NLTC_POS

S_16_373

4/25/2022

Load the libraries

```
library(tidyverse)
library(survival)
library(ggfortify)
library(survminer)
library(rms)
library(dynpred)
library(MASS)
library(CPE)
library(scatterplot3d)
library(plotly)
```

Read the data sets

```
Home <- read.csv("~/Research/Research---ST426/Data/NewData/Merge/Home/Home.csv")
Overseas <- read.csv("~/Research/Research---ST426/Data/NewData/Merge/Overseas/Overseas.csv")
```

Merge two data sets

```
data_merge <- rbind(Home , Overseas)
```

Filter the data of NLTC Perera

```
bt12 <- data_merge %>% filter(grepl('NLTC Perera' , Striker))
```

1)Analysis for different batting position

```
bt12_ps <- bt12 %>% filter(bat_position == c(7,8))

NLTC_diff_pos <- survdiff(Surv(cum_balls , wicket) ~ bat_position, data = bt12_ps)
NLTC_diff_pos
```

```
## Call:
## survdiff(formula = Surv(cum_balls, wicket) ~ bat_position, data = bt12_ps)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## bat_position=7 313        6     9.71      1.42       7.7
## bat_position=8 123        6     2.29      6.00       7.7
##
##   Chisq= 7.7  on 1 degrees of freedom, p= 0.006
```

The p value is less than 0.05. Therefore there is a significant difference between the batting positions 7 and 8.

batting position 7

Filter the first bat_position

```
NLTC_P_7<- bt12 %>% filter(bat_position == 7)
```

Cox model

```
NLTC_P_7_cut <- survSplit(NLTC_P_7 , cut = 20 , end = "cum_balls" , event = "wicket" ,
                          start = "start" , id = "id")

NLTC_P_7_cut$hv1 <- NLTC_P_7_cut$cum_runs * (NLTC_P_7_cut$start < 20)
NLTC_P_7_cut$hv2 <- NLTC_P_7_cut$cum_runs * (NLTC_P_7_cut$start >= 20)


cox_NLTC_P_7_heav <- coxph(Surv(start,cum_balls , wicket) ~
                            hv1 +hv2 + cum_boundaries  + ridge(cum_dotBalls) +
                            cluster (id) , data = NLTC_P_7_cut)

summary(cox_NLTC_P_7_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     cum_boundaries + ridge(cum_dotBalls), data = NLTC_P_7_cut,
##     cluster = id)
##
##   n= 872, number of events= 24
##
##                    coef    se(coef) se2     Chisq DF p
## hv1                -0.2564 0.08617  0.07576  8.85 1  2.9e-03
## hv2                -0.1793 0.04943  0.05317 13.16 1  2.9e-04
```

2

```
## cum_boundaries        0.6313 0.29134  0.29115  4.69 1  3.0e-02
## ridge(cum_dotBalls) -0.2811 0.04625  0.05606 36.94 1  1.2e-09
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## hv1                     0.7739     1.2922    0.6536    0.9163
## hv2                     0.8359     1.1964    0.7587    0.9209
## cum_boundaries          1.8800     0.5319    1.0621    3.3277
## ridge(cum_dotBalls)     0.7550     1.3245    0.6896    0.8266
##
## Iterations: 2 outer, 10 Newton-Raphson
## Degrees of freedom for terms= 1.0 1.0 1.0 0.6
## Concordance= 0.916  (se = 0.02 )
## Likelihood ratio test= 67.97  on 3.56 df,   p=3e-14
```

Test pH assumptions

```
test_NLTC_7 <- cox.zph(cox_NLTC_P_7_heav)
test_NLTC_7
```

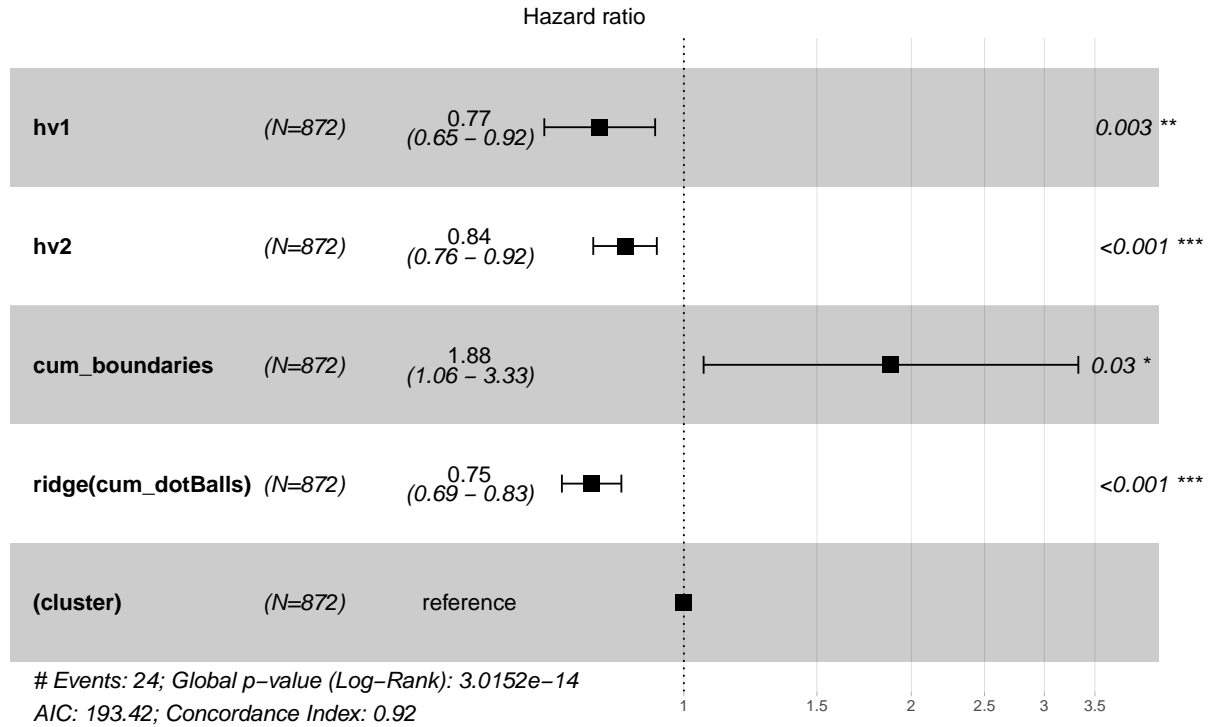```
chisq   df   p
hv1   0.64332 1.00 0.42
hv2   0.75295 0.97 0.37
cum_boundaries   1.02021 1.00 0.31
ridge(cum_dotBalls) 0.00187 0.60 0.86
GLOBAL   5.62764 3.56 0.18
```

**Plot the survival curve**

```
cox_plot_NLTC_P_7 <- ggadjustedcurves(cox_NLTC_P_7_heav , data = NLTC_P_7_cut , size = 2,
                    ggtheme =theme_survminer()) + geom_hline(yintercept = c(0.25,0.5,0.75) ,
                    linetype = "dashed") +ggtitle("Survival curve for NLTC Perera for the 7th
                    batting position") + theme(axis.title = element_text(size = 15),axis.text =
                    element_text(size = 14))
```

**coeffient plot**

```
coef_plot_NLTC_P_7 <- ggforest(cox_NLTC_P_7_heav , data = NLTC_P_7_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_P_7
```

Hazard ratio

hv1	*(N=872)*	0.77 *(0.65 – 0.92)*	■	*0.003* **
hv2	*(N=872)*	0.84 *(0.76 – 0.92)*	■	*<0.001* ***
cum_boundaries	*(N=872)*	1.88 *(1.06 – 3.33)*	■	*0.03* *
ridge(cum_dotBalls)	*(N=872)*	0.75 *(0.69 – 0.83)*	■	*<0.001* ***
(cluster)	*(N=872)*	reference	■	

 # Events: 24; Global p–value (Log–Rank): 3.0152e–14
AIC: 193.42; Concordance Index: 0.92

1   1.5   2   2.5   3   3.5

Area under the curve

```
AUC(Surv(start,cum_balls , wicket) ~
     hv1 +hv2 + cum_boundaries  + ridge(cum_dotBalls) + cluster (id) , data = NLTC_P_7_cut)
```
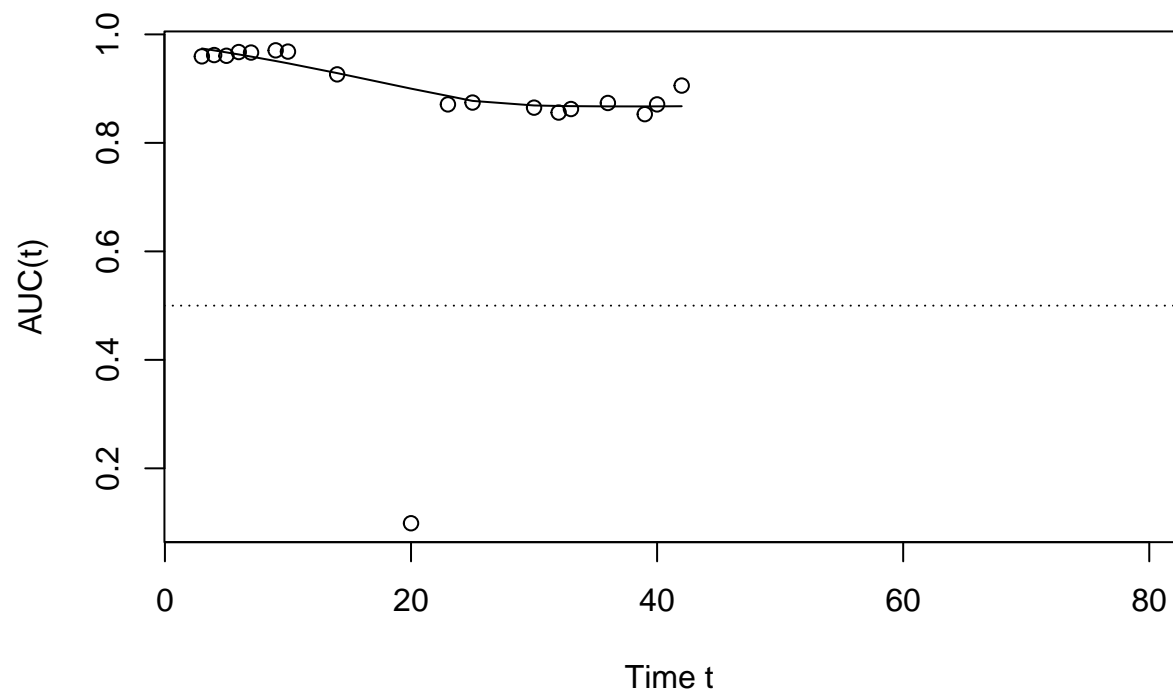
```
## $AUCt
##    time        AUC
## 1      3 0.95950920
## 2      4 0.96188056
## 3      5 0.96057819
## 4      6 0.96747967
## 5      7 0.96652720
## 6      9 0.97058824
## 7     10 0.96827795
## 8     14 0.92629816
## 9     20 0.09881423
## 10    23 0.87096774
## 11    25 0.87434555
## 12    30 0.86496350
## 13    32 0.85593220
## 14    33 0.86238532
## 15    36 0.87356322
## 16    39 0.85294118
## 17    40 0.87096774
## 18    42 0.90566038
## 19    79        NaN
##
## $AUC
## [1] 0.8879535
```

8th batting position

Filter the second bat_position

```
NLTC_P_8 <- bt12 %>% filter(bat_position == 8)
```

Cox model

```
NLTC_P_8_cut <- survSplit(NLTC_P_8 , cut = 20 , end = "cum_balls" , event = "wicket" ,
                          start = "start" , id = "id")

NLTC_P_8_cut$hv1 <- NLTC_P_8_cut$cum_runs * (NLTC_P_8_cut$start < 20)
NLTC_P_8_cut$hv2 <- NLTC_P_8_cut$cum_runs * (NLTC_P_8_cut$start >= 20)


cox_NLTC_P_8_heav <- coxph(Surv(start,cum_balls , wicket) ~
                           hv1 + hv2 +  cum_boundaries + ridge(cum_dotBalls) + cluster (id) ,
                           data = NLTC_P_8_cut)
summary(cox_NLTC_P_8_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     cum_boundaries + ridge(cum_dotBalls), data = NLTC_P_8_cut,
##     cluster = id)
##
##   n= 309, number of events= 15
##
##                       coef    se(coef) se2     Chisq DF p
## hv1                  -0.2211 0.09918  0.09076 4.97  1  0.026
## hv2                  -0.0801 0.08114  0.06085 0.97  1  0.320
## cum_boundaries       -0.7599 0.77963  0.46783 0.95  1  0.330
## ridge(cum_dotBalls)  -0.1701 0.11963  0.07934 2.02  1  0.160
##
##                       exp(coef) exp(-coef) lower .95 upper .95
## hv1                     0.8016    1.247     0.6600    0.9736
## hv2                     0.9230    1.083     0.7873    1.0821
## cum_boundaries          0.4677    2.138     0.1015    2.1558
## ridge(cum_dotBalls)     0.8436    1.185     0.6673    1.0665
##
## Iterations: 2 outer, 9 Newton-Raphson
## Degrees of freedom for terms= 1.0 0.8 0.9 0.5
## Concordance= 0.907  (se = 0.049 )
## Likelihood ratio test= 31.78  on 3.2 df,    p=8e-07
```

Test pH assumptions

```
test_NLTC_P_8 <- cox.zph(cox_NLTC_P_8_heav)
test_NLTC_P_8
```

```
##                       chisq   df     p
## hv1               3.018976 0.99 0.081
## hv2               0.257400 0.78 0.511
## cum_boundaries    1.803029 0.88 0.155
## ridge(cum_dotBalls) 0.000871 0.55 0.867
## GLOBAL            1.249272 3.20 0.773
```

Plot the survival curve

```
cox_plot_NLTC_P_8 <- ggadjustedcurves(cox_NLTC_P_8_heav , data = NLTC_P_8_cut , size = 2,ggtheme =
                      theme_survminer()) + geom_hline(yintercept = c(0.25,0.5,0.75) ,
                      linetype = "dashed") +ggtitle("Survival curve for NLTC Perera for
                      the 8th batting position") + theme(axis.title = element_text(size = 15),
                      axis.text = element_text(size = 14))
```
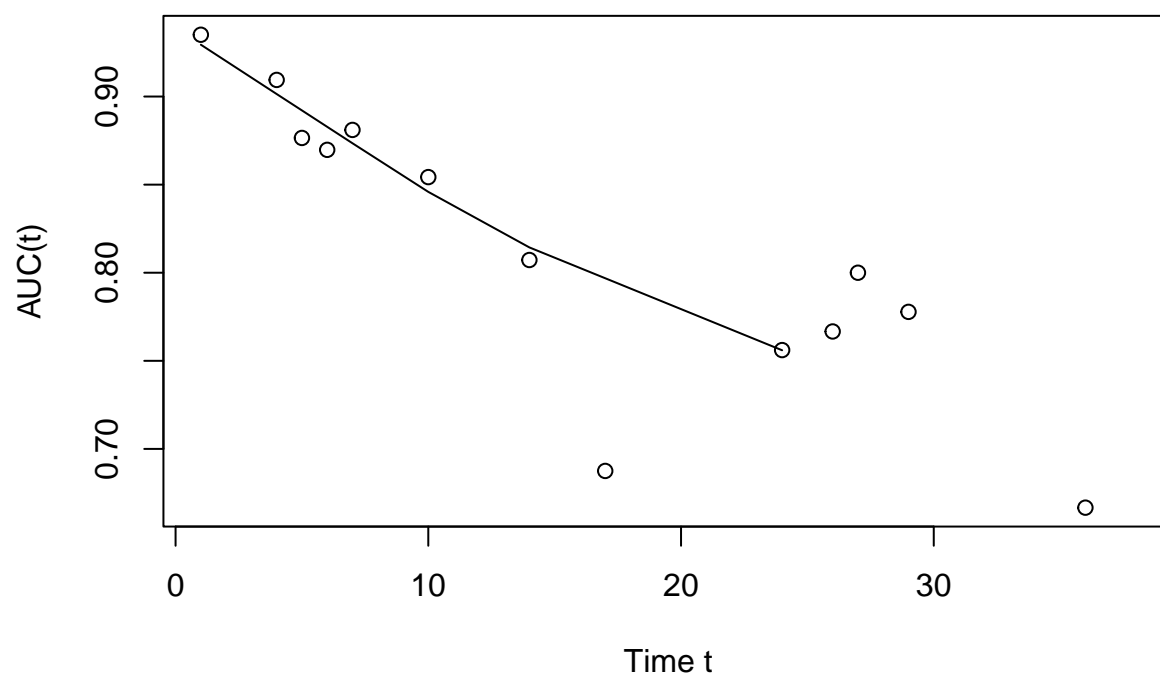
coeffient plot

```
coef_plot_NLTC_P_8 <- ggforest(cox_NLTC_P_8_heav , data = NLTC_P_8_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_P_8
```
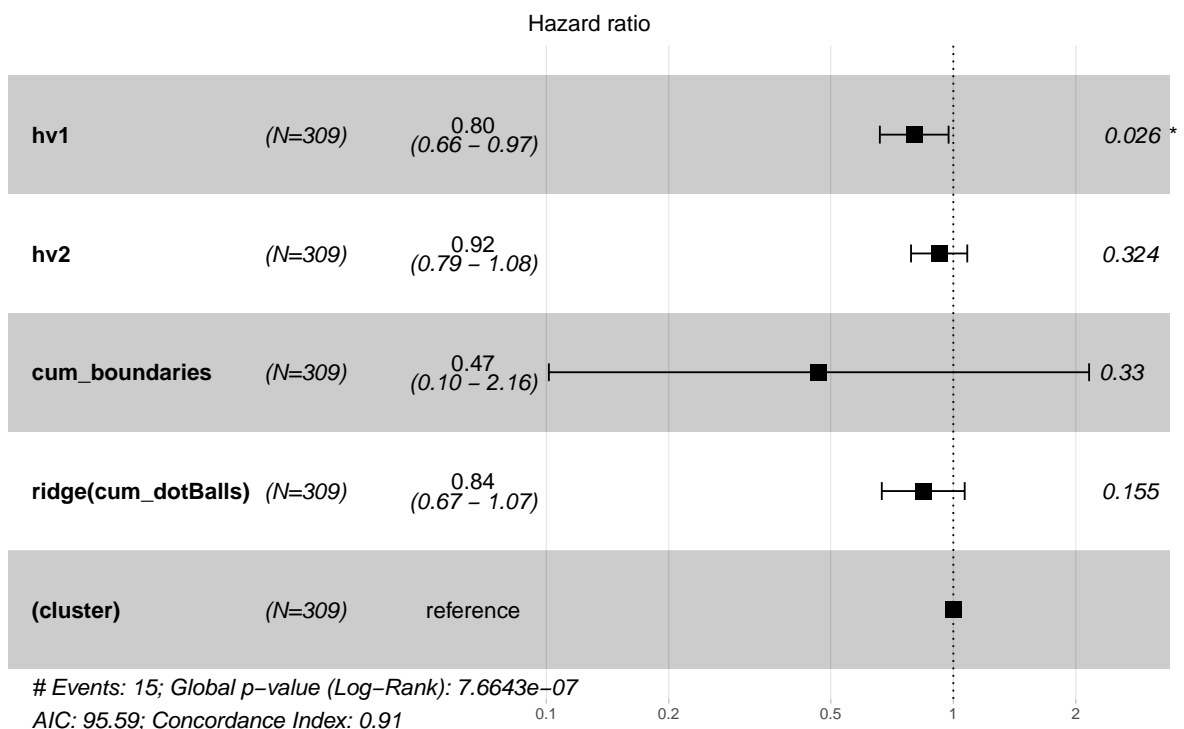

Area under the curve

```
AUC(Surv(start,cum_balls , wicket) ~
     hv1 + hv2 +  cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = NLTC_P_8_cut)
```
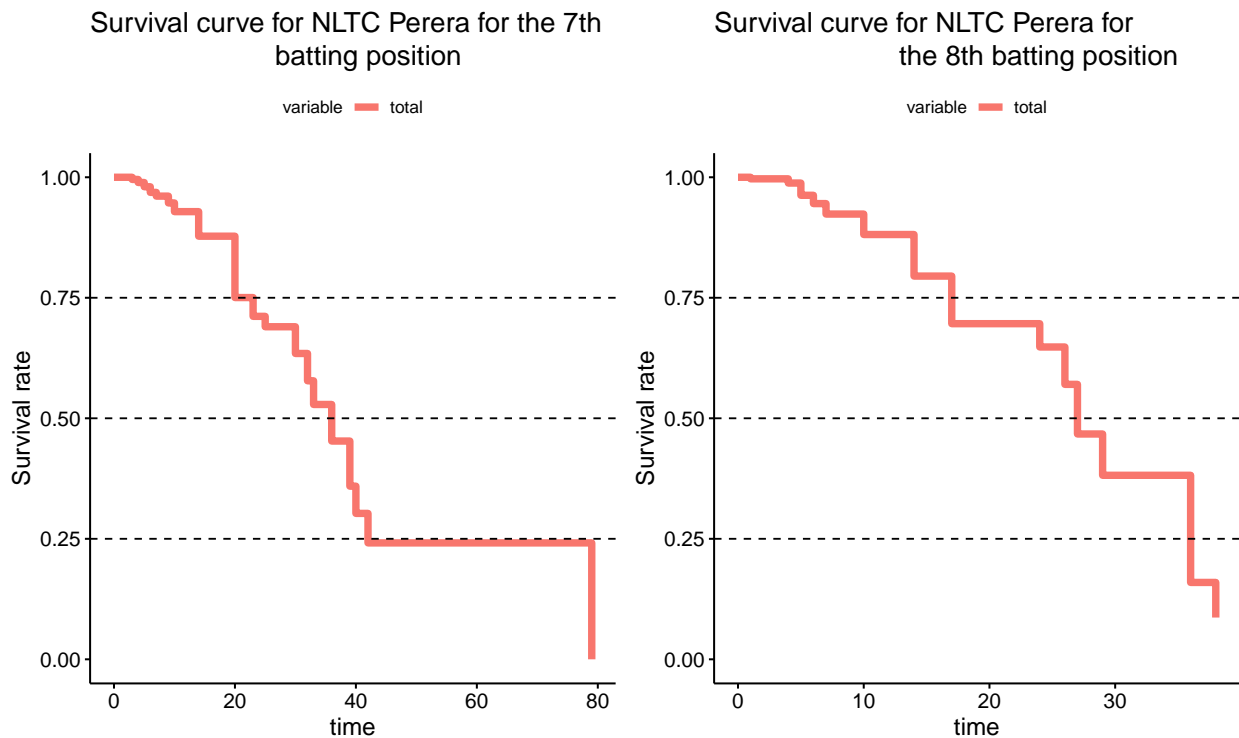


```
## $AUCt
##    time       AUC
## 1     1 0.9350649
## 2     4 0.9094340
## 3     5 0.8764940
## 4     6 0.8697479
## 5     7 0.8810573
## 6    10 0.8542714
## 7    14 0.8072289
## 8    17 0.6875000
## 9    24 0.7560976
## 10   26 0.7666667
## 11   27 0.8000000
## 12   29 0.7777778
## 13   36 0.6666667
## 14   38       NaN
##
## $AUC
## [1] 0.8610966
```

Survival curve

8

```
NLTCSurvPlotPos <- ggarrange(cox_plot_NLTC_P_7,cox_plot_NLTC_P_8,ncol = 2)
NLTCSurvPlotPos
```


Upload to texStudio

```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"NLTCSurvPlotPos.eps"),width = 10,height = 7,horizontal = FALSE)
NLTCSurvPlotPos
dev.off()
```
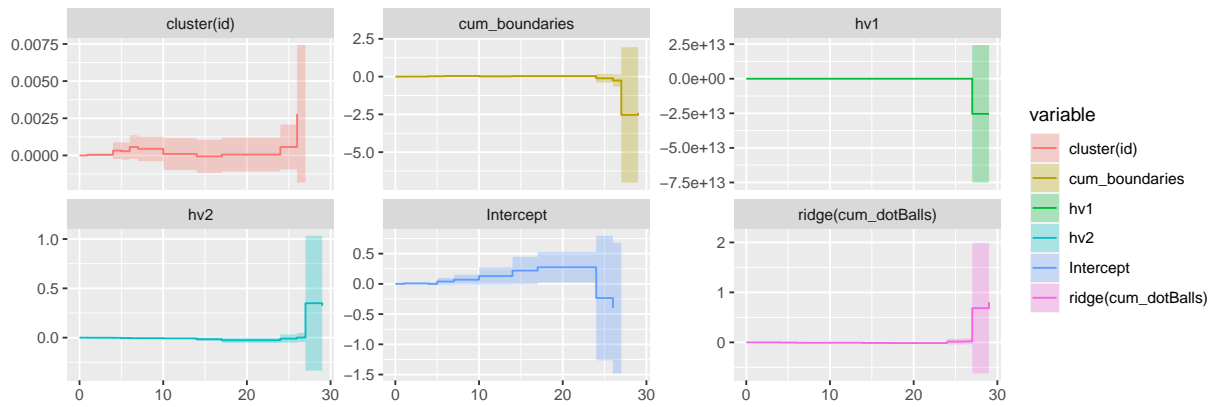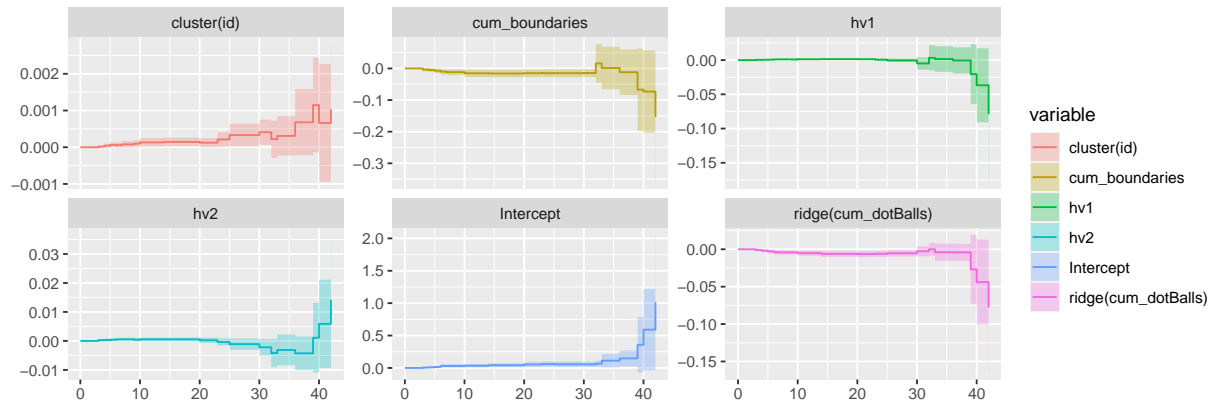
Aalen model

```
NLTC_p7 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
    hv1 + hv2 +  cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = NLTC_P_7_cut))

NLTC_p8 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
    hv1 + hv2 +  cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = NLTC_P_8_cut))

NLTCAalenPos <- ggarrange(NLTC_p7,NLTC_p8, nrow = 2)
NLTCAalenPos
```

```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"NLTCAalenPos.eps"),width = 10,height = 7,horizontal = FALSE)
NLTCAalenPos
dev.off()
```