

Package: MultiEFM (via r-universe)

September 27, 2026

Type Package

Title Robust Estimation for Multi-Study Elliptical Factor Model

Version 0.1.2

Date 2026-04-20

Description We introduce a high-dimensional multi-study elliptical factor model (MultiEFM), which learns latent features and accounts for the heterogeneity among sources. It provides robust estimation for heterogeneous high-dimensional data analysis, particularly suitable for multi-study RNA sequencing data and spatial multi-omics. The package implements efficient algorithms for model configuration, identifiability conditions, and parameter estimation.

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

LazyData true

LinkingTo Rcpp, RcppArmadillo

VignetteBuilder knitr

RoxygenNote 7.3.3

Suggests cowplot, ggplot2, knitr, rmarkdown, VIMSFA

Repository <https://rruihan-zhang.r-universe.dev>

Date/Publication 2026-05-07 03:33:57 UTC

RemoteUrl <https://github.com/rruihan-zhang/multiefm>

RemoteRef HEAD

RemoteSha 733fafad2c9cc55e428d7f36ffd39da7d792cbc4

Contents

gendata_simu_robust	2
MultiEFM	3
selectFac.MultiEFM	4

Index	6
--------------	----------

gendata_simu_robust	<i>Generate Simulated Data for Robust Multi-Study Elliptical Factor Model</i>
---------------------	---

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 = 1,
  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
)
```

Arguments

seed	Integer, default = 1. Random seed for reproducibility.
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").

Value

A list containing the simulated data and true parameter values:

- `Xlist`: List of data matrices ($n_s \times p$) 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 = 4,
  sample_pairs = 0,
  epsObj = 1e-06,
  maxIter = 20,
  verbose = TRUE,
  seed = 1
)
```

Arguments

<code>XList</code>	A length-S list, where each component represents a data matrix for a specific study.
<code>q</code>	an integer, specify the number of study-shared factors.
<code>qs_vec</code>	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 1.

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

```
## Not run:
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)

## End(Not run)
```

selectFac.MultiEFM	<i>Select the number of factors for MultiEFM</i>
--------------------	--

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 = 4,
  sample_pairs = 0,
  verbose = TRUE,
  seed = 1
)
```

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 4.
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; default as 1.

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`, [4](#)