

# Package ‘COR’

May 7, 2026

**Title** The COR for Optimal Subset Selection in Distributed Estimation

**Date** 2024-12-10

**Version** 0.2.0

## Description

An algorithm of optimal subset selection, related to Covariance matrices, observation matrices and Response vectors (COR) to select the optimal subsets in distributed estimation. The philosophy of the package is described in Guo G. (2024) <[doi:10.1007/s11222-024-10471-z](https://doi.org/10.1007/s11222-024-10471-z)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**Imports** stats

**NeedsCompilation** no

**Author** Guangbao Guo [aut, cre] (ORCID:  
<<https://orcid.org/0000-0002-4115-6218>>),  
Haoyue Song [aut],  
Lixing Zhu [aut]

**Maintainer** Guangbao Guo <[ggb11111111@163.com](mailto:ggb11111111@163.com)>

**Depends** R (>= 3.5.0)

**Repository** CRAN

**Date/Publication** 2024-12-16 10:20:02 UTC

## Contents

beta_AD . . . . .	2
beta_cor . . . . .	3
beta_LW . . . . .	4
communities . . . . .	5
COR . . . . .	9
ethylene_CO . . . . .	10
LICbeta . . . . .	11
LICnew . . . . .	12

MSEbeta	13
MSEcom	13
MSEver	14

<b>Index</b>	<b>16</b>
--------------	-----------

---

beta_AD	<i>Calculate the estimators of beta on the A-opt and D-opt</i>
---------	--

---

### Description

Calculate the estimators of beta on the A-opt and D-opt

### Usage

```
beta_AD(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

### Arguments

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

### Value

A list containing:

betaA	The estimator of beta on the A-opt.
betaD	The estimator of beta on the D-opt.

### References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). [doi:10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

### Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%%beta+e;
beta_AD(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

---

beta\_cor

*Calculate the estimator of beta on the COR*

---

### Description

Calculate the estimator of beta on the COR

### Usage

```
beta_cor(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

### Arguments

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

### Value

A list containing:

betaC	The estimator of beta on the COR.
-------	-----------------------------------

### References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). [doi:10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

### Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1)))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%%beta+e;
beta_cor(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

---

beta_LW	<i>Calculate the estimators of beta on the LEV-opt#'</i>
---------	--

---

### Description

Calculate the estimators of beta on the LEV-opt#'

### Usage

beta\_LW(X, Y, K, nk)

### Arguments

X	is the observation matrix
Y	is the response vector
K	is the number of subsets
nk	is the length of subsets

### Value

A list containing:

betalev	The estimator of beta on the LEV-opt subset.
betam	The mean of the beta estimators across all K subsets.
AMSE	The Average Mean Squared Error (AMSE) for the estimator.
WMSE	The Weighted Mean Squared Error (WMSE) for the estimator.
MSElevb	The Mean Squared Error (MSE) of the LEV-opt estimator compared to the true beta.
MSEb	The Mean Squared Error (MSE) of the mean estimator (betam) compared to the true beta.
MSEyleva	The Mean Squared Error (MSE) of the LEV-opt estimator on the subset with the maximum hat value (Xleva).
MSEyleviy	The Mean Squared Error (MSE) of the LEV-opt estimator on the subset with the minimum hat value (Xlevi).
MSEW	The Mean Squared Error (MSE) of the weighted estimator (Wbeta) compared to the true beta.
MSEw	The Mean Squared Error (MSE) of the weighted estimator (wbeta) compared to the true beta.

### References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). [doi:10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

---

communities

*The communities and crime data set*

---

**Description**

A data set about the communities and crime

**Usage**

```
data("communities")
```

**Format**

A data frame with 1994 observations on the following 128 variables.

V1 a numeric vector  
V2 a numeric vector  
V3 a numeric vector  
V4 a character vector  
V5 a numeric vector  
V6 a numeric vector  
V7 a numeric vector  
V8 a numeric vector  
V9 a numeric vector  
V10 a numeric vector  
V11 a numeric vector  
V12 a numeric vector  
V13 a numeric vector  
V14 a numeric vector  
V15 a numeric vector  
V16 a numeric vector  
V17 a numeric vector  
V18 a numeric vector  
V19 a numeric vector  
V20 a numeric vector  
V21 a numeric vector  
V22 a numeric vector  
V23 a numeric vector  
V24 a numeric vector  
V25 a numeric vector

V26 a numeric vector  
V27 a numeric vector  
V28 a numeric vector  
V29 a numeric vector  
V30 a numeric vector  
V31 a numeric vector  
V32 a numeric vector  
V33 a numeric vector  
V34 a numeric vector  
V35 a numeric vector  
V36 a numeric vector  
V37 a numeric vector  
V38 a numeric vector  
V39 a numeric vector  
V40 a numeric vector  
V41 a numeric vector  
V42 a numeric vector  
V43 a numeric vector  
V44 a numeric vector  
V45 a numeric vector  
V46 a numeric vector  
V47 a numeric vector  
V48 a numeric vector  
V49 a numeric vector  
V50 a numeric vector  
V51 a numeric vector  
V52 a numeric vector  
V53 a numeric vector  
V54 a numeric vector  
V55 a numeric vector  
V56 a numeric vector  
V57 a numeric vector  
V58 a numeric vector  
V59 a numeric vector  
V60 a numeric vector  
V61 a numeric vector  
V62 a numeric vector

- V63 a numeric vector
- V64 a numeric vector
- V65 a numeric vector
- V66 a numeric vector
- V67 a numeric vector
- V68 a numeric vector
- V69 a numeric vector
- V70 a numeric vector
- V71 a numeric vector
- V72 a numeric vector
- V73 a numeric vector
- V74 a numeric vector
- V75 a numeric vector
- V76 a numeric vector
- V77 a numeric vector
- V78 a numeric vector
- V79 a numeric vector
- V80 a numeric vector
- V81 a numeric vector
- V82 a numeric vector
- V83 a numeric vector
- V84 a numeric vector
- V85 a numeric vector
- V86 a numeric vector
- V87 a numeric vector
- V88 a numeric vector
- V89 a numeric vector
- V90 a numeric vector
- V91 a numeric vector
- V92 a numeric vector
- V93 a numeric vector
- V94 a numeric vector
- V95 a numeric vector
- V96 a numeric vector
- V97 a numeric vector
- V98 a numeric vector
- V99 a numeric vector

V100 a numeric vector  
V101 a numeric vector  
V102 a numeric vector  
V103 a numeric vector  
V104 a numeric vector  
V105 a numeric vector  
V106 a numeric vector  
V107 a numeric vector  
V108 a numeric vector  
V109 a numeric vector  
V110 a numeric vector  
V111 a numeric vector  
V112 a numeric vector  
V113 a numeric vector  
V114 a numeric vector  
V115 a numeric vector  
V116 a numeric vector  
V117 a numeric vector  
V118 a numeric vector  
V119 a numeric vector  
V120 a numeric vector  
V121 a numeric vector  
V122 a numeric vector  
V123 a numeric vector  
V124 a numeric vector  
V125 a numeric vector  
V126 a numeric vector  
V127 a numeric vector  
V128 a numeric vector

### Source

UCI repository

### References

Redmond, M. A. and A. Baveja: A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments. *European Journal of Operational Research* 141 (2002) 660-678.

### Examples

```
data(communities)
## maybe str(communities) ; plot(communities) ...
```

COR

*Calculate the optimal subset lengths on the COR***Description**

Calculate the optimal subset lengths on the COR

**Usage**

```
COR(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

**Arguments**

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

**Value**

A list containing:

seqL	The index of the subset with the minimum L value.
seqN	The index of the subset with the minimum N value.
lWMN	The optimal subset lengths on the COR.

**References**

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). [doi:10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

**Examples**

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1)))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%%beta+e;
COR(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

---

ethylene\_CO

*The chemical sensor data set*

---

**Description**

A data set about chemical sensor

**Usage**

```
data("ethylene_CO")
```

**Format**

A data frame with 4001 observations on the following 19 variables.

V1 a character vector

V2 a character vector

V3 a character vector

V4 a character vector

V5 a character vector

V6 a character vector

V7 a character vector

V8 a character vector

V9 a character vector

V10 a character vector

V11 a character vector

V12 a character vector

V13 a character vector

V14 a character vector

V15 a character vector

V16 a character vector

V17 a character vector

V18 a character vector

V19 a character vector

**Details**

We selected the first 4001 rows on the original data set about 1048576 observations on 19 variables.

**Source**

UCI Repository

## References

Wang, H. Y., Zhu, R., and Ma, P. (2018). Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association*, 113(522), 829-844.

## Examples

```
data(ethylene_CO)
## maybe str(ethylene_CO) ; plot(ethylene_CO) ...
```

---

LICbeta

*Calculate the LIC estimator for linear regression*

---

## Description

This function estimates the coefficients of a linear regression model using a design matrix ‘X’ and a response vector ‘Y’. It implements an A-optimal and D-optimal design criteria to choose optimal subsets of observations.

## Usage

```
LICbeta(X, Y, alpha, K, nk)
```

## Arguments

X	The observation matrix (n x p)
Y	The response vector (n x 1)
alpha	The significance level for computing confidence intervals
K	The number of subsets
nk	The number of observations per subset

## Value

A list containing:

E5	The LIC estimator for linear regression.
----	--

## References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). [doi:10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

---

LICnew	<i>Calculate the LIC estimator based on A-optimal and D-optimal criterion</i>
--------	---

---

### Description

Calculate the LIC estimator based on A-optimal and D-optimal criterion

### Usage

```
LICnew(X, Y, alpha, K, nk)
```

### Arguments

X	A matrix of observations (design matrix) with size $n \times p$
Y	A vector of responses with length $n$
alpha	The significance level for confidence intervals
K	The number of subsets to consider
nk	The size of each subset

### Value

A list containing:

E5	The LIC estimator based on A-optimal and D-optimal criterion.
----	---

### References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:[10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

### Examples

```
p = 6; n = 1000; K = 2; nk = 200; alpha = 0.05; sigma = 1
e = rnorm(n, 0, sigma); beta = c(sort(c(runif(p, 0, 1)))));
data = c(rnorm(n * p, 5, 10)); X = matrix(data, ncol = p);
Y = X %*% beta + e;
LICnew(X = X, Y = Y, alpha = alpha, K = K, nk = nk)
```

---

MSEbeta	<i>Calculate MSE values for different beta estimation methods</i>
---------	---

---

**Description**

Calculate MSE values for different beta estimation methods

**Usage**

MSEbeta(X, Y, alpha, K, nk)

**Arguments**

X	The design matrix (observations).
Y	The response vector.
alpha	The significance level.
K	The number of subsets.
nk	The length of subsets (number of observations in each subset).

**Value**

A list containing:

MSECOR	The MSE of the COR beta estimator.
MSEAopt	The MSE of the A-optimal beta estimator.
MSEDopt	The MSE of the D-optimal beta estimator.
MSElic	The MSE of the LIC beta estimator.

**References**

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

---

MSEcom	<i>Calculate the MSE values of the COR criterion in simulation</i>
--------	--

---

**Description**

Calculate the MSE values of the COR criterion in simulation

**Usage**

MSEcom(K = K, nk = nk, alpha = alpha, X = X, y = y)

**Arguments**

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

**Value**

A list containing:

MSEx	The Mean Squared Error between the true beta and the estimate betax based on the COR.
MSEA	The Mean Squared Error between the true beta and the estimate betaA based on the least squares estimate for subset A.
MSEc	The Mean Squared Error between the true beta and the estimate betac based on the COR-selected subset.
MSEm	The Mean Squared Error between the true beta and the median estimator betamm across all subsets.
MSEa	The Mean Squared Error between the true beta and the mean estimator betaa across all subsets.

**References**

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:[10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

**Examples**

```
p=6;n=1000;K=2;nk=500;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1)))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%%beta+e;
MSEcom(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

---

MSEver

*Calculate the MSE values of the COR criterion for redundant data in simulation*

---

**Description**

Calculate the MSE values of the COR criterion for redundant data in simulation

**Usage**

```
MSEver(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

**Arguments**

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

**Value**

A list containing:

minE	The minimum value of the error variance estimator.
Mcor	The MSE of the COR estimator.
Mx	The MSE of the estimator based on the subset with the maximum M.
MA	The MSE of the estimator based on the subset with the minimum W.

**References**

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:[10.1007/s1122202410471z](https://doi.org/10.1007/s1122202410471z)

**Examples**

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1)))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X**beta+e;
MSEver(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

# Index

## \* datasets

communities, [5](#)  
ethylene\_CO, [10](#)

beta\_AD, [2](#)  
beta\_cor, [3](#)  
beta\_LW, [4](#)

communities, [5](#)  
COR, [9](#)

ethylene\_CO, [10](#)

LICbeta, [11](#)  
LICnew, [12](#)

MSEbeta, [13](#)  
MSEcom, [13](#)  
MSEver, [14](#)