Package 'mexhaz'

September 14, 2021

Type Package

Title Mixed Effect Excess Hazard Models

Version 2.0

Date 2021-09-14

Author Hadrien Charvat, Aurelien Belot

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

Description Fit flexible (excess) hazard regression models with the possibility of including non-proportional effects of covariables and of adding a random effect at the cluster level (corresponding to a shared frailty). A detailed description of the package functionalities is provided in Charvat and Belot (2021) <doi:10.18637/jss.v098.i14>.

Depends survival

Imports statmod, MASS, numDeriv, splines, lamW

Suggests utils, rstpm2

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2021-09-14 13:30:02 UTC

R topics documented:

---

mexhaz-package   *Mixed effect parametric excess hazard models*

Description

Fit an (excess) hazard regression model using different shapes for the baseline hazard (Weibull, piecewise constant, exponential of a B-spline of degree 1 to 3, exponential of a restricted cubic spline), with the possibility to include time-dependent effects of variable(s) and a random effect defined at the cluster level. Follow-up time can be entered in the right-censored or counting process input style. The latter allows the modelling of survival data with delayed entries. The time-dependent effect of a covariable is modelled by adding interaction terms between the covariable 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 or restricted cubic spline functions). The random effect is assumed to be normally distributed with mean 0 and standard deviation sigma. The optimisation process uses adaptive Gaussian quadrature to calculate the cluster-specific marginal likelihoods. The logarithm of the full marginal likelihood, defined as the sum of the logarithms of the cluster-specific marginal likelihoods, is then maximised using an optimisation routine such as `nlm` (default) or `optim`. Functions to compute and plot the predicted (excess) hazard and (net) survival (possibly with cluster-specific predictions in the case of a random effect model) are provided.

Details

Package:	mexhaz
Type:	Package
Version:	2.0
Date:	2021-09-14
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; CENSUR Working Survival Group. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. Stat Med 2016;35(18):3066-3084 (doi: 10.1002/sim.6881)

Charvat H, Belot A. An R package for fitting flexible hazard-based regression models for overall and excess mortality with a random effect. J Stat Softw 2021;98(14):1-36 (doi: 10.18637/jss.v098.i14)

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")


## Examples of syntax for various models (not run)

## 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=simdatn1, base="weibull",
# expected="popmrate", verbose=1000, fnoptim="optim",
# method="BFGS")


## 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 <- 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",
# 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),
# verbose=1000)


## Fit of a fixed-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_2 <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
# degree=3, knots=c(1,5), expected="popmrate", 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 and with
## effects of age (agecr), deprivation index (depindex) and sex (IsexH)

# 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",
# 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)
## 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=simdatn1,
# base="exp.bs", degree=3, knots=c(1,5), expected="popmrate",
# random="clust", verbose=1000)
```

---

fixef   *Method for extracting fixed effects*

Description

This is a generic function.

Usage

```
fixef(x, ...)
```

Arguments

x	a fitted object from which fixed effects can be extracted.
...	may be required by some methods using this generic function.

Value

see the documentation for the specific class.

See Also

[ranef](#)

Examples

```
## See the specific class documentation.
```

---

fixef.mexhaz *Method for extracting fixed effects from a mexhaz object*

Description

Display a vector containing the fixed effects of a mexhaz model.

Usage

```
## S3 method for class 'mexhaz'
fixef(x, ...)
```

Arguments

x	an object of class mexhaz.
...	not used.

Value

A vector containing the fixed effect estimates.

See Also

[mexhaz](#)

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")

fixef(Mod_weib_mix)
```

---

lines.predMexhaz   *Lines method for a predMexhaz object*

Description

Function for adding to an already existing graphical window the predicted (excess) hazard or (net) survival based on a predMexhaz object.

Usage

```
## S3 method for class 'predMexhaz'
lines(x, which=c("surv", "hazard"), conf.int=TRUE,
lty.pe="solid", lty.ci="dashed", ...)
```

Arguments

x	an object of class predMexhaz, corresponding to (excess) hazard and (net) survival predictions based on a survival model fitted with the mexhaz function. Predictions can be obtained for multiple times for one vector of covariables ("multitime") or for several vectors of covariables at one time point ("multiobs"). The lines() function only applies to the "multitime" type of predictions.
which	type of curve to be plotted. Selection can be made between "surv" (default value) for the (net) survival curve and "hazard" for the (excess) hazard.
conf.int	logical values allowing the user to decide whether to plot the confidence limits of the survival (or hazard).
lty.pe	type of line used for drawing the hazard/survival estimate.
lty.ci	type of line used for drawing the confidence limits.
...	additional parameters that are directly passed to the lines function. These parameters, if not already used internally by the lines.predMexhaz function, will apply simultaneously to the point estimate and confidence limit curves.

See Also

predict.mexhaz, plot.predMexhaz

Examples

```
data(simdatn1)

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

Mod_bs2_2 <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
degree=2, knots=c(1,5), verbose=0)
```

```
## Prediction at several time points for one vector of covariates
Pred_Modbs2_2A <- predict(Mod_bs2_2, time.pts=seq(0,10,by=0.1),
data.val=data.frame(agecr=0,depindex=0.5,IsexH=1))

plot(Pred_Modbs2_2A, which="hazard", col="red")
lines(Pred_Modbs2_2A, which="hazard", conf.int=FALSE)
```

---

mexhaz	*mexhaz function*

Description

Fit an (excess) hazard regression model using different shapes for the baseline hazard (Weibull, piecewise constant, exponential of a B-spline of degree 1 to 3, exponential of a restricted cubic spline), with the possibility to include time-dependent effects of variable(s) and a random effect defined at the cluster level. The function accepts right-censored and counting process input styles for the follow-up time. The latter allows the modelling of survival data with delayed entries. The time-dependent effect of a covariable is modelled by adding interaction terms between the covariable 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 or restricted cubic spline functions). The random effect is assumed to be normally distributed with mean 0 and standard deviation sigma. The optimisation process uses adaptive Gaussian quadrature to calculate the cluster-specific marginal likelihoods. The logarithm of the full marginal likelihood, defined as the sum of the logarithms of the cluster-specific marginal likelihoods, is then maximised using an optimisation routine such as nlm (default) or optim.

Usage

```
mexhaz(formula, data, expected = NULL, base = c("weibull",
"exp.bs", "exp.ns", "pw.cst"), degree = 3, knots = NULL, bound = NULL,
n.gleg = 20, init = NULL, random = NULL, n.aghq = 10,
fnoptim = c("nlm", "optim"), verbose = 0, method = "Nelder-Mead",
iterlim = 10000, numHess = FALSE, print.level = 1,
exactGradHess = TRUE, gradtol = ifelse(exactGradHess, 1e-8, 1e-6), ...)
```

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) for right censored data or Surv(time,time2,event) for counting process style data. The linear predictor accepts a special instruction nph() for specifying variables for which a time-dependent effect should be modelled (if several variables are modelled with time-dependent effects, separate these variables inside the nph() instruction with a + sign).

In case time takes the value 0 for some observations when data are entered in the right censored style, it is assumed that these observations refer to events/censoring

that occurred on the first day of follow-up. Consequently, a value of 1/730.5 (half a day) is substituted in order to make computations possible. However, it should be stressed that this is just a convention and that it does not make much sense if the time scale is not expressed in years. We therefore advise the analyst to deal with 0 time values during the dataset preparation stage.

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), `"exp.ns"` for a hazard described by the exponential of a restricted cubic spline (also called 'natural spline'), `"pw.cst"` for a piecewise constant hazard. By default, `base="weibull"`.

For the Weibull hazard model, the cumulative hazard is given by the following relationship:

$H(t,x,z) = lambda*exp(x'b)*t^(rho*exp(z'g))$

where lambda and rho are the parameters of the Weibull baseline hazard, x represent variables with proportional effect (with corresponding regression coefficients 'b') and z represent variables with time-dependent effects (with corresponding regression coefficients 'g'). The `mexhaz()` function does not estimate directly lambda and rho but their logarithms (in the output of the function, these are named respectively 'logLambda' and 'logRho').

For the spline-based hazards, it should be noted that the B-spline and restricted cubic spline bases created internally in the `mexhaz()` function are identical to those obtained by the use of, respectively, the `bs()` and `ns()` functions from the `splines` package.

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"` or `"exp.ns"`, knots is the vector of interior knots of the spline. If `base="pw.cst"`, knots is the vector defining the endpoints of the time intervals on which the hazard is assumed to be constant. By default, `knots=NULL` (that is, it produces a B-spline with no interior knots if `base="exp.bs"`, a linear B-spline with no interior knots if `base="exp.ns"`, or a constant hazard over the whole follow-up period if `base="pw.cst"`).
bound	a vector of two numerical values corresponding to the boundary knots of the spline functions. If `base="exp.bs"` or `base="exp.ns"`, computation of the B-spline basis requires that boundary knots be given. The bound argument allows the user to specify these boundary knots. If `base="exp.bs"`, the interval defined by the boundary knots must at least include the interval `c(0,max(time))` (otherwise, there could be problems with ill-conditioned bases). If `base="exp.ns"`, the boundary knots correspond to the knots beyond which the spline is constrained to be linear (in that case, the boundary knots can be values contained in `c(0,max(time))`). By default, the boundary knots are set to `c(0,max(time))`.

n.gleg	if base="exp.bs" and degree is equal to 2 or 3, or if base="exp.ns", the cumulative hazard is computed via Gauss-Legendre quadrature and n.gleg is the number of quadrature nodes to be used to compute the cumulative hazard. By default, n.gleg=20.
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 exactGradHess=TRUE (except for the excess hazard random effect model for which this argument is ignored), the intercept is set to 0.5*log(Number of events/Person-years of observation) and all other parameters set to 0. In case of failed convergence, several trials are run with an adaptation of the value of the intercept.
	if exactGradHess=FALSE, the following values are used:
	- if base="weibull", the logarithm of the scale and shape parameters is set to 0.1;
	- if base="exp.bs", the parameters of the B-spline are all set to -1;
	- if base="exp.ns", the parameters of the restricted cubic 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 covariables are all set to 0;
	the parameter representing the standard deviation of the random effect is set to 0.1 (if exactGradHess=TRUE, several trials are run with an adaptation of the value in case of failed convergence).
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 fits 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" (by default) and "optim". Note: if exactGradHess=TRUE, this argument will be ignored (fnoptim will be set automatically to "nlm").
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 updates 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 (default), nothing is displayed.
method	if fnoptim="optim", method represents the optimisation method to be used by optim. By default, method="Nelder-Mead". This parameter is not used if fnoptim="nlm".
iterlim	if fnoptim="nlm", iterlim represents the maximum number of iterations before the nlm optimisation procedure is terminated. By default, iterlim is set to

10000. This parameter is not used if fnoptim="optim" (in this case, the maximum number of iterations must be given as part of a list of control parameters via the control argument: see the help page of optim for further details).

numHess   logical value allowing the user to choose between the Hessian returned by the optimization algorithm (default) or the Hessian estimated by the hessian function from the numDeriv package. The latter might be more accurate but its estimation is more time-consuming. We suggest to use the default Hessian estimation procedure during model building and estimate the numDeriv-based Hessian only on the final model. Note: if exactGradHess=TRUE, this argument is ignored.

print.level   this argument is only used if fnoptim="nlm". It determines the level of printing during the optimisation process. The default value (for the mexhaz function) is set to '1' which means that details on the initial and final step of the optimisation procedure are printed (see the help page of nlm for further details).

exactGradHess   logical value allowing the user to decide whether maximisation of the likelihood should be based on the analytic gradient and Hessian computed internally (default, corresponding to exactGradHess=TRUE). In that case, optimisation is performed with the nlm function. Note: even if set to TRUE, this argument is ignored when the user wants to fit an excess hazard model including a random effect because in that case, there is no simple way to obtain the analytic gradient and Hessian.

gradtol   this argument is only used if fnoptim="nlm". It corresponds to the tolerance at which the scaled gradient is considered close enough to zero to terminate the algorithm. The default value depends on the value of the argument exactGradHess.

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

Value

An object of class mexhaz containing the following elements:

dataset   name of the dataset used to fit the model.

call   function call on which the model is based.

formula   formula part of the call.

expected   name of the variable corresponding to the expected hazard (takes the value NA for standard, i.e., 'non-excess' hazard models).

xlevels   information concerning the levels of the categorical variables used in the model (used by predict.mexhaz).

n.obs.tot   total number of observations in the dataset.

n.obs   number of observations used to fit the model (after exclusion of missing values).

n.events   number of events (after exclusion of missing values).

n.clust   number of clusters.

n.time.0   number of observations for which the observed follow-up time was equal to 0 (only for right censored type data).

base   function used to model the baseline hazard.

max.time	maximal observed time in the dataset.
boundary.knots	vector of boundary values used to define the B-spline (or natural spline) bases.
degree	degree of the B-spline used to model the logarithm of the baseline hazard.
knots	vector of interior knots used to define the B-spline (or natural spline) bases.
names.ph	names of the covariables with a proportional effect.
random	name of the variable defining cluster membership (set to NA in the case of a purely fixed effects model).
init	a vector containing the initial values of the parameters.
coefficients	a vector containing the parameter estimates.
std.errors	a vector containing the standard errors of the parameter estimates.
vcov	the variance-covariance matrix of the estimated parameters.
gradient	the gradient of the log-likelihood function evaluated at the estimated parameters.
hessian	the Hessian of the log-likelihood function evaluated at the estimated parameters.
mu.hat	a data.frame containing the estimated cluster-specific random effects (shrinkage estimators).
var.mu.hat	the covariance matrix of the cluster-specific shrinkage estimators.
vcov.fix.mu.hat	
	a matrix containing the covariances between the fixed effect and the cluster-specific shrinkage estimators. More specifically, the i-th line of the matrix represents the covariances between the shrinkage estimator of the i-th cluster and the fixed effect estimates. This matrix is used by the function predict.mexhaz to make cluster-specific predictions.
n.par	number of estimated parameters.
n.gleg	number of Gauss-Legendre quadrature points used to calculate the cumulative (excess) hazard (only relevant if a B-spline of degree 2 or 3 or a cubic restricted spline was used to model the logarithm of the baseline hazard).
n.aghq	number of adaptive Gauss-Hermite quadrature points used to calculate the cluster-specific marginal likelihoods (only relevant if a multi-level model is fitted).
fnoptim	name of the R optimisation procedure used to maximise the likelihood.
method	optimisation method used by optim.
code	code (integer) indicating the status of the optimisation process (this code has a different meaning for nlm and for optim: see their respective help page for details).
loglik	value of the log-likelihood at the end of the optimisation procedure.
iter	number of iterations used in the optimisation process.
eval	number of evaluations used in the optimisation process.
time.elapsed	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; CENSUR Working Survival Group. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. Stat Med 2016;35(18):3066-3084 (doi: 10.1002/sim.6881)

Charvat H, Belot A. An R package for fitting flexible hazard-based regression models for overall and excess mortality with a random effect. J Stat Softw 2021;98(14):1-36 (doi: 10.18637/jss.v098.i14)

See Also

`fixef.mexhaz`, `predict.mexhaz`, `print.mexhaz`, `ranef.mexhaz`, `summary.mexhaz`, `update.mexhaz`, `vcov.mexhaz`

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")


## More complex examples (not run)

## 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", verbose=1000)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a restricted cubic spline with two knots at the
## tertiles of the distribution of follow-up times for individuals who
## presented the event and with effects of age (agecr) and sex (IsexH)

# Mod_ns3_2mix_nph <- mexhaz(formula=Surv(time=timesurv,
# event=vstat)~agecr+nph(agecr), data=simdatn1, base="exp.ns", degree=3,
# knots=quantile(simdatn1[simdatn1$vstat==1,]$timesurv, probs=c(1:2/3)),
# expected="popmrate", random="clust", verbose=1000)
```

---

plot.predMexhaz   *Plot method for a predMexhaz object*

Description

Function for plotting the predicted (excess) hazard or (net) survival based on a predMexhaz object.

Usage

```
## S3 method for class 'predMexhaz'
plot(x, which=c("surv", "hazard"), conf.int=TRUE,
lty.pe="solid", lty.ci="dashed", ...)
```

Arguments

x	an object of class predMexhaz, corresponding to (excess) hazard and (net) survival predictions based on a survival model fitted with the mexhaz function. Predictions can be obtained for multiple times for one vector of covariables ("multitime") or for several vectors of covariables at one time point ("multiobs"). The plot() function only applies to the "multitime" type of predictions.
which	type of curve to be plotted. Selection can be made between "surv" (default value) for the (net) survival curve and "hazard" for the (excess) hazard.
conf.int	logical value allowing the user to decide whether to plot the confidence limits of the survival (or hazard).
lty.pe	type of line used for drawing the hazard/survival estimate.
lty.ci	type of line used for drawing the confidence limits.
...	additional parameters that are directly passed to the plot function. These parameters will apply simultaneously to the point estimate and confidence limit curves.

See Also

predict.mexhaz, points.predMexhaz, lines.predMexhaz

Examples

```
data(simdatn1)

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

Mod_bs2_2 <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
degree=2, knots=c(1,5), verbose=0)
```

```
## Prediction at several time points for one vector of covariates
Pred_Modbs2_2A <- predict(Mod_bs2_2, time.pts=seq(0,10,by=0.1),
data.val=data.frame(agecr=0,depindex=0.5,IsexH=1))

plot(Pred_Modbs2_2A, which="hazard")
```

---

points.predMexhaz   *Points method for a predMexhaz object*

Description

Function for adding to an already existing graphical window the predicted (excess) hazard or (net)
survival based on a predMexhaz object.

Usage

```
## S3 method for class 'predMexhaz'
points(x, which=c("surv", "hazard"), conf.int=TRUE,
lty.pe="solid", lty.ci="dashed", ...)
```

Arguments

x	an object of class predMexhaz, corresponding to (excess) hazard and (net) survival predictions based on a survival model fitted with the mexhaz function. Predictions can be obtained for multiple times for one vector of covariables ("multitime") or for several vectors of covariables at one time point ("multiobs"). The points() function only applies to the "multitime" type of predictions.
which	type of curve to be plotted. Selection can be made between "surv" (default value) for the (net) survival curve and "hazard" for the (excess) hazard.
conf.int	logical values allowing the user to decide whether to plot the confidence limits of the survival (or hazard).
lty.pe	type of line used for drawing the hazard/survival estimate (used when 'type="l"').
lty.ci	type of line used for drawing the confidence limits (used when 'type="l"').
...	additional parameters that are directly passed to the points function. These parameters, if not already used internally by the points.predMexhaz function, will apply simultaneously to the point estimate and confidence limit curves.

See Also

[predict.mexhaz](), [plot.predMexhaz]()

Examples

```
data(simdatn1)

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

Mod_bs2_2 <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
degree=2, knots=c(1,5), verbose=0)

## Prediction at several time points for one vector of covariates
Pred_Modbs2_2A <- predict(Mod_bs2_2, time.pts=seq(0,10,by=0.1),
data.val=data.frame(agecr=0,depindex=0.5,IsexH=1))

plot(Pred_Modbs2_2A, which="hazard", col="red")
points(Pred_Modbs2_2A, which="hazard", type="l", conf.int=FALSE)
```

---

predict.mexhaz	*Predictions based on a mexhaz model*

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 covariables. If the survival model was fitted with an expected hazard (excess hazard model), the estimates obtained are excess hazard and net survival estimates. Corresponding variance estimates are based on the Delta Method or Monte Carlo simulation (based on the assumption of multivariate normality of the model parameter estimates). When the model includes a random effect, predictions can be made for a particular cluster (using the corresponding shrinkage estimate) or for the value 0 of the random effect. This function allows the computation of the hazard and the survival at one time point for several vectors of covariables or for one vector of covariables at several time points.

Usage

```
## S3 method for class 'mexhaz'
predict(object, time.pts, data.val = data.frame(.NotUsed=NA),
cluster=NULL, conf.int=c("delta","simul","none"), level=0.95,
delta.type.h=c("log", "plain"), delta.type.s=c("log-log", "log",
"plain"), nb.sim = 10000, include.gradient=FALSE, ...)
```

Arguments

object   an object of class mexhaz, corresponding to a hazard-based regression model
                fitted with the mexhaz function.

time.pts   a vector of numerical values representing the time points at which predictions
                are requested. Time values greater than the maximum follow-up time on which
                the model estimation was based are discarded.

data.val   a data.frame containing the values of the covariables at which predictions
                should be calculated.

cluster   a single number or character string corresponding to the name of the cluster for
                which predictions should be calculated. In that case, standard errors of predic-
                tion are obtained with the approximation of the conditional mean squared error
                of prediction (without the second-order bias correction term) described in Booth
                and Hobert. This argument is not used if the model is a fixed effect model. The
                default value is NULL: this corresponds to prediction for the value 0 of the ran-
                dom effect.

conf.int   method to be used to compute confidence limits. Selection can be made between
                the following options: "delta" for the Delta Method (default value); "simul"
                for Monte Carlo simulations (can be time-consuming, especially for models us-
                ing B-splines for the logarithm of the baseline hazard); "none" indicates absence
                of confidence limits estimation.

level   a number in (0,1) specifying the level of confidence for computing the confi-
                dence intervals of the hazard and the survival. By default, level=0.95.

delta.type.h   type of confidence limits for the hazard when using the Delta Method. If delta.type.h="log"
                (default value), the confidence limits are based on a Wald-type confidence in-
                terval for the logarithm of the hazard, otherwise they are based directly on a
                Wald-type CI for the hazard.

delta.type.s   type of confidence limits for the survival when using the Delta Method. If
                delta.type.s="log-log" (default value), the confidence limits are based on
                a Wald-type confidence interval for the logarithm of the cumulative hazard; if
                delta.type.s="log", they are based on a Wald-type CI for the cumulative
                hazard; otherwise they are based directly on a Wald-type CI for the survival.

include.gradient
                logical value allowing the function to return the components of the gradient of
                the logarithm of the hazard and of the logarithm of the cumulative hazard for
                each prediction. This argument is used only if conf.int="delta". The default
                value is FALSE.

nb.sim   integer value representing the number of simulations used to estimate the con-
                fidence limits for the (excess) hazard and the (net) survival. This argument is
                used only if conf.int="simul".

...   for potential additional parameters.

Value

An object of class predMexhaz that can be used by the functions plot.predMexhaz, points.predMexhaz
and lines.predMexhaz to produce graphics of the (excess) hazard and the (net) survival. It con-
tains the following elements:

call   the mexhaz function call on which the model is based.

results   a `data.frame` consisting of: the time points at which the (excess) hazard and the (net) survival have been calculated; the values of the covariables used to estimate the (excess) hazard and the (net) survival; the (excess) hazard values with their confidence limits; and the (net) survival values with their confidence limits.

variances   a `data.frame` consisting of two columns: the variance of the logarithm of the (excess) hazard and the variance of the (excess) cumulative hazard for each time points or each vector of covariables. The object `variances` is produced only when `conf.int="delta"`.

grad.loghaz   a `data.frame` consisting of the components of the gradient of the logarithm of the (excess) hazard for each time points or each vector of covariables. The number of columns corresponds to the number of model parameters. This object is produced only when `conf.int="delta"` and `include.gradient=TRUE`.

grad.logcum   a `data.frame` consisting of the components of the gradient of the logarithm of the (excess) cumulative hazard for each time points or each vector of covariables. The number of columns corresponds to the number of model parameters. This object is produced only when `conf.int="delta"` and `include.gradient=TRUE`.

vcov   a matrix corresponding to the covariance matrix used to compute the confidence intervals.

type   the type of predictions produced. Can take the value "multitime" (computation of the hazard and the survival at at several time points for one vector of covariables) or `multiobs` (computation of the hazard and the survival at at one time point for several vectors of covariables). This value is used by `plot.predMexhaz` and `points.predMexhaz`.

ci.method   the method used to compute confidence limits.

level   level of confidence used to compute confidence limits.

delta.type   type of confidence limits for the hazard and the survival when using the Delta Method.

nb.sim   number of simulations used to estimate the confidence limits when `ci.method = "simul"`.

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A; CENSUR Working Survival Group. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. Stat Med 2016;35:3066-3084 (doi: 10.1002/sim.6881)

Booth JG, Hobert JP. Standard errors of prediction in generalized linear mixed models. J Am Stat Assoc 1998;93:262-272 (doi: 10.2307/2669622).

See Also

print.predMexhaz, plot.predMexhaz, points.predMexhaz

Examples

```
data(simdatn1)

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

Mod_bs1_2 <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
degree=1, knots=c(1,5), verbose=0)

## Prediction at several time points for one vector of covariates
Pred_Modbs1_2A <- predict(Mod_bs1_2, time.pts=seq(0.1,10,by=0.1),
data.val=data.frame(agecr=0,depindex=0.5,IsexH=1))

## Prediction for several vectors of covariates at one time point
Pred_Modbs1_2B <- predict(Mod_bs1_2, time.pts=10,
data.val=data.frame(agecr=c(-0.2,-0.1,0), depindex=c(0.5,0.5,0.5),
IsexH=c(1,1,1)))

## Prediction for all individuals of the study population at one time point
Pred_Modbs1_2C <- predict(Mod_bs1_2, time.pts=10,
data.val=simdatn1)


# Example of cluster-specific prediction (not run)

## 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

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

## Prediction at several time points for an individual in cluster 15
## with a specific vector of covariables
# Pred_Modbs3_2A <- predict(Mod_bs3_2mix,
# time.pts=seq(0.1,10,by=0.1), data.val=data.frame(agecr=0.2, IsexH=1),
# cluster=15)
```

---

print.mexhaz   *Print method for a mexhaz object*

Description

Display the model call as well as the values of the estimated model parameters.

Usage

```
## S3 method for class 'mexhaz'
print(x, ...)
```

Arguments

x	an object of class mexhaz.
...	represents additional parameters directly passed to print.

See Also

[mexhaz](), [summary.mexhaz]()

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")

print(Mod_weib_mix)
```

---

print.predMexhaz   *Print method for a predMexhaz object*

Description

Display the first lines of the data.frame containing the predictions provided by the predMexhaz function.

Usage

```
## S3 method for class 'predMexhaz'
print(x, ...)
```

Arguments

x	an object of class predMexhaz.
...	represents additional parameters directly passed to print.

See Also

[predict.mexhaz]()

Examples

```
data(simdatn1)

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

Mod_bs1_2 <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~agecr+depindex+IsexH, data=simdatn1, base="exp.bs",
degree=1, knots=c(1,5), verbose=0)

## Prediction at several time points for one vector of covariates
Pred_Modbs1_2A <- predict(Mod_bs1_2, time.pts=seq(0.1,10,by=0.1),
data.val=data.frame(agecr=0,depindex=0.5,IsexH=1), conf.int="delta")

print(Pred_Modbs1_2A)
```

---

print.summary.mexhaz   *Printing method for a summary.mexhaz object*

Description

Display the model call, the values of the estimated model parameters, as well as the corresponding hazard ratios (only for proportional effects).

Usage

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

Arguments

x	an object of class summary.mexhaz.
...	represents additional parameters directly passed to print.

See Also

mexhaz, print.mexhaz, summary.mexhaz

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib <- mexhaz(formula=Surv(time=timesurv,
```

```
event=vstat)~1, data=simdatn1, base="weibull", verbose=0)

summary(Mod_weib)
```

---

ranef   *Method for extracting random effects*

Description

This is a generic function.

Usage

```
ranef(x, ...)
```

Arguments

x	a fitted object from which random effects can be extracted.
...	might be used by some methods using this generic function.

Value

see the documentation for the specific class.

See Also

[fixef](#)

Examples

```
## See the specific class documentation.
```

---

ranef.mexhaz   *Method for extracting random effects from a mexhaz object*

Description

Display a data frame containing the cluster-specific random effects with their standard errors.

Usage

```
## S3 method for class 'mexhaz'
ranef(x, ...)
```

Arguments

x   an object of class mexhaz.

...   not used.

Value

A data frame with three columns containing the cluster names, the random effect estimates, and their standard errors.

See Also

mexhaz

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")

ranef(Mod_weib_mix)
```

---

  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.

---

simdatn2   *Simulated dataset*

Description

The simdatn2 dataset has 4000 rows and 8 columns. The IsexH variable is simulated with a non-proportional effect.

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.

---

summary.mexhaz   *Summary method for a mexhaz object*

Description

Produces a summary of a mexhaz object.

Usage

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

Arguments

object   an object of class mexhaz.

...   represents additional parameters directly passed to summary.

See Also

[mexhaz](), [print.mexhaz](), [print.summary.mexhaz]()

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull", verbose=0)

summary(Mod_weib)
```

---

update.mexhaz	*Update of a mexhaz model*

Description

Function allowing the user to update an existing mexhaz model. All the arguments of the model can be updated. If the argument 'init' is not provided, the function uses the estimated values of the existing model as starting values for the corresponding parameters of the new model.

Usage

```
## S3 method for class 'mexhaz'
update(object, formula, data, expected = NULL,
base = c("weibull","exp.bs", "exp.ns", "pw.cst"), degree = 3, knots =
NULL, bound = NULL, n.gleg = 20, init = NULL, random = NULL, n.aghq =
10, fnoptim = c("nlm", "optim"), verbose = 0, method = "Nelder-Mead",
iterlim=10000, numHess=FALSE, print.level=1, exactGradHess = TRUE,
gradtol = ifelse(exactGradHess, 1e-8, 1e-6), ...)
```

Arguments

object	an object of class mexhaz, corresponding to a survival model fitted with the mexhaz function.
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) for right censored data or Surv(time,time2,event) for counting process style data. The linear predictor accepts a special instruction nph() for specifying variables for which a time-dependent effect should be modelled (if several variables are modelled with time-dependent effects, separate these variables inside the nph() instruction with a + sign).
	In case time takes the value 0 for some observations when data are entered in the right censored style, it is assumed that these observations refer to events/censoring that occurred on the first day of follow-up. Consequently, a value of 1/730.5 (half a day) is substituted in order to make computations possible. However, it should be stressed that this is just a convention and that it does not make much

	sense if the time scale is not expressed in years. We therefore advise the analyst to deal with 0 time values during the dataset preparation stage.
	See [update.formula](#) for more details.
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), "exp.ns" for a hazard described by the exponential of a restricted cubic B-spline (also called 'natural spline'), "pw.cst" for a piecewise constant hazard. By default, base="weibull".
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" or "exp.ns", knots is the vector of interior knots of the spline. If base="pw.cst", knots is the vector defining the endpoints of the time intervals on which the hazard is assumed to be constant. By default, knots=NULL (that is, it produces a B-spline with no interior knots if base="exp.bs", a linear B-spline with no interior knots if base="exp.ns", or a constant hazard over the whole follow-up period if base="pw.cst").
bound	If base="exp.bs" or base="exp.ns", computation of the B-spline basis requires that boundary knots be given. The bound argument allows the user to specify these boundary knots. If base="exp.bs", the interval defined by the boundary knots must at least include the interval c(0,max(time)) (otherwise, there could be problems with ill-conditioned bases). If base="exp.ns", the boundary knots correspond to the knots beyond which the spline is contrained to be linear (in that case, the boundary knots can be values contained in c(0,max(time))). By default, the boundary knots are set to c(0,max(time)).
n.gleg	if base="exp.bs" and degree is equal to 2 or 3, or if base="exp.ns", the cumulative hazard is computed via Gauss-Legendre quadrature and n.gleg is the number of quadrature nodes to be used to compute the cumulative hazard. By default, n.gleg=20.
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 exactGradHess=TRUE (except for the excess hazard random effect model for which this argument is ignored), the intercept is set to 0.5*log(Number of events/Person-years of observation) and all other parameters set to 0. In case of failed convergence, several trials are run with an adaptation of the value of the intercept.
	if exactGradHess=FALSE, the following values are used:
	- 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="exp.ns", the parameters of the restricted cubic 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 deviation of the random effect is set to 0.1. (if exactGradHess=TRUE, several trials are run with an adaptation of the value in case of failed convergence).

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 fits 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" (by default) and "optim". Note: if exactGradHess=TRUE, this argument will be ignored (fnoptim will be set automatically to "nlm").

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

iterlim   if fnoptim="nlm", iterlim represents the maximum number of iterations before the nlm optimisation procedure is terminated. By default, iterlim is set to 10000. This parameter is not used if fnoptim="optim" (in this case, the maximum number of iterations must be given as part of a list of control parameters via the control argument: see the help page of optim for further details).

numHess   logical value allowing the user to choose between the Hessian returned by the optimization algorithm (default) or the Hessian estimated by the hessian function from the numDeriv package. The latter might be more accurate but its estimation is more time-consuming. We suggest to use the default Hessian estimation procedure during model building and estimate the numDeriv-based Hessian only on the final model. Note: if exactGradHess=TRUE, this argument is ignored.

print.level   this argument is only used if fnoptim="nlm". It determines the level of printing during the optimisation process. The default value (for the mexhaz function) is set to '1' which means that details on the initial and final step of the optimisation procedure are printed (see the help page of nlm for further details).

exactGradHess   logical value allowing the user to decide whether maximisation of the likelihood should be based on the analytic gradient and Hessian computed internally

(default, corresponding to exactGradHess=TRUE). In that case, optimisation is performed with the nlm function. Note: even if set to TRUE, this argument is ignored when the user wants to fit an excess hazard model including a random effect because in that case, there is no simple way to obtain the analytic gradient and Hessian.

gradtol   this argument is only used if fnoptim="nlm". It corresponds to the tolerance at which the scaled gradient is considered close enough to zero to terminate the algorithm. The default value depends on the value of the argument exactGradHess.

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

Value

An object of class mexhaz. See mexhaz for more details.

Author(s)

Hadrien Charvat, Aurelien Belot

References

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A; CENSUR Working Survival Group. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. Stat Med 2016;35:3066-3084 (doi: 10.1002/sim.6881)

See Also

mexhaz

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull", verbose=0)

## Add an effect of gender
Mod_weib_2 <- update(Mod_weib, formula=~.+IsexH)
```

---

vcov   *Method for extracting the covariance matrix*

Description

This is a generic function.

Usage

```
vcov(object, ...)
```

Arguments

object	a fitted object from which the covariance matrix can be extracted.
...	may be required by some methods using this generic function.

Value

see the documentation for the specific class.

Examples

```
## See the specific class documentation.
```

---

vcov.mexhaz   *Method for extracting the covariance matrix of the fixed effects*

Description

Display a the covariance matrix of the fixed effects of a mexhaz model.

Usage

```
## S3 method for class 'mexhaz'
vcov(object, ...)
```

Arguments

object	an object of class mexhaz.
...	not used.

Value

The estimated covariance matrix of the fixed effects.

See Also

[mexhaz](#)

Examples

```
data(simdatn1)

## Fit of a mixed-effect excess hazard model, with the baseline hazard
## described by a Weibull distribution (without covariables)

Mod_weib_mix <- mexhaz(formula=Surv(time=timesurv,
event=vstat)~1, data=simdatn1, base="weibull",
expected="popmrate", verbose=0, random="clust")

vcov(Mod_weib_mix)
```

Index