

Package ‘GrFA’

May 7, 2026

Type Package

Title Group Factor Analysis

Date 2025-12-05

Version 0.2.2

Description Implements statistical methods for group factor analysis, focusing on estimating the number of global and local factors and extracting them. Several algorithms are implemented, including Canonical Correlation-based Estimation by Choi et al. (2021) <[doi:10.1016/j.jeconom.2021.09.008](https://doi.org/10.1016/j.jeconom.2021.09.008)>, Generalised Canonical Correlation Estimation by Lin and Shin (2023) <[doi:10.2139/ssrn.4295429](https://doi.org/10.2139/ssrn.4295429)>, Circularly Projected Estimation by Chen (2022) <[doi:10.1080/07350015.2022.2051520](https://doi.org/10.1080/07350015.2022.2051520)>, and the Aggregated Projection Method by Hu et al. (2025) <[doi:10.1080/01621459.2025.2491154](https://doi.org/10.1080/01621459.2025.2491154)>.

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Depends R (>= 3.5.0)

Imports mvtnorm

NeedsCompilation no

Author Jiaqi Hu [cre, aut],
Ting Li [aut],
Xueqin Wang [aut]

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

Repository CRAN

Date/Publication 2025-12-07 00:40:25 UTC

Contents

APM	2
CCA	3
CP	5
est_num	6

FA	7
GCC	8
gendata	9
print.GFA	10
print.GFD	11
TraceRatio	11
UShouseprice	12

Index	14
--------------	-----------

APM	<i>Aggregated Projection Method</i>
-----	-------------------------------------

Description

Aggregated Projection Method for Group Factor Model.

Usage

```
APM(
  y,
  rmax = 8,
  r0 = NULL,
  r = NULL,
  localfactor = FALSE,
  weight = TRUE,
  method = "ic",
  type = "IC3"
)
```

Arguments

y	A list of the observation data, each element is a data matrix of each group with dimension $T \times N_m$.
rmax	The maximum factor numbers of all groups. Default is 8.
r0	The number of global factors. Default is NULL, the algorithm will automatically estimate the number of global factors. If you have prior information about the true number of global factors, you can set it manually.
r	The number of local factors in each group. Default is NULL, the algorithm will automatically estimate the number of local factors. If you have prior information, set it manually as an integer vector of length M (the number of groups).
localfactor	Logical. If FALSE (default), local factors are not estimated. If TRUE, local factors will be estimated.
weight	The weight of each projection matrix. If TRUE (default), weights are $w_m = N_m/N$. If FALSE, the mean of all projection matrices is calculated (equal weights). Can also be a numeric vector of length M specifying custom weights.

method	The method used in the algorithm. Default is "ic", can also be "gap".
type	The method used in estimating the factor numbers in each group initially. Default is "IC3".

Value

An object of class "GFA" containing:

r0hat	The estimated number of global factors.
rhat	The estimated number of local factors (if localfactor = TRUE).
rho	The first rmax eigenvalues of the weighted projection matrix.
Ghat	The estimated global factors.
loading_G	A list consisting of the estimated global factor loadings.
Fhat	The estimated local factors (if localfactor = TRUE).
loading_F	A list consisting of the estimated local factor loadings (if localfactor = TRUE).
residual	A list consisting of the residuals (if localfactor = TRUE).
threshold	The threshold used in determining the number of global factors (only for method = "ic").

References

Aggregated Projection Method: A New Approach for Group Factor Model. Jiaqi Hu, Ting Li, Xueqin Wang (2025). Journal of the American Statistical Association, doi:10.1080/01621459.2025.2491154

Examples

```
## Not run:
dat <- GrFA::gendata()
APM(dat$y, rmax = 8, localfactor = TRUE, method = "ic")
APM(dat$y, rmax = 8, localfactor = TRUE, method = "gap")

## End(Not run)
```

Description

Canonical Correlation Estimation for Group Factor Model.

Usage

```
CCA(
  y,
  rmax = 8,
  r0 = NULL,
  r = NULL,
  localfactor = FALSE,
  method = "CCD",
  type = "IC3"
)
```

Arguments

<code>y</code>	A list of the observation data, each element is a data matrix of each group with dimension $T \times N_m$.
<code>rmax</code>	The maximum factor numbers of all groups. Default is 8.
<code>r0</code>	The number of global factors. Default is NULL, the algorithm will automatically estimate the number of global factors. If you have prior information about the true number of global factors, you can set it manually.
<code>r</code>	The number of local factors in each group. Default is NULL, the algorithm will automatically estimate the number of local factors. If you have prior information, set it manually as an integer vector of length M (the number of groups).
<code>localfactor</code>	Logical. If FALSE (default), local factors are not estimated. If TRUE, local factors will be estimated.
<code>method</code>	The method used in the algorithm. Default is "CCD", can also be "MCC".
<code>type</code>	The method used in estimating the factor numbers in each group initially. Default is "IC3".

Value

An object of class "GFA" containing:

<code>r0hat</code>	The estimated number of global factors.
<code>rhat</code>	The estimated number of local factors (if <code>localfactor = TRUE</code>).
<code>rho</code>	The vector of average canonical correlations (eigenvalues).
<code>Ghat</code>	The estimated global factors.
<code>Fhat</code>	The estimated local factors (if <code>localfactor = TRUE</code>).
<code>loading_G</code>	A list consisting of the estimated global factor loadings.
<code>loading_F</code>	A list consisting of the estimated local factor loadings (if <code>localfactor = TRUE</code>).
<code>residual</code>	A list consisting of the residuals (if <code>localfactor = TRUE</code>).
<code>threshold</code>	The threshold used in determining the number of global factors (only for <code>method = "MCC"</code>).

References

Choi, I., Lin, R., & Shin, Y. (2021). Canonical correlation-based model selection for the multilevel factors. *Journal of Econometrics*.

Examples

```
dat <- GrFA::gendata()
CCA(dat$y, rmax = 8, localfactor = TRUE, method = "CCD")
CCA(dat$y, rmax = 8, localfactor = TRUE, method = "MCC")
```

 CP

Circularly Projected Estimation

Description

Circularly Projected Estimation for Group Factor Model.

Usage

```
CP(y, rmax = 8, r0 = NULL, r = NULL, localfactor = FALSE, type = "IC3")
```

Arguments

y	A list of the observation data, each element is a data matrix of each group with dimension $T \times N_m$.
rmax	The maximum factor numbers of all groups. Default is 8.
r0	The number of global factors. Default is NULL, the algorithm will automatically estimate the number of global factors. If you have prior information about the true number of global factors, you can set it manually.
r	The number of local factors in each group. Default is NULL, the algorithm will automatically estimate the number of local factors. If you have prior information, set it manually as an integer vector of length M (the number of groups).
localfactor	Logical. If FALSE (default), local factors are not estimated. If TRUE, local factors will be estimated.
type	The method used in estimating the local factor numbers in each group after projecting out the global factors. Default is "IC3".

Value

An object of class "GFA" containing:

r0hat	The estimated number of global factors.
rhat	The estimated number of local factors (if localfactor = TRUE).
rho	The eigenvalues of the circular projection matrix.

Ghat	The estimated global factors.
Fhat	The estimated local factors (if <code>localfactor = TRUE</code>).
loading_G	A list consisting of the estimated global factor loadings.
loading_F	A list consisting of the estimated local factor loadings (if <code>localfactor = TRUE</code>).
residual	A list consisting of the residuals (if <code>localfactor = TRUE</code>).

References

Chen, M. (2023). Circularly Projected Common Factors for Grouped Data. *Journal of Business & Economic Statistics*, 41(2), 636-649.

Examples

```
dat <- GrFA::gendata()
CP(dat$y, rmax = 8, localfactor = TRUE)
```

est_num	<i>Estimate Factor Numbers</i>
---------	--------------------------------

Description

Estimates the number of factors using various Information Criteria (IC) and Eigenvalue Ratio tests.

Usage

```
est_num(X, kmax = 8, type = "BIC3")
```

Arguments

X	The observation data matrix of dimension $T \times N$.
kmax	The maximum number of factors to consider. Default is 8.
type	The criterion used in determining the number of factors. Default is "BIC3". Options: "PC1", "PC2", "PC3", "IC1", "IC2", "IC3", "AIC3", "BIC3", "ER", "GR".

Value

rhat	The estimated number of factors (an integer).
------	---

References

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191-221.

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3), 1203-1227.

Examples

```
## Not run:  
X <- matrix(rnorm(100*20), 100, 20)  
est_num(X, kmax = 8, type = "BIC3")  
est_num(X, kmax = 8, type = "ER")  
  
## End(Not run)
```

FA

Factor Analysis

Description

Performs Factor Analysis using Principal Component Analysis (PCA) to extract factors and loadings.

Usage

```
FA(X, r)
```

Arguments

X The observation data matrix of dimension $T \times N$.
r The number of factors to estimate.

Value

A list containing:

F The estimated factors matrix of dimension $T \times r$.
L The estimated factor loadings matrix of dimension $N \times r$.

Author(s)

Jiaqi Hu

References

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191-221.

Examples

```
X <- matrix(rnorm(100*20), 100, 20)  
res <- FA(X, r = 2)  
head(res$F)  
head(res$L)
```

GCC

*Generalised Canonical Correlation***Description**

Generalised Canonical Correlation Estimation for Group Factor Model.

Usage

```
GCC(y, rmax = 8, r0 = NULL, r = NULL, localfactor = FALSE, type = "IC3")
```

Arguments

<code>y</code>	A list of the observation data, each element is a data matrix of each group with dimension $T \times N_m$.
<code>rmax</code>	The maximum factor numbers of all groups. Default is 8.
<code>r0</code>	The number of global factors. Default is NULL, the algorithm will automatically estimate the number of global factors. If you have prior information about the true number of global factors, you can set it manually.
<code>r</code>	The number of local factors in each group. Default is NULL, the algorithm will automatically estimate the number of local factors. If you have prior information, set it manually as an integer vector of length M (the number of groups).
<code>localfactor</code>	Logical. If FALSE (default), local factors are not estimated. If TRUE, local factors will be estimated.
<code>type</code>	The method used in estimating the factor numbers in each group initially. Default is "IC3".

Value

An object of class "GFA" containing:

<code>r0hat</code>	The estimated number of global factors.
<code>rhat</code>	The estimated number of local factors (if <code>localfactor = TRUE</code>).
<code>rho</code>	The ratio of the singular values used to estimate the number of global factors.
<code>Ghat</code>	The estimated global factors.
<code>Fhat</code>	The estimated local factors (if <code>localfactor = TRUE</code>).
<code>loading_G</code>	A list consisting of the estimated global factor loadings.
<code>loading_F</code>	A list consisting of the estimated local factor loadings (if <code>localfactor = TRUE</code>).
<code>residual</code>	A list consisting of the residuals (if <code>localfactor = TRUE</code>).

References

Lin, R., & Shin, Y. (2023). Generalised Canonical Correlation Estimation of the Multilevel Factor Model. Available at SSRN 4295429.

Examples

```
dat <- GrFA::gendata()
GCC(dat$y, rmax = 8, localfactor = TRUE)
```

gendata	<i>Generate the grouped data</i>
---------	----------------------------------

Description

Generate the grouped data for simulation studies.

Usage

```
gendata(
  seed = 1,
  T = 50,
  N = rep(30, 5),
  r0 = 2,
  r = rep(2, 5),
  Phi_G = 0.5,
  Phi_F = 0.5,
  Phi_e = 0.5,
  W_F = 0.5,
  beta = 0.2,
  kappa = 1,
  case = 1
)
```

Arguments

<code>seed</code>	The seed used in <code>set.seed</code> . Default is 1.
<code>T</code>	The number of time points. Default is 50.
<code>N</code>	A vector representing the number of variables in each group. Default is <code>rep(30, 5)</code> .
<code>r0</code>	The number of global factors. Default is 2.
<code>r</code>	A vector representing the number of the local factors. Notice, the length of <code>r</code> is the same as the length of <code>N</code> (which implies the number of groups M). Default is <code>rep(2, 5)</code> .
<code>Phi_G</code>	Hyperparameter of the global factors (AR(1) coefficient). Default is 0.5. The value should be between 0 and 1.
<code>Phi_F</code>	Hyperparameter of the local factors (AR(1) coefficient). Default is 0.5. The value should be between 0 and 1.
<code>Phi_e</code>	Hyperparameter of the errors. Default is 0.5. The value should be between 0 and 1.

W_F	Hyperparameter of the correlation of local factors. Only applicable when case = 3. The value should be between 0 and 1. Default is 0.5.
beta	Hyperparameter of the errors (spatial correlation). Default is 0.2.
kappa	Hyperparameter of signal to noise ratio. Default is 1.
case	The case of the data-generating process. Default is 1. It can also be 2 or 3.

Value

An object of class "GFD" containing:

y	A list of the generated data matrices.
G	The global factors matrix.
F	A list of the local factors.
loading_G	A list of the global factor loadings.
loading_F	A list of the local factor loadings.
T	The number of time points.
N	The vector of variables per group.
M	The number of groups.
r0	The number of global factors.
r	The vector of local factors.
case	The generation case used.

References

Aggregated Projection Method: A New Approach for Group Factor Model. Jiaqi Hu, Ting Li, Xueqin Wang (2025). Journal of the American Statistical Association, doi:10.1080/01621459.2025.2491154

Examples

```
dat <- gendata()
print(dat)
```

```
print.GFA
```

Print Method for GFA Objects

Description

Print the summarized results of the estimated group factor model, such as the estimated global and local factor numbers and reference statistics.

Usage

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

Arguments

x The GFA object returned from the estimation algorithms (e.g., APM, CCA, GCC, CP).
 ... Additional arguments passed to methods.

Value

No return value, called for side effects.

print.GFD *Print Method for GFD Objects*

Description

Print the summary of the generated grouped data.

Usage

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

Arguments

x The GFD object returned from [gendata](#).
 ... Additional arguments passed to methods.

Value

No return value, called for side effects.

TraceRatio *Trace Ratio*

Description

Evaluation of the estimated factors by trace ratios. The value is between 0 and 1; higher values indicate better estimation accuracy.

Usage

```
TraceRatio(G, Ghat)
```

Arguments

G The true factors matrix.
 Ghat The estimated factors matrix.

Value

A numeric value representing the trace ratio, defined as: $TR = \text{tr}(\mathbf{G}'\hat{\mathbf{G}}(\hat{\mathbf{G}}'\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\mathbf{G})/\text{tr}(\mathbf{G}'\mathbf{G})$.

Examples

```
G <- matrix(rnorm(100 * 2), 100, 2)
Ghat <- G + matrix(rnorm(100 * 2, sd = 0.1), 100, 2)
TraceRatio(G, Ghat)
```

UShouseprice

Housing price data for 16 states in the U.S.

Description

This dataset contains the Zillow Home Value Index (ZHVI) at the county level for single-family residences and condos with 1, 2, 3, 4, or 5+ bedrooms. It focuses on the middle tier of home values (33rd to 67th percentile) and features smoothed, seasonally adjusted values presented on a monthly basis. The data spans 16 U.S. states from January 2000 to April 2023. Within each state, the data is organized as a matrix, and the data for all states is compiled into a list.

Usage

```
data(UShouseprice)
```

Format

The dataset is structured as a list containing 16 elements, with each element corresponding to a state. Each element is a matrix where the columns represent time series data for house prices at the county level.

Rows Each time series has a length of 280, representing monthly data points from January 2000 to April 2023.

Columns The number of columns in each matrix varies, ranging from 90 to 250, depending on the number of counties and bedroom categories in the state.

Labels The columns are labeled with the county name and bedroom count (e.g., "Pulaski County bd1" for one-bedroom homes or "Garland County bd5" for homes with five or more bedrooms).

Details

The column names of the data matrix represent county names combined with bedroom counts. For example, "Pulaski County bd1" indicates the house price in Pulaski County for one-bedroom homes, while "Garland County bd5" refers to the house price in Garland County for homes with more than five bedrooms.

The abbreviations and full names of these 16 states are as follows:

- AR: Arkansas
- CA: California
- CO: Colorado
- FL: Florida
- GA: Georgia
- KY: Kentucky
- MD: Maryland
- MI: Michigan
- NC: North Carolina
- NJ: New Jersey
- NY: New York
- OH: Ohio
- OK: Oklahoma
- PA: Pennsylvania
- TN: Tennessee
- VA: Virginia

Source

The original data is downloaded from the website of Zillow.

References

Aggregated Projection Method: A New Approach for Group Factor Model. Jiaqi Hu, Ting Li, Xueqin Wang (2025). Journal of the American Statistical Association, doi:10.1080/01621459.2025.2491154

Examples

```
data(UShouseprice)

# Helper function to calculate log differences and scale
log_diff <- function(x) {
  T <- nrow(x)
  res <- log(x[2:T, ] / x[1:(T - 1), ]) * 100
  scale(res, center = TRUE, scale = TRUE)
}

# Apply to all states
UShouseprice1 <- lapply(UShouseprice, log_diff)
```

Index

* datasets

UShouseprice, [12](#)

APM, [2](#)

CCA, [3](#)

CP, [5](#)

est_num, [6](#)

FA, [7](#)

GCC, [8](#)

gendata, [9](#), [11](#)

print.GFA, [10](#)

print.GFD, [11](#)

TraceRatio, [11](#)

UShouseprice, [12](#)