Package 'mexhaz'

November 6, 2015

Type Package

Title Mixed effect excess hazard models

Version 1.0

Date 2015-10-21

Author Hadrien Charvat, Aurelien Belot

Maintainer Hadrien Charvat <chadrien@ncc.go.jp>

Description Fit excess hazard regression models with a random effect defined at the cluster level.

Depends survival

Imports statmod, MASS

License GPL (>=2)

Archs i386, x64

R topics documented:

`mexhaz-package` *Mixed effect excess hazard model*

Description

Fit an (excess) hazard regression model using different shapes for the baseline hazard (Weibull, piecewise constant and B-splines), with the possibility to include time-dependent and/or non-linear effect(s) of variable(s) and a random effect defined at the cluster level. The time-dependent effect of a covariate is modelled by adding interaction terms between the covariate and a function of time of the same class as the one used for the baseline hazard (in particular, with the same knots for piecewise constant hazards; and with the same degree and the same knots for B-spline functions). The random effect is assumed to follow a normal distribution with mean 0 and standard deviation sigma. The optimisation process uses the adaptive Gaussian quadrature to calculate the cluster

specific marginal likelihood. The full (log) marginal likelihood, defined as the sum of the (log) cluster-specific marginal likelihood, is then maximised using optimisation routine such as nlm or optim. Functions to compute and plot the predicted (excess) hazard and (net) survival are provided.

Details

Package:	mexhaz
Type:	Package
Version:	1.0
Date:	2015-11-06
License:	GPL (>=2)

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A. A multi-level excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates.

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a cubic B-spline with two knots at 1 and 5 year, with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH)
## and with a random effect for the cluster (clust)

# Mod_bs3_2mix <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
# degree=3, knots=c(1,5), expected="popmrate", random="clust", pl.nlm=0,
# verbose=1000)

## Fit of a fixed-effect overall hazard model, with the baseline hazard
## described by a piecewise constant function with the following vector
## of knots (defining the endpoints of the intervals on which the hazard
## is constant): (1,3,5,8), and with effects of age (agecr), deprivation
## index (depindex) and sex (IsexH)

# Mod_pw <- mexhaz(formula=Surv(time=timesurv, event=vstat)~
# agecr+depindex+IsexH, data= simdatn1, base="pw.cst", knots=c(1,3,5,8),
# pl.nlm=0, verbose=1000)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a cubic B-spline with two knots at 1 and 5 year, with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH),
## with a random effect for the cluster (clust), and with a
## time-dependent effect for age (agecr) and sex (IsexH).
```

```
# Mod_bs3_2mixnph <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH + nph(agecr+IsexH), data=mydata,
# base="exp.bs", degree=3, knots=c(1,5), expected="popmrate",
# random="clust", pl.nlm=0, verbose=1000)

## Fit of a fixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution and with effects of age (agecr),
## deprivation index (depindex) and sex (IsexH) using the optim
## procedure and the BFGS method (see help of optim).

# Mod_weib <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH, data=mydata, base="weibull",
# expected="popmrate", pl.nlm=0, verbose=1000, fnoptim="optim",
# method="BFGS")
```

`graph.mexhaz` *graph.mexhaz function*

Description

Function for plotting the predicted (excess) hazard or (net) survival obtained thanks to the `pred.mexhaz` function.

Usage

```
graph.mexhaz(pred, type = c("hazard", "survival", "both"),
ylim.H = c(0,1), ylim.S = c(0,1), points = FALSE, col = "black", ...)
```

Arguments

`pred`	an object of class `pred.mxh`, corresponding to predicted hazard and survival values obtained with the `pred.mexhaz` function for a user-specified set of covariates (see details of the `pred.mexhaz` function).
`type`	a character string indicating whether to plot the hazard, the survival or both.
`ylim.H`	a vector of size 2 corresponding to the limits of the y-axis to be used when plotting the hazard. The default value is set to c(0,1).
`ylim.S`	a vector of size 2 corresponding to the limits of the y-axis to be used when plotting the survival. The default value is set to c(0,1).
`points`	a logical value indicating whether the function should use an existing graphical window or create a new one. `FALSE` is the default value, meaning that a new graphical window will be used.
`col`	a character string indicating the colour to be used to plot the hazard and survival.
`...`	additional parameters that are directly passed to the `plot` function.

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates.

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a cubic B-spline with two knots at 1 and 5 year and with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH)
## and with a random effect for the cluster (clust)

 Mod_bs3_2mix <- mexhaz(formula=Surv(time=timesurv,
 event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
 degree=3, knots=c(1,5), expected="popmrate", random="clust", pl.nlm=0,
 verbose=1000)

## Prediction of the excess hazard and the net survial up to 10 years
## for men of 70 years old and with a deprivation index equal to 0.5
 Pred_Modbs3_2mix <- pred.mexhaz(Mod_bs3_2mix,t=10,
 data.val=data.frame(agecr=0,depindex=0.5,IsexH=1),
 nb.time.pts=1000)

## Plot of the excess hazard and the net survial up to 10 years
## for men of 70 years old and with a deprivation index equal to 0.5
 graph.mexhaz(Pred_Modbs3_2mix,type="both",ylim.H=c(0,2))
```

---

mexhaz *mexhaz function*

Description

Fit an (excess) hazard regression model using different shapes for the baseline hazard (Weibull, piecewise constant and B-splines), with the possibility to include time-dependent and/or non-linear effect(s) of variable(s) and a random effect defined at the cluster level. The time-dependent effect of a covariate is modelled by adding interaction terms between the covariate and a function of time of the same class as the one used for the baseline hazard (in particular, with the same knots for piecewise constant hazards; and with the same degree and the same knots for B-spline functions). The random effect is assumed to follow a normal distribution with mean 0 and standard deviation sigma. The optimisation process uses the adaptive Gaussian quadrature to calculate the cluster specific marginal likelihood. The full (log) marginal likelihood, defined as the sum of the (log) cluster-specific marginal likelihood, is then maximised using optimisation routine such as nlm or optim.

Usage

```
mexhaz(formula, data, expected = NULL, base = c("weibull",
"exp.bs", "pw.cst"), degree = 3, knots = NULL, bound = NULL, n.gleg =
100, init = NULL, random = NULL, n.aghq = 10, fnoptim = c("nlm",
"optim"), verbose = 100, method = "Nelder-Mead", pl.nlm = 2, ...)
```

Arguments

formula	a formula object, with the response on the left of the ~ operator, and the linear predictor on the right. The response must be of the form Surv(time, event). The linear predictor accepts a special instruction nph() for specifying variables for which a time-dependent effect should be modelled.
data	a data.frame containing the variables referred to in the formula, as well as in the expected and random arguments if these arguments are used.
expected	name of the variable (must be given in quotes) representing the population (i.e. expected) hazard. By default, expected = NULL, which means that the function estimates the overall hazard (and not the excess hazard).
base	functional form that should be used to model the baseline hazard. Selection can be made between the following options: "weibull" for a Weibull hazard, "exp.bs" for a hazard described by the exponential of a B-spline (only B-splines of degree 1, 2 or 3 are accepted), "pw.cst" for a piecewise constant hazard.
degree	if base = "exp.bs", degree represents the degree of the B-spline used. Only integer values between 1 and 3 are accepted, and 3 is the default.
knots	if base = "exp.bs", knots is the vector of interior knots of the B-spline. If base = "pw.cst", knots is the vector defining the endpoints of the time intervals on which the hazard is constant. By default, knots = NULL (that is, it produces a B-spline with no interior knots if base = "exp.bs" or a constant hazard over all the follow-up if base = "pw.cst").
bound	if base = "exp.bs", bound is a vector of length 2 representing the boundary knots of the B-spline. By default, bound=NULL and the boundaries are set to c(0, max(time) + 0.001).
n.gleg	if base = "exp.bs" and degree is 2 or 3, the cumulative hazard is computed via Gauss-Legendre quadrature. n.gleg is the number of quadrature nodes to be used to compute the cumulative hazard. By default, n.gleg = 100.
init	vector of initial values. By default init = NULL and the initial values are internally set to the following values: for the baseline hazard: if base = "weibull", the scale and shape parameters are set to 0.1; if base = "exp.bs", the parameters of the B-spline are all set to -1; if base = "pw.cst", the logarithm of the piecewise-constant hazards are set to -1; the parameters describing the effects of the covariates are all set to 0; the parameter representing the standard error of the random effect is set to 0.1.
random	name of the variable to be entered as a random effect (must be given between quotes), representing the cluster membership. By default, random = NULL which means that the function fit a fixed effects model.
n.aghq	number of quadrature points to be used for estimating the cluster-specific marginal likelihoods by adaptive Gauss-Hermite quadrature. By default, n.aghq = 10.
fnoptim	name of the R optimisation procedure used to maximise the likelihood. Selection can be made between "nlm" and "optim".
verbose	integer parameter representing the frequency at which the current state of the optimisation process is displayed. Internally, an 'evaluation' is defined as an estimation of the log-likelihood for a given vector of parameters. This means that the number of evaluations is increased each time the optimisation procedure changes the value of any of the parameters to be estimated. If verbose = <n> (with n an integer), the function will display the current values of the parameters,

the log-likelihood and the time elapsed every n evaluations. If `verbose = 0`, nothing is displayed.

`method` if `fnoptim = "optim"`, method represents the method to be used by optim. By default,`method = "Nelder-Mead"`. This parameter is not used if `fnoptim = "nlm"`.

`pl.nlm` integer parameter taking values between 0 and 2, and used only when `fnoptim = "nlm"`. The `nlm` function provides a `print.level` option which allows one to control the display of the current state of the optimisation process (in terms of iterations and not of evaluations, that is, the iteration counter is incremented only when `nlm` has chosen a new direction in the parameter space, i.e., after several evaluations). The `pl.nlm` argument is directly passed to the `print.level` argument of `nlm`. If `pl.nlm = 0`, nothing is displayed; if `pl.nlm = 1`, the first and final iterations are displayed; and if `pl.nlm = 2`, `nlm` displays the state of the optimisation process for each iteration (log-likelihood, current value of the parameters, current value of the gradient). Because the `mexhaz` function already has a mechanism for printing the number of function evaluations (see above the details concerning the `verbose` argument), the `verbose` option is discarded when `fnoptim = "nlm"` and `pl.nlm = 2`, in order to avoid printing too much information on screen.

`...` represents additional parameters directly passed to `nlm` or `optim` to control the optimisation process.

Value

An object of class `model.mxh` containing the following elements:

`data` a `data.frame` with the name of the file used, the number of total observation and the number of observation used (may be different in case of missing data), the name of the variable associated with the random effect and the number of different cluster.

`formula` the formula part of the code used for the model.

`baseline` a `data.frame` with the parameter defining the baseline, mainly useful when a B-spline is used to get the boundaries and the degree of the spline.

`knots` a vector with the position of the knot(s), when a B-spline is used for the baseline.

`coefficients` a `data.frame` with the parameter estimates and their standard errors.

`varcov` the variance-covariance matrix of the estimated parameters.

`mu.hat` a `data.frame` with the shrinkage estimates predicted for each cluster.

`details` a `data.frame` with the number of iteration and evaluation calculated in the optimisation process, the number of estimated parameters, the baseline used, the number of quadrature points, the optimisation function and method, the code (integer) indicating why the optimisation process terminated, the final value of the log-likelihood and the total time required to reach convergence.

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates.

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a cubic B-spline with two knots at 1 and 5 year and with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH)

# Mod_bs3_2mix_nph <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH+nph(agecr), data=simdatn1,
# base="exp.bs", degree=3, knots=c(1,5), expected="popmrate",
# random="clust", pl.nlm=0, verbose=1000)
```

`pred.mexhaz` *pred.mexhaz function*

Description

Function for predicting the (excess) hazard and the corresponding (net) survival from a model fitted with the `mexhaz` function for a particular vector of covariates. If the survival model was fitted with an expected hazard, the estimates obtained are excess hazard and net survival estimates. When the model includes a random effect, the predicted values are obtained for the value 0 of the random effect. Confidence intervals can be obtained by Monte-Carlo simulation (for all types of baseline hazard) and by the Delta Method (only for B-splines of degree 2 or 3).

Usage

```
pred.mexhaz(model, t, data.val = data.frame(.NotUsed=NA),
nb.time.pts = 100, nb.sim = 10000, delta = TRUE)
```

Arguments

`model`	an object of class `model.mxh`, corresponding to a survival model fitted with the `mexhaz` function.
`t`	a numerical value representing the maximum time of prediction.
`data.val`	a `data.frame` containing the values of the covariates at which the predictions should be calculated.
`nb.time.pts`	integer value representing the number of time points into which the interval (0,t) should be divided.
`nb.sim`	integer value representing the number of simulations used to estimate the confidence intervals for the (excess) hazard and the (net) survival. When the logarithm of the baseline hazard is modelled by a B-spline of degree 2 or 3, this argument is used only if `delta = FALSE`.
`delta`	logical value indicating whether confidence intervals should be estimated by the Delta Method. This option operates only for models in which the logarithm of the baseline hazard is modelled by a B-spline of degree 2 or 3. Indeed, in this case, obtaining confidence intervals by simulations can be time-consuming.

Value

An object of class `pred.mxh` that can be used by the function `graph.mexhaz` to produce graphics of the (excess) hazard and the (net) survival. It contains the following elements:

`results`	a `data.frame` consisting of six columns: the time points at which the (excess) hazard and the (net) survival have been calculated, the (excess) hazard values with their confidence limits, and the (net) survival values with their confidence limits.
`formula`	the formula used in the model.
`data.val`	a `data.frame` containing the values of the covariates used to calculate the (excess) hazard and the (net) survival.

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates.

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a cubic B-spline with two knots at 1 and 5 year and with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH)
## and with a random effect for the cluster (clust)

# Mod_bs3_2mix <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
# degree=3, knots=c(1,5), expected="popmrate", random="clust", pl.nlm=0,
# verbose=1000)

## Prediction of the excess hazard and the net survial up to 10 years
## for men of 70 years old and with a deprivation index equal to 0.5

# Pred_Modbs3_2mix <- pred.mexhaz(Mod_bs3_2mix,t=10,
# data.val=data.frame(agecr=0,depindex=0.5,IsexH=1),
# nb.time.pts=1000)
```

`simdatn1` *Simulated dataset*

Description

The `simdatn1` dataset has 4000 rows and 8 columns.

Format

This dataset contains the following columns:

age Age at diagnosis (continuous).

agecr Centred and rescaled age variable (age-70)/100.

depindex Deprivation index (continuous).

IsexH Sex (0 = Female, 1 = Male).

clust ID number of the cluster.

vstat Vital status (0 = Alive, 1 = Dead).

timesurv Follow-up time (years).

popmrate Population (expected) mortality rate at the time of censoring.

Index