Package ‘spmoran’

July 24, 2019

Type Package

Title Moran Eigenvector-Based Spatial Regression Models

Version 0.1.7.1

Date 2019-07-24

Author Daisuke Murakami

Maintainer Daisuke Murakami <dmuraka@ism.ac.jp>

Description Functions for estimating Moran's eigenvector-based spatial regression models.
For details see Murakami (2019) <arXiv:1703.04467>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports fields, vegan, Matrix, doParallel, foreach, ggplot2, spdep, rARPACK, methods

Suggests R.rsp, rgdal

VignetteBuilder R.rsp

NeedsCompilation no

Repository CRAN

Date/Publication 2019-07-24 21:30:02 UTC

R topics documented:

`besf` *Memory-free implementation of the RE-ESF model for very large datasets*

Description

This function estimates the RE-ESF model through a parallel computation whose memory consumption is independent of the sample size.

Usage

```
besf( y, x = NULL, coords,  method = "reml", covmodel="exp",
      enum = 200, bsize = 3000, cl=NULL)
```

Arguments

`y`	Vector of explained variables (N x 1)
`x`	Matrix of explanatory variables (N x K). Default is NULL
`coords`	Matrix of spatial point coordinates (N x 2)
`method`	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"
`covmodel`	Type of kernel to model spatial dependence. The currently available options are "exp" for the exponential kernel, "gau" for the Gaussian kernel, and "sph" for the spherical kernel
`enum`	Number of eigenvectors and eigenvalues to be extracted (scalar). Default is 200
`bsize`	Block/badge size. bsize x bsize elements are iteratively processed during the parallelized computation. Default is 3000
`cl`	Number of cores used for the parallel computation. If cl=NULL, which is the default, the number of available cores is detected and used

Value

`b`	Matrix with columns for the estimated coefficients on x, their standard errors, t-values, and p-values (K x 4)
`s`	Vector of estimated variance parameters (2 x 1). The first and the second elements denote the standard error and the Moran's I value of the estimated spatially dependent component, respectively. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked

`e` Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)

`r` Vector of estimated random coefficients on Moran's eigenvectors (L x 1)

`sf` Vector of estimated spatial dependent component (N x 1)

`pred` Vector of predicted values (N x 1)

`resid` Vector of residuals (N x 1)

`other` List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Griffith, D. A. (2003). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.

Murakami, D. and Griffith, D.A. (2015) Random effects specifications in eigenvector spatial filtering: a simulation study. Journal of Geographical Systems, 17 (4), 311-331.

Murakami, D. and Griffith, D.A. (2019) A memory-free spatial additive mixed modeling for big spatial data. ArXiv.

See Also

`besf`, `meigen`, `meigen_f`

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE",
                         "DIS" ,"RAD", "TAX", "PTRATIO", "B", "LSTAT")]
xgroup  <- boston.c[,"TOWN"]
coords  <- boston.c[,c("LAT","LON")]
# res   <- besf(y = y, x = x, coords=coords)
# res$b
# res$s
# res$e
```

besf_vc *Spatially varying coefficient modeling for very large datasets*

Description

This function estimates a spatially varying coefficient model based on the random effects eigenvector spatial filtering (RE-ESF) approach. Spatially varying coefficients are selected to stablize the estimates.

Usage

```
besf_vc(y, x=NULL, xconst=NULL, coords, method="reml", penalty="bic",
        allsvc=FALSE, maxiter=30, sizelimit=2000, covmodel="exp",
        enum = 200, bsize = 3000, cl=NULL)
```

Arguments

`y`	Vector of explained variables (N x 1)
`x`	Matrix of explanatory variables whose coefficients are allowed to vary across geographical space (N x K). Default is NULL
`xconst`	Matrix of explanatory variables whose coefficients are assumed constant across space (N x K_const). Default is NULL
`coords`	Matrix of spatial point coordinates (N x 2)
`method`	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"
`penalty`	Penalty to select varying coefficients and stablize the estimates. The current options are "bic" for the Baysian information criterion-type penalty (N x log(K)) and "aic" for the Akaike information criterion (2K) (see Muller et al., 2013). Default is "bic"
`allsvc`	If it is TRUE, SVCs are selected using the penalty. If it is FALSE, all the coefficients on the explanatory variables in x are SVCs. Default is FALSE
`maxiter`	Maximum number of iterations. Default is 30
`sizelimit`	Maximum size of matrix being inverted. Roughly speaking, this value contraints K + KL, where K and L are the numbers of the explanatory variables and eigenpairs, respectively. If K + KL exceeds sizelimit, L is reduced to fullfill the contraint. Default is 2000
`covmodel`	Type of kernel to model spatial dependence. The currently available options are "exp" for the exponential kernel, "gau" for the Gaussian kernel, and "sph" for the spherical kernel
`enum`	Number of eigenvectors and eigenvalues to be extracted (scalar). Default is 200
`bsize`	Block/badge size. bsize x bsize elements are iteratively processed during the parallelized computation. Default is 3000
`cl`	Number of cores used for the parallel computation. If cl=NULL, which is the default, the number of available cores is detected and used

Value

`b`	Matrix with columns for the estimated coefficients on xconst, their standard errors, t-values, and p-values (K_const x 4)
`s`	Matrix of estimated variance parameters (2 x K). The (1, k)-th element denotes the standard error of the k-th SVCs, while the (2, k)-th element denotes the Moran's I value of the SVCs. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked
`e`	Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)
`b_vc`	Matrix of estimated spatially varying coefficients (SVCs) on x (N x K)
`bse_vc`	Matrix of estimated standard errors for the SVCs (N x k)
`t_vc`	Matrix of estimated t-values for the SVCs (N x K)
`p_vc`	Matrix of estimated p-values for the SVCs (N x K)
`pred`	Vector of predicted values (N x 1)
`resid`	Vector of residuals (N x 1)
`vc`	Vector indicating whether spatial variations are found or not from each coefficients (K x 1)
`other`	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Muller, S., Scealy, J.L., and Welsh, A.H. (2013) Model selection in linear mixed models. Statistical Science, 28 (2), 136-167.

Murakami, D., Yoshida, T., Seya, H., Griffith, D.A., and Yamagata, Y. (2017) A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. Spatial Statistics, 19, 68-89.

Murakami, D., and Griffith, D.A. (2019). Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions. Spatial Statistics, 30, 39-64.

Murakami, D. and Griffith, D.A. (2019) A memory-free spatial additive mixed modeling for big spatial data. ArXiv.

See Also

`resf_vc`

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV"]
x       <- boston.c[,c("ZN", "INDUS" ,"LSTAT")]
xconst  <- boston.c[,c("CRIM", "NOX", "CHAS", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "R
coords  <- boston.c[,c("LAT","LON")]

######## Without the penalty-based SVC selection (just like the usual GWR)
# res   <- besf_vc(y=y,x=x,xconst=xconst,coords=coords,allsvc=TRUE)

######## With the penalty-based SVC selection
# res2 <- besf_vc(y=y,x=x,xconst=xconst,coords=coords)

# res$b
# res$s
# res$e
# res$vc
# res$b_vc[1:10,]
# res$bse_vc[1:10,]
# res$t_vc[1:10,]
# res$p_vc[1:10,]

###Plot the first SVC (INDUS)
#
# require(ggplot2)
# ggplot(mapping = aes(x = coords$LON, y = coords$LAT)) +
#   geom_point(aes(colour = res$b_vc[,3])) +
#   scale_color_gradientn(colours=rev(rainbow(4)))
```

`esf` *Spatial regression with eigenvector spatial filtering*

Description

This function estimates the linear eigenvector spatial filtering (ESF) model. The eigenvectors are selected by a forward stepwise method.

Usage

```
esf( y, x = NULL, vif = NULL, meig, fn = "r2" )
```

Arguments

y	Vector of explained variables (N x 1)
x	Matrix of explanatory variables (N x K). Default is NULL

`vif` Maximum acceptable value for the variance inflation factor (VIF) (scalar). For example, if vif = 10, eigenvectors are selected so that the maximum VIF value among explanatory variables and eigenvectors is equal to or less than 10. Default is NULL

`meig` Moran's eigenvectors and eigenvalues. Output from `meigen` or `meigen_f`

`fn` Objective function for the stepwise eigenvector selection. The adjusted R2 ("r2"), AIC ("aic"), or BIC ("bic") are available. Alternatively, all the eigenvectors in meig are use if fn = "all". This is acceptable for large samples (see Murakami and Griffith, 2018). Default is "r2"

Value

`b` Matrix with columns for the estimated coefficients on x, their standard errors, t-values, and p-values (K x 4)

`s` Vector of the statistics for the estimated spatial component (2 x 1). The first and the second elements denote the standard deviation and the Moran's I value of the estimated spatially dependent component, respectively. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked

`r` Matrix with columns for the estimated coefficients on Moran's eigenvectors, their standard errors, t-values, and p-values (L x 4)

`vif` Vector of variance inflation factors of the explanatory variables (N x 1)

`e` Vector whose elements are residual standard error (resid_SE), adjusted R2 (adjR2), log-likelihood (logLik), AIC, and BIC

`sf` Vector of estimated spatial dependent component ($E\gamma$) (N x 1)

`pred` Vector of predicted values (N x 1)

`resid` Vector of residuals (N x 1)

`other` List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Griffith, D. A. (2003). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.

Tiefelsdorf, M., and Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. Environment and Planning A, 39 (5), 1193-1221.

Murakami, D. and Griffith, D.A. (2018) Eigenvector spatial filtering for large data sets: fixed and random effects approaches. Geographical Analysis, doi: 10.1111/gean.12156.

See Also

`resf`

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE")]
coords  <- boston.c[,c("LAT","LON")]

#########Distance-based ESF
meig  <- meigen(coords=coords)
esfD <- esf(y=y,x=x,meig=meig, vif=5)
esfD$vif
esfD$b
esfD$s
esfD$e

#########Fast approximation
meig_f<- meigen_f(coords=coords)
esfD <- esf(y=y,x=x,meig=meig_f, vif=10, fn="all")

############################Not run
#########Topoligy-based ESF (it is commonly used in regional science)
#
#cknn <- knearneigh(coordinates(coords), k=4) #4-nearest neighbors
#cmat <- nb2mat(knn2nb(cknn), style="B")
#meig <- meigen(cmat=cmat, threshold=0.25)
#esfT <- esf(y=y,x=x,meig=meig)
```

`lsem` *Low rank spatial error model (LSEM) estimation*

Description

This function estimates the low rank spatial error model.

Usage

```
lsem( y, x, weig, method = "reml" )
```

Arguments

`y`	Vector of explained variables (N x 1)
`x`	Matrix of explanatory variables (N x K)
`weig`	eigenvectors and eigenvalues of a spatial weight matrix. Output from `weigen`
`method`	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"

Value

- `b` — Matrix with columns for the estimated coefficients on x, their standard errors, t-values, and p-values (K x 4)
- `s` — Vector of estimated shrinkage parameters (2 x 1). The first and the second elements denote the estimated rho parameter (sp_lambda), quantfying the scale of spatial dependence, and the standard error of the spatial dependent component (sp_SE), respectively.
- `e` — Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)
- `r` — Vector of estimated random coefficients on Moran's eigenvectors (L x 1)
- `pred` — Vector of predicted values (N x 1)
- `resid` — Vector of residuals (N x 1)

Author(s)

Daisuke Murakami

References

Murakami, D., Seya, H. and Griffith, D.A. (2018) Low rank spatial econometric models. Arxiv.

See Also

`meigen`, `meigen_f`

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE",
                        "DIS" ,"RAD", "TAX", "PTRATIO", "B", "LSTAT")]
coords  <- boston.c[,c("LAT","LON")]
weig  <- weigen( coords )
res <- lsem(y=y,x=x,weig=weig)
res$b
res$s
res$e
```

`lslm` *Low rank spatial lag model (LSLM) estimation*

Description

This function estimates the low rank spatial lag model.

Usage

```
lslm( y, x, weig, method = "reml", boot = FALSE, iter = 200 )
```

Arguments

`y`	Vector of explained variables (N x 1)
`x`	Matrix of explanatory variables (N x K)
`weig`	eigenvectors and eigenvalues of a spatial weight matrix. Output from `weigen`
`method`	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"
`boot`	If it is TRUE, confidence intervals for the spatial dependence parameters (s), the mean direct effects (de), and the mean indirect effects (ie), are estimated by a parametric bootstrapping. Default is FALSE
`iter`	The number of bootstrap replications. Default is 200

Value

`b`	Matrix with columns for the estimated coefficients on x, their standard errors, t-values, and p-values (K x 4)
`s`	Vector of estimated shrinkage parameters (2 x 1). The first and the second elements denote the estimated rho parameter (sp_rho), quantfying the scale of spatial dependence, and the standard error of the spatial dependent component (sp_SE), respectively. If boot = TRUE, their 95 percent confidence intervals and the resulting p-values are also provided
`e`	Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)
`de`	Matrix with columns for the estimated mean direct effects on x. If boot = TRUE, their 95 percent confidence intervals and the resulting p-values are also provided
`ie`	Matrix with columns for the estimated mean indirect effects on x. If boot = TRUE, their 95 percent confidence intervals and the resulting p-values are also provided
`r`	Vector of estimated random coefficients on the spatial eigenvectors (L x 1)
`pred`	Vector of predicted values (N x 1)
`resid`	Vector of residuals (N x 1)

Author(s)

Daisuke Murakami

References

Murakami, D., Seya, H. and Griffith, D.A. (2018) Low rank spatial econometric models. Arxiv.

See Also

`weigen`, `lsem`

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE",
                       "DIS" ,"RAD", "TAX", "PTRATIO", "B", "LSTAT")]
coords  <- boston.c[,c("LAT","LON")]
weig    <- weigen(coords)
res     <- lslm(y=y,x=x,weig=weig)
## res  <- lslm(y=y,x=x,weig=weig, boot=TRUE)
res$b
res$s
res$de
res$ie
res$e
```

`meigen` *Extraction of Moran's eigenvectors*

Description

This function calculates Moran's eigenvectors and their corresponding eigenvalues.

Usage

```
meigen( coords, model = "exp", threshold = 0, enum = NULL, cmat = NULL )
```

Arguments

`coords`	Matrix of spatial point coordinates (N x 2)
`model`	Type of kernel to model spatial dependence. The currently available options are "exp" for the exponential kernel, "gau" for the Gaussian kernel, and "sph" for the spherical kernel

`threshold`	Threshold for the eigenvalues (scalar). Suppose that λ_1 is the maximum eigenvalue, eigenvectors whose corresponding eigenvalues that are equal or greater than [threshold x λ_1] are extracted. threshold must be a value between 0 and 1. Default is zero (see Details)
`cmat`	Optional. A user-specified spatial connectivity matrix (N x N). It must be provided when the user wants to use a spatial connectivity matrix other than the default matrices
`enum`	Optional. The muxmum acceptable mumber of eigenvectors to be extracted (scalar)

Details

If cmat is not provided and model = "exp" (default), this function extracts Moran's eigenvectors from MCM, where M = I - 11'/N is a centering operator. C is a N x N connectivity matrix whose (i, j)-th element equals exp(-d(i,j)/h), where d(i,j) is the Euclidean distance between the sample sites i and j, and h is given by the maximum length of the minimum spanning tree connecting sample sites (see Dray et al., 2006). If cmat is provided, this function performs the same calculation after C is replaced with cmat.

If threshold is not provided (default), all eigenvectors corresponding to positive eigenvalues are extracted. It implies to consider all the elements describing positive spatial dependence. If threshold is provided, eigenvectors whose corresponding eigenvalues are equal to or greater than [threshold x λ_1] are extracted. threshold = 0.00 or 0.25 are standard assumptions (see Griffith, 2003; Murakami and Griffith, 2015).

Value

`sf`	Matrix of the first L eigenvectors (N x L)
`ev`	Vector of the first L eigenvalues (L x 1)
`ev_full`	Vector of all eigenvalues (N x 1)
`other`	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Dray, S., Legendre, P., and Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196 (3), 483-493.

Griffith, D.A. (2003) Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.

Murakami, D. and Griffith, D.A. (2015) Random effects specifications in eigenvector spatial filtering: a simulation study. Journal of Geographical Systems, 17 (4), 311-331.

See Also

`meigen_f` for fast eigen-decomposition

`meigen0` *Nystrom extension of Moran's eigenvectors*

Description

This function estimates Moran's eigenvectors at predicted sites using the Nystrom extension.

Usage

```
meigen0( meig, coords0 )
```

Arguments

`coords0`	Matrix of spatial point coordinates of predicted sites (N_0 x 2)
`meig`	Moran's eigenvectors and eigenvalues. Output from `meigen` or `meigen_f`

Value

`sf`	Matrix of the first L eigenvectors at predicted sites (N_0 x L)
`ev`	Vector of the first L eigenvalues (L x 1)
`ev_full`	Vector of all eigenvalues (N x 1)

Author(s)

Daisuke Murakami

References

Drineas, P. and Mahoney, M.W. (2005) On the Nystrom method for approximating a gram matrix for improved kernel-based learning. Journal of Machine Learning Research, 6 (2005), 2153-2175.

See Also

`meigen`, `meigen_f`

meigen_f *Fast approximation of Moran's eigenvectors*

Description

This function performs a fast approximation of Moran's eigenvectors and their corresponding eigenvalues.

Usage

```
meigen_f( coords, model = "exp", enum = 200 )
```

Arguments

coords	Matrix of spatial point coordinates (N x 2)
model	Type of kernel to model spatial dependence. The currently available options are "exp" for the exponential kernel, "gau" for the Gaussian kernel, and "sph" for the spherical kernel
enum	Number of eigenvectors and eigenvalues to be extracted (scalar). Default is 200

Details

This function extracts approximated Moran's eigenvectors from MCM. M = I - 11'/N is a centering operator, and C is a spatial connectivity matrix whose (i, j)-th element is given by exp(-d(i,j)/h), where d(i,j) is the Euclidean distance between the sample sites i and j, and h is a range parameter given by the maximum length of the minimum spanning tree connecting sample sites (see Dray et al., 2006).

Following a simulation result that 200 eigenvectors are sufficient for accurate approximation of ESF models (Murakami and Griffith, 2018), this function approximates the first 200 eigenvectors by default (i.e., enum = 200). If enum is given by a smaller value like 100, the computation time will be shorter, but with greater approximation error. Following `meigen`, eigenvectors corresponding to negative eigenvalues are omitted among the 200 eigenvectors.

Value

sf	Matrix of the first L approximated eigenvectors (N x L)
ev	Vector of the first L approximated eigenvalues (L x 1)
ev_full	Vector of all approximated eigenvalues (enum x 1)
other	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Dray, S., Legendre, P., and Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196 (3), 483-493.

Murakami, D. and Griffith, D.A. (2018) Eigenvector spatial filtering for large data sets: fixed and random effects approaches. Geographical Analysis, doi: 10.1111/gean.12156.

See Also

`meigen`

---

plot_qr — *Plot quantile regression coefficients estimated from SF-UQR*

Description

This function plots regression coefficients estimated from the spatial filter unconditional quantile regression (SF-UQR) approach.

Usage

```
plot_qr( mod, pnum = 1, par = "b", cex.main = 28, cex.lab = 26, cex.axis = 24, lwd
```

Arguments

`mod`	Object produced by the `resf_qr` function
`pnum`	A number specifying which parameter is plotted. If par = "b", coefficients for the pnum-th explanatory variable are plotted (intercepts are plotted if pnum = 1). On the other hand, the estimated standard errors for the reidual spatial process are plotted if par = "s" and pnum = 1, whereas the scale/range parameter for the process is plotted if par = "s" and pnum = 2
`par`	If it is "b", regression coefficeints are plotted. If it is "s", shrinkage (variance) parameters for the residual spatial process are plotted. Default is "b"
`cex.main`	Graphical parameter specifying the size of the main title
`cex.lab`	Graphical parameter specifying the size of the x and y axis labels
`cex.axis`	Graphical parameter specifying the size of the tick label numbers
`lwd`	Graphical parameters specifying the width of the line drawing the coefficient estimates

Note

See `par` for the graphical parameters

See Also

`resf_qr`

`predict0` *Spatial prediction using eigenvector spatial filtering (ESF) or random effects ESF*

Description

This function predicts explained variables using eigenvector spatial filtering (ESF) or random effects ESF. The Nystrom extension is used to minimize the expected prediction error

Usage

```
predict0( mod, meig0, x0 = NULL, xgroup0 = NULL )
```

Arguments

`mod`	ESF or RE-ESF model estimates. Output from `esf` or `resf`
`meig0`	Moran's eigenvectors at predicted sites. Output from `meigen0`
`x0`	Matrix of explanatory variables at predicted sites (N_0 x K). Default is NULL
`xgroup0`	Matrix of group indeces that may be group IDs (integers) or group names (N_0 x K_group). Default is NULL

Value

`pred`	Matrix with the first column for the predicted values. The second and the third columns are the trend component and the spatial component in the predicted values. The remining columns are group effects (if any)

References

Drineas, P. and Mahoney, M.W. (2005) On the Nystrom method for approximating a gram matrix for improved kernel-based learning. Journal of Machine Learning Research, 6 (2005), 2153-2175.

See Also

`meigen0`, `predict0_vc`

Examples

```
require(spdep)
data(boston)
samp    <- sample( dim( boston.c )[ 1 ], 400)

d       <- boston.c[ samp, ]  ## Data at observed sites
y       <- d[, "CMEDV"]
x       <- d[,c("ZN","INDUS", "NOX","RM", "AGE", "DIS")]
coords  <- d[,c("LAT","LON")]

d0      <- boston.c[-samp, ]  ## Data at unobserved sites
```

```
y0     <- d0[, "CMEDV"]
x0      <- d0[,c("ZN","INDUS", "NOX","RM", "AGE", "DIS")]
coords0 <- d0[,c("LAT","LON")]

############ Model estimation
meig    <- meigen( coords = coords )
mod     <- resf(y=y, x=x, meig=meig)
## or
# mod   <- esf(y=y,x=x,meig=meig)

############ Spatial prediction

meig0  <- meigen0( meig = meig, coords0 = coords0 )
pred0   <- predict0( mod = mod, x0 = x0, meig0 = meig0 )
pred0[1:10,]
```

`predict0_vc` *Prediction of explained variables and spatially varying coefficients*

Description

This function predicts explained variables and spatially varying coefficients using a Moran's eigenvector-based spatially varying coefficient model. The Nystrom extension is used to minimize the expected prediction error

Usage

```
predict0_vc( mod, meig0, x0 = NULL, xgroup0 = NULL, xconst0 = NULL )
```

Arguments

`mod`	spatially varying coefficient model estimates. Output from `resf_vc`
`meig0`	Moran's eigenvectors at predicted sites. Output from `meigen0`
`x0`	Matrix of explanatory variables at predicted sites whose coefficients are allowed to vary across geographical space (N_0 x K). Default is NULL
`xgroup0`	Matrix of group indeces that may be group IDs (integers) or group names (N_0 x K_group). Default is NULL
`xconst0`	Matrix of explanatory variables at predicted sites whose coefficients are assumed constant across space (N_0 x K_const). Default is NULL

Value

`pred`	Matrix with the first column for the predicted values. The second and the third columns are the trend component (i.e., component explained by x0 and xconst0) and the spatial component in the predicted values (N x 3). The output has anoteher column including the estimated group effects if group effects are considered
`b_vc`	Matrix of estimated spatially varying coefficients (SVCs) on x0 (N_0 x K)

bse_vc	Matrix of estimated standard errors for the SVCs (N_0 x K)
t_vc	Matrix of estimated t-values for the SVCs (N_0 x K)
p_vc	Matrix of estimated p-values for the SVCs (N_0 x K)

References

Drineas, P. and Mahoney, M.W. (2005) On the Nystrom method for approximating a gram matrix for improved kernel-based learning. Journal of Machine Learning Research, 6 (2005), 2153-2175.

Murakami, D., Yoshida, T., Seya, H., Griffith, D.A., and Yamagata, Y. (2017) A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. Spatial Statistics, 19, 68-89.

See Also

meigen0, predict0

Examples

```
require(spdep)
data(boston)
samp    <- sample( dim( boston.c )[ 1 ], 300)

d       <- boston.c[ samp, ]    ## Data at observed sites
y       <- d[, "CMEDV"]
x       <- d[,c("ZN", "LSTAT")]
xconst  <- d[,c("CRIM", "NOX", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "RM")]
coords  <- d[,c("LAT","LON")]

d0      <- boston.c[-samp, ]    ## Data at unobserved sites
x0      <- d0[,c("ZN", "LSTAT")]
xconst0 <- d0[,c("CRIM", "NOX", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "RM")]
coords0 <- d0[,c("LAT","LON")]

############ Model estimation
meig    <- meigen( coords = coords )
mod     <- resf_vc(y=y, x=x, xconst=xconst, meig=meig )

############ Spatial prediction of y and spatially varying coefficients
meig0  <- meigen0( meig = meig, coords0 = coords0 )
pred0   <- predict0_vc( mod = mod, x0 = x0, xconst0=xconst0, meig0 = meig0 )

pred0$pred[1:10,]
pred0$b_vc[1:10,]
pred0$bse_vc[1:10,]
pred0$t_vc[1:10,]
pred0$p_vc[1:10,]

############ or spatial prediction of spatially varying coefficients only
# pred00  <- predict0_vc( mod = mod, meig0 = meig0 )
# pred00$b_vc[1:10,]
# pred00$bse_vc[1:10,]
```

```
# pred00$t_vc[1:10,]
# pred00$p_vc[1:10,]
```

`resf` *Spatial regression with random effects eigenvector spatial filtering*

Description

This function estimates the random effects eigenvector spatial filtering (RE-ESF) model.

Usage

```
resf( y, x = NULL, xgroup = NULL, meig, method = "reml" )
```

Arguments

`y`	Vector of explained variables (N x 1)
`x`	Matrix of explanatory variables (N x K). Default is NULL
`xgroup`	Matrix of group indexes. The indeces may be group IDs (numbers) or group names (N x K_group). Default is NULL
`meig`	Moran's eigenvectors and eigenvalues. Output from `meigen` or `meigen_f`
`method`	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"

Value

`b`	Matrix with columns for the estimated coefficients on x, their standard errors, t-values, and p-values (K x 4)
`b_g`	List of K_group matrices with columns for the estimated group effects, their standard errors, and t-values
`s`	Vector of estimated variance parameters (2 x 1). The first and the second elements denote the standard error and the Moran's I value of the estimated spatially dependent component, respectively. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked
`s_g`	Vector of estimated standard errors of the group effects
`e`	Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)
`r`	Vector of estimated random coefficients on Moran's eigenvectors (L x 1)
`sf`	Vector of estimated spatial dependent component (N x 1)
`pred`	Vector of predicted values (N x 1)
`resid`	Vector of residuals (N x 1)
`other`	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Murakami, D. and Griffith, D.A. (2015) Random effects specifications in eigenvector spatial filtering: a simulation study. Journal of Geographical Systems, 17 (4), 311-331.

Griffith, D. A. (2003). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.

See Also

besf, meigen, meigen_f

Examples

```
require(spdep);require(Matrix)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE",
                              "DIS" ,"RAD", "TAX", "PTRATIO", "B", "LSTAT")]
xgroup<- boston.c[,"TOWN"]
coords<- boston.c[,c("LAT","LON")]
meig  <- meigen(coords=coords)
res   <- resf(y = y, x = x, xgroup = xgroup, meig = meig)
res$b
res$b_g
res$s
res$s_g
res$e

#########Fast approximation
meig_f <- meigen_f(coords=coords)
res2   <- resf(y=y,x=x, xgroup = xgroup,meig=meig_f)
```

---

resf_qr *Spatial filter unconditional quantile regression*

Description

This function estimates the spatial filter unconditional quantile regression (SF-UQR) model.

Usage

```
resf_qr( y, x = NULL, meig, tau = NULL, boot = TRUE, iter = 200, cl=NULL )
```

Arguments

- `y` Vector of explained variables (N x 1)
- `x` Matrix of explanatory variables (N x K). Default is NULL
- `meig` Moran's eigenvectors and eigenvalues. Output from `meigen` or `meigen_f`
- `tau` The quantile(s) to be modeled. It must be a number (or a vector of numbers) strictly between 0 and 1. By default, tau = c(0.1, 0.2, ..., 0.9)
- `boot` If it is TRUE, confidence intervals for regression coefficients are estimated by a semiparametric bootstrapping. Default is TRUE
- `iter` The number of bootstrap replications. Default is 200
- `cl` Number of cores used for the parallel computation. If cl=NULL, which is the default, the number of available cores is detected and used

Value

- `b` Matrix of estimated regression coefficients (K x Q), where Q is the number of quantiles (i.e., the length of tau)
- `r` Matrix of estimated random coefficients on Moran's eigenvectors (L x Q)
- `s` Vector of estimated variance parameters (2 x 1). The first and the second elements denote the standard error and the Moran's I value of the estimated spatially dependent component, respectively. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked
- `e` Vector whose elements are residual standard error (resid_SE) and adjusted quasi conditional R2 (quasi_adjR2(cond))
- `B` Q matrices (K x 4) summarizing bootstrapped estimates for the regression coefficients Columns of these matrices consist of the estimated coefficients, the lower and upper bounds for the 95 percent confidencial intervals, and p-values. It is returned if boot = TRUE
- `S` Q matrices (2 x 3) summarizing bootstrapped estimates for the variance parameters. Columns of these matrices consist of the estimated parameters, the lower and upper bounds for the 95 percent confidencial intervals. It is returned if boot = TRUE
- `B0` List of Q matrices (K x iter) summarizing bootstrapped coefficients. The q-th matrix consists of the coefficients on the q-th quantile. It is returned if boot = TRUE
- `S0` List of Q matrices (2 x iter) summarizing bootstrapped variance parameters. The q-th matrix consists of the parameters on the q-th quantile. It is returned if boot = TRUE

Author(s)

Daisuke Murakami

References

Murakami, D. and Seya, H. (2017) Spatially filtered unconditional quantile regression. ArXiv.

See Also

plot_qr

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV" ]
x <- boston.c[,c("CRIM","ZN","INDUS", "CHAS", "NOX","RM", "AGE",
                 "DIS" ,"RAD", "TAX", "PTRATIO", "B", "LSTAT")]
coords  <- boston.c[,c("LAT","LON")]
meig    <- meigen(coords=coords)
res     <- resf_qr(y=y,x=x,meig=meig, boot=FALSE)
res$b
res$s
res$e
plot_qr(res,1)     # The first explanatory vafriable (intercept)
plot_qr(res,2)     # The second explanatory variable
plot_qr(res,1,"s") # spcomp_SE
plot_qr(res,2,"s") # spcomp_Moran.I/max(Moran.I)

###Not run
#res <- resf_qr(y=y,x=x,meig=meig, boot=TRUE)
#res$B
#res$S
#res$e
#plot_qr(res,1)     # The first explanatory vafriable (intercept)
#plot_qr(res,2)     # The second explanatory variable
#plot_qr(res,1,"s") # spcomp_SE
#plot_qr(res,2,"s") # spcomp_Moran.I/max(Moran.I)
```

---

resf_vc — *Spatially varying coefficient modeling with/without automatic coefficient selection*

Description

This function estimates a spatially varying coefficient model based on the random effects eigenvector spatial filtering (RE-ESF) approach. Spatially varying coefficients are selected to stablize the estimates.

Usage

```
resf_vc(y, x=NULL, xgroup = NULL, xconst=NULL, meig, method="reml",
        penalty="bic", allsvc=FALSE, maxiter=30, sizelimit=2000 )
```

Arguments

y	Vector of explained variables (N x 1)
x	Matrix of explanatory variables whose coefficients are allowed to vary across geographical space (N x K). Default is NULL
xgroup	Matrix of group indexes. The indeces may be group IDs (numbers) or group names (N x K_group). Default is NULL
xconst	Matrix of explanatory variables whose coefficients are assumed constant across space (N x K_const). Default is NULL
meig	Moran's eigenvectors and eigenvalues. Output from `meigen` or `meigen_f`
method	Estimation method. Restricted maximum likelihood method ("reml") and maximum likelihood method ("ml") are available. Default is "reml"
penalty	Penalty to select varying coefficients and stablize the estimates. The current options are "bic" for the Baysian information criterion-type penalty (N x log(K)) and "aic" for the Akaike information criterion (2K) (see Muller et al., 2013). Default is "bic"
allsvc	If it is TRUE, SVCs are selected using the penalty. If it is FALSE, all the coefficients on the explanatory variables in x are SVCs. Default is FALSE
maxiter	Maximum number of iterations. Default is 30
sizelimit	Maximum size of matrix being inverted. Roughly speaking, this value contraints K + KL, where K and L are the numbers of the explanatory variables and eigenpairs, respectively. If K + KL exceeds sizelimit, L is reduced to fullfill the contraint. Default is 2000

Value

b	Matrix with columns for the estimated coefficients on xconst, their standard errors, t-values, and p-values (K_const x 4)
b_g	List of K_group matrices with columns for the estimated group effects, their standard errors, and t-values
s	Matrix of estimated variance parameters (2 x K). The (1, k)-th element denotes the standard error of the k-th SVCs, while the (2, k)-th element denotes the Moran's I value of the SVCs. The Moran's I value is scaled to take a value between 0 (no spatial dependence) and 1 (the maximum possible spatial dependence). Based on Griffith (2003), the scaled Moran'I value is interpretable as follows: 0.25-0.50:weak; 0.50-0.70:moderate; 0.70-0.90:strong; 0.90-1.00:marked
s_g	Vector of estimated standard errors of the group effects
e	Vector whose elements are residual standard error (resid_SE), adjusted conditional R2 (adjR2(cond)), restricted log-likelihood (rlogLik), Akaike information criterion (AIC), and Bayesian information criterion (BIC). When method = "ml", restricted log-likelihood (rlogLik) is replaced with log-likelihood (logLik)
b_vc	Matrix of estimated spatially varying coefficients (SVCs) on x (N x K)
bse_vc	Matrix of estimated standard errors for the SVCs (N x k)

t_vc	Matrix of estimated t-values for the SVCs (N x K)
p_vc	Matrix of estimated p-values for the SVCs (N x K)
pred	Vector of predicted values (N x 1)
resid	Vector of residuals (N x 1)
vc	Vector indicating whether spatial variations are found or not from each coefficients (K x 1)
other	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Muller, S., Scealy, J.L., and Welsh, A.H. (2013) Model selection in linear mixed models. Statistical Science, 28 (2), 136-167.

Murakami, D., Yoshida, T., Seya, H., Griffith, D.A., and Yamagata, Y. (2017) A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. Spatial Statistics, 19, 68-89.

Murakami, D., and Griffith, D.A. (2019). Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions. Spatial Statistics, 30, 39-64.

Griffith, D. A. (2003). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.

See Also

resf,besf_vc

Examples

```
require(spdep)
data(boston)
y <- boston.c[, "CMEDV"]
x       <- boston.c[,c("ZN", "LSTAT")]
xconst  <- boston.c[,c("CRIM", "NOX", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "RM")]
xgroup  <- boston.c[,"TOWN"]
coords  <- boston.c[,c("LAT","LON")]
meig    <- meigen(coords=coords)
# meig_f  <- meigen_f(coords=coords)  ## for fast computation


######## Without the penalty-based SVC selection (just like the usual GWR)
res     <- resf_vc(y=y,x=x,xconst=xconst,meig=meig,allsvc=TRUE)

######## With the penalty-based SVC selection
# res2 <- resf_vc(y=y,x=x,xconst=xconst,meig=meig)

######## SVC modeling with group effects
```

```
# res3 <- resf_vc(y=y,x=x,xconst=xconst,xgroup=xgroup,meig=meig)

res$b
res$s
res$e

res$vc
res$b_vc[1:10,]
res$bse_vc[1:10,]
res$t_vc[1:10,]
res$p_vc[1:10,]

###Plot the first SVC (INDUS)
#
#require(ggplot2)
#ggplot(mapping = aes(x = coords$LON, y = coords$LAT)) +
#  geom_point(aes(colour = res$b_vc[,2])) +
#  scale_color_gradientn(colours=rev(rainbow(4)))
```

---

weigen *Extraction of eigenvectors from a spatial weight matrix*

Description

This function extracts eigenvectors and their corresponding eigenvalues from a spatial weight matrix.

Usage

```
weigen( x = NULL, type = "knn", k = 4, threshold = 0.25, enum = NULL )
```

Arguments

x	Matrix of spatial point coordinates (N x 2), ShapePolygons object (N spatial units), or an user-specified spatial weight matrix (N x N) (see Details)
type	Type of spatial weights. The currently available options are "knn" for the k-nearest neighbor-based weights, and "tri" for the Delaunay triangulation-based weights. If ShapePolygons are provided for x, type is ignored, and the rook-based neighborhood matrix is created
k	Number of nearest neighbors. It is used if type ="knn"
threshold	Threshold for the eigenvalues (scalar). Suppose that λ_1 is the maximum eigenvalue, eigenvectors whose eigenvalues that are equal or greater than [threshold x λ_1] are extracted. It must be a value between 0 and 1. Default is 0.25 (see Details)
enum	Optional. The muxmum acceptable mumber of eigenvectors to be extracted (scalar)

Details

If a user-specified spatial weight matrix is provided for x, this function returns the eigen-pairs of the matrix. Otherwise, if a SpatialPolygons object is provided to x, the rook-type neighborhood matrix is created from this polygon, and the matrix is eigen-decomposed. Otherwise, if point coordinats are provided to x, a spatial weight matrix is created according type, abd it is eigen-decomposed.

By default, the ARPACK routine is implemented for fast eigen-decomposition.

threshold = 0.25 (default) is a standard setting for a topology-based ESF (see Tiefelsdorf and Griffith, 2007) while threshold = 0.00 is a usual setting for a distance-based ESF, which the meigen function assumes.

Value

`sf`	Matrix of the first L eigenvectors (N x L)
`ev`	Vector of the first L eigenvalues (L x 1)
`other`	List of other outcomes, which are internally used

Author(s)

Daisuke Murakami

References

Tiefelsdorf, M. and Griffith, D.A. (2007) Semiparametric filtering of spatial autocorrelation: the eigenvector approach. Environment and Planning A, 39 (5), 1193-1221.

Murakami, D. and Griffith, D.A. (2018) Low rank spatial econometric models. Arxiv.

See Also

`meigen`, `meigen_f`

Examples

```
require(spdep);library(rgdal)
data(boston)

########## Rook adjacency-based W
poly     <- readOGR(system.file("shapes/boston_tracts.shp",package="spData")[1])
weig1     <- weigen( poly )

########## knn-based W
coords    <- boston.c[,c("LAT", "LON")]
weig2     <- weigen( coords, type = "knn" )

########## Delaunay triangulation-based W
coords    <- boston.c[,c("LAT", "LON")]
weig3     <- weigen( coords, type = "tri")

########## User-specified W
dmat      <- as.matrix(dist(coords))
```

```
cmat       <- exp(-dmat)
diag(cmat)<- 0
weig4      <- weigen( cmat, threshold = 0 )
```