Package: HSPOR (via r-universe)

October 25, 2024

Title Hidden Smooth Polynomial Regression for Rupture Detection

Version 1.1.9

Description Several functions that allow by different methods to infer a piecewise polynomial regression model under regularity constraints, namely continuity or differentiability of the link function. The implemented functions are either specific to data with two regimes, or generic for any number of regimes, which can be given by the user or learned by the algorithm. A paper describing all these methods will be submitted soon. The reference will be added to this file as soon as available.

License LGPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Imports stats, corpcor, npregfast, graphics

NeedsCompilation no

Author Florine Greciet [aut, cre], Romain Azais [aut]

Maintainer Florine Greciet <florine.greciet@gmail.com>

Date/Publication 2019-09-03 07:30:11 UTC

Repository https://florine-greciet.r-universe.dev

RemoteUrl https://github.com/cran/HSPOR

RemoteRef HEAD

RemoteSha e1321dcf76146d67bc055a95505fba8f6952edd8

Contents

H2SPOR *Inference method for two regimes*

Description

H2SPOR is an inference method that estimates, under regularity constraint, the parameters of a polynomial regression model with 2 regimes.

Usage

```
H2SPOR(X, Y, deg, constraint = 1, EM = TRUE, TimeTrans_Prop = c(),
  plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.
Y	A numerical vector corresponding to the variable to be explain. It should contain two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains dupplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	The degree of polynomials. The size of X and Y must be greater than 2(deg+2) + 1.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.
EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
TimeTrans_Prop	A numerical vector. This vector is empty by default. If you want to estimate the model parameters for a fixed jump time value, you can propose this value here.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the H2SPOR function are plotted.

Value

A dataframe that contains the estimated parameters of the polynomial regression model at two regimes: the jump time, the coefficients of the polynomials and the variances of the two regimes. If plotG = TRUE, the data (X,Y) and the estimated model will be plotted.

Examples

```
#generated data with two regimes
set.seed(1)

xgrid1 = seq(0,10,length.out=6)
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR(xgrid,ygrid,2,1,EM=FALSE,c())


set.seed(1)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR(xgrid,ygrid,2,1,EM=FALSE,c())
#Executed time : 9.69897 secs (intel core i7 processor)
```

H2SPOR_DynProg *Inference method that does not require a priori knowledge of the number of regimes and uses dynamic programming*

Description

H2SPOR_DynProg is an inference method implemented as a binary segmentation algorithm. This method makes it possible to estimate, using dynamic programming and under regularity assumption, the parameters of a piecewise polynomial regression model when we have no a priori knowledge of the number of regimes.

Usage

```
H2SPOR_DynProg(X, Y, deg, constraint = 1, EM = TRUE, plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.
Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains dupplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	Degree of the polynomials. The size of X and Y must be greater than 2(deg+2) + 1.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.
EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the H2SPOR_DynProg function are plotted.

Value

A dataframe which contains the estimated parameters of the polynomial regression model at an estimated number of regimes: the times of jump, the polynomials coefficients and the variances of an estimated number of regimes. If plotG = TRUE, the data(X,Y) and the estimated model will be plotted.

Examples

```
set.seed(1)
#generated data with two regimes
xgrid1 = seq(0,10,length.out = 6)
```

```
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)
# Inference of a piecewise polynomial regression model on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR_DynProg(xgrid,ygrid,2,1,EM=FALSE)


set.seed(1)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
datX = c(xgrid1,xgrid2,xgrid3)
datY = c(ygrid1,ygrid2,ygrid3)
#Inference of a piecewise polynomial regression model on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR_DynProg(datX,datY,2,1)
#Executed time : 2.349685 mins (intel core i7 processor)
```

HKSPOR *Inference method for any number K of regimes*

Description

HKSPOR is an inference method that estimates, under regularity constraint, the parameters of a polynomial regression model for a number K of regimes given by the user.

Usage

```
HKSPOR(X, Y, deg, K, constraint = 1, EM = TRUE, TimeTrans_Prop = c(),
  plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.

Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains dupplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	Degree of the polynomials. The size of X and Y must be greater than K(deg+2) + K.
K	The number of regimes. The size of X and Y must be greater than K(deg+2) + K.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.
EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
TimeTrans_Prop	A numerical vector. This vector is empty by default. If you want to estimate the model parameters for fixed jump time values, you can propose these values here. Warning, the size of this vector must be equal to K-1.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the HKSPOR function are plotted.

Value

A dataframe which contains the estimated parameters of the polynomial regression model at K regimes: the times of transition, the polynomials coefficients and the variances of the K regimes. If plotG = TRUE, the data (X,Y) and the estimated model will be plotted.

Examples

```
set.seed(3)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
xgrid = c(xgrid1,xgrid2,xgrid3)
ygrid = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
```

```
# under continuity constraint when the times of jump are fixed to 10 and 20.

HKSPOR(xgrid,ygrid,2,3,1,EM = FALSE,c(10,20))



set.seed(3)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
xgrid = c(xgrid1,xgrid2,xgrid3)
ygrid = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three regimes (K=3) on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR(xgrid,ygrid,2,3,1)
#Executed time : 49.70051 mins (intel core i7 processor)
```

HKSPOR_DynProg *Inference method for any number K of regimes using dynamic programming*

Description

HKSPOR_DynProg is an inference method implemented in the form of a Bellman algorithm that estimates, under the assumption of regularity, the parameters of a polynomial regression model for a number K of regimes given by the user..

Usage

```
HKSPOR_DynProg(X, Y, deg, K, constraint = 1, smoothing = TRUE,
  verbose = FALSE, plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.

Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains dupplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	The degree of the polynomials. The size of X and Y must be greater than K(deg+2) + K.
K	The number of regimes. The size of X and Y must be greater than K(deg+2) + K.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, in this dynamic programming method, to ensure that the number of constraints is not greater that the number of parameters to be estimated, the degree of the polynomials must be at least equal to 3 to be able to use the differentiability assumption.
smoothing	A Boolean. If TRUE (default), the method will estimate the parameters of a piecewise polynomial regression model with latent variable by maximizing the log-likelihood weighted by the probability of being in the latent variable regime. If FALSE, the method will estimate the parameters of the piecewise polynomial regression model.
verbose	A Boolean. If FALSE (default) the HKSPOR_Dynprog function will return only one dataframe containing the parameter estimates obtained for a model at K regimes. If TRUE, the function will return all the results obtained for a model with 1 regime up to K regimes.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the HKSPOR_DynProg function are plotted.

Value

One or more dataframes depend on the verbose value. If verbose = False, the output table will contain the estimated parameters of the polynomial regression model at K regimes: jump times, polynomial coefficients and variances of K regimes. If verbose = True then there will be K dataframes in output. Each table will contain the results of the estimated parameters obtained for each value of k (k=1,...,k=K). If plotG = TRUE, the data (X,Y) and the estimated model(s) will be plotted.

Examples

```
#generated data with three regimes
set.seed(1)
xgrid1 = seq(0,10,length.out=6)
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,4)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,4)
```

```
datX = c(xgrid1,xgrid2)
datY = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes (K=2) on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR_DynProg(datX,datY,2,2)


set.seed(2)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
datX = c(xgrid1,xgrid2,xgrid3)
datY = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three (K=3) regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR_DynProg(datX,datY,2,3)
#Executed time : 3.658121 mins (intel core i7 processor)
```

Index