Package 'NSAE'

May 27, 2022

Type Package

Title Nonstationary Small Area Estimation

Version 0.4.0

Maintainer Saurav Guha <saurav.iasri@gmail.com>

Description Executes nonstationary Fay-Herriot model and nonstationary generalized linear mixed model for small area estimation.The empirical best linear unbiased predictor (EBLUP) under stationary and nonstationary Fay-Herriot models and empirical best predictor (EBP) under nonstationary generalized linear mixed model along with the mean squared error estimation are included. EBLUP for prediction of non-sample area is also included under both stationary and nonstationary Fay-Herriot models. This extension to the Fay-Herriot model that accounts for the presence of spatial nonstationarity was developed by Hukum Chandra, Nicola Salvati and Ray Chambers (2015) <doi:10.1093/jssam/smu026> and nonstationary generalized linear mixed model was developed by Hukum Chandra, Nicola Salvati and Ray Chambers (2017) <doi:10.1016/j.spasta.2017.01.004>. This package is dedicated to the memory of Dr. Hukum Chandra who passed away while the package creation was in progress.

License GPL-3

Encoding UTF-8

LazyData true

Imports rlist,cluster,MASS,lattice,Matrix,numDeriv,nlme,spgwr,SemiPar

RoxygenNote 7.1.2

Depends R (>= 3.5.0)

NeedsCompilation no

Author Hukum Chandra [aut],
Nicola Salvati [aut],
Ray Chambers [aut],
Saurav Guha [aut, cre]

Repository CRAN

Date/Publication 2022-05-27 12:40:20 UTC

R topics documented:

`eblupFH1` *EBLUP under stationary Fay-Herriot model for sample area*

Description

This function gives the EBLUP and the estimate of mean squared error (mse) based on a stationary Fay-Herriot model for sample area.

Usage

```
eblupFH1(formula, vardir, method = "REML", MAXITER, PRECISION, data)
```

Arguments

`formula`	an object of class list of formula, describe the model to be fitted
`vardir`	a vector of sampling variances of direct estimators for each small area
`method`	type of fitting method, default is "REML" method
`MAXITER`	number of iterations allowed in the algorithm. Default is 100 iterations
`PRECISION`	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
`data`	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

eblup a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, eblup, mse, standard error (SE) and coefficient of variation (CV)

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- goodness : goodness of fit statistics
- randomeffect : a data frame with the values of the random effect estimators

Examples

```
# Load data set
data(paddysample)
# Fit Fay-Herriot model using sample part of paddy data
result <- eblupFH1(y ~ x1+x2, var, "REML", 100, 1e-04,paddysample)
result
```

`eblupFH2` *EBLUP under stationary Fay-Herriot model for sample and non-sample area*

Description

This function gives the EBLUP and the estimate of mean squared error (mse) based on a stationary Fay-Herriot model for both sample and non-sample area.

Usage

```
eblupFH2(formula, vardir, indicator, method = "REML", MAXITER, PRECISION, data)
```

Arguments

`formula`	an object of class list of formula, describe the model to be fitted
`vardir`	a vector of sampling variances of direct estimators for each small area
`indicator`	a vector indicating the sample and non-sample area
`method`	type of fitting method, default is "REML" methods
`MAXITER`	number of iterations allowed in the algorithm. Default is 100 iterations
`PRECISION`	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
`data`	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

eblup a vector with the values of the estimators for each sample area

eblup.out a vector with the values of the estimators for each non-sample area

mse a vector of the mean squared error estimates for each sample area

mse.out a vector of the mean squared error estimates for each non-sample area

sample a matrix consist of area code, eblup, mse, SE and CV for sample area

nonsample a matrix consist of area code, eblup, mse, SE and CV for non-sample area

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta),their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- goodness : goodness of fit statistics
- randomeffect : a data frame with the values of the random effect estimators

Examples

```
# Load data set
data(paddy)
# Fit Fay-Herriot model using sample and non-sample part of paddy data
result <- eblupFH2(y ~ x1+x2, var, indicator ,"REML", 100, 1e-04,paddy)
result
```

---

eblupNSFH1 *EBLUP under nonstationary Fay-Herriot model for sample area*

Description

This function gives the EBLUP and the estimate of mean squared error (mse) based on a nonstationary Fay-Herriot model for sample area.

Usage

```
eblupNSFH1(
  formula,
  vardir,
  lat,
  long,
  method = "REML",
  MAXITER,
  PRECISION,
  data
)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
lat	a vector of latitude for each small area
long	a vector of longitude for each small area
method	type of fitting method, default is "REML" methods
MAXITER	number of iterations allowed in the algorithm. Default is 100 iterations
PRECISION	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
data	a data frame comprising the variables named in formula, vardir, lat and long

Value

The function returns a list with the following objects:

eblup a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, eblup, mse, SE and CV

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta),their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- spatialcorr : spatial correlation parameter
- randomeffect : a data frame with the values of the random effect estimators
- goodness : goodness of fit statistics

Examples

```
# Load data set
data(paddysample)
# Fit nonstationary Fay-Herriot model using sample part of paddy data
result <- eblupNSFH1(y ~ x1+x2, var, latitude, longitude, "REML", 100, 1e-04,paddysample)
result
```

`eblupNSFH2` *EBLUP under nonstationary Fay-Herriot model for sample and non-sample area*

Description

This function gives the EBLUP and the estimate of mean squared error (mse) based on a nonstationary Fay-Herriot model for both sample and non-sample area.

Usage

```
eblupNSFH2(
  formula,
  vardir,
  lat,
  long,
  indicator,
  method = "REML",
  MAXITER,
  PRECISION,
  data
)
```

Arguments

`formula`	an object of class list of formula, describe the model to be fitted
`vardir`	a vector of sampling variances of direct estimators for each small area
`lat`	a vector of latitude for each small area
`long`	a vector of longitude for each small area
`indicator`	a vector indicating the sample and non-sample area
`method`	type of fitting method, default is "REML" methods
`MAXITER`	number of iterations allowed in the algorithm. Default is 100 iterations
`PRECISION`	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
`data`	a data frame comprising the variables named in formula, vardir, lat and long

Value

The function returns a list with the following objects:

eblup a vector with the values of the estimators for each sample area

eblup.out a vector with the values of the estimators for each non-sample area

mse a vector of the mean squared error estimates for each sample area

mse.out a vector of the mean squared error estimates for each non-sample area

sample a matrix consist of area code, eblup, mse, SE and CV for sample area

nonsample a matrix consist of area code, eblup, mse, SE and CV for non-sample area

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta),their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- spatialcorr : estimated spatial correlation parameter
- randomeffect : a data frame with the values of the random effect estimators
- goodness : goodness of fit statistics

Examples

```
# Load data set
data(paddy)
# Fit nonstationary Fay-Herriot model using sample and non-sample part of paddy data
result <- eblupNSFH2(y ~ x1+x2, var, latitude, longitude, indicator , "REML", 100, 1e-04,paddy)
result
```

---

ebp *EBP for proportion under generalized linear mixed model*

Description

This function gives the ebp and the estimate of mean squared error (mse) for proportion based on a generalized linear mixed model.

Usage

```
ebp(
  formula,
  vardir,
  Ni,
  ni,
  method = "REML",
  maxit = 100,
  precision = 1e-04,
  data
)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
Ni	a vector of population size for each small area

ni	a vector of sample size for each small area
method	type of fitting method, default is "REML" method
maxit	number of iterations allowed in the algorithm. Default is 100 iterations
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
data	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

ebp a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, ebp, mse, standard error (SE) and coefficient of variation (CV)

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- randomeffect : a data frame with the values of the random effect estimators
- loglike : value of the loglikelihood
- deviance : value of the deviance
- loglike1 : value of the restricted loglikelihood

Examples

```
# Load data set
data(headcount)
# Fit generalized linear mixed model using HCR data
result <- ebp(y~x1, var, N, n,"REML",100,1e-04, headcount)
result
```

---

ebpNP *Nonparametric ebp using spatial spline for proportion under generalized linear mixed model*

Description

This function gives the nonparametric ebp and the estimate of mean squared error (mse) for proportion based on a nonstationary generalized linear mixed model.

Usage

```
ebpNP(
  formula,
  vardir,
  n.knot,
  Ni,
  ni,
  lat,
  lon,
  method = "REML",
  maxit = 100,
  precision = 1e-04,
  data
)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
n.knot	number of knot in spatial splines. Default is 25 knot
Ni	a vector of population size for each small area
ni	a vector of sample size for each small area
lat	a vector of latitude for each small area
lon	a vector of longitude for each small area
method	type of fitting method, default is "REML" method
maxit	number of iterations allowed in the algorithm. Default is 100 iterations
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
data	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

ebp a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, ebp, mse, standard error (SE) and coefficient of variation (CV)

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance

- lambda : estimated spatial intensity paprameter
- randomeffect : a data frame with the values of the area specific random effect
- gamma : a data frame with the values of the spatially correlated random effect
- variance : a covariance matrix of estimated variance components

Examples

```
# Load data set
data(headcount)
# Fit a nonparametric generalized linear mixed model using headcount data
result <- ebpNP(y~x1, var,25, N, n,  lat, long, "REML", 100, 1e-04,headcount)
result
```

---

ebpNS *Nonstationary ebp for proportion under generalized linear mixed model*

Description

This function gives the nonstationary ebp and the estimate of mean squared error (mse) for proportion based on a generalized linear mixed model.

Usage

```
ebpNS(
  formula,
  vardir,
  Ni,
  ni,
  lat,
  lon,
  method = "REML",
  maxit = 100,
  precision = 1e-04,
  data
)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
Ni	a vector of population size for each small area
ni	a vector of sample size for each small area
lat	a vector of latitude for each small area
lon	a vector of longitude for each small area
method	type of fitting method, default is "REML" method

maxit	number of iterations allowed in the algorithm. Default is 100 iterations
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
data	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

ebp a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, ebp, mse, standard error (SE) and coefficient of variation (CV)

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- lambda : estimated spatial intensity parameter
- randomeffect : a data frame with the values of the area specific random effect
- gamma : a data frame with the values of the spatially correlated random effect
- variance : a covariance matrix of estimated variance components
- loglike : value of the loglikelihood
- deviance : value of the deviance
- loglike1 : value of the restricted loglikelihood

Examples

```
# Load data set
data(headcount)
# Fit a nonstationary generalized linear mixed model using headcount data
result <- ebpNS(y~x1, var, N, n, lat, long, "REML", 100, 1e-04, headcount)
result
```

---

ebpSP *Spatial ebp for proportion under generalized linear mixed model*

Description

This function gives the spatial ebp and the estimate of mean squared error (mse) for proportion based on a generalized linear mixed model.

Usage

```
ebpSP(
  formula,
  vardir,
  Ni,
  ni,
  proxmat,
  method = "REML",
  maxit = 100,
  precision = 1e-04,
  data
)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
Ni	a vector of population size for each small area
ni	a vector of sample size for each small area
proxmat	a D*D proximity matrix of D small areas. The matrix must be row-standardized.
method	type of fitting method, default is "REML" method
maxit	number of iterations allowed in the algorithm. Default is 100 iterations
precision	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
data	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with the following objects:

ebp a vector with the values of the estimators for each small area

mse a vector of the mean squared error estimates for each small area

sample a matrix consist of area code, ebp, mse, standard error (SE) and coefficient of variation (CV)

fit a list containing the following objects:

- estcoef : a data frame with the estimated model coefficients in the first column (beta), their asymptotic standard errors in the second column (std.error), the t statistics in the third column (tvalue) and the p-values of the significance of each coefficient in last column (pvalue)
- refvar : estimated random effects variance
- rho : estimated spatial correlation
- randomeffect : a data frame with the values of the area specific random effect
- variance : a covariance matrix of estimated variance components
- loglike : value of the loglikelihood
- deviance : value of the deviance

Examples

```
# Load data set
data(headcount)
# Fit a generalized linear mixed model with SAR spcification using headcount data
result <- ebpSP(ps~x1, var, N, n, Wmatrix, "REML", 100, 1e-04, headcount)
result
```

`headcount` *Head count data*

Description

Dataset on head count used by Chandra et al. (2017).

Usage

```
data(headcount)
```

Format

A data frame with 71 observations on the following 11 variables:

Area Small area code

lat Latitude of each small areas

long Longitude of each small areas

N Sample size of each small areas

n Sample size of each small areas

y Head count (direct estimates for the small areas)

ps proportion of head count

var Estimated variance

x1 First covariate used by Chandra et al. (2017)

x2 Second covariate used by Chandra et al. (2017)

x3 Second covariate used by Chandra et al. (2017)

Reference

Chandra, H., Salvati, N., & Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. Spatial Statistics. 20. 30-56. DOI:10.1016/j.spasta.2017.01.004.

Examples

```
data(headcount)
y <- headcount$y
summary(y)
```

NS.test *Parametric bootstrap-based spatial nonstationarity test for Fay-Herroit model*

Description

This function performs a parametric bootstrap-based test procudure for testing spatial nonstationarity in the data.

Usage

```
NS.test(formula, vardir, lat, long, iter = 100, data)
```

Arguments

formula	an object of class list of formula, describe the model to be fitted
vardir	a vector of sampling variances of direct estimators for each small area
lat	a vector of latitude for each small area
long	a vector of longitude for each small area
iter	number of iterations allowed in the algorithm. Default is 100 iterations
data	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with class "htest" containing the following components:

method a character string indicating what type of test was performed.

p.value the p-value for the test.

data.name a character string giving the name of the data.

Examples

```
# Load data set
data(paddysample)
# Testing spatial nonstationarity of the data
result <- NS.test(y ~ x1+x2, var, latitude, longitude, iter=50, data = paddysample[1:10,])
result
```

NSAE *NSAE : Nonstationary Small Area Estimation*

Description

Executes nonstationary Fay-Herriot model and nonstationary generalized linear mixed model for small area estimation. It produces empirical best linear unbiased predictor (EBLUP) and empirical best predictor (EBP) under stationary and nonstationary Fay-Herriot models. Functions give EBLUP and EBP estimators along with their mean squared error (MSE) estimator for each model. The nonstationary Fay-Herriot model was developed by Hukum Chandra, Nicola Salvati and Ray Chambers (2015) <doi:10.1093/jssam/smu026> and the nonstationary generalized linear mixed model was developed by Hukum Chandra, Nicola Salvati and Ray Chambers (2017) <doi:10.1016/j.spasta.2017.01.004>.

Author(s)

Hukum Chandra, Nicola Salvati, Ray Chambers, Saurav Guha

Maintainer: Saurav Guha <saurav.iasri@gmail.com>

Functions

`eblupFH1` Provides the EBLUPs and MSE under stationary Fay-Herriot model for sample area

`eblupFH2` Provides the EBLUPs and MSE under stationary Fay-Herriot model for sample and non-sample area

`eblupNSFH1` Provides the EBLUPs and MSE under nonstationary Fay-Herriot model for sample area

`eblupNSFH2` Provides the EBLUPs and MSE under nonstationary Fay-Herriot model for sample and non-sample area

`NS.test` Provides a p-value for testing spatial nonstationarity in the data under Fay-Herriot model.

`ebp` Provides the EBPs and MSE under stationary generalized linear mixed model.

`ebpNS` Provides the EBPs and MSE under nonstationary generalized linear mixed model.

`ebpSP` Provides the EBPs and MSE under a spatially correlated generalized linear mixed model.

`ebpNP` Provides the EBPs and MSE under nonparametric generalized linear mixed model.

`NSglm.test` Provides a p-value for testing spatial nonstationarity in the data under generalized linear mixed model.

Reference

- Chandra, H., Salvati, N., & Chambers, R. (2015). A spatially nonstationary fay-herriot model for small area estimation. Journal of survey statistics and methodology. 3. 109-135. DOI:10.1093/jssam/smu026.
- Chandra, H., Salvati, N., & Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. Spatial Statistics. 20. 30-56. DOI:10.1016/j.spasta.2017.01.004.

- Chandra, H., Salvati, N., & Chambers, R. (2018). Small area estimation under a spatially non-linear model. Computational Statistics and Data Analysis. 126. 19-38. DOI:10.1016/j.csda.2018.04.002.
- Fay, R. E. & Herriot, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association. 74. 269-277. DOI:10.2307/2286322.
- Rao, J.N.K & Molina. (2015). Small Area Estimation 2nd Edition. New York: John Wiley and Sons, Inc.

`NSglm.test` *Parametric bootstrap-based spatial nonstationarity test for generalized linear mixed model*

Description

This function performs a parametric bootstrap-based test procudure for testing spatial nonstationarity in the data.

Usage

```
NSglm.test(
  formula,
  vardir,
  Ni,
  ni,
  lat,
  lon,
  method = "REML",
  maxit = 100,
  precision = 1e-04,
  data
)
```

Arguments

`formula`	an object of class list of formula, describe the model to be fitted
`vardir`	a vector of sampling variances of direct estimators for each small area
`Ni`	a vector of population size for each small area
`ni`	a vector of sample size for each small area
`lat`	a vector of latitude for each small area
`lon`	a vector of longitude for each small area
`method`	type of fitting method, default is "REML" method
`maxit`	number of iterations allowed in the algorithm. Default is 100 iterations
`precision`	convergence tolerance limit for the Fisher-scoring algorithm. Default value is 1e-04
`data`	a data frame comprising the variables named in formula and vardir

Value

The function returns a list with class "htest" containing the following components:

method a character string indicating what type of test was performed.

p.value the p-value for the test.

data.name a character string giving the name of the data.

Examples

```
# Load data set
data(headcount)
# Testing spatial nonstationarity of the data
result <- NSglm.test(y~x1, var, N,n,lat,long, "REML", 10, 1e-04, headcount[1:10,])
result
```

---

paddy *Yield data of paddy*

Description

Dataset on paddy yield used by Chandra et al. (2016).

Usage

```
data(paddy)
```

Format

A data frame with 70 observations on the following 9 variables:

D Small area code

latitude Latitude of each small areas

longitude Longitude of each small areas

n Sample size of each small areas

y Average yield data of paddy crop for the year 2009-10 (direct estimates for the small areas)

var Estimated variance of y

x1 First covariate (average household size) used by Chandra et al. (2016)

x2 Second covariate (female population of marginal household) used by Chandra et al. (2016)

indicator Index for sample and non-sample area

Reference

Chandra, H., salvati, N., chambers, R. and Sud, U. C. (2016). A Spatially Nonstationary Fay-Herriot Model for Small Area Estimation - An Application to Crop Yield Estimation. Seventh International Conference on Agricultural Statistics. Rome. DOI:10.1481/icasVII.2016.f35.

Examples

```
data(paddy)
yield <- paddy$y
summary(yield)
```

`paddysample` *Yield data of paddy for sample area*

Description

Dataset on paddy yield for sample area used by Chandra et al. (2016).

Usage

```
data(paddysample)
```

Format

A data frame with 58 observations on the following 8 variables:

D Small area code

latitude Latitude of each small areas

longitude Longitude of each small areas

n Sample size of each small areas

y Average yield data of paddy crop for the year 2009-10 (direct estimates for the small areas)

var Estimated variance of y

x1 First covariate (average household size) used by Chandra et al. (2016)

x2 Second covariate (female population of marginal household) used by Chandra et al. (2016)

Reference

Chandra, H., salvati, N., chambers, R. and Sud, U. C. (2016). A Spatially Nonstationary Fay-Herriot Model for Small Area Estimation - An Application to Crop Yield Estimation. Seventh International Conference on Agricultural Statistics. Rome. DOI:10.1481/icasVII.2016.f35.

Examples

```
data(paddysample)
yield <- paddysample$y
summary(yield)
```

Wmatrix *Proximity matrix*

Description

Proximity matrix for the areas included in data set of Chandra et al. (2017)

Usage

```
data(Wmatrix)
```

Format

A 71*71 proximity matrix of the areas. It must be in row-standerdized form

Reference

Chandra, H., Salvati, N., & Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. Spatial Statistics. 20. 30-56. DOI:10.1016/j.spasta.2017.01.004.

Examples

```
data(Wmatrix)
```

Index