

# Package ‘POET’

January 20, 2025

**Type** Package

**Title** Principal Orthogonal ComplEment Thresholding (POET) Method

**Version** 2.0

**Date** 2016-06-26

**Author** Jianqing Fan, Yuan Liao, Martina Mincheva

**Maintainer** Martina Mincheva <m.z.mincheva@gmail.com>

**Description** Estimate large covariance matrices in approximate factor models by thresholding principal orthogonal complements.

**License** GPL-2

**LazyLoad** yes

**Repository** CRAN

**Date/Publication** 2016-06-29 02:33:30

**NeedsCompilation** no

## Contents

POET . . . . .	1
POETCmin . . . . .	3
POETKhat . . . . .	5

<b>Index</b>	<b>7</b>
--------------	----------

---

POET	<i>Principal Orthogonal ComplEment Thresholding (POET) method</i>
------	---

---

## Description

Estimates large covariance matrices in approximate factor models by thresholding principal orthogonal complements.

**Usage**

POET(Y, K, C, thres, matrix)

**Arguments**

- Y** p by n matrix of raw data, where p is the dimensionality, n is the sample size. It is recommended that Y is de-meant, i.e., each row has zero mean.
- K** number of factors. K is pre-determined by the users. Default value is set at the average value obtained from the Hallin&Liska and Bai&Ng methods. Suggestions on choosing K:  
 A simple way of determining K is to count the number of very spiked (much larger than others) eigenvalues of the p by p sample covariance matrix of Y.  
 A formal data-driven way of determining K is described in Bai and Ng (2002): "Determining the number of factors in approximate factor models", *Econometrica*, 70, 191-221. This procedure requires a one-dimensional optimization.  
 POET is very robust to over-estimating K. But under-estimating K can result to VERY BAD performance. Therefore we strongly recommend choosing a relatively large K (normally less than 8) to avoid missing any important common factor.  
 K=0 corresponds to thresholding the sample covariance directly.
- C** the positive constant for thresholding, user-specified. Default value is set at C=0.5 Our experience shows that C=0.5 performs quite well for soft thresholding.
- thres** choice of thresholding. Users can choose from three thresholding methods:  
 'soft': soft thresholding  
 'hard' hard thresholding  
 'scad': scad thresholding  
 'lasso': adaptive lasso thresholding  
 Default value is set at thres='soft'.  
 Details are found in Rothman et al. (2009): "Generalized thresholding of large covariance matrices." *JASA*, 104, 177-186
- matrix** the option of thresholding either correlation or covariance matrix. Users can choose from:  
 'cor': threshold the error correlation matrix then transform back to covariance matrix  
 'vad': threshold the error covariance matrix directly.  
 Default value is set at matrix='cor'.

**Details**

This function is for POET, proposed by Fan, Liao and Mincheva (2012) "Large Covariance Estimation by Thresholding Principal Orthogonal Complements", manuscript of Princeton University

Model:  $Y_t = Bf_t + u_t$ , where B,  $f_t$  and  $u_t$  represent factor loading matrix, common factors and idiosyncratic error respectively. Only  $Y_t$  is observable.  $t=1, \dots, n$ . Dimension of  $Y_t$  is p. The goal is to estimate the covariance matrices of  $Y_t$  and  $u_t$ .

Note: (1) POET is optimization-free, so no initial value, tolerant, or maximum iterations need to be specified as inputs.

(2) We can apply the adaptive thresholding (Cai and Liu 2011, JASA) on either the correlation matrix or the covariance matrix, specified by the option 'matrix'.

(3) If no factor structure is assumed, i.e., no common factors exist and  $\text{var}(Y_t)$  itself is sparse, set  $K=0$ .

### Value

SigmaY: estimated p by p covariance matrix of  $y_t$   
 SigmaU: estimated p by p covariance matrix of  $u_t$   
 factors: estimated unobservable factors in a K by T matrix form  
 loadings: estimated factor loadings in a p by K matrix form

### Author(s)

Jianqing Fan, Yuan Liao, Martina Mincheva

### References

Fan, Liao and Mincheva (2012) "Large Covariance Estimation in Approximate Factor Models by Thresholding Principal Orthogonal Complements", manuscript of Princeton University, arXiv: 1201.0175

### Examples

```
p=100
n=100
Y<-array(rnorm(p*n),dim=c(p,n))
Sy<-POET(Y,3,0.5,'soft','vad')$SigmaY
Su<-POET(Y,3,0.5,'soft','vad')$SigmaU
F<-POET(Y,3,0.5,'soft','vad')$factors
B<-POET(Y,3,0.5,'soft','vad')$loadings
```

---

POETCmin

*Cmin - Minimum threshold constant*

---

### Description

This function is for determining the minimum constant in the threshold that guarantees the positive definiteness of the POET estimator.

### Usage

```
POETCmin(Y, K, thres, matrix)
```

**Arguments**

Y	p by n matrix of raw data, where p is the dimensionality, n is the sample size. It is recommended that Y is de-meant, i.e., each row has zero mean.
K	number of factors. K is pre-determined by the users. Suggestions on choosing K: (1) A simple way of determining K is to count the number of very spiked (much larger than others) eigenvalues of the p by p sample covariance matrix of Y. (2) A formal data-driven way of determining K is described in Bai and Ng (2002): "Determining the number of factors in approximate factor models", <i>Econometrica</i> , 70, 191-221. This procedure requires a one-dimensional optimization. (3) POET is very robust to over-estimating K. But under-estimating K can result to VERY BAD performance. Therefore we strongly recommend choosing a relatively large K (normally less than 8) to avoid missing any important common factor. (4) K=0 corresponds to threshoding the sample covariance directly.
thres	choice of thresholding. Users can choose from three thresholding methods: 'soft': soft thresholding 'hard': hard thresholding 'scad': scad thresholding 'lasso': adaptive lasso thresholding Details are found in Rothman et al. (2009): "Generalized thresholding of large covariance matrices." <i>JASA</i> , 104, 177-186
matrix	the option of thresholding either correlation or covairance matrix. Users can choose from: 'cor': threshold the error correlation matrix then transform back to covariance matrix 'vad': threshold the error covariance matrix directly.

**Details**

Model:  $Y_t = Bf_t + u_t$ , where B,  $f_t$  and  $u_t$  represent factor loading matrix, common factors and idiosyncratic error respectively. Only  $Y_t$  is observable.  $t=1, \dots, n$ . Dimension of  $Y_t$  is p. The goal is to estimate the covariance matrices of  $Y_t$  and  $u_t$ .

Note: (1) POET is optimization-free, so no initial value, tolerant, or maximum iterations need to be specified as inputs.

(2) We can apply the adaptive thresholding (Cai and Liu 2011, *JASA*) on either the correlation matrix or the covariance matrix, specified by the option 'matrix'.

(3) If no factor structure is assumed, i.e., no common factors exist and  $\text{var}(Y_t)$  itself is sparse, set  $K=0$ .

**Value**

Outputs:

SigmaY: estimated p by p covariance matrix of  $y_t$

SigmaU: estimated p by p covariance matrix of  $u_t$

**Author(s)**

Jianqing Fan, Yuan Liao, Martina Mincheva

**References**

Fan, Liao and Mincheva (2012) "Large Covariance Estimation in Approximate Factor Models by Thresholding Principal Orthogonal Complements", manuscript of Princeton University, arXiv: 1201.0175

**Examples**

```
p=100
n=50
Y<-array(rnorm(p*n),dim=c(p,n))
C<-POETCmin(Y,3,'soft','vad')
```

---

 POETKhat

---

*Khat - number of factors in approximate factor model*


---

**Description**

This function is for calculating the optimal number of factors in an approximate factor model.

**Usage**

```
POETKhat(Y)
```

**Arguments**

Y                      p by n matrix of raw data, where p is the dimensionality, n is the sample size. It is recommended that Y is de-meaned, i.e., each row has zero mean.

**Details**

This method was proposed by Bai & Ng (2002) and Hallin & Liska (2007). They propose two penalty functions and in turn minimize the corresponding information criteria. Notice that this method may underestimate K. POET is very robust to over-estimating K. But under-estimating K can result to VERY BAD performance. Therefore we strongly recommend choosing a relatively large K (normally less than 8) to avoid missing any important common factor.

**Value**

K1HL                    estimated number of factors based on the first information criterion using Hallin & Liska method

K2HL                    estimated number of factors based on the second information criterion using Hallin & Liska method

K1BN	estimated number of factors based on the first information criterion using Bai & Ng method
K2BN	estimated number of factors based on the second information criterion using Bai & Ng method

**Author(s)**

Jianqing Fan, Yuan Liao, Martina Mincheva

**References**

Bai,Ng,2002.Determining the number of factors in approximate factor models. *Econometrica* 70,191-221.

Hallin,Liska,2007.Determining the number of factors in the general dynamic factor model.*JASA* 102,603-617.

Alessi,Barigozzi,Capasso,2010. Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters* 80, 1806-1813.

**Examples**

```
p=100
n=100
Y<-array(rnorm(p*n),dim=c(p,n))
K<-POETKhat(Y)
```

# Index

POET, 1  
POETCmin, 3  
POETKhat, 5