Package: MultiEFM (via r-universe)

October 18, 2026

Type Package

Title Robust Estimation for Multi-Study High-Dimensional Elliptical Factor Analytics

Version 0.1.4

Date 2026-08-03

Description We introduce a multi-study high-dimensional factor analysis toolbox 'MultiEFM' based on the elliptical factor model framework, which learns latent heterogeneous features and accounts for cross-study variation among sources. It provides robust estimation algorithms for heterogeneous datasets, particularly tailored for multi-study RNA sequencing arrays and complex spatial multi-omics layers. The package implements highly efficient initialization strategies, identifiability constraints alignment, and computationally scalable parameter estimation paradigms.

License GPL-3

Depends R (>= 3.5.0)

Imports MASS, irlba, mvtnorm, Rcpp (>= 1.0.8.3), stats

URL https://RRuihan-Zhang.github.io/MultiEFM/

BugReports https://github.com/RRuihan-Zhang/MultiEFM/issues

Encoding UTF-8

LinkingTo Rcpp, RcppArmadillo

VignetteBuilder knitr

Suggests cowplot, ggplot2, knitr, rmarkdown

Config/roxygen2/version 8.0.0

NeedsCompilation yes

Author Ruihan Zhang [aut, cre], Wei Liu [aut]

Maintainer Ruihan Zhang <ruihanzhang6166@gmail.com>

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

Date/Publication 2026-08-04 14:56:58 UTC

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

RemoteRef HEAD

RemoteSha abc72c6f6304b511ede81deffc82eba9c90ab3e5

Contents

gendata_simu_robust 2
MultiEFM 3
selectFac.MultiEFM 5

Index **7**

`gendata_simu_robust` *Generate Simulated Data for Robust Multi-Study Elliptical Factor Model*

Description

Generate simulated data for high-dimensional multi-study factor analysis under robust elliptical distributions (such as Multivariate t-distribution). The data follows a factor model with common factors (shared across studies) and study-specific factors, plus noise.

Usage

```
gendata_simu_robust(
  seed = NULL,
  nvec = c(100, 300),
  p = 50,
  q = 3,
  qs = rep(2, length(nvec)),
  fac.type = c("gaussian", "mvt"),
  err.type = c("gaussian", "mvt"),
  rho = c(1, 1),
  sigma2_eps = 0.1,
  nu = 1,
  verbose = FALSE
)
```

Arguments

`seed`	Integer specifying the random seed, or NULL; default as NULL.
`nvec`	Numeric vector (length >= 2). Sample sizes of each study.
`p`	Integer, default = 50. Number of variables (features) in the data.
`q`	Integer, default = 3. Number of common factors (shared across all studies).
`qs`	Numeric vector. Number of study-specific factors for each study.

fac.type	Character, default = "gaussian". Factor distribution type, one of "gaussian" or "mvt".
err.type	Character, default = "gaussian". Error distribution type, one of "gaussian" or "mvt".
rho	Numeric vector of length 2, default = ‘c(1,1)‘. Scaling factors for common and specific factor loadings.
sigma2_eps	Numeric, default = 0.1. Variance of the error term.
nu	Integer, default = 1. Degrees of freedom for the multivariate t-distribution ("mvt").
verbose	Logical, default = FALSE. Whether to output the information of data generation.

Value

A list containing the simulated data and true parameter values:

- Xlist: List of data matrices (ns × p) for each study.
- mu0: True mean vector matrix.
- A0: True common factor loadings matrix.
- Blist0: List of true study-specific factor loadings matrices.
- Flist: List of true common factor score matrices.
- Hlist: List of true study-specific factor score matrices.
- q: Number of common factors used.
- qs: Number of study-specific factors used.

MultiEFM *Fit the high-dimensional multi-study elliptical factor model*

Description

Fit the high-dimensional multi-study elliptical factor model which learns latent features and accounts for the heterogeneity among sources.

Usage

```
MultiEFM(
  XList,
  q,
  qs_vec,
  loads_method = c("Residual", "Ortho"),
  fac_est_method = c("Huber", "LS"),
  n_threads = 2,
  sample_pairs = 0,
  epsObj = 1e-06,
  maxIter = 20,
  verbose = TRUE,
  seed = NULL
)
```

Arguments

XList	A length-S list, where each component represents a data matrix for a specific study.
q	an integer, specify the number of study-shared factors.
qs_vec	an integer vector with length S, specify the number of study-specified factors for each study.
loads_method	a character string specifying the method for loading estimation. Options are "Residual" or "Ortho".
fac_est_method	a character string specifying the method for factor estimation. Options are "Huber" or "LS" (Least Squares).
n_threads	an integer, specify the number of threads for parallel computing; default as 4.
sample_pairs	an integer, specify the number of sample pairs used for estimation; default as 0.
epsObj	a positive value, tolerance for convergence, default as 1e-6.
maxIter	the maximum iteration of the algorithm. The default is 20.
verbose	a logical value, whether to output the information in iteration.
seed	an integer, specify the random seed for reproducibility in initialization; default as NULL.

Value

return a list including the following components:

A	the estimated loading matrix corresponding to study-shared factors;
F	a list composed by the estimation of study-shared factor matrix for each study;
B	a list composed by the loading matrices corresponding to the study-specified factors;
H	a list composed by the estimation of study-specified factor matrix for each study.

Examples

```
p <- 100
nvec <- c(150, 200)
qs <- c(2, 2)
datList <- gendata_simu_robust(seed=1, nvec=nvec, p=p, q=3, qs=qs, rho=c(5,5),
                               err.type='mvt', sigma2_eps = 1, nu=3)
XList <- datList$Xlist
res <- MultiEFM(XList, q=3, qs_vec=qs)
str(res)
```

`selectFac.MultiEFM` *Select the number of factors for MultiEFM*

Description

Automatically estimate the number of study-shared factors and study-specific factors based on the eigenvalue ratio method.

Usage

```
selectFac.MultiEFM(
  XList,
  q_max = 20,
  qs_max = 20,
  qd = NULL,
  n_threads = 2,
  sample_pairs = 0,
  verbose = TRUE,
  seed = NULL
)
```

Arguments

`XList`	A length-S list, where each component represents a data matrix for a specific study.
`q_max`	an integer, specify the maximum number of study-shared factors to consider; default as 20.
`qs_max`	an integer, specify the maximum number of study-specified factors to consider; default as 20.
`qd`	an optional integer, specify the true number of combined factors if known in advance; default is NULL.
`n_threads`	an integer, specify the number of threads for parallel computing; default as 2.
`sample_pairs`	an integer, specify the number of sample pairs used for estimation; default as 0.
`verbose`	a logical value, whether to output the estimation information.
`seed`	an integer, specify the random seed for reproducibility in initialization; default as NULL.

Value

return a list including the following components:

`hq`	the estimated number of study-shared factors;
`hqs_vec`	an integer vector representing the estimated number of study-specific factors for each study;
`eigvals_A`	the eigenvalues used for determining the shared factor number;

`eig_ratio_A`	the eigenvalue ratios used for determining the shared factor number;
`eigvals_B`	a list of eigenvalues used for determining the specific factor numbers;
`eig_ratio_B`	a list of eigenvalue ratios used for determining the specific factor numbers.

Index

gendata_simu_robust, 2

MultiEFM, 3

selectFac.MultiEFM, 5