of the test statistic within its eta * G-environment for the common component is deemed as a change point estimator. Also the bottom-up merging across the multiple bandwidths G.seq depends on this parameter

Details

See Algorithm 1 of Cho, Eckley, Fearnhead and Maeng (2022) for further details.

Value

a list containing the following fields:

est.cp	a matrix containing the change point estimators in the first column and the finest bandwidth at which each is detected in the second column
G.seq	an integer vector of bandwidths
thr	a vector of thresholds which is of the same length as G.seq
est.cp.list	a list containing the following fields: • cp change point estimators • G bandwidth • ll kernel window size for spectral density estimation • norm.stat a matrix containing test statistic values at Fourier frequencies • stat a vector containing test statistic values across multiple frequencies
mean.x	if center = TRUE, returns a vector containing row-wise sample means of x; if center = FALSE, returns a vector of zeros

References

Cho, H., Eckley, I., Fearnhead, P. & Maeng, H. (2022) High-dimensional time series segmentation via factor-adjusted vector autoregressive modelling. arXiv preprint arXiv: TODO

Examples

```
## Not run:
out <- sim.data(n = 2000, p = 50, q = 2, d = 1,
cp.common = 1:3/4, den.common = .5, type.common = 'ma',
cp.idio = c(3, 5)/8, seed = 123)
cs <- common.seg(out$x)
cs$est.cp

## End(Not run)
```

`fvar.seg` *Segment factor-adjusted VAR process*

Description

Segment high-dimensional time series using the two-stage segmentation method proposed in Cho, Eckley, Fearnhead and Maeng (2022). It first detects change points from the factor-driven common component, then from the idiosyncratic VAR process.

Usage

```
fvar.seg(
  x,
  center = TRUE,
  q = NULL,
  d = 1,
  eta = 0.5,
  common.args = list(G.seq = NULL, thr = NULL, tt.by = floor(2 * log(dim(x)[2]))),
  idio.args = list(G.seq = NULL, thr = NULL),
  cv.args = list(path.length = 10, n.folds = 1, do.cv = FALSE)
)
```

Arguments

`x`	input time series matrix, with each row representing a variable
`center`	whether to de-mean the input `x` row-wise
`q`	an integer specifying the number of factors. If `q = NULL`, the factor number is estimated by an information criterion-based approach of Hallin and Liška (2007) for each segment
`d`	an integer specifying the VAR order
`eta`	a constant between `0` and `1`; each local maximiser of the test statistic within its `eta * G`-environment for the common component is deemed as a change point estimator. Also the bottom-up merging across the multiple bandwidths `G.seq` depends on this parameter
`common.args`	a list specifying the tuning parameters required for segmenting the factor-driven common component, see also common.seg. It contains • `G.seg` an integer vector of bandwidth; see `fvarseg`[common.seg] for the default choice when `G.seq = NULL` • `thr` a vector of thresholds which is of the same length as `G.seq`; if `thr = NULL`, a default choice based on simulations is used

- `tt.by` an integer specifying the grid over which the test statistic is computed, which is `round(seq(G,dim(x)[2]-G,by = tt.by))` for each bandwidth `G`

`idio.args` a list specifying the tuning parameters required for segmenting the idiosyncratic VAR process, see also idio.seg. It contains

- `G.seg` an integer vector of bandwidth; see `fvarseg`[idio.seg] for the default choice when `G.seq = NULL`
- `thr` a vector of thresholds which is of the same length as `G.seq`; if `thr = NULL`, a default choice based on simulations is used

`cv.args` a list specifying the tuning parameters required for Dantzig selector tuning parameter selection via cross-validation. It contains:

- `n.folds` number of folds
- `path.length` number of regularisation parameter values to consider; a sequence is generated automatically based in this value
- `do.cv` if `do.cv = FALSE`, a fixed value is selected from a sequence of 10 values chosen in a data-driven way

Details

See Cho, Eckley, Fearnhead and Maeng (2022) for further details.

Value

a list containing the following fields:

`common.out, idio.out`
output from common.seg and idio.seg

- `est.cp` a matrix containing the change point estimators in the first column and the finest bandwidth at which each is detected in the second column
- `G.seq` an integer vector of bandwidths
- `thr` a vector of thresholds which is of the same length as `G.seq`
- `est.cp.list` a list containing various quantities related to the segmentation; see common.seg and idio.seg for further details

`mean.x` if `center = TRUE`, returns a vector containing row-wise sample means of `x`; if `center = FALSE`, returns a vector of zeros

References

Cho, H., Eckley, I., Fearnhead, P. & Maeng, H. (2022) High-dimensional time series segmentation via factor-adjusted vector autoregressive modelling. arXiv preprint arXiv: TODO

Hallin, M. & Liška, R. (2007) Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association, 102(478), 603–617.

Examples

```
## Not run:
out <- sim.data(n = 2000, p = 50, q = 2, d = 1,
cp.common = 1:3/4, den.common = .5, type.common = 'ma',
cp.idio = c(3, 5)/8, seed = 123)
fs <- fvar.seg(out$x, q = NULL, d = 1)
fs$common.out$est.cp
```

```
fs$idio.out$est.cp

## End(Not run)
```

`idio.seg` *Segment idiosyncratic VAR process*

Description

Segment idiosyncratic VAR process

Usage

```
idio.seg(
  x,
  center = TRUE,
  common.out = NULL,
  q = NULL,
  d = 1,
  G.seq = NULL,
  thr = NULL,
  eta = 0.5,
  cv.args = list(path.length = 10, n.folds = 1, do.cv = FALSE)
)
```

Arguments

`x`	input time series matrix, with each row representing a variable
`center`	whether to de-mean the input `x` row-wise
`common.out`	output from common.seg; if `common.out = NULL`, `x` is regarded as a piecewise stationary VAR process
`q`	an integer specifying the number of factors. If `q = NULL`, the factor number is estimated by an information criterion-based approach of Hallin and Liška (2007) for each segment
`d`	an integer specifying the VAR order
`G.seq`	an integer vector of bandwidth; if `G.seq = NULL`, a default choice `G.seq = round(seq(2.5 * p,n / min(4,n/(3 * p))` is used when common component is present and `G.seq = round(seq(2 * p,n / min(5,n/(2 * p))` when it is absent
`thr`	a vector of thresholds which is of the same length as `G.seq`; if `thr = NULL`, a default choice based on simulations is used
`eta`	a constant between `0` and `1`; the bottom-up merging across the multiple bandwidths `G.seq` depends on this parameter
`cv.args`	a list specifying the tuning parameters required for Dantzig selector tuning parameter selection via cross-validation. It contains: • `n.folds` number of folds • `path.length` number of regularisation parameter values to consider; a sequence is generated in a data-driven way based in this value • `do.cv` if `do.cv = FALSE`, a fixed value is selected from a sequence of 10 values chosen in a data-driven way

Details

See Algorithm 2 of Cho, Eckley, Fearnhead and Maeng (2022) for further details.

Value

a list containing the following fields:

`est.cp`	a matrix containing the change point estimators in the first column and the finest bandwidth at which each is detected in the second column
`G.seq`	an integer vector of bandwidths
`thr`	a vector of thresholds which is of the same length as `G.seq`
`est.cp.list`	a list containing the following fields: • `cp` change point estimators • `G` bandwidth • `stat` a vector containing test statistic values • `check.cp` a vector of integers indicating where the test statistic exceeds the threshold locally
`mean.x`	if `center = TRUE`, returns a vector containing row-wise sample means of `x`; if `center = FALSE`, returns a vector of zeros

References

Cho, H., Eckley, I., Fearnhead, P. & Maeng, H. (2022) High-dimensional time series segmentation via factor-adjusted vector autoregressive modelling. arXiv preprint arXiv: TODO

Hallin, M. & Liška, R. (2007) Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association, 102(478), 603–617.

Examples

```
## Not run:
out <- sim.data(n = 2000, p = 50, q = 2, d = 1,
cp.common = 1:3/4, den.common = .5, type.common = 'ma',
cp.idio = c(3, 5)/8, seed = 123
cs <- common.seg(out$x)
cs$est.cp
is <- idio.seg(out$x, common.out = cs, d = 1)
is$est.cp

## End(Not run)
```

`sim.data` *Simulate a piecewise stationary factor-adjusted VAR process*

Description

Generate time series used in the simulation studies of Cho, Eckley, Fearnhead and Maeng (2022) for further details.

Usage

```
sim.data(
  n,
  p,
  q = 2,
  d = 1,
  cp.common = c(1:3)/4,
  den.common = 0.5,
  type.common = c("ma", "ar"),
  ma.order = 2,
  cp.idio = c(3, 5)/8,
  size.idio = 1,
  do.scale = TRUE,
  seed = NULL
)
```

Arguments

n	sample size
p	number of variables
q	number of dynamic factors
d	VAR order
cp.common	a vector specifying the re-scaled locations of the change points between 0 and 1 in the common component; possible to set cp.common = c() (no change point)
den.common	a value between 0 and 1 specifying the cross-sectional density of each change point
type.common	if type.common = 'ma', factors are loaded as innovations of a moving average process with order ma.order; if type.common = 'ar', factors are loaded as innovations of an autoregressive process of order 1
ma.order	order of the factor-driven moving average process; used when type.common = 'ma'
cp.idio	a vector specifying the re-scaled locations of the change points between 0 and 1 in the idiosyncratic component; possible to set cp.idio = c() (no change point)
size.idio	at each change point, each of VAR parameter matrices has its sign changed and is multiplied by size.idio
do.scale	if do.scale = TRUE, each variable of the common component is scaled to have the same sample variance as the corresponding idiosyncratic variable
seed	an integer setting the seed of the random number generator

Value

a list containing

x	generated piecewise stationary factor-adjusted vector autoregressive process
xi	generated piecewise stationary vector autoregressive process
A.list	a list containing the VAR parameter matrices over the segments
cp.common	input parameter
cp.idio	input parameter

References

Cho, H., Eckley, I., Fearnhead, P. & Maeng, H. (2022) High-dimensional time series segmentation via factor-adjusted vector autoregressive modelling. arXiv preprint arXiv: TODO

Examples

```
out <- sim.data(n = 2000, p = 50, q = 2, d = 1,
cp.common = 1:3/4, den.common = .5, type.common = 'ma',
cp.idio = c(3, 5)/8, seed = 123)
```

Index