Package ‘threg’

July 2, 2014

Type Package

Title Threshold Regression

Version 1.0.2

Date 2012-11-26

Author Tao Xiao

Maintainer Tao Xiao <taoxiao1@gmail.com>

Depends survival, Formula

Description Fit a threshold regression model based on the first-hitting-time of a boundary by the sample path of a Wiener diffusion process. The threshold regression methodology is well suited to applications involving survival and time-to-event data.

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2013-04-18 07:29:52

R topics documented:

`bmt` *Bone Marrow Transplantation Data*

Description

Survival of 137 acute leukemia patients treated with bone marrow transplants which are a standard treatment for acute leukemia.

Usage

```
bmt
```

Format

time:	time to relapse, death or end of study (in days)
indicator:	censoring indicator variable - 1 = dead or replapse, 0 = otherwise
recipient_age:	patient age
group:	risk categories based on their status at the time of transplantation
fab:	French-American-British (FAB) classification based on standard morphological criteria

References

Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. 2 edition, 2003 Springer-Verlag New York, Inc.

`lkr` *Leukemia Remission Data*

Description

Data of a leukemia remission study. There are 42 patients in the dataset, who were monitored for whether they relapsed and for how long (in weeks) they remained in remission.

Usage

```
lkr
```

Format

weeks:	the time that a patient was remained in remission (in weeks)
relapse:	whether the patient was relapsed - 1 = yes, 0 = no
treatment1:	1=drug A, 0 = standard drug for treatment1
treatment2:	1=drug B, 0 = standard drug for treatment2
wbc3cat:	white blood cell count, recorded in three categories - 1 = normal, 2= moderate, 3 = high

References

Garrett JM. sbe14: Odds Ratios and Confidence Intervals for Logistic Regression Models with Effect Modification. Stata Technical Bulletin, 36, 15-22. Reprinted in Stata Technical Bulletin Reprints, vol. 6, pp. 104-114, 1997 College Station, TX: Stata Press

---

threg — *fit a threshold regression model*

Description

This function can be used to fit a threshold regressio model based on the first-hitting-time of a boundary by the sample path of a Wiener diffusion process. It uses maximum likelihood estimation method for calculating regression coefficient estimates, asymptotic standard errors and p-values. The outputs can also provide the estimates of hazard ratios at selected time points for specified scenarios (based on given categories or value settings of covariates) and the plots for curves of estimated hazard functions, survival functions and probability density functions of the first-hitting-time.

Usage

```
threg(formula, data, hr, timevalue, scenario, graph, nolegend=0, nocolor=0)
```

Arguments

`formula` a formula object, with the response on the left of a ~ operator, and the independent variables on the right. The response must be a survival object as returned by the Surv function. On the right of the ~ operator, a | operator must be used: on the left of the | operator, users specify independent variables that will be used in the linear regression function for $\ln y_0$ in the threshold regression model; on the right of the | operator, users specify independent variables that will be used in the linear regression function for μ in the threshold regression model. If users just want to use a constant $\ln y_0$ or μ, he or she can put 0 or 1 as a placeholder on the left or right of the | operator, instead of listing the independent variables for $\ln y_0$ or μ.

`data` input dataset. Such dataset must be a survival dataset including at least the survival time variable and censoring variable. For the censoring variable, 1 should be used to indicate the subjects with failure observed, and 0 should be used to indicate the subjects that are right censored. The dataset can also include other independent variables that will be used in the threshold regression model.

`hr` specifies the categorical variable for the calculation of hazard ratios. Such categorical variable must be a factor variable.

`scenario` specifies a scenario where the hazard ratios are calculated.

`timevalue` specifies a value of time at which the hazard ratios are calculated.

`graph` specifies the type of curves to be generated. The “hz” option is to plot hazard function curves, the “sv” option is to plot survival function curves, and the “ds” option is to plot density function curves.

`nolegend` set the “nolegend” argument to 1 if users do not want the “threg” package to generate legends for the picture. Note that even if “nolegend” is set to 1, users can still generate legends by themselves after the picture is generated, by using the “legend” function in R.

`nocolor` set the “nocolor” argument to 1 if users want to depict all curves in black.

Details

Threshold regression is a recently developed regression methodology to analyze time to event data. For a review of this regression model, see Lee and Whitmore (2006, 2010). A unique feature of threshold regression is that the event time is considered as the time when an underlying stochastic process first hits a boundary threshold. In the context of survival data, for example, the event can be death. The death time of an individual is considered as the time when his/her latent health status decreases to the zero boundary.

In the `threg` package, a Wiener process $Y(t)$ is used to model the latent health status process. An event is observed when $Y(t)$ reaches 0 for the first time. Three parameters of the Wiener process are involved: μ, y_0 and σ. Parameter μ, called the drift of the Wiener process, is the rate per unit time at which the level of the sample path is changing. The sample path approaches the threshold if $\mu < 0$. Parameter y_0 is the initial value of the process and is taken as positive. Parameter σ represents the variability per unit time of the process (Lee and Whitmore 2006). The first hitting time (FHT) of a Wiener process with μ, y_0 and σ is an inverse Gaussian distribution with probability density function (p.d.f):

$$f(t|\mu, \sigma^2, y_0) = \frac{y_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(y_0+\mu t)^2}{2\sigma^2 t}\right],$$

where $-\infty < \mu < \infty, \sigma^2 > 0$, and $y_0 > 0$. The p.d.f. is proper if $\mu \leq 0$. The cumulative distribution function of the FHT is:

$$F(t|\mu, \sigma^2, y_0) = \Phi\left[-\frac{(y_0+\mu t)^2}{\sqrt{\sigma^2 t}}\right] + \exp\left(-\frac{2y_0\mu}{\sigma^2}\right)\Phi\left[\frac{\mu t - y_0}{\sqrt{\sigma^2 t}}\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Note that if $\mu > 0$, the Wiener process may never hit the boundary at zero and hence there is a probability that the FHT is ∞, that is, $P(FHT = \infty) = 1 - \exp(-2y_0\mu/\sigma^2)$.

Since the health status process is usually latent (i.e., unobserved), an arbitrary unit can be used to measure such a process. Hence the variance parameter σ^2 of the process is set to 1 in the `threg` package to fix the measurement unit of the process. Then we can regress the other two process parameters, y_0 and μ on covariate data. We assume that μ and $\ln(y_0)$ are linear in regression coefficients.

Suppose that the covariate vector is $\boldsymbol{Z'} = (1, Z_1, \cdots, Z_k)$, where $Z_1, \cdots, Z_k$ are covariates and the leading 1 in $\boldsymbol{Z'}$ allows for a constant term in the regression relationship. Then $\ln(y_0)$ can be linked to the covariates with the following regression form:

$$\ln(y_0) = \gamma_0 + \gamma_1 Z_1 + \cdots + \gamma_k Z_k = \boldsymbol{Z'\gamma}$$

and μ can be linked to the covariates with the following regression form:

$$\mu = \beta_0 + \beta_1 Z_1 + \cdots + \beta_k Z_k = \boldsymbol{Z'\beta}$$

Vectors γ and β are regression coefficients for $\ln(y_0)$ and μ, respectively, with $\gamma' = (\gamma_0, \cdots, \gamma_k)$ and $\beta' = (\beta_0, \cdots, \beta_k)$. Note that researchers can set some elements in γ or β to zero if they feel the corresponding covariates are not important in predicting $\ln(y_0)$ or μ. For example, if covariate Z_1 in the vector Z' is considered not important to predict $\ln(y_0)$, we can remove the Z_1 term by setting γ_1 to zero.

Author(s)

Tao Xiao

Maintainer: Tao Xiao <taoxiao1@gmail.com>

References

Lee, M-L. T., Whitmore, G. A. (2010) Proportional hazards and threshold regression: their theoretical and practical connections., *Lifetime Data Analysis* 16, 2: 196-214.

Lee, M-L. T., Whitmore, G. A. (2006) Threshold regression for survival analysis: modeling event times by a stochastic process, *Statistical Science* 21: 501-513.

Klein, J. P., Moeschberger, M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. 2 edition. Springer-Verlag New York, Inc.

Garrett, J. M. (1997) Odds ratios and confidence intervals for logistic regression models with effect modification. *Stata Technical Bulletin*, 36, 1522. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 104-114. College Station, TX: Stata Press.

Examples

```
#load the data "lkr"
data("lkr")

#Transform the "treatment2" variable into factor variable "f.treatment2" .
lkr$f.treatment2=factor(lkr$treatment2)

#fit the threshold regression model on the factor variable "f.treatment2",
#calculate the hazard ratio of the drug B group v.s. the standard group at
#week 20 (this hazard ratio is calculated as 0.12),
```

```
#and generate the predicted survival curves for the drug B group and
#the standard group.
threg(Surv(weeks, relapse)~ f.treatment2|f.treatment2,data = lkr,
      hr=f.treatment2,timevalue=20,graph=sv,
      nolegend=1,nocolor=1)
legend(20, 1, c("Standard","Drug B"), lty = 1:2)

#fit the threshold regression model on the factor variable "f.treatment2",
#and calculate the hazard ratio of the drug B group v.s. the standard group at
#week 5 (this hazard ratio is calculated as 2.08).
threg(Surv(weeks, relapse)~ f.treatment2|f.treatment2,data = lkr,
      hr=f.treatment2,timevalue=5)



#As a comparison, fit the Cox proportion hazards model on "f.treatment2",
#and the Cox model gives a constant hazard ratio, 0.73.
summary(coxph(Surv(weeks, relapse) ~ f.treatment2, data = lkr))


#load the data "bmt"
data("bmt")

#Transform the "group" and "fab" variables into factor variables
#"f.group" and "f.fab".
bmt$f.group=factor(bmt$group)
bmt$f.fab=factor(bmt$fab)

#fit a threshold regression model on the "bmt" dataset, by using "recipient_age" and
#"f.fab" as the predictors for ln(y0), and "f.group" and "f.fab" as predictors for mu.
threg(Surv(time, indicator)~ recipient_age+f.fab|f.group+f.fab, data = bmt)

#fit the same model as above, but additionally calculate the hazard ratio for
#"f.group" for the specified scenario that "the patient age is 18 years old and
#the FAB classification is 0", at the time ``500 days''.
threg(Surv(time, indicator)~ recipient_age+f.fab|f.group+f.fab, data = bmt,
hr=f.group,timevalue=500,scenario=recipient_age(18)+f.fab(0))



#fit the same model as above, but additionally overlay curves of survival functions
#corresponding to different levels of "f.group'.
threg(Surv(time, indicator)~ recipient_age+f.fab|f.group+f.fab, data = bmt,
hr=f.group,timevalue=500,scenario=recipient_age(18)+f.fab(0),graph=sv,nocolor=1)


#fit the same model as above, but additionally overlay curves of hazard functions
#corresponding to different levels of "f.group'.
threg(Surv(time, indicator)~ recipient_age+f.fab|f.group+f.fab, data = bmt,
hr=f.group,timevalue=500,scenario=recipient_age(18)+f.fab(0),graph=hz,nocolor=1)

#fit the same model as above, but additionally overlay curves of probability density
#functions corresponding to different levels of "f.group'.
```

```
threg(Surv(time, indicator)~ recipient_age+f.fab|f.group+f.fab, data = bmt,
hr=f.group,timevalue=500,scenario=recipient_age(18)+f.fab(0),graph=ds,nocolor=1)
```

Index