Package ‘HSAR’

October 18, 2017

Type Package
Title Hierarchical Spatial Autoregressive Model
Version 0.4.2
Date 2017-10-16
Author Guanpeng Dong, Richard Harris, Angelos Mimis
Maintainer Angelos Mimis <mimis@panteion.gr>
Description A Hierarchical Spatial Autoregressive Model (HSAR), based on a Bayesian Markov Chain Monte Carlo (MCMC) algorithm (Dong and Harris (2014) <doi:10.1111/gean.12049>). The creation of this package was supported by the Economic and Social Research Council (ESRC) through the Applied Quantitative Methods Network: Phase II, grant number ES/K006460/1.
License GPL (>= 2)
Depends R (>= 2.7.0)
Imports spdep, Rcpp
LinkingTo Rcpp, RcppArmadillo
Suggests maptools, rgdal, rgeos, RColorBrewer, classInt, knitr
VignetteBuilder knitr
NeedsCompilation yes
Repository CRAN
Date/Publication 2017-10-18 13:36:07 UTC

R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSAR-package</td>
<td>2</td>
</tr>
<tr>
<td>Beijingdistricts</td>
<td>3</td>
</tr>
<tr>
<td>hsr</td>
<td>3</td>
</tr>
<tr>
<td>landprice</td>
<td>8</td>
</tr>
<tr>
<td>landSPDF</td>
<td>9</td>
</tr>
<tr>
<td>sar</td>
<td>10</td>
</tr>
<tr>
<td>summary.mcmc_hsar</td>
<td>12</td>
</tr>
<tr>
<td>summary.mcmc_hsar_lambda_0</td>
<td>13</td>
</tr>
</tbody>
</table>
Hierarchical Spatial Autoregressive Model

Description

Implements a Hierarchical Spatial Simultaneous Autoregressive Model (HSAR) or a multi-scale spatial econometrics model, with inference in a Bayesian setting using Markov chain Monte Carlo (MCMC) simulation. The approach is developed for modelling geographic data with a hierarchical/nested structure, for example, houses nesting into districts or fine-grained spatial units into more aggregated units. The HSAR model brings together the spatial econometrics and multilevel models and thus suitable for a simultaneous capturing the potential spatial dependence (autocorrelations) at each level of the data hierarchy arising from geographical proximity effect and the contextual effect (or group dependence effect) from higher-level units upon lower-level units. The creation of this package was supported by the Economic and Social Research Council (ESRC) through the Applied Quantitative Methods Network: Phase II, grant number ES/K006460/1.

Details

Package: HSAR
Type: Package
Version: 0.4.2
Date: 2017-10-16
License: GPL (>= 2)

Author(s)

Guanpeng Dong, Richard Harris, Angelos Mimis <mimis@panteion.gr>

References

Beijingdistricts

**Beijingdistricts**  
*Boundaries of districts (SpatialPolygonsDataFrame) in Beijing*

**Description**

The geographic boundaries of districts (SpatialPolygonsDataFrame) in Beijing. This is only a subset of districts in Beijing where our land parcel samples fall into.

**Usage**

```r
data(Beijingdistricts)
```

**See Also**

`landprice, landSPDF`

**Examples**

```r
data(Beijingdistricts)

library(spdep)
plot(Beijingdistricts, border="light grey")
# extract the area of each district
library(rgeos)
library(classInt)
library(RColorBrewer)

Beijingdistricts$geo.area <- gArea(Beijingdistricts, byid=TRUE) / 1000000
x <- Beijingdistricts$geo.area
breaks <- classIntervals(x, 4, "fisher")$brks
groups <- cut(x, breaks, include.lowest=TRUE, labels=FALSE)
palette <- brewer.pal(4, "Blues")
plot(Beijingdistricts, col=palette[groups], border="grey")

# extract the district level spatial weights matrix
nb.list <- poly2nb(Beijingdistricts, queen=FALSE)
mat.list <- nb2mat(nb.list, style="W")
M <- as(mat.list, "dgCMatrix")
```

**hsar**

*Hierarchical SAR model estimation*
Description

The specification of a HSAR model is as follows:

\[ y_{i,j} = \rho \ast W_i \ast y + x'_{i,j} \ast \beta + z'_{j} \ast \gamma + \theta_j + \epsilon_{i,j} \]

\[ \theta_j = \lambda \ast M_j \ast \theta + \mu_j \]

\[ \epsilon_{i,j} \sim N(0, \sigma^2_e), \quad \mu_j \sim N(0, \sigma^2_u) \]

where \( i = 1, 2, ..., n \) and \( j = 1, 2, ..., J \) are indicators of lower- and higher-level spatial units. \( n_j \) is the number of lower-level units in the \( j \)-th higher level unit and \( \sum_{j=1}^J = N \). \( x'_{i,j} \) and \( z'_{j} \) represent vectors of lower- and higher-level independent variables. \( \beta \) and \( \gamma \) are regression coefficients to estimate. \( \theta \), a \( N \times J \) vector of higher-level random effects, also follows a simultaneous autoregressive process. \( W \) and \( M \) are two spatial weights matrices (or neighbourhood connection matrices) at the lower and higher levels, defining how spatial units at each level are connected. \( \rho \) and \( \lambda \) are two spatial autoregressive parameters measuring the strength of the dependencies/correlations at the two spatial scales.

A succinct matrix formulation of the model is,

\[ y = \rho \ast W \ast y + X \ast \beta + Z \ast \gamma + \Delta \ast \theta + \epsilon \]

\[ \theta = \lambda \ast M \ast \theta + \mu \]

It is also useful to note that the HSAR model nests a standard (random intercept) multilevel model model when \( \rho \) and \( \lambda \) are both equal to zero and a standard spatial econometric model when \( \lambda \) and \( \sigma^2_u \) are both equal to zero.

Usage

hsar(formula, data = NULL, W=NULL, M=NULL, Delta, burnin=5000, Nsim=10000, thinning = 1, parameters.start = NULL)

Arguments

formula A symbolic description of the model to fit. A formula for the covariate part of the model using the syntax of the lm() function fitting standard linear regression models. Neither the response variable nor the explanatory variables are allowed to contain NA values.

data A data frame containing variables used in the formula object.

w The N by N lower-level spatial weights matrix or neighbourhood matrix where N is the total number of lower-level spatial units. The formulation of W could be based on geographical distances separating units or based on geographical contiguity. To ensure the maximum value of the spatial autoregressive parameter \( \rho \) less than 1, W should be row-normalised before running the HSAR model. As
in most cases, spatial weights matrix is very sparse, therefore W here should be converted to a sparse matrix before imported into the hsar() function to save computational burden and reduce computing time. More specifically, W should be a column-oriented numeric sparse matrices of a dgCMatrix class defined in the Matrix package. The conversion between a dense numeric matrix and a sparse numeric matrix is made quite convenient through the Matrix library.

\( M \)  
The J by J higher-level spatial weights matrix or neighbourhood matrix where J is the total number of higher-level spatial units. Similar with W, the formulation of M could be based on geographical distances separating units or based on geographical contiguity. To ensure the maximum value of the spatial autoregressive parameter \( \lambda \) less than 1, M is also row-normalised before running the HSAR model. As with W, M should also be a column-oriented numeric sparse matrices.

\( \Delta \)  
The N by J random effect design matrix that links the J by 1 higher-level random effect vector back to the N by 1 response variable under investigation. It is simply how lower-level units are grouped into each high-level units with columns of the matrix being each higher-level units. As with W and M, \( \delta \) should also be a column-oriented numeric sparse matrices.

\text{burnin}  
The number of MCMC samples to discard as the burnin period.

\text{nsim}  
The total number of MCMC samples to generate.

\text{thinning}  
MCMC thinning factor.

\text{parameters.start}  
A list with names "rho", "lambda", "sigma2e", "sigma2u" and "beta" corresponding to initial values for the model parameters \( \rho, \lambda, \sigma^2_e, \sigma^2_u \) and the regression coefficients respectively.

Details

In order to use the hsar() function, users need to specify the two spatial weights matrices W and M and the random effect design matrix \( \delta \). However, it is very easy to extract such spatial weights matrices from spatial data using the package \texttt{spdep}. Geographic distance-based or contiguity-based spatial weights matrix for both spatial points data and spatial polygons data are available in the \texttt{spdep} package.

Before the extraction of W and M, it is better to first sort the data using the higher-level unit identifier. Then, the random effect design matrix can be extracted simply (see the following example) and so are the two spatial weights matrices. Make sure the order of higher-level units in the weights matrix M is in line with that in the \( \delta \) matrix.

Two simpler versions of the HSAR model can also be fitted using the hsar() function. The first is a HSAR model with \( \lambda \) equal to zero, indicating an assumption of independence in the higher-level random effect \( \theta \). The second is a HSAR with \( \rho \) equal to zero, indicating an independence assumption in the outcome variable conditioning on the higher-level random effect. This model is useful in situations where we are interested in the neighbourhood/contextual effect on individual’s outcomes and have good reasons to suspect the effect from geographical contexts upon individuals to be dependent. Meanwhile we have no information on how lower-level units are connected.
Value

A list object containing:

- `mbetas` A vector of estimated mean values of regression coefficients.
- `sdbetas` The standard deviations of estimated regression coefficients.
- `mrho` The estimated mean of the lower-level spatial autoregressive parameter $\rho$.
- `sdrho` The standard deviation of the estimated lower-level spatial autoregressive parameter.
- `mlambda` The estimated mean of the higher-level spatial autoregressive parameter $\lambda$.
- `sdlambda` The standard deviation of the estimated higher-level spatial autoregressive parameter.
- `msigmaRe` The estimated mean of the lower-level variance parameter $\sigma_e^2$.
- `sdsigmaRe` The standard deviation of the estimated lower-level variance parameter $\sigma_e^2$.
- `msigmaRu` The estimated mean of the higher-level variance parameter $\sigma_u^2$.
- `sdsigmaRu` The standard deviation of the estimated higher-level variance parameter $\sigma_u^2$.
- `mus` Mean values of $\theta$
- `s dus` Standard deviation of $\theta$
- `DIC` The deviance information criterion (DIC) of the fitted model.
- `pd` The effective number of parameters of the fitted model.
- `log_likelihood` The log-likelihood of the fitted model.
- `R_Squared` A pseudo R square model fit indicator.
- `impact_direct` Summaries of the direct impact of a covariate effect on the outcome variable.
- `impact_idirect` Summaries of the indirect impact of a covariate effect on the outcome variable.
- `impact_total` Summaries of the total impact of a covariate effect on the outcome variable.

Author(s)

Guanpeng Dong, Richard Harris, Angelos Mimis <mimis@panteion.gr>

References


See Also

`sar`, `landprice`, `Beijingdistricts`, `landSPDF`
Examples

library(spdep)

# Running the hsar() function using the Beijing land price data
data(landprice)

data(Beijingdistricts)
data(landSPDF)

plot(Beijingdistricts,border="green")
plot(landSPDF,add=TRUE,col="red",pch=16,cex=0.8)

# Define the random effect matrix
model.data <- landprice[order(landprice$district.id),]
head(model.data,50)

# the number of individuals within each neighbourhood
MM <- as.data.frame(table(model.data$district.id))
# the total number of neighbourhood, 100
Utotal <- dim(MM)[1]
Unum <- MM[2]
Uid <- rep(c(1:Utotal),Unum)

n <- nrow(model.data)
Delta <- matrix(0,nrow=n,ncol=Utotal)
for(i in 1:Utotal) {
  Delta[Uid==i,i] <- 1
}

# Delta[1:50,1:10]
Delta <- as(Delta,"dgCMatrix")

# extract the district level spatial weights matrix using the queen's rule
nb.list <- poly2nb(Beijingdistricts)
mat.list <- nb2mat(nb.list,style="W")
M <- as(mat.list,"dgCMatrix")

# extract the land parcel level spatial weights matrix
nb.25 <- dnearneigh(landSPDF,0,2500)
# to a weights matrix
dist.25 <- nb2dists(nb.25,landSPDF)
dist.25 <- lapply(dist.25,function(x) exp(-0.5 * (x / 2500)^2))
mat.25 <- nb2mat(nb.25,glist=dist.25,style="W")
W <- as(mat.25,"dgCMatrix")

## run the hsar() function
res.formula <- lnprice ~ lnarea + lndcbd + dsubway + dpark + dele +
popden + crimerate + as.factor(year)

betas= coef(lm(formula=res.formula,data=landprice))
landprice

Leased residential land parcels, from 2003 to 2009 in Beijing, China

Description

The residential land parcel data leased to real estate developers from the government during 2003 to 2009. The data contains price information and a range of locational and neighbourhood characteristics for each land parcel. There are 1117 samples in the data after dropping those without price information.

Usage

data(landprice)

Format

A data frame with 1117 observations on the following 11 variables.

obs  An unique identifier for each land parcel.
lnprice  The log of the leasing price per square metre of each residential land parcel (unit: RMB, Chinese yuan)
dsubway  The log of the distance of each land parcel to the nearest railway station (unit: meters)
dele  The log of the distance of each land parcel to the nearest elementary school (unit: meters)
dpark  The log of the distance of each land parcel to the nearest green park (unit: meters)
lnarea  The log of the size of each land parcel (unit: square meters).
lncbd  The log of the distance of each land parcel to the CBD (centre business district) in Beijing (unit: meters)
year The year when each land parcel was leased with values of 0,1,2,3,4,5,6 representing year 2003,2004,2005,2006,2007,2008,2009

popden The population density of each district (unit: 1000 persons per square kilometers)

crimerate The number of reported serious crimes committed in each district per 1000 persons.

district.id The identifier of the district where each land parcel is located.

References


See Also

*Beijingdistricts, landSPDF*

Examples

data(landprice)
head(landprice)

---

**landSPDF**

*Beijing land price data*

**Description**

The spatial locations (SpatialPointsDataFrame) of the Beijing land price data. It is used to extract the spatial weights matrix among land parcels.

**Usage**

data(landSPDF)

**See Also**

*landprice, Beijingdistricts*
**Examples**

```r
data(landSPDF)
library(spdep)
plot(landSPDF,col="red",pch=16,cex=0.7)
data(landprice)
library(classInt)
library(RColorBrewer)

# link the variables in the landprice to the spatial data
index.match <- match(landSPDF$obs,landprice$obs)
landSPDF$data <- data.frame(landSPDF$data,landprice[index.match,])

par(mar=c(0,0,0,0))
x <- landSPDF$lnprice
breaks <- classIntervals(x,4,"fisher")$brks
groups <- cut(x,breaks,include.lowest=TRUE,labels=FALSE)
palette <- brewer.pal(4, "Blues")
plot(landSPDF,pch=19,col=palette[groups],cex=0.8)

# extract a spatial weights matrix based on the distances between pairs of land parcels
nb.25 <- dnearneigh(landSPDF,0,2500)
# to a weights matrix
dist.25 <- nbdists(nb.25,landSPDF)
dist.25 <- lapply(dist.25,function(x) exp(-0.5 * (x / 2500)^2))
mat.25 <- nb2mat(nb.25,glist=dist.25,style="W")
W <- as(mat.25,"dgCMatrix")
```

---

**sar**

**SAR model estimation**

**Description**

The `sar()` function implements a standard spatial econometrics model (SAR) or a spatially lagged dependent variable model using the Markov chain Monte Carlo (McMC) simulation approach.

**Usage**

```r
sar(formula, data = NULL, W, burnin=5000, Nsim=10000, thinning=1, parameters.start = NULL )
```

**Arguments**

- `formula`: A symbolic description of the model to fit. A formula for the covariate part of the model using the syntax of the `lm()` function fitting standard linear regression models. Neither the response variable nor the explanatory variables are allowed to contain NA values.
A data frame containing variables used in the formula object.

The N by N spatial weights matrix or neighbourhood matrix where N is the number of spatial units. The formulation of W could be based on geographical distances separating units or based on geographical contiguity. To ensure the maximum value of the spatial autoregressive parameter \( \rho \) less than 1, W is usually row-normalised before implementing the SAR model. As in most cases, spatial weights matrix is very sparse, therefore W here should be converted to a sparse matrix before imported into the sar() function to save computational burden and reduce computing time. More specifically, W should be a column-oriented numeric sparse matrices of a dgCMatrix class defined in the Matrix package. The conversion between a dense numeric matrix and a sparse numeric matrix is made quite convenient through the Matrix library.

The number of MCMC samples to discard as the burnin period.

The total number of MCMC samples to generate.

MCMC thinning factor.

A list with names "rho", "sigma2e", and "beta" corresponding to initial values for the model parameters \( \rho, \sigma^2_e \) and the regression coefficients respectively.

A vector of estimated mean values of regression coefficients.

The standard deviations of estimated regression coefficients.

The estimated mean of the lower-level spatial autoregressive parameter \( \rho \).

The standard deviation of the estimated lower-level spatial autoregressive parameter.

The estimated mean of the lower-level variance parameter \( \sigma^2_e \).

The standard deviation of the estimated lower-level variance parameter \( \sigma^2_e \).

The deviance information criterion (DIC) of the fitted model.

The effective number of parameters of the fitted model.

The log-likelihood of the fitted model.

A pseudo R square model fit indicator.

Summaries of the direct impact of a covariate effect on the outcome variable.

Summaries of the indirect impact of a covariate effect on the outcome variable.

Summaries of the total impact of a covariate effect on the outcome variable.

Guanpeng Dong, Richard Harris, Angelos Mimis <mimis@panteion.gr>
## References


## See Also

`hsar, landprice, Beijingdistricts, landSPDF`

## Examples

```r
data(landprice)
head(landprice)

data(landSPDF)

# extract the land parcel level spatial weights matrix
library(spdep)

nb25 <- dnearneigh(landSPDF, 0, 2500)

# to a weights matrix

dist25 <- nb2dists(nb25, landSPDF)
dist25 <- lapply(dist25, function(x) exp(-0.5 * (x / 2500)^2))

mat25 <- nb2mat(nb25, glist = dist25, style = "W")

W <- as(mat25, "dgCMatrix")

## run the sar() function
res.formula <- lnprice ~ lnarea + lnbdcbd + ds subway + dpark + dele + popden + crimerate + as.factor(year)

betas <- coef(lm(formula = res.formula, data = landprice))

pars <- list(rho = 0.5, sigma2e = 2.0, betas = betas)

## Not run:
res <- sar(res.formula, data = landprice, W = W,
            burnin = 5000, Nsim = 10000, thinning = 1,
            parameters.start = pars)

summary(res)

## End(Not run)
```

## Description

Methods for presenting the result of a fitted HSAR model.
**summary.mcmc_hsar_lambda_0**

**Usage**

```
## S3 method for class 'mcmc_hsar'
summary(object, ...)
## S3 method for class 'mcmc_hsar'
print(x, ...)
```

**Arguments**

- `object` `mcmc_hsar` An `mcmc_hsar` object returned from the `hsar` function
- `x` `mcmc_hsar` An `mcmc_hsar` object returned from the `hsar` function
- `...` Other arguments passed through

**See Also**

- `hsar`

---

**summary.mcmc_hsar_lambda_0**

*summary method for class mcmc_hsar_lambda_0*

---

**Description**

Methods for presenting the result of a fitted HSAR model with $\lambda$ equal to 0. This is a model without the consideration of the possible interdependency between higher-level spatial units.

**Usage**

```
## S3 method for class 'mcmc_hsar_lambda_0'
summary(object, ...)
## S3 method for class 'mcmc_hsar_lambda_0'
print(x, ...)
```

**Arguments**

- `object` `mcmc_hsar_lambda_0` An `mcmc_hsar_lambda_0` object returned from the `hsar` function
- `x` `mcmc_hsar_lambda_0` An `mcmc_hsar_lambda_0` object returned from the `hsar` function
- `...` Other arguments passed through

**See Also**

- `hsar`
Summary

Methods for presenting the result of a fitted HSAR model with $\rho$ equal to 0. This is a model without the consideration of the possible interdependency between lower-level spatial units.

Usage

```r
## S3 method for class 'mcmc_hsar_rho_0'
summary(object, ...) # S3 method for class 'mcmc_hsar_rho_0'
```

Arguments

- `object`: mcmc_hsar_rho_0 An mcmc_hsar_rho_0 object returned from the hsar function
- `x`: mcmc_hsar_rho_0 An mcmc_hsar_rho_0 object returned from the hsar function
- `...`: Other arguments passed through

See Also

- hsar

Summary

Methods for presenting the result of a fitted standard spatial econometric model or a spatially lagged dependent variable model

Usage

```r
## S3 method for class 'mcmc_sar'
summary(object, ...) # S3 method for class 'mcmc_sar'
```

Arguments

- `object`: mcmc_sar A mcmc_sar object returned from the mc_sar function
- `x`: mcmc_sar A mcmc_sar object returned from the mc_sar function
- `...`: Other arguments passed through

See Also

- mc_sar
Arguments

object
  mcmc_sar An mcmc_sar object returned from the sar function
x
  mcmc_sar An mcmc_sar object returned from the sar function
... Other arguments passed through

See Also

sar
Index

*Topic **hierarchical spatial data**
  Beijingdistricts, 3
  landprice, 8
  landSPDF, 9

*Topic **spatial econometrics**
  hsar, 3
  HSAR-package, 2
  sar, 10
    summary.mcmc_hsar, 12
    summary.mcmc_hsar_lambda_0, 13
    summary.mcmc_hsar_rho_0, 14
    summary.mcmc_sar, 14

Beijingdistricts, 3, 6, 9, 12

HSAR (HSAR-package), 2
  hsar, 3, 12–14
  HSAR-package, 2

landprice, 3, 6, 8, 9, 12
landSPDF, 3, 6, 9, 9, 12

print.mcmc_hsar (summary.mcmc_hsar), 12
print.mcmc_hsar_lambda_0
  (summary.mcmc_hsar_lambda_0), 13
print.mcmc_hsar_rho_0
  (summary.mcmc_hsar_rho_0), 14
print.mcmc_sar (summary.mcmc_sar), 14

sar, 6, 10, 15
summary.mcmc_hsar, 12
summary.mcmc_hsar_lambda_0, 13
summary.mcmc_hsar_rho_0, 14
summary.mcmc_sar, 14