Package 'dacc'

September 20, 2024

Title Detection and Attribution Analysis of Climate Change

Version 0.0-5

Date 2024-09-20

Description Conduct detection and attribution of climate change using methods including optimal fingerprinting via
generalized total least squares or estimating equation approach from Ma et al. (2023) <doi:10.1175/JCLI-D-22-0681.1>.
Provide shrinkage estimators for covariance matrix from Ledoit and Wolf (2004) <doi:10.1016/S0047-259X(03)00096-4>,
and Ledoit and Wolf (2017) <doi:10.2139/ssrn.2383361>.

Depends R (>= 3.5.0)

Imports MASS, stats, utils, Iso, pracma, janitor, magrittr, sp,
CFtime, ncdf4

License GPL (>= 3)

Encoding UTF-8

URL https://github.com/LiYanStat/dacc

BugReports https://github.com/LiYanStat/dacc/issues

Repository CRAN

RoxygenNote 7.3.1

NeedsCompilation no

Author Yan Li [aut, cre],
Kun Chen [aut],
Jun Yan [aut]

Maintainer Yan Li <yan.4.li@uconn.edu>

Date/Publication 2024-09-20 18:40:05 UTC

Contents

Index   **8**

Covest   *Regularized estimators for covariance matrix.*

Description

This function estimate the covariance matrix under l2 loss and minimum variance loss, provide
linear shrinkage estimator under l2 loss and nonlinear shrinkage estimator under minimum variance
loss.

Usage

```
Covest(Z, method = c("mv", "l2"), bandwidth = NULL)
```

Arguments

Z	n*p matirx with sample size n and dimension p. Replicates for computing the covariance matrix, should be centered.
method	methods used for estimating the covariance matrix.
bandwidth	bandwidth for the "mv" estimator, default value are set to be list in (0.2, 0.5).

Value

regularized estimate of covariance matrix.

Author(s)

Yan Li

References

- Olivier Ledoit and Michael Wolf (2004), A well-conditioned estimator for large-dimensional
  covariance matrices, *Journal of multivariate analysis*, 88(2), 365–411.

- Olivier Ledoit and Michael Wolf (2017), Direct nonlinear shrinkage estimation of large-
  dimensional covariance matrices, *Working Paper No. 264, UZH*.

- Li et al (2023), Regularized fingerprinting in detection and attribution of climate change with
  weight matrix optimizing the efficiency in scaling factor estimation, *Ann. Appl. Stat.* 17(1),
  225–239.

Examples

```
## randomly generate a n * p matrix where n = 50, p = 100
Z <- matrix(rnorm(50 * 100), nrow = 50, 100)
## linear shrinkage estimator under l2 loss
Cov.est <- Covest(Z, method = "l2")$output
## nonlinear shrinkage estimator under minimum variance loss
Cov.est <- Covest(Z, method = "mv", bandwidth = 0.35)$output
```

---

fingerprint	*Optimal Fingerprinting via total least square regression.*

Description

This function estimates the signal factors and corresponding confidence interval via the estimating equation or total least squares.

Usage

```
fingerprint(
  Xtilde,
  Y,
  mruns,
  ctlruns.sigma,
  ctlruns.bhvar,
  S,
  T,
  B = 0,
  Proj = diag(ncol(Xtilde)),
  method = c("EE", "PBC", "TS"),
  cov.method = c("l2", "mv"),
  conf.level = 0.9,
  missing = FALSE,
  cal.a = TRUE,
  ridge = 0
)
```

Arguments

Xtilde	$n \times p$ matrix, signal pattern to be detected.
Y	$n \times 1$ matrix, length $S \times T$, observed climate variable.
mruns	number of ensembles to estimate the corresponding pattern. It is used as the scale of the covariance matrix for X_i.
ctlruns.sigma	$m \times n$ matrix, a group of m independent control runs for estimating covariance matrix, which is used in point estimation of the signal factors.
ctlruns.bhvar	$m \times n$ matrix, another group of m independent control runs for estimating the corresponding confidence interval of the signal factors, in EE or PBC approach should be same as ctlruns.sigma.

S	number of locations for the observed responses.
T	number of time steps for the observed responses.
B	number of replicates in bootstrap procedure, mainly for the PBC and TS methods, can be specified in "EE" method but not necessary. By default B = 0 as the default method is "EE".
Proj	The projection matrix for computing for scaling factors of other external forcings with the current input when using EE. For example, when ALL and NAT are used for modeling, specifying the Proj matrix to return the results for ANT and NAT.
method	for estimating the scaling factors and corresponding confidence interval
cov.method	method for estimation of covariance matrix in confidence interval estimation of PBC method. (only for PBC method).
conf.level	confidence level for confidence interval estimation.
missing	indicator for whether missing values present in Y.
cal.a	indicator for calculating the a value, otherwise use default value a = 1. (only for EE method)
ridge	shrinkage value for adjusting the method for missing observations if missing = TRUE. (only for EE method)

Value

a list of the fitted model including point estimate and interval estimate of coefficients and corresponding estimate of standard error.

Author(s)

Yan Li

References

- Gleser (1981), Estimation in a Multivariate "Errors in Variables" Regression Model: Large Sample Results, *Ann. Stat.* 9(1) 24–44.

- Golub and Laon (1980), An Analysis of the Total Least Squares Problem, *SIAM J. Numer. Anal.* 17(6) 883–893.

- Pesta (2012), Total least squares and bootstrapping with applications in calibration, *Statistics* 47(5), 966–991.

- Li et al (2021), Uncertainty in Optimal Fingerprinting is Underestimated, *Environ. Res. Lett.* 16(8) 084043.

- Sai et al (2023), Optimal Fingerprinting with Estimating Equations, *Journal of Climate* 36(20), 7109—7122.

- Li et al (2024), Detection and Attribution Analysis of Temperature Changes with Estimating Equations, *Submitted to Journal of Climate*.

Examples

```
## load the example dataset
data(simDat)
Cov <- simDat$Cov[[1]]
ANT <- simDat$X[, 1]
NAT <- simDat$X[, 2]

## generate the simulated data set
## generate regression observation
Y <- MASS::mvrnorm(n = 1, mu = ANT + NAT, Sigma = Cov)
## generate the forcing responses
mruns <- c(1, 1)
Xtilde <- cbind(MASS::mvrnorm(n = 1, mu = ANT, Sigma = Cov / mruns[1]),
                MASS::mvrnorm(n = 1, mu = NAT, Sigma = Cov / mruns[2]))
## control runs
ctlruns <- MASS::mvrnorm(100, mu = rep(0, nrow(Cov)), Sigma = Cov)
## ctlruns.sigma for the point estimation and ctlruns.bhvar for the interval estimation
ctlruns.sigma <- ctlruns.bhvar <- ctlruns
## number of locations
S <- 25
## number of year steps
T <- 10

## call the function to estimate the signal factors via EE
fingerprint(Xtilde, Y, mruns,
         ctlruns.sigma, ctlruns.bhvar,
         S, T,
         ## B = 0, by default
         method = "EE",
         conf.level = 0.9,
         cal.a = TRUE,
         missing = FALSE, ridge = 0)
```

---

fpPrep	*Process netCDF4 Gridded Data into Format of the fingerprint() Function*

Description

This function detects the signal factors on the observed data via total least square linear regression model.

Usage

```
fpPrep(
  datafile,
  variable,
  region = "GL",
  target.year,
```

```
    average = 5,
    reference = c(1961, 1990),
    regridding = NULL
)
```

Arguments

`datafile`	path to the netCDF4 gridded datafile to be processed
`variable`	the climate variable to be extracted
`region`	the longitude and latitude boundary for selected region, should match the format of IPCC AR6 regions, the lon and lat of the vertices
`target.year`	vector of length 2, the starting and ending year of the selected time period for D&A analysis
`average`	number of years for average on each gridbox, default is 5-year average
`reference`	vector of length 2, the starting and ending year of reference time period for computing anomalies
`regridding`	whether the grid box should be regridded. Specify the size of the grid box, e.g., c(40, 30) for $40° \times 30°$ grid box. If no regridding, leave empty

Value

a dataset of the processed gridded climate variables for Y, Xtilde or control runs

Author(s)

Yan Li

---

globalDat	*Sample Dataset Used in Numerical Studies of "Detection and Attribution Analysis of Temperature Changes with Estimating Equations".*

Description

A list of the observations and expected responses to different external forcings with name `Y`, `X`, `ctlruns`, `nruns.X` and `Xtilde` where

- `Y`: the $5° \times 5°$ gridded observations on global scale
- `X`: a data matrix of the expected responses to external forcing;
- `ctlruns`: replicates of control runs from pre-industrial simulations
- `nruns.X`: number of runs for the estimated responses to external forcings
- `Xtilde`: the selected estimated responses to external forcing ANT and NAT

Usage

```
data(globalDat)
```

Format

A data list with the observed and simulated data on global scale.

Examples

```
data(globalDat)
```

---

simDat	*Sample Dataset in Simulation Studies of "Regularized Fingerprinting in Detection and Attribution of Climate Change with Weight Matrix Optimizing the Efficiency in Scaling Factor Estimation".*

Description

A data list of designed covariance matrix and the expected responses to the two forcings ANT and NAT with name Cov and X where

- Cov: a list of the true covariance matrices;
- X: a data matrix of the expected responses to external forcing;

Usage

```
data(simDat)
```

Format

A data list with two separate data sets.

Examples

```
data(simDat)
```

Index