Package ‘dacc’

January 9, 2024

Title Detection and Attribution Analysis of Climate Change

Version 0.0-3

Date 2024-01-08

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

License GPL (>= 3)

Encoding UTF-8

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

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

Repository CRAN

RoxygenNote 7.2.3

NeedsCompilation no

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

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

Date/Publication 2024-01-09 08:20:20 UTC

R topics documented:

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

Usage

```
fingerprint(
  X,
  Y,
  cov,
  nruns.X,
  ctlruns,
  precision = FALSE,
  conf.level = 0.9,
  conf.method = c("none", "TSB", "PBC", "both"),
  cov.method = c("l2", "mv"),
  B = 1000
)
```

Arguments

X	signal pattern to be detected.
Y	observed data.
cov	Weight matrix used in prewhitening process, can be estimate of covariance matrix or precision matrix.
nruns.X	number of ensembles to estimate the corresponding pattern. It is used as the scale of the covariance matrix for Xi.
ctlruns	a group of independent control runs for estimating covariance matrix, which is used in two stage bootstrap and the parametric bootstrap calibration.
precision	indicator for precision matrix, if precision matrix estimate is used, precision should be set to TRUE.
conf.level	confidence level for confidence interval estimation.
conf.method	method for calibrating the confidence intervals, including no calibration (none), two stage bootstrap (TSB), and parametric bootstrap calibration (PBC).
cov.method	method for estimation of covariance matrix in confidence interval procedure. It should be consistent to the method to get cov. (only valid if TSB or PBC is considered.)
B	number of replicates in two stage bootstrap and/or parametric bootstrap calibration, default value is 1000. (only valid if TSB or PBC is considered.)

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.

Examples

```
data(simDat)
## set the true covariance matrix and expected pattern
Cov <- simDat$Cov[[1]]
ANT <- simDat$X[, 1]
NAT <- simDat$X[, 2]
## estimate the covariance matrix
Z <- MASS::mvrnorm(100, mu = rep(0, nrow(Cov)), Sigma = Cov)
## linear shrinkage estimator under l2 loss
Cov.est <- Covest(Z, method = "l2")$output
## generate regression observation and pattern
nruns.X <- c(1, 1)
Y <- MASS::mvrnorm(n = 1, mu = ANT + NAT, Sigma = Cov)
X <- cbind(MASS::mvrnorm(n = 1, mu = ANT, Sigma = Cov),
           MASS::mvrnorm(n = 1, mu = NAT, Sigma = Cov))
fingerprint(X, Y, Cov.est, nruns.X, ctlruns = Z, conf.method = "TSB", B = 5)
```

`fingerprintCEE` *Optimal Fingerprinting with Estimating Equations.*

Description

This function estimates the signal factors and corresponding confidence interval via the estimating equation methods.

Usage

```
fingerprintCEE(
  Xtilde,
  Y,
  mruns,
  ctlruns.1,
  ctlruns.2,
  nS,
  nT,
  nB = 0,
  conf.level = 0.9,
  cal.a = TRUE,
  missing = FALSE,
  ridge = 0
)
```

Arguments

`Xtilde`	$n \times p$ matrix, signal pattern to be detected.
`Y`	$n \times 1$ matrix, length $nS \times nT$, observed data.
`mruns`	number of ensembles to estimate the corresponding pattern. It is used as the scale of the covariance matrix for Xi.
`ctlruns.1`	$m \times n$ matrix, a group of m independent control runs for estimating covariance matrix, which is used in point estimation of the signal factors.
`ctlruns.2`	$m \times n$ matrix, another group of m independent control runs for estimating the corresponding confidence interval of the signal factors, default same as ctlruns.1.
`nS`	number of locations for the observed responses.
`nT`	number of time steps for the observed responses.
`nB`	number of replicates in bootstrap for the bootstrapped variance and confidence interval estimations.
`conf.level`	confidence level for confidence interval estimation.
`cal.a`	indicator for calculating the a value, otherwise use default value a = 1.
`missing`	indicator for whether missing values present in Y.
`ridge`	shrinkage value for adjusting the method for missing observations if missing = TRUE.

Value

a list of the fitted model including point estimate and interval estimate of beta and a coefficients and corresponding estimate of variance, together with the residuals for model diagnostics.

Author(s)

Yan Li

References

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

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)
## ctlruns1 for the point estimation and ctlruns2 for the interval estimation
ctlrun.1 <- ctlrun.2 <- ctlruns
## number of locations
nS <- 25
## number of year steps
nT <- 10
## call the function to estimate the signal factors via CEE
fingerprintCEE(Xtilde, Y, mruns,
               ctlrun.1, ctlrun.2,
               nS, nT, nB = 10,
               conf.level = 0.9,
               cal.a = TRUE,
               missing = FALSE, ridge = 0)
```

`simDat` *Sample Data Used 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