Time-dependent Covariate with R: Recidivism*

Data from the `RcmdrPlugin.survival` package

A data frame with 432 observations on the following 62 variables.

- week: Week of first arrest after release, or censoring; all censored observations are censored at 52 weeks.
- arrest: 1 if arrested, 0 if not arrested.
- fin: Financial aid: 0=No, 1=Yes.
- age: Age in years at time of release.
- race: Black or other.
- wexp: Full-time work experience before incarceration: 0=No, 1=Yes.
- mar: marital status at time of release: married or not married.
- paro: Released on parole? 0=No, 1=Yes.
- prio: number of convictions prior to current incarceration.
- educ: level of education: 2 = 6th grade or less; 3 = 7th to 9th grade; 4 = 10th to 11th grade; 5 = 12th grade; 6 = Some college.
- emp1: Employment status in the first week after release: 0=No, 1=Yes.
- emp2: as above.

  .

  .

  .
- emp52: as above.

```
> rm(list=ls())
> library(survival)
> Rossi = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/Rossi.data.txt",
+                    header=TRUE)
> dim(Rossi)
[1] 432  63
>
```

* Copyright information is on the last page.

```
> Rossi[1:10,c(1:15,61,62)] # First 10 rows, and columns 1-16, 61, 62
   id week arrest fin age  race wexp        mar paro prio educ emp1 emp2 emp3 emp4
1   1   20      1   0  27 black    0 notmarried    1    3    3    0    0    0    0
2   2   17      1   0  18 black    0 notmarried    1    8    4    0    0    0    0
3   3   25      1   0  19 other    1 notmarried    1   13    3    0    0    0    0
4   4   52      0   1  23 black    1    married    1    1    5    0    0    0    0
5   5   52      0   0  19 other    1 notmarried    1    3    3    0    0    0    0
6   6   52      0   0  24 black    1 notmarried    0    2    4    0    0    0    0
7   7   23      1   0  25 black    1    married    1    0    4    1    1    1    1
8   8   52      0   1  21 black    1 notmarried    1    4    3    0    0    0    0
9   9   52      0   0  22 black    0 notmarried    0    6    3    0    0    0    0
10 10   52      0   0  20 black    1 notmarried    0    0    5    0    1    1    1
   emp50 emp51
1     NA    NA
2     NA    NA
3     NA    NA
4      1     1
5      0     0
6      0     0
7     NA    NA
8      0     0
9      1     1
10     0     0

> # Fix it up a bit within the data frame
> # Make education and other variables into factors (better for display)
> # Do emp1-emp52 in a loop
> Rossi = within(Rossi,
+ {
+ educ = factor(educ,labels = c("6th grade or less", # 2-6
+                                "7th to 9th grade",
+                                "10th to 11th grade",
+                                "12th grade",
+                                "Some college"))
+ # Convert 0-1 to No-Yes
+ arrest = factor(arrest,labels = c("No","Yes") )
+ fin = factor(fin,labels = c("No","Yes") )
+ wexp = factor(wexp,labels = c("No","Yes") )
+ paro = factor(paro,labels = c("No","Yes") )
+
+ for(j in 1:52)
+     {
+     num = as.character(j)
+     command = paste("emp",num, "=factor(emp",num,",labels=c('No','Yes'))",  sep="")
+     print(command,quote=F)
+     eval(parse(text=command))
+     } # End loop over j
+
+ rm(command,num,j) # Cleaning up
+
+ } # End modifications of the data frame
+ ) # End within statement

[1] emp1=factor(emp1,labels=c('No','Yes'))
[1] emp2=factor(emp2,labels=c('No','Yes'))
[1] emp3=factor(emp3,labels=c('No','Yes'))

      . . . skipping . . .

[1] emp50=factor(emp50,labels=c('No','Yes'))
[1] emp51=factor(emp51,labels=c('No','Yes'))
[1] emp52=factor(emp52,labels=c('No','Yes'))
>
```

```
> summary(Rossi)
```

```
       id              week          arrest     fin           age          race
 Min.   :  1.0   Min.   : 1.00   No :318   No :216   Min.   :17.0   black:379
 1st Qu.:108.8   1st Qu.:50.00   Yes:114   Yes:216   1st Qu.:20.0   other: 53
 Median :216.5   Median :52.00                       Median :23.0
 Mean   :216.5   Mean   :45.85                       Mean   :24.6
 3rd Qu.:324.2   3rd Qu.:52.00                       3rd Qu.:27.0
 Max.   :432.0   Max.   :52.00                       Max.   :44.0
  wexp            mar         paro          prio                      educ
 No :185   married   : 53   No :165   Min.   : 0.000   6th grade or less : 24
 Yes:247   notmarried:379   Yes:267   1st Qu.: 1.000   7th to 9th grade  :239
                                      Median : 2.000   10th to 11th grade:119
                                      Mean   : 2.984   12th grade        : 39
                                      3rd Qu.: 4.000   Some college      : 11
                                      Max.   :18.000
  emp1       emp2        emp3        emp4        emp5        emp6        emp7
 No :372   No  :317   No  :293   No  :274   No  :258   No  :250   No  :249
 Yes: 60   Yes :114   Yes :137   Yes :155   Yes :170   Yes :177   Yes :177
           NA's:  1   NA's:  2   NA's:  3   NA's:  4   NA's:  5   NA's:  6
```

. . . skipping . . .

```
  emp43       emp44       emp45       emp46       emp47       emp48       emp49
 No  :166   No  :164   No  :162   No  :160   No  :155   No  :154   No  :158
 Yes :179   Yes :177   Yes :177   Yes :177   Yes :178   Yes :178   Yes :172
 NA's: 87   NA's: 91   NA's: 93   NA's: 95   NA's: 99   NA's:100   NA's:102


  emp50       emp51       emp52
 No  :158   No  :156   No  :156
 Yes :167   Yes :166   Yes :166
 NA's:107   NA's:110   NA's:110
```

```
> # Same data in start-stop format
> recid = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/Rossi-ss.data.txt")

> dim(recid)
[1] 19809    13
```

```
> recid[c(1:37,242:293),] # Subjects 1, 2 and 8
```

	id	fin	age	race	wexp	mar	paro	prio	educ	tstart	tstop	emp	busted
1	1	0	27	black	0	notmarried	1	3	3	0	1	0	0
2	1	0	27	black	0	notmarried	1	3	3	1	2	0	0
3	1	0	27	black	0	notmarried	1	3	3	2	3	0	0
4	1	0	27	black	0	notmarried	1	3	3	3	4	0	0
5	1	0	27	black	0	notmarried	1	3	3	4	5	0	0
6	1	0	27	black	0	notmarried	1	3	3	5	6	0	0
7	1	0	27	black	0	notmarried	1	3	3	6	7	0	0
8	1	0	27	black	0	notmarried	1	3	3	7	8	0	0
9	1	0	27	black	0	notmarried	1	3	3	8	9	0	0
10	1	0	27	black	0	notmarried	1	3	3	9	10	0	0
11	1	0	27	black	0	notmarried	1	3	3	10	11	0	0
12	1	0	27	black	0	notmarried	1	3	3	11	12	0	0
13	1	0	27	black	0	notmarried	1	3	3	12	13	0	0
14	1	0	27	black	0	notmarried	1	3	3	13	14	0	0
15	1	0	27	black	0	notmarried	1	3	3	14	15	0	0
16	1	0	27	black	0	notmarried	1	3	3	15	16	0	0
17	1	0	27	black	0	notmarried	1	3	3	16	17	0	0
18	1	0	27	black	0	notmarried	1	3	3	17	18	0	0
19	1	0	27	black	0	notmarried	1	3	3	18	19	0	0
20	1	0	27	black	0	notmarried	1	3	3	19	20	0	1
21	2	0	18	black	0	notmarried	1	8	4	0	1	0	0
22	2	0	18	black	0	notmarried	1	8	4	1	2	0	0
23	2	0	18	black	0	notmarried	1	8	4	2	3	0	0
24	2	0	18	black	0	notmarried	1	8	4	3	4	0	0
25	2	0	18	black	0	notmarried	1	8	4	4	5	0	0
26	2	0	18	black	0	notmarried	1	8	4	5	6	0	0
27	2	0	18	black	0	notmarried	1	8	4	6	7	0	0
28	2	0	18	black	0	notmarried	1	8	4	7	8	0	0
29	2	0	18	black	0	notmarried	1	8	4	8	9	0	0
30	2	0	18	black	0	notmarried	1	8	4	9	10	0	0
31	2	0	18	black	0	notmarried	1	8	4	10	11	1	0
32	2	0	18	black	0	notmarried	1	8	4	11	12	1	0
33	2	0	18	black	0	notmarried	1	8	4	12	13	1	0
34	2	0	18	black	0	notmarried	1	8	4	13	14	1	0
35	2	0	18	black	0	notmarried	1	8	4	14	15	1	0
36	2	0	18	black	0	notmarried	1	8	4	15	16	0	0
37	2	0	18	black	0	notmarried	1	8	4	16	17	0	1
242	8	1	21	black	1	notmarried	1	4	3	0	1	0	0
243	8	1	21	black	1	notmarried	1	4	3	1	2	0	0
244	8	1	21	black	1	notmarried	1	4	3	2	3	0	0
245	8	1	21	black	1	notmarried	1	4	3	3	4	0	0
246	8	1	21	black	1	notmarried	1	4	3	4	5	0	0
247	8	1	21	black	1	notmarried	1	4	3	5	6	0	0
248	8	1	21	black	1	notmarried	1	4	3	6	7	0	0
249	8	1	21	black	1	notmarried	1	4	3	7	8	0	0
250	8	1	21	black	1	notmarried	1	4	3	8	9	0	0
251	8	1	21	black	1	notmarried	1	4	3	9	10	0	0
252	8	1	21	black	1	notmarried	1	4	3	10	11	0	0
253	8	1	21	black	1	notmarried	1	4	3	11	12	0	0
254	8	1	21	black	1	notmarried	1	4	3	12	13	0	0
255	8	1	21	black	1	notmarried	1	4	3	13	14	0	0
256	8	1	21	black	1	notmarried	1	4	3	14	15	0	0
257	8	1	21	black	1	notmarried	1	4	3	15	16	0	0
258	8	1	21	black	1	notmarried	1	4	3	16	17	0	0
259	8	1	21	black	1	notmarried	1	4	3	17	18	0	0
260	8	1	21	black	1	notmarried	1	4	3	18	19	0	0
261	8	1	21	black	1	notmarried	1	4	3	19	20	0	0
262	8	1	21	black	1	notmarried	1	4	3	20	21	1	0
263	8	1	21	black	1	notmarried	1	4	3	21	22	1	0
264	8	1	21	black	1	notmarried	1	4	3	22	23	0	0
265	8	1	21	black	1	notmarried	1	4	3	23	24	0	0

```
266   8    1   21 black     1 notmarried     1     4     3      24     25    1        0
267   8    1   21 black     1 notmarried     1     4     3      25     26    1        0
268   8    1   21 black     1 notmarried     1     4     3      26     27    0        0
269   8    1   21 black     1 notmarried     1     4     3      27     28    0        0
270   8    1   21 black     1 notmarried     1     4     3      28     29    0        0
271   8    1   21 black     1 notmarried     1     4     3      29     30    0        0
272   8    1   21 black     1 notmarried     1     4     3      30     31    0        0
273   8    1   21 black     1 notmarried     1     4     3      31     32    0        0
274   8    1   21 black     1 notmarried     1     4     3      32     33    0        0
275   8    1   21 black     1 notmarried     1     4     3      33     34    0        0
276   8    1   21 black     1 notmarried     1     4     3      34     35    0        0
277   8    1   21 black     1 notmarried     1     4     3      35     36    0        0
278   8    1   21 black     1 notmarried     1     4     3      36     37    0        0
279   8    1   21 black     1 notmarried     1     4     3      37     38    0        0
280   8    1   21 black     1 notmarried     1     4     3      38     39    0        0
281   8    1   21 black     1 notmarried     1     4     3      39     40    0        0
282   8    1   21 black     1 notmarried     1     4     3      40     41    0        0
283   8    1   21 black     1 notmarried     1     4     3      41     42    0        0
284   8    1   21 black     1 notmarried     1     4     3      42     43    0        0
285   8    1   21 black     1 notmarried     1     4     3      43     44    0        0
286   8    1   21 black     1 notmarried     1     4     3      44     45    0        0
287   8    1   21 black     1 notmarried     1     4     3      45     46    0        0
288   8    1   21 black     1 notmarried     1     4     3      46     47    0        0
289   8    1   21 black     1 notmarried     1     4     3      47     48    0        0
290   8    1   21 black     1 notmarried     1     4     3      48     49    0        0
291   8    1   21 black     1 notmarried     1     4     3      49     50    0        0
292   8    1   21 black     1 notmarried     1     4     3      50     51    0        0
293   8    1   21 black     1 notmarried     1     4     3      51     52    0        0
```

```
> # Make factors within the recid data frame
> recid = within(recid,
+ {
+ educ = factor(educ,labels = c("6th grade or less", # 2-6
+                                "7th to 9th grade",
+                                "10th to 11th grade",
+                                "12th grade",
+                                "Some college"))
+ # Convert 0-1 to No-Yes
+ fin = factor(fin,labels = c("No","Yes") )
+ wexp = factor(wexp,labels = c("No","Yes") )
+ paro = factor(paro,labels = c("No","Yes") )
+ emp = factor(emp, labels = c("No","Yes"))
+ } # End modifications of the data frame
+ ) # End within statement
>
> head(recid)
```

```
  id fin age  race wexp        mar paro prio             educ tstart tstop emp busted
1  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      0     1  No      0
2  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      1     2  No      0
3  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      2     3  No      0
4  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      3     4  No      0
5  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      4     5  No      0
6  1  No  27 black   No notmarried  Yes    3 7th to 9th grade      5     6  No      0
```

```
> jail = coxph(Surv(tstart,tstop,busted) ~ fin + age + race + wexp + mar + paro + prio
+ educ + emp, data=recid)
> summary(jail)

Call:
coxph(formula = Surv(tstart, tstop, busted) ~ fin + age + race +
    wexp + mar + paro + prio + educ + emp, data = recid)

  n= 19809, number of events= 114

                          coef exp(coef) se(coef)      z Pr(>|z|)
finYes                 -0.37607   0.68656  0.19240 -1.955 0.050631 .
age                    -0.04438   0.95659  0.02214 -2.005 0.045003 *
raceother              -0.36876   0.69159  0.31243 -1.180 0.237884
wexpYes                -0.04248   0.95841  0.21368 -0.199 0.842407
marnotmarried           0.33096   1.39230  0.38330  0.863 0.387886
paroYes                -0.08282   0.92052  0.19570 -0.423 0.672163
prio                    0.07477   1.07764  0.02971  2.517 0.011843 *
educ7th to 9th grade    0.47772   1.61239  0.52315  0.913 0.361163
educ10th to 11th grade  0.24206   1.27387  0.54627  0.443 0.657689
educ12th grade         -0.23286   0.79227  0.67648 -0.344 0.730681
educSome college       -0.59589   0.55107  1.12346 -0.530 0.595831
empYes                 -0.77817   0.45924  0.21864 -3.559 0.000372 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                       exp(coef) exp(-coef) lower .95 upper .95
finYes                    0.6866     1.4565   0.47087    1.0010
age                       0.9566     1.0454   0.91596    0.9990
raceother                 0.6916     1.4459   0.37489    1.2758
wexpYes                   0.9584     1.0434   0.63047    1.4569
marnotmarried             1.3923     0.7182   0.65686    2.9512
paroYes                   0.9205     1.0863   0.62727    1.3509
prio                      1.0776     0.9280   1.01668    1.1422
educ7th to 9th grade      1.6124     0.6202   0.57831    4.4955
educ10th to 11th grade    1.2739     0.7850   0.43665    3.7163
educ12th grade            0.7923     1.2622   0.21040    2.9833
educSome college          0.5511     1.8146   0.06094    4.9831
empYes                    0.4592     2.1775   0.29919    0.7049

Concordance= 0.689  (se = 0.027 )
Rsquare= 0.003   (max possible= 0.066 )
Likelihood ratio test= 52.59  on 12 df,   p=4.868e-07
Wald test            = 47.17  on 12 df,   p=4.353e-06
Score (logrank) test = 50.7  on 12 df,   p=1.053e-06
```

The test of significance for financial aid has huge political consequences.

Maybe say “For a released prisoner who got financial aid, the risk of re-arrest within one year is estimated to be between 47% as great and 100% as great (just the same), compared to a comparable prisoner who did not get financial aid.”

```
> # Test financial aid with a (partial) likelihood ratio test.
> # We are fishing, and that's questionable. On the other hand, likelihood ratio
> # tests are better in general.
>
> # There are no missing values in this data set, but in the real world that never
> # happens. With real data, always fit the restricted model on a data frame that
> # has no missing values for the full model,
>
> nomiss1 = na.omit(recid)
> dim(recid); dim(nomiss1)

[1] 19809    13
[1] 19809    13

> nohelp = update(jail, . ~ . - fin, data=nomiss1) # Remove var(s) being tested
> anova(nohelp,jail)

Analysis of Deviance Table
 Cox model: response is  Surv(tstart, tstop, busted)
 Model 1: ~ age + race + wexp + mar + paro + prio + educ + emp
 Model 2: ~ fin + age + race + wexp + mar + paro + prio + educ + emp
   loglik  Chisq Df P(>|Chi|)
1 -651.02
2 -649.08 3.8723  1   0.04909 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Moral dilemma

```
> # Test education
> noedu = update(jail, . ~ . - educ, data=nomiss1) # Remove var(s) being tested
> anova(noedu,jail)

Analysis of Deviance Table
 Cox model: response is  Surv(tstart, tstop, busted)
 Model 1: ~ fin + age + race + wexp + mar + paro + prio + emp
 Model 2: ~ fin + age + race + wexp + mar + paro + prio + educ + emp
   loglik  Chisq Df P(>|Chi|)
1 -651.57
2 -649.08 4.9762  4    0.2897

>
> # Wald test of education controlling for other variables.
> # function(L,Tn,Vn,h=0) # H0: L theta = h
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> beta_hat = jail$coefficients; Vn_hat = vcov(jail)
> length(beta_hat); dim(Vn_hat)
[1] 12
[1] 12 12
> LL = rbind(c(0,0,0,0,0,0,0,1,0,0,0,0),
+            c(0,0,0,0,0,0,0,0,1,0,0,0),
+            c(0,0,0,0,0,0,0,0,0,1,0,0),
+            c(0,0,0,0,0,0,0,0,0,0,1,0) )
> Wtest(LL,beta_hat,Vn_hat)

        W        df   p-value
4.2957530 4.0000000 0.3674572
```

```
> # Test financial aid two more ways, ignoring all other vars, and after dropping
> # the non-significant ones.
>
> summary(coxph(Surv(tstart,tstop,busted) ~ fin, data=recid))

Call:
coxph(formula = Surv(tstart, tstop, busted) ~ fin, data = recid)

  n= 19809, number of events= 114

          coef exp(coef) se(coef)      z Pr(>|z|)
finYes -0.3691    0.6914   0.1897 -1.945   0.0517 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

       exp(coef) exp(-coef) lower .95 upper .95
finYes    0.6914      1.446    0.4767     1.003

Concordance= 0.546  (se = 0.024 )
Rsquare= 0   (max possible= 0.066 )
Likelihood ratio test= 3.84  on 1 df,   p=0.05013
Wald test            = 3.78  on 1 df,   p=0.05174
Score (logrank) test = 3.83  on 1 df,   p=0.05042

>
> baltimore = coxph(Surv(tstart,tstop,busted) ~ fin + age + prio + emp, data=recid)
> summary(baltimore)

Call:
coxph(formula = Surv(tstart, tstop, busted) ~ fin + age + prio +
    emp, data = recid)

  n= 19809, number of events= 114

           coef exp(coef) se(coef)      z Pr(>|z|)
finYes -0.33801   0.71319  0.19014 -1.778 0.075454 .
age    -0.05570   0.94582  0.02073 -2.687 0.007210 **
prio    0.08841   1.09244  0.02760  3.203 0.001360 **
empYes -0.81403   0.44307  0.21632 -3.763 0.000168 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

       exp(coef) exp(-coef) lower .95 upper .95
finYes    0.7132     1.4022    0.4913     1.035
age       0.9458     1.0573    0.9082     0.985
prio      1.0924     0.9154    1.0349     1.153
empYes    0.4431     2.2570    0.2900     0.677

Concordance= 0.665  (se = 0.027 )
Rsquare= 0.002   (max possible= 0.066 )
Likelihood ratio test= 44.71  on 4 df,   p=4.558e-09
Wald test            = 40.82  on 4 df,   p=2.93e-08
Score (logrank) test = 43.56  on 4 df,   p=7.923e-09
```

```
>
> nomiss2 = na.omit(recid[,c(1,2,3,8,10:13)]); dim(nomiss2)

[1] 19809     8

> head(nomiss2) # Make sure start,stop,busted and id are included.

  id fin age prio tstart tstop emp busted
1  1  No  27    3      0     1  No      0
2  1  No  27    3      1     2  No      0
3  1  No  27    3      2     3  No      0
4  1  No  27    3      3     4  No      0
5  1  No  27    3      4     5  No      0
6  1  No  27    3      5     6  No      0

> nohelp2 = update(baltimore, . ~ . - fin, data=nomiss2)
> anova(nohelp2,baltimore)

Analysis of Deviance Table
 Cox model: response is  Surv(tstart, tstop, busted)
 Model 1: ~ age + prio + emp
 Model 2: ~ fin + age + prio + emp
   loglik  Chisq Df P(>|Chi|)
1 -654.62
2 -653.02 3.2019  1   0.07355 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```
