Package: factree (via r-universe)

November 23, 2026

Type Package

Title Factor-Augmented Clustering Tree

Date 2026-10-30

Version 0.2.0

Maintainer Jiaqi Hu <hujiaqi@mail.ustc.edu.cn>

Description Implements the Factor-Augmented Clustering Tree (FACT) for clustering time series using covariates and group factor models. The tree compares child-node factor spaces through projector eigenvalues and uses PCA residual energy to determine node-specific stopping tolerances. Working ranks are selected by PC2 (default), BIC1 or BIC3, and comparison ranks are selected automatically at each node or supplied as a fixed value. The package also recovers global and local factors and their loadings, provides factor-space distance measures and a data generator. Candidate splits can be evaluated in parallel.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.5.0)

Imports irlba, foreach, doParallel, parallel, ggplot2 (>= 3.4.0), igraph, grid

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Jiaqi Hu [cre, aut], Ting Li [ctb] (Assisted with methodology), Zidan Luo [ctb] (Assisted with methodology), Xueqin Wang [ctb] (Assisted with methodology)

Repository https://cran.r-universe.dev

Date/Publication 2026-11-23 14:59:11 UTC

RemoteUrl https://github.com/cran/factree

RemoteRef HEAD

RemoteSha 143a9d2235b74282ff0675da9704057fcab250c5

Contents

FACT 2
gendata 3
plot.FACT 5
print.FACT 5
projector_distance 6
recover_factors 6

Index **8**

FACT *Factor-Augmented Clustering Tree*

Description

Recursively clusters time series using covariate splits. The comparison rank is selected separately at every node from the extrema of averaged-projector eigenvalues, or supplied as a fixed positive integer in `control$r_c`. In automatic mode, an empty qualifying rank set stops the node.

Usage

```
FACT(X, Y, control = list(), rank_method = c("PC2", "BIC1", "BIC3"))
```

Arguments

`X`	Finite numeric covariate matrix with dimension $N \times p$.
`Y`	Finite numeric time series matrix with dimension $T \times N$.
`control`	A list with the following optional entries:
	`rank_max` Maximum candidate factor rank; default `12`.
	`r_c` Comparison rank: `NULL` for automatic selection, or a fixed positive integer.
	`minsplit` Minimum parent size for splitting; default `90`.
	`minbucket` Minimum child size; default `30`.
	`sep` Candidate split stride: a positive integer or `"auto"`; default `1`.
	`parallel` Parallel split search; default `FALSE`.
	`n_cores` Number of workers; `NULL` selects automatically.
	`verbose` Print progress; default `FALSE`.
`rank_method`	Working-rank criterion: `"PC2"` (default, Bai–Ng PC_{p2}), `"BIC1"`, or Bai–Ng `"BIC3"`.

Details

Each node uses one working rank for both candidate children, selected by PC2 by default. Splits compare their factor spaces through projector eigenvalues. The stopping tolerance uses the PCA residual energy at the selected rank. Automatic mode selects the smallest qualifying comparison rank and stops when none qualifies. A positive `control$r_c` uses a fixed rank instead.

Value

An object of class "FACT" containing frame (node and split information), diagnostics (candidate spectra and scores), membership (terminal node labels), control, terms, factor estimates G, F, and loading estimates Gamma, Lambda. Local estimates are lists named by terminal node ID. Absent factors and loadings are zero-column matrices.

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191–221. doi:10.1111/14680262.00273.

See Also

recover_factors, print.FACT, plot.FACT

Examples

```
dat <- gendata(seed = 1, T = 60, N = rep(50, 4), p = 3)
fit <- FACT(dat$X, dat$Y, control = list(sep = 2))
table(dat$group, fit$membership)
dim(fit$G)
lapply(fit$F, dim)
plot(fit)
```

---

gendata *Generate Covariate-Defined Group Factor Data*

Description

Generates the four-group simulation design, with unequal group sizes, correlated local factors, heavy-tailed or heteroscedastic errors, and additional irrelevant covariates.

Usage

```
gendata(
  seed = 1,
  T = 100,
  N = c(100, 100, 100, 100),
  r0 = 2,
  r = rep(2, length(N)),
  M = length(N),
  sigma = 1,
  p = 10,
  mu = 3,
  type_F = "Independent",
  type_X = "Uniform",
  type_noise = "Gaussian",
```

```
  local_corr = if (type_F == "Correlated") 0.4 else 0
)
```

Arguments

seed	Random seed; use NULL to continue the current RNG stream.
T	Number of time periods, at least two.
N	Positive integer vector of four group sizes.
r0	Nonnegative number of global factors.
r	Nonnegative integer vector of four local factor counts.
M	Number of groups, inferred from N; must be four.
sigma	Positive noise standard deviation; under heteroscedasticity, σ^2 is the average variance within each group.
p	Number of covariates, at least three; only the first three define groups.
mu	Separation parameter for type_X = "Gaussian".
type_F	"Independent" or "Correlated", determining the default local_corr. Both use AR(0.5) factors.
type_X	"Uniform" for the manuscript design, or "Gaussian" for overlapping, group-dependent normal covariates.
type_noise	"Gaussian", "t3", or "heteroscedastic".
local_corr	Common off-diagonal correlation of stacked local-factor innovations, in $[0, 1)$. An explicit value overrides type_F.

Details

Factors follow AR(0.5) processes with normal loadings. Errors may be Gaussian, standardized t3, or heteroscedastic Gaussian. The covariate design uses four groups; heteroscedastic errors require even group sizes.

Value

A list with data X, Y, group labels group, factors G, F, loadings Gamma, Lambda, errors E, factor counts r0, r, and noise_variance. Local-factor columns are ordered by group. Unused loading columns are zero. G0 is an alias of G.

See Also

FACT

Examples

```
dat <- gendata(seed = 12345, T = 100, N = c(100, 100, 200, 400), sigma = 2)
correlated <- gendata(seed = 12345, local_corr = 0.4)
heavy_tail <- gendata(seed = 12345, type_noise = "t3")
high_p <- gendata(seed = 12345, p = 100)
heterogeneous <- gendata(seed = 12345, type_noise = "heteroscedastic")
```

plot.FACT *Plot a FACT Tree Object*

Description

Draws a vertical tree with a rounded root box, split variables at internal nodes, thresholds on branches, and narrow terminal-node boxes. Only terminal nodes display sample counts. No caption is added, and outer margins are small.

Usage

```
## S3 method for class 'FACT'
plot(x, ..., cex = 0.8, compact = FALSE, labels = NULL)
```

Arguments

x	An object of class FACT, returned by FACT().
...	Additional arguments (not used).
cex	Text-size multiplier; the default 0.8 uses the standard sizes.
compact	Use shorter terminal-node labels.
labels	Optional named character vector mapping covariate names to short display labels. Unmapped names are shown as supplied in the tree.

Value

Invisibly returns the displayed ggplot object.

print.FACT *Print a FACT Tree Object*

Description

Provides a concise textual representation of the FACT tree structure.

Usage

```
## S3 method for class 'FACT'
print(x, ...)
```

Arguments

x	An object of class ‘FACT‘, typically the result of a call to ‘FACT()‘.
...	Additional arguments (not used).

Value

Invisibly returns the original ‘FACT‘ object.

projector_distance *Factor-Space Distance*

Description

Factor-Space Distance

Usage

```
projector_distance(Ghat, G)
```

Arguments

Ghat	Estimated factors, scaled as in FACT so that $Ghat'Ghat = TI$.
G	Reference factors with the same number of rows and full column rank.

Value

Spectral-norm distance between the two orthogonal projectors. Zero-column matrices represent absent factors.

Examples

```
G <- diag(4)[, 1:2]
projector_distance(2 * G, G)
```

recover_factors *Recover Global and Local Factors*

Description

Recover Global and Local Factors

Usage

```
recover_factors(Y, membership, rank_max = 12)
```

Arguments

Y	Numeric time series matrix, with series in columns.
membership	Group label for each series.
rank_max	Maximum candidate factor rank; default 12.

Details

Group ranks use PC2. Averaged projections recover global factors; PCA of the remaining data recovers local factors. With one group, all factors are treated as global.

Value

A list with factor estimates `G`, `F` and loading estimates `Gamma`, `Lambda`. `F` and `Lambda` are lists named by group label. Loading rows follow the original series order within each group; `Gamma` follows the full input order. Absent factors and their loadings are zero-column matrices.

See Also

`FACT`

Index

FACT, 2, *4*, *7*

gendata, 3

plot.FACT, *3*, 5
print.FACT, *3*, 5
projector_distance, 6

recover_factors, *3*, 6