# Package ‘BAYSTAR’

May 6, 2026

**Version** 0.2-10

**Date** 2022-5-15

**Title** On Bayesian Analysis of Threshold Autoregressive Models

**Author** Cathy W. S. Chen, Edward M.H. Lin, F.C. Liu, and Richard Gerlach

**Maintainer** Edward M.H. Lin <ed.mhlin@gmail.com>

**Depends** R (>= 3.0.1), mvtnorm, coda

**Description**
Fit two-regime threshold autoregressive (TAR) models by Markov chain Monte Carlo methods.

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**RoxygenNote** 7.2.0

**Date/Publication** 2022-05-14 23:10:09 UTC

## Contents

BAYSTAR *Threshold Autoregressive model: Bayesian approach*

## Description

Bayesian estimation and one-step-ahead forecasting for two-regime TAR model, as well as monitoring MCMC convergence. One may want to allow for higher-order AR models in the different regimes. Parsimonious subset AR could be assigned in each regime in the BAYSTAR function rather than a full AR model (i.e. the autoregressive orders could be not a continuous series).

## Usage

```
BAYSTAR(x, lagp1, lagp2, Iteration, Burnin, constant, d0,
        step.thv, thresVar, mu01, v01, mu02, v02, v0, lambda0, refresh,tplot)
```

## Arguments

| | |
|---|---|
| `x` | A vector of time series. |
| `lagp1` | A vector of non-zero autoregressive lags for the lower regime (regime one). For example, an AR model with p1=3 in lags 1,3, and 5 would be set as `lagp1<-c(1,3,5)`. |
| `lagp2` | A vector of non-zero autoregressive lags for the upper regime (regime two). |
| `Iteration` | The number of MCMC iterations. |
| `Burnin` | The number of burn-in iterations for the sampler. |
| `constant` | The intercepts include in the model for each regime, if `constant`=1. Otherwise, if `constant`=0. (Default: `constant`=1) |
| `d0` | The maximum delay lag considered. (Default: `d0` = 3) |
| `step.thv` | Step size of tuning parameter for the Metropolis-Hasting algorithm. |
| `thresVar` | A vector of time series for the threshold variable. (if missing, the series x is used.) |
| `mu01` | The prior mean of $\phi$ in regime one. This setting can be a scalar or a column vector with dimension equal to the number of $\phi$. If this sets a scalar value, then the prior mean for all of $\phi$ are this value. (Default: a vector of zeros) |
| `v01` | The prior covariance matrix of $\phi$ in regime one. This setting can either be a scalar or a square matrix with dimensions equal to the number of $\phi$. If this sets a scalar value, then prior covariance matrix of $\phi$ is that value times an identity matrix. (Default: a diagonal matrix are set to 0.1) |
| `mu02` | The prior mean of $\phi$ in regime two. This setting can be a scalar or a column vector with dimension equal to the number of $\phi$. If this sets a scalar value, then the prior mean for all of $\phi$ are this value. (Default: a vector of zeros) |
| `v02` | The prior covariance matrix of $\phi$ in regime two. This setting can either be a scalar or a square matrix with dimensions equal to the number of $\phi$. If this sets a scalar value, then prior covariance matrix of $\phi$ is that value times an identity matrix. (Default: a diagonal matrix are set to 0.1) |

| | |
|---|---|
| `v0` | `v0`/2 is the shape parameter for Inverse-Gamma prior of $\sigma^2$. (Default: `v0` = 3) |
| `lambda0` | `lambda0`*`v0`/2 is the scale parameter for Inverse-Gamma prior of $\sigma^2$. (Default: `lambda0` = the residual mean squared error of fitting an AR(p1) model to the data.) |
| `refresh` | Each `refresh` iteration for monitoring MCMC output. (Default: `refresh=Iteration/2`) |
| `tplot` | Trace plots and ACF plots for all parameter estimates. (Default: `tplot=FALSE` ) |

## Details

Given the maximum AR orders p1 and p2, the two-regime SETAR(2:p1;p2) model is specified as:

$$x_t = (\phi_0^{(1)} + \phi_1^{(1)} x_{t-1} + \ldots + \phi_{p1}^{(1)} x_{t-p1} + a_t^{(1)}) I(z_{t-d} <= th) + (\phi_0^{(2)} + \phi_1^{(2)} x_{t-1} + \ldots + \phi_{p2}^{(2)} x_{t-p2} + a_t^{(2)} I(z_{t-d} > th)$$

where $th$ is the threshold value for regime switching; $z_t$ is the threshold variable; $d$ is the delay lag of threshold variable; and the error term $a_t^{(j)}$, $j, (j = 1, 2)$, for each regime is assumed to be an i.i.d. Gaussian white noise process with mean zero and variance $sigma_j^2, j = 1, 2$. I(A) is an indicator function. Event A will occur if I(A)=1 and otherwise if I(A)=0. One may want to allow parsimonious subset AR model in each regime rather than a full AR model.

## Value

A list of output with containing the following components:

| | |
|---|---|
| `mcmc` | All MCMC iterations. |
| `posterior` | The initial `Burnin` iterations are discarded as a burn-in sample, the final sample of (`Iteration-Burnin`) iterates is used for posterior inference. |
| `coef` | Summary Statistics of parameter estimation based on the final sample of (`Iteration-Burnin`) iterates. |
| `residual` | Residuals from the estimated model. |
| `lagd` | The mode of time delay lag of the threshold variable. |
| `DIC` | The deviance information criterion (DIC); a Bayesian method for model comparison (Spiegelhalter et al, 2002) |

## Author(s)

Cathy W. S. Chen, Edward M.H. Lin, F.C. Liu, and Richard Gerlach

## Examples

```
set.seed(989981)
## Set the true values of all parameters
nob<- 200                ## No. of observations
lagd<- 1                  ## delay lag of threshold variable
r<- 0.4                   ## r is the threshold value
sig.1<- 0.8; sig.2<- 0.5  ## variances of error distributions for two regimes
p1<- 2; p2<- 1            ## No. of covariate in two regimes
ph.1<- c(0.1,-0.4,0.3)    ## mean coefficients for regime 1
```

```
ph.2<- c(0.2,0.6)     ## mean coefficients for regime 2
lagp1<-1:2
lagp2<-1:1

yt<- TAR.simu(nob,p1,p2,ph.1,ph.2,sig.1,sig.2,lagd,r,lagp1,lagp2)


## Total MCMC iterations and burn-in iterations
Iteration <- 500
Burnin    <- 200


## A RW (random walk) MH algorithm is used in simulating the threshold value
## Step size for the RW MH
step.thv<- 0.08

out <- BAYSTAR(yt,lagp1,lagp2,Iteration,Burnin,constant=1,step.thv=step.thv,tplot=TRUE)
```

---

TAR.coeff *Estimate AR coefficients*

---

## Description

We assume a normal prior for the AR coefficients and draw AR coefficients from a multivariate normal posterior distribution. Parsimonious subset AR could be assigned in each regime in the BAYSTAR function rather than a full AR model.

## Usage

```
TAR.coeff(reg, ay, p1, p2, sig, lagd,
          thres, mu0, v0, lagp1, lagp2, constant = 1, thresVar)
```

## Arguments

A list containing:

| | |
|---|---|
| reg | The regime is assigned. (equal to one or two) |
| ay | The real data set. (input) |
| p1 | Number of AR coefficients in regime one. |
| p2 | Number of AR coefficients in regime two. |
| sig | The error terms of TAR model. |
| lagd | The delay lag parameter. |
| thres | The threshold parameter. |
| mu0 | Mean vector of conditional prior distribution in mean equation. |
| v0 | Covariance matrix of conditional prior distribution in mean equation. |

| | |
|---|---|
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |
| `thresVar` | Exogenous threshold variable. (if missing, the self series are used) |

### Author(s)

Cathy W.S. Chen, F.C. Liu

---

`TAR.lagd` *Identification of lag order of threshold variable*

---

### Description

The delay d has a discrete uniform prior over the integers: 1,2,..., d0, where d0 is a set maximum delay. We draw the delay lag of threshold variable from a multinomial distribution.

### Usage

```
TAR.lagd(ay, p1, p2, ph.1, ph.2, sig.1, sig.2,
         thres, lagp1, lagp2, constant = 1, d0, thresVar)
```

### Arguments

A list containing:

| | |
|---|---|
| `ay` | The real data set. (input) |
| `p1` | Number of AR coefficients in regime one. |
| `p2` | Number of AR coefficients in regime two. |
| `ph.1` | The vector of AR parameters in regime one. |
| `ph.2` | The vector of AR parameters in regime two. |
| `sig.1` | The error terms of AR model in the regime one. |
| `sig.2` | The error terms of AR model in the regime two. |
| `thres` | The threshold parameter. |
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |
| `d0` | The maximum delay lag could be selected. |
| `thresVar` | Exogenous threshold variable. (if missing, the series x is used) |

## Author(s)

Cathy W.S. Chen, Edward Lin

---

`TAR.lik` *Log-likelihood function*

---

## Description

To calculate the model log-likelihood function.

## Usage

```
TAR.lik(ay, p1, p2, ph.1, ph.2, sig.1, sig.2,
        lagd, thres, lagp1, lagp2, constant = 1, thresVar)
```

## Arguments

A list containing:

| | |
|---|---|
| `ay` | The real data set. (input) |
| `p1` | Number of AR coefficients in regime one. |
| `p2` | Number of AR coefficients in regime two. |
| `ph.1` | The vector of AR parameters in regime one. |
| `ph.2` | The vector of AR parameters in regime two. |
| `sig.1` | The error terms of AR model in the regime one. |
| `sig.2` | The error terms of AR model in the regime two. |
| `lagd` | The delay lag parameter. |
| `thres` | The threshold parameter. |
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |
| `thresVar` | Exogenous threshold variable. (if missing, the series x is used) |

## Author(s)

Cathy W.S. Chen, F.C. Liu

## TAR.sigma *To draw the variance of error distribution.*

### Description

We employ a conjugate prior, Inverse-Gamma distribution, for sigma squared in regime j, j=1,2. To draw the variance of error distribution from an Inverse-Gamma posterior distribution.

### Usage

```
TAR.sigma(reg, ay, thres, lagd, p1, p2, ph, v,
          lambda, lagp1, lagp2, constant = 1, thresVar)
```

### Arguments

A list containing:

| | |
|---|---|
| `reg` | The regime is assigned. (equal to one or two) |
| `thres` | The threshold parameter. |
| `lagd` | The delay lag parameter. |
| `p1` | Number of AR coefficient in regime one. |
| `p2` | Number of AR coefficient in regime two. |
| `ph` | The vector of AR parameters in regime `reg`. |
| `ay` | The real data set. (input) |
| `v`, `lambda` | The hyper-parameter of Inverse Gamma distribution for priors of variance. (i.e. IG(v/2,lambda/2)) |
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |
| `thresVar` | Exogenous threshold variable. (if missing, the series x is used) |

### Author(s)

Cathy W.S. Chen, Edward Lin

TAR.simu *Simulated data from TAR model*

## Description

To generate the simulated data from TAR(2;p1,p2) model.

## Usage

```
TAR.simu(nob, p1, p2, ph.1, ph.2, sig.1, sig.2, lagd, thres, lagp1, lagp2)
```

## Arguments

| | |
|---|---|
| nob | Number of observations that we want to simulate. |
| p1 | Number of AR coefficient in regime one. |
| p2 | Number of AR coefficient in regime two. |
| ph.1 | The vector of AR parameters in regime one. |
| ph.2 | The vector of AR parameters in regime two. |
| sig.1 | The error terms in regime one. |
| sig.2 | The error terms in regime two. |
| lagd | The delay lag parameter. |
| thres | The threshold parameter. |
| lagp1 | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| lagp2 | The vector of non-zero autoregressive lags for the upper regime. (regime two) |

## Author(s)

Cathy W.S. Chen, Edward Lin

## Examples

```
## Set the true values of all parameters
nob<- 2000                 ## No. of observations
lagd<- 1                   ## delay lag of threshold variable
r<- 0.4                    ## r is the threshold value
sig.1<- 0.8; sig.2<- 0.5  ## variances of error distributions for two regimes
p1<- 2; p2<- 2             ## No. of covariate in two regimes
ph.1<- c(0.1,-0.4,0.3)     ## mean coefficients for regime 1
ph.2<- c(0.2,0.3,0.3)      ## mean coefficients for regime 2
lagp1<-1:2
lagp2<-1:2

yt<- TAR.simu(nob,p1,p2,ph.1,ph.2,sig.1,sig.2,lagd,r,lagp1,lagp2)
```

## TAR.summary — *Calculate summary statistics for all parameters*

### Description

A summary of the MCMC output can be obtained via the function `TAR.summary`. `TAR.summary` returns the posterior mean, median, standard deviation and the lower and upper bound of the 95% Bayes posterior interval for all parameters, all botained from the sampling period only, after burn-in.

### Usage

```
TAR.summary(x, lagp1, lagp2, constant = 1)
```

### Arguments

A list containing:

| | |
|---|---|
| `x` | A matrix of the MCMC output of estimater parameters. |
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |

### Author(s)

Cathy W.S. Chen, F.C. Liu

## TAR.thres — *To draw a threshold value.*

### Description

The prior for the threshold parameter $thres$, follows a uniform prior on a range (l,u), where l and u can be set as relevant percentiles of the observed threshold variable. This prior could be considered to correspond to an empirical Bayes approach, rather than a fully Bayesian one. The posterior distribution of $thres$ is not of a standard distributional form, thus requiring us to use the Metropolis-Hastings (MH) method to achieve the desired sample for $thres$.

### Usage

```
TAR.thres(ay, p1, p2, ph.1, ph.2, sig.1, sig.2, lagd, thres,
          step.r = 0.02, bound, lagp1, lagp2, constant = 1, thresVar)
```

### Arguments

A list containing:

| | |
|---|---|
| `ay` | The real data set. (input) |
| `p1` | Number of AR coefficients in regime one. |
| `p2` | Number of AR coefficients in regime two. |
| `ph.1` | The vector of AR parameters in regime one. |
| `ph.2` | The vector of AR parameters in regime two. |
| `sig.1` | The error terms of AR model in the regime one. |
| `sig.2` | The error terms of AR model in the regime two. |
| `lagd` | The delay lag parameter. |
| `thres` | The threshold parameter. |
| `step.r` | Step size of threshold variable for the MH algorithm are controlled the proposal variance. |
| `bound` | The bound of threshold parameter. |
| `lagp1` | The vector of non-zero autoregressive lags for the lower regime. (regime one); e.g. An AR model with p1=3, it could be non-zero lags 1,3, and 5 would set lagp1<-c(1,3,5). |
| `lagp2` | The vector of non-zero autoregressive lags for the upper regime. (regime two) |
| `constant` | Use the CONSTANT option to fit a model with/without a constant term (1/0). By default CONSTANT=1. |
| `thresVar` | Exogenous threshold variable. (if missing, the series x is used) |

### Author(s)

Cathy W.S. Chen, F.C. Liu

---

`unemployrate` *U.S. monthly civilian unemployment rate*

---

### Description

U.S. monthly civilian unemployment rate, seasonally adjusted and measured in percentage, from January 1948 to March 2004. The data set is available in Tsay (2005).

### Usage

```
data(unemployrate)
```

### Format

unemployrate is a data frame with 675 observations

## Source

The data are obtained from the Bureau of Labor statistics, Department of Labor.

## References

Tsay, R.S. (2005) Analysis of Financial Time Series. Second ed. Wiley, Hoboken.

## Examples

```
data(unemployrate)
plot.ts(unemployrate)
```

# Index