NLTC_VEN

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 two different Venue

```
NLTC_diff_ven <- survdiff(Surv(cum_balls , wicket) ~ Venue, data = bt12)
NLTC_diff_ven
```

```
## Call:
## survdiff(formula = Surv(cum_balls, wicket) ~ Venue, data = bt12)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## Venue=Home      463       19     19.7    0.0257    0.0449
## Venue=Overseas 584       30     29.3    0.0173    0.0449
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.8
```

The p value is greater than 0.05. Therefore we cannot see a significant difference between the Venue.

Home

Filter the first Venue

```
NLTC_H <- bt12 %>% filter(Venue == "Home")
```

Cox model

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

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


cox_NLTC_H_heav <- coxph(Surv(start,cum_balls , wicket) ~
                           hv1 + hv2 +cum_singles+ridge(cum_dotBalls) + cluster (id) ,
                         data = NLTC_H_cut)

summary(cox_NLTC_H_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     cum_singles + ridge(cum_dotBalls), data = NLTC_H_cut, cluster = id)
##
##   n= 618, number of events= 19
##
##                       coef     se(coef) se2     Chisq DF p
## hv1                   -0.23871 0.04803  0.06712 24.70 1  6.7e-07
## hv2                   -0.08563 0.04450  0.04971  3.70 1  5.4e-02
## cum_singles           -0.84134 0.14679  0.13450 32.85 1  9.9e-09
## ridge(cum_dotBalls) -0.61303 0.07559  0.08963 65.77 1  5.1e-16
##
##                       exp(coef) exp(-coef) lower .95 upper .95
```

```
## hv1                    0.7876      1.270    0.7169    0.8654
## hv2                    0.9179      1.089    0.8413    1.0016
## cum_singles            0.4311      2.319    0.3233    0.5749
## ridge(cum_dotBalls)    0.5417      1.846    0.4671    0.6282
##
## Iterations: 2 outer, 9 Newton-Raphson
## Degrees of freedom for terms= 0.9 0.8 0.8 0.5
## Concordance= 0.98  (se = 0.008 )
## Likelihood ratio test= 94.55  on 3.04 df,   p=<2e-16
```

Test pH assumptions

```
test_NLTC_H <- cox.zph(cox_NLTC_H_heav)
test_NLTC_H
```

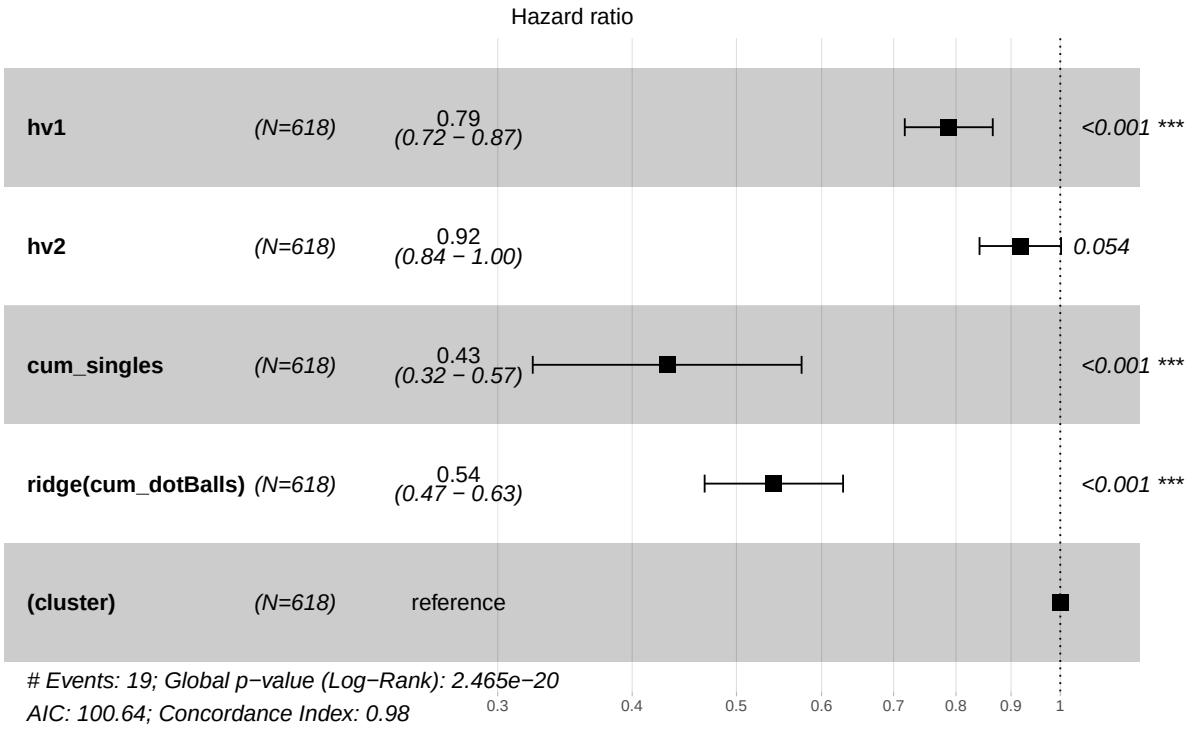
```
##                       chisq   df     p
## hv1                  0.8831 0.89 0.310
## hv2                  1.4383 0.81 0.180
## cum_singles          2.7542 0.83 0.076
## ridge(cum_dotBalls) 0.0195 0.51 0.659
## GLOBAL               5.3159 3.04 0.153
```

Plot the survival curve

```
cox_plot_NLTC_H <- ggadjustedcurves(cox_NLTC_H_heav , data = NLTC_H_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 Venue Home") +
                          theme(axis.title = element_text(size = 15),
                                axis.text = element_text(size = 14))
```

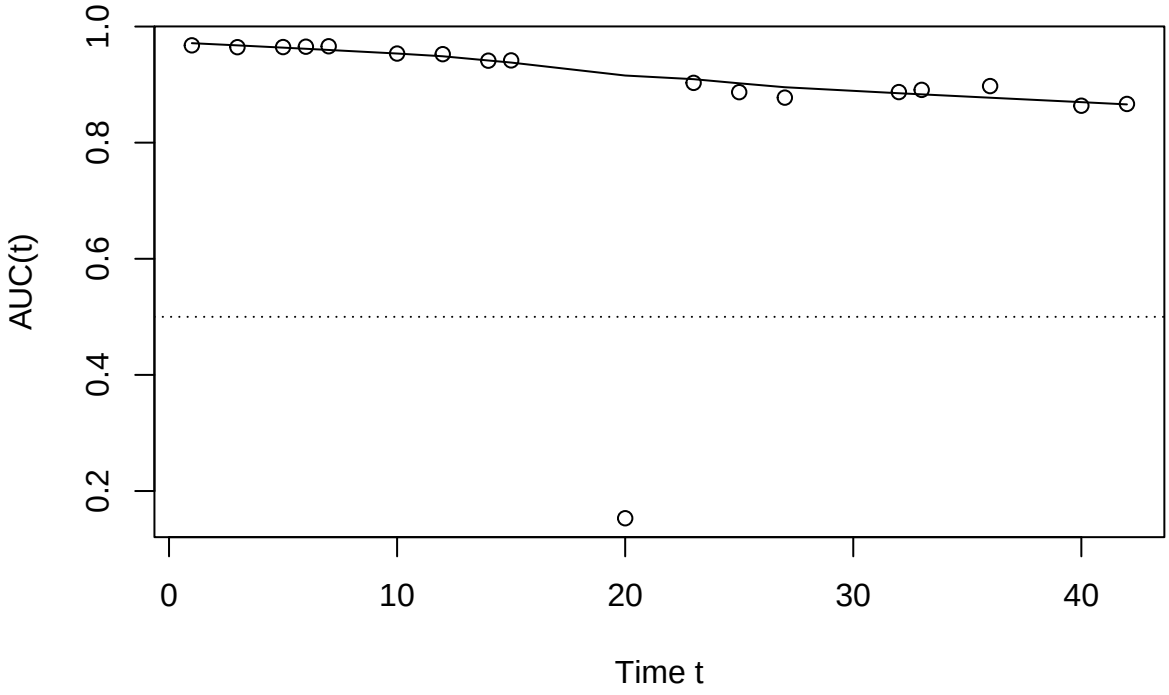
coeffient plot

```
coef_plot_NLTC_H <- ggforest(cox_NLTC_H_heav , data = NLTC_H_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_H
```



Area under the curve

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




```
## $AUCt
##    time       AUC
## 1     1 0.9675851
## 2     3 0.9644097
## 3     5 0.9646182
## 4     6 0.9652510
## 5     7 0.9660000
## 6    10 0.9534368
## 7    12 0.9523810
## 8    14 0.9411765
## 9    15 0.9416446
## 10   20 0.1531250
## 11   23 0.9029851
## 12   25 0.8869565
## 13   27 0.8775510
## 14   32 0.8870968
## 15   33 0.8909091
## 16   36 0.8974359
## 17   40 0.8636364
## 18   42 0.8666667
##
## $AUC
## [1] 0.9025157
```

Second Venue

Filter the second Venue

```
NLTC_O <- bt12 %>% filter(Venue == "Overseas")
```

Cox model

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

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


cox_NLTC_O_heav <- coxph(Surv(start,cum_balls , wicket) ~
                           hv1 + hv2 +cum_singles+ridge(cum_dotBalls) + cluster (id) ,
                         data = NLTC_O_cut)
summary(cox_NLTC_O_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     cum_singles + ridge(cum_dotBalls), data = NLTC_O_cut, cluster = id)
##
##   n= 786, number of events= 30
##
##                       coef    se(coef) se2     Chisq DF p
## hv1                  -0.1638 0.08559  0.06411  3.66 1  5.6e-02
## hv2                  -0.2018 0.04444  0.04841 20.62 1  5.6e-06
## cum_singles          -0.4085 0.11177  0.10508 13.36 1  2.6e-04
## ridge(cum_dotBalls) -0.3706 0.04950  0.05317 56.06 1  7.0e-14
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## hv1                     0.8489      1.178    0.7178    1.0040
## hv2                     0.8173      1.224    0.7491    0.8916
## cum_singles             0.6646      1.505    0.5339    0.8274
## ridge(cum_dotBalls)     0.6903      1.449    0.6265    0.7606
##
## Iterations: 2 outer, 11 Newton-Raphson
## Degrees of freedom for terms= 1.0 1.0 0.9 0.4
## Concordance= 0.966  (se = 0.007 )
## Likelihood ratio test= 120  on 3.33 df,   p=<2e-16
```

Test pH assumptions

```
test_NLTC_O <- cox.zph(cox_NLTC_O_heav)
test_NLTC_O
```

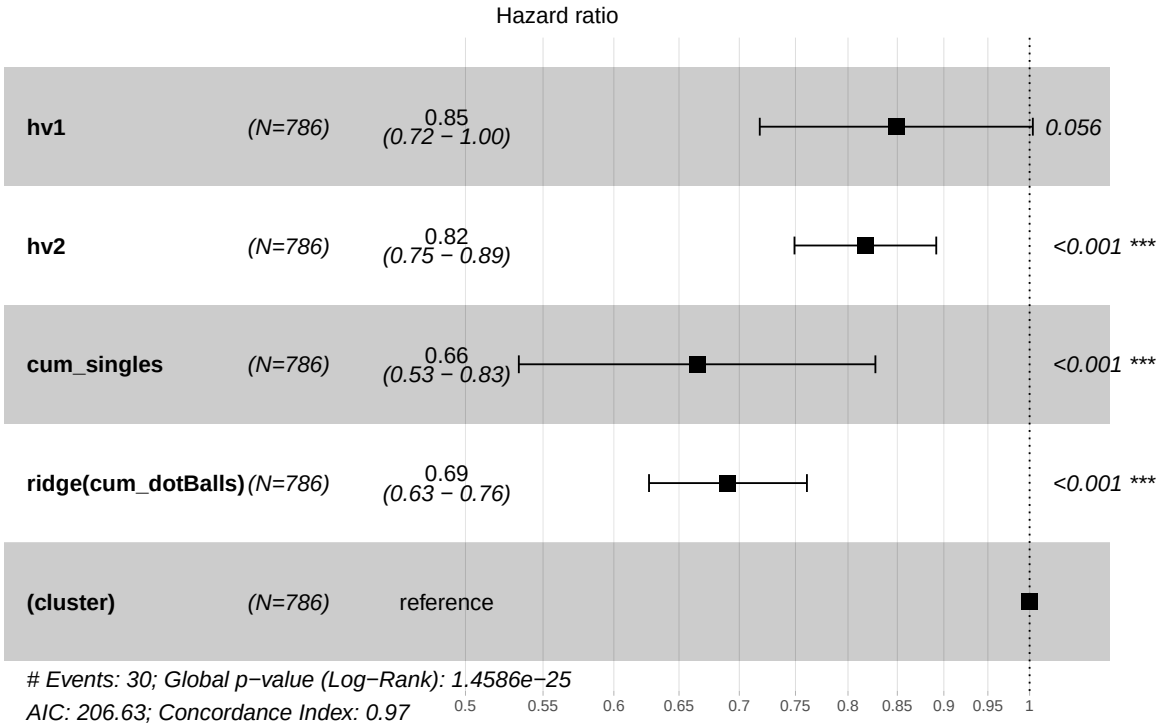
```
##                          chisq   df     p
## hv1                   1.36e+00 0.99 0.240
## hv2                   3.05e-02 0.98 0.854
## cum_singles           3.55e+00 0.92 0.053
## ridge(cum_dotBalls) 5.67e-05 0.44 0.892
## GLOBAL                3.23e+00 3.33 0.411
```

Plot the survival curve

```
cox_plot_NLTC_O <- ggadjustedcurves(cox_NLTC_O_heav , data = NLTC_O_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 Venue Overseas") +
                            theme(axis.title = element_text(size = 15),
                                  axis.text = element_text(size = 14))
```

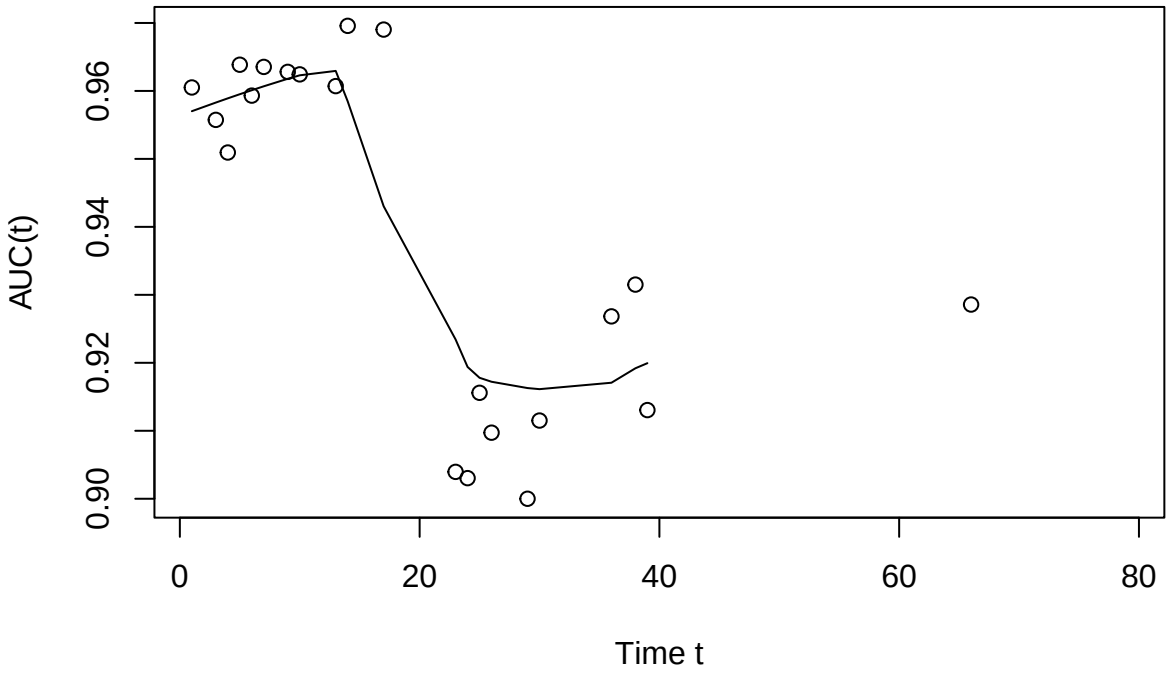
coeffient plot

```
coef_plot_NLTC_O <- ggforest(cox_NLTC_O_heav , data = NLTC_O_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_O
```



Area under the curve

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

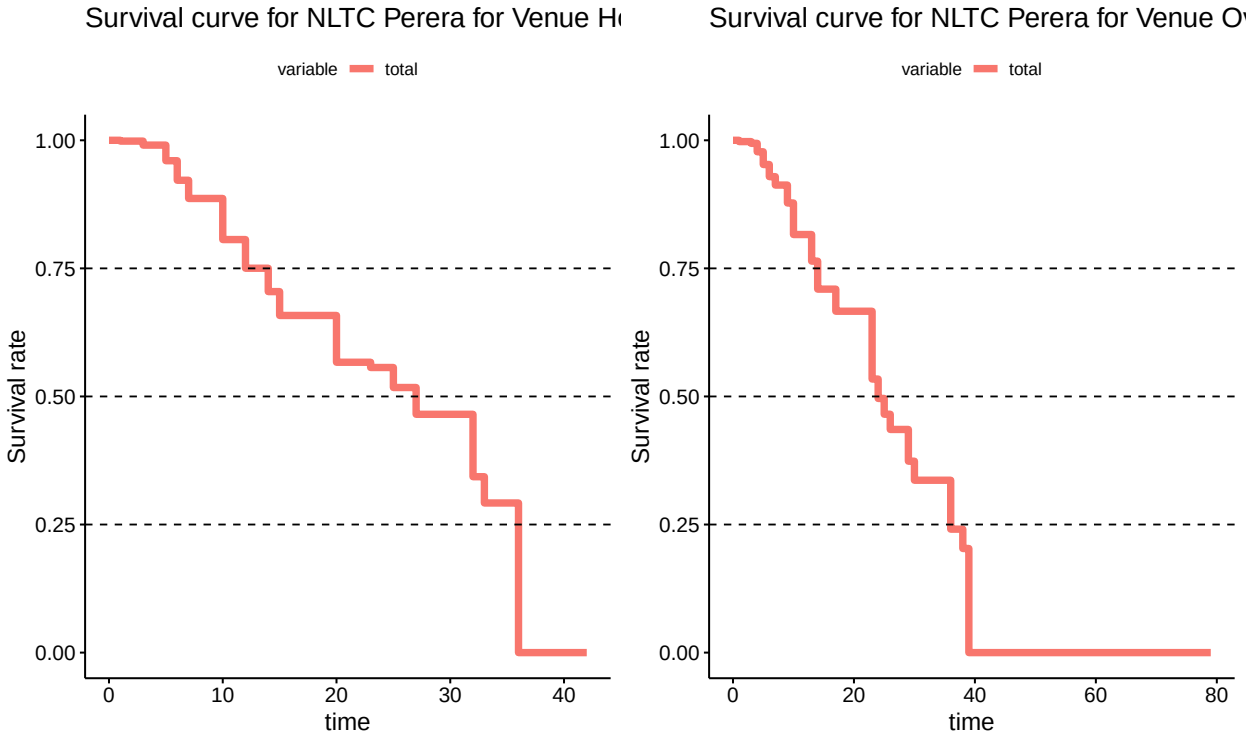



```
## $AUCt
##    time       AUC
## 1     1 0.9605096
## 2     3 0.9557400
## 3     4 0.9509380
## 4     5 0.9638554
## 5     6 0.9593114
## 6     7 0.9635332
## 7     9 0.9628028
## 8    10 0.9624329
## 9    13 0.9607073
## 10   14 0.9695740
## 11   17 0.9690265
## 12   23 0.9039548
## 13   24 0.9030303
## 14   25 0.9155844
## 15   26 0.9097222
## 16   29 0.9000000
## 17   30 0.9115044
## 18   36 0.9268293
## 19   38 0.9315068
## 20   39 0.9130435
```

```
## 21   66 0.9285714
## 22   79       NaN
##
## $AUC
## [1] 0.9539818
```

Survival curve

```
NLTCSurvPlotVen <- ggarrange(cox_plot_NLTC_H,cox_plot_NLTC_O,ncol = 2)
NLTCSurvPlotVen
```



Upload to texStudio

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

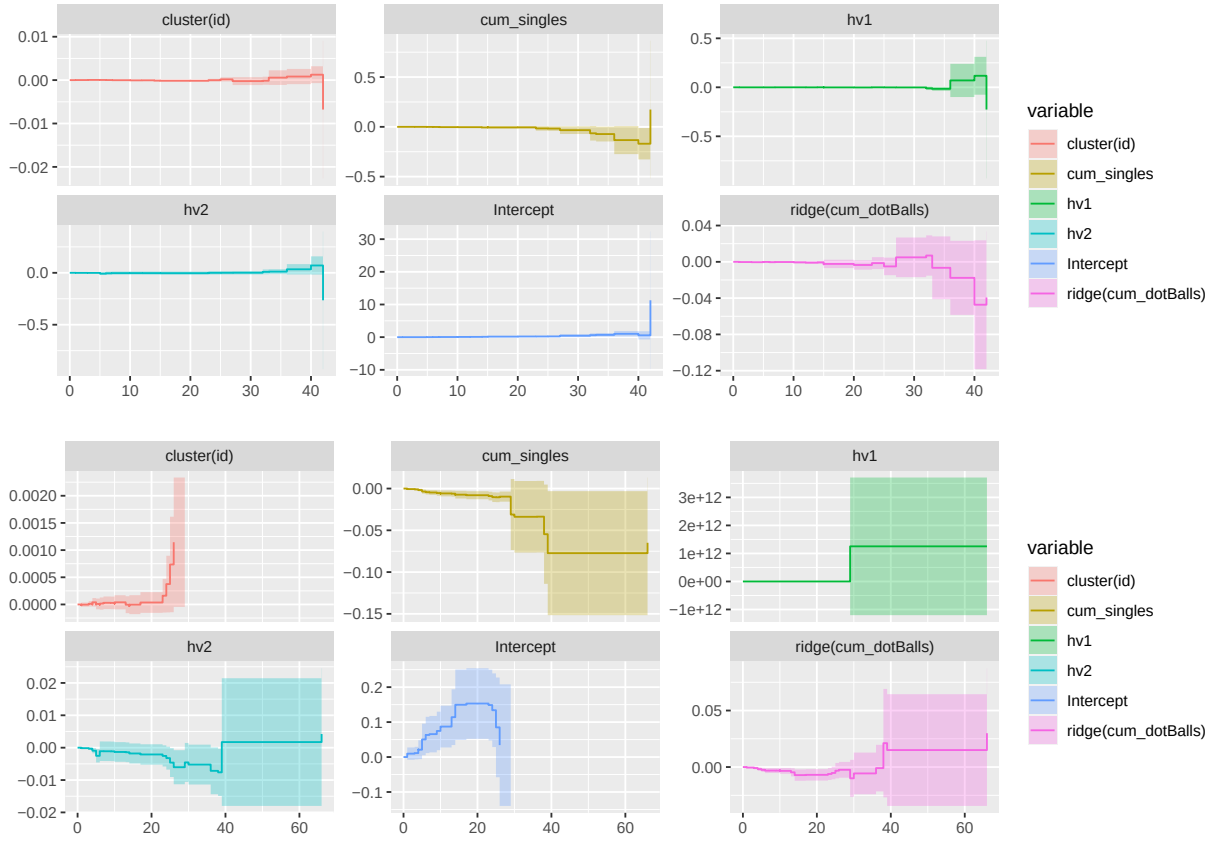
Aalen model

```
NLTC_pH <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
      hv1 + hv2 +cum_singles+ridge(cum_dotBalls) + cluster (id) , data = NLTC_H_cut))

NLTC_pO <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
```

```
      hv1 + hv2 +cum_singles+ridge(cum_dotBalls) + cluster (id) , data = NLTC_O_cut))

NLTCAalenVen <- ggarrange(NLTC_pH,NLTC_pO, nrow = 2)
NLTCAalenVen
```




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