Package 'landmulti'

July 22, 2025

Type Package

Title Landmark Prediction with Multiple Short-Term Events

Version 0.5.0

Maintainer Wen Li <wen.li@uth.tmc.edu>

Description Contains functions for a flexible varying-coefficient landmark model by incorporating multiple short-term events into the prediction of long-term survival probability. For more information about landmark prediction please see Li, W., Ning, J., Zhang, J., Li, Z., Savitz, S.I., Tahanan, A., Rahbar.M.H., (2023+). ``Enhancing Long-term Survival Prediction with Multiple Short-term Events: Landmarking with A Flexible Varying Coefficient Model''.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 4.0.0)

Imports survival, landpred, NMOF, emdbook, snow

RoxygenNote 7.2.3

Suggests knitr, rmarkdown

VignetteBuilder knitr

Collate 'HelperFunctions.R' 'dat.R' 'globals.R' 'multipredict.R'

NeedsCompilation no

Author Wen Li [aut, cre],
Qian Wang [aut]

Repository CRAN

Date/Publication 2023-11-10 07:20:05 UTC

Contents

1

---

multipredict   *Landmark prediction with multiple short-term events*

Description

Landmark prediction with multiple short-term events

Usage

```
multipredict(
  data,
  formula,
  t0,
  L,
  SE = FALSE,
  SE.gs = FALSE,
  s1_beta1 = NULL,
  s2_beta2 = NULL,
  s1s2_beta3 = NULL,
  grid1 = seq(0.01, 5, length.out = 20),
  grid2 = seq(0.01, 5, length.out = 20),
  grid3 = list(seq(0.01, 5, length.out = 20), seq(0.01, 5, length.out = 20)),
  folds.grid = 8,
  reps.grid = 3,
  c01 = 0.1,
  c02 = 0.1,
  c03 = 0.05,
  B = 500,
  gs.method = "loop",
  gs.cl = NULL,
  gs.seed = NULL
)
```

Arguments

data
: Input dataset

formula
: a `formula` object, with a `Surv()` object, such as `Surv(time, event)`, on the left of a ~ operator, and the terms on the right. On the right-hand-side, the time to the occurrence of short-term event 1 and short-term event 2 should be called by statement `s1()` and `s2()`, respectively. The details of model specification are given under 'Details'

t0
: Landmark time

L
: Length of time into the future (starting from the landmark time) for which we want to make a risk prediction. This is called the 'prediction horizon' in the dynamic prediction literature

SE	Logical. 'True' if user wants to estimate SE for the coefficient using the perturbation-resampling method
SE.gs	Logical. 'True' if user wants to conduct grid search for the bandwidth in each perturbation. It is expected to give more accurate results but will consume longer time. 'False' if user wants to use the same bandwidth found in the point estimation for all perturbations
s1_beta1	A scalar or a vector. Time to the occurrence of short-term event 1 for the estimation of the regression coefficient beta1 in group 2. If a Null is given, then the coefficients for group 2 will NOT be estimated
s2_beta2	A scalar or a vector. Time to the occurrence of short-term event 2 for the estimation of the regression coefficient beta2 in group 3. If a Null is given, then the coefficients for group 3 will NOT be estimated
s1s2_beta3	A matrix or a dataframe with two columns. The first column should be s1 and the second should be s2. Time to the occurrence of short-term event 1 & 2 for the estimation of the regression coefficient beta3 in group 4. If a Null is given, then the coefficients for group 4 will NOT be estimated.
grid1	A prespecified grid for bandwidth search for group2
grid2	A prespecified grid for bandwidth search for group3
grid3	A list with prespecified grids for bandwidth search for group4
folds.grid	The number of folds in cross-validation
reps.grid	The number of repetitions of cross-validation
c01	A constant to shrink the bandwidth for group2
c02	A constant to shrink the bandwidth for group3
c03	A constant to shrink the bandwidth for group4
B	Number of perturbations for estimating SE
gs.method	Method used by gridsearch. Default is 'loop'. Use 'snow' will implement parallel computing and will speed up the calculation
gs.cl	Default is Null. Number of clusters used in parallel computing in gridsearch. Specify when gs.method = 'snow'
gs.seed	An integer to set the seed for parallel computing to ensure reproducible outcome, or 'NULL' if not to set reproducible outcome

Details

The multipredict function fits time-fixed model and univariate/bivariate varying-coefficient models using the data from subgroups formed based on the information on the short-term outcomes (such as HF hospitalization and CHD hospitalization) before landmark time t0, among those who haven't experienced the long-term outcome (such as death) at t0. In this way the short-term outcome information are incorporated into the prediction of long-term survival outcomes, and the risk prediction can vary based on the event times of the short-term outcomes.

The +s1() statement specified the column that determines the occurrence time of the first short-term outcome. The +s2() statement specified the column that determines the occurrence time of the second short-term outcome.

User may set the statement gs.method = 'True'.

By default the regression coefficients for group 1 is calculated in each run of this function.

Currently, parameter estimates from parallel computing are slightly different in each run because of the different (uncontrolled) random numbers used in the estimation. This will be solved in the near future.

Value

returns estimated coefficients for each short-term outcome and the long-term outcome:

coefficients   A named vector of the estimated regression coefficients

SE   The standard error of coefficients estimated by perturbation resampling

Author(s)

Wen Li, Qian Wang

References

Li, Wen. (2023), "Landmarking Using A Flexible Varying Coefficient Model to Improve Prediction Accuracy of Long-term Survival Following Multiple Short-term Events An Application to the Atherosclerosis Risk in Communities (ARIC) Study", *Statistics in Medicine* **90**(7) 1-29. doi:10.18637/jss.v090.i07

Parast, Layla, Su-Chun Cheng, and Tianxi Cai. (2012), "Landmark Prediction of Long Term Survival Incorporating Short Term Event Time Information", *J Am Stat Assoc* **107**(500) 1492-1501. doi: 10.1080/01621459.2012.721281

"Incorporating short-term outcome information to predict long-term survival with discrete markers". *Biometrical Journal* **53.2** (2011): 294-307. doi: 10.1080/01621459.2012.721281

Examples

```
library(survival)
library(emdbook)
library(NMOF)
library(landpred)
library(snow)
set.seed(1234)
res <- multipredict(data = simulation, formula = Surv(time, outcome) ~ age + s1(st1) + s2(st2),
                t0 = 5, L = 20, SE = FALSE,
                gs.method = "loop", gs.cl = 2, SE.gs = FALSE, B = 200, gs.seed = 100,
                s1_beta1 = 1.5, grid1 = seq(0.01, 5, length.out=20),
                s2_beta2 = 1.5, grid2 = seq(0.01, 5, length.out=20),
                s1s2_beta3 = NULL, grid3=list(seq(0.01, 5, length.out=20),
                                                seq(0.01, 5, length.out=20)))
print(res)
```

---

s1 *Identifier of the first short-term outcome*

Description

Identifier of the first short-term outcome

Usage

```
s1(x)
```

Arguments

x   the first short-term outcome

Details

This is a helperfunction used by `multipredict()` to identify the first short-term outcome. If used directly it will only return the input. Therefore it should not be called directly and only be used in the 'formula' argument in `multipredict()` to identify the variable that is the first short-term outcome. See 'Examples' section of `multipredict()` for more details.

Value

returns the input if this function is used directly. See 'details' for more explanation

---

s2 *Identifier of the second short-term outcome*

Description

Identifier of the second short-term outcome

Usage

```
s2(x)
```

Arguments

x   the second short-term outcome

Details

This is a helperfunction used by `multipredict()` to identify the second short-term outcome. It should not be called directly. If used directly it will only return the input. Therefore it should not be called directly and only be used in the 'formula' argument in `multipredict()` to identify the variable that is the second short-term outcome. See 'example' section of `multipredict()` for more details.

Value

returns the input if this function is used directly. See 'details' for more explanation

---

simulation   *Simulated data for landmulti package*

Description

This is a simulated data used as an example and test for `multipredict()` function. The data is randomly generated by computer and has no practical meanning.

Usage

```
data(simulation)
```

Format

An object of class `data.frame` with 3000 rows and 6 columns.

Index