BIO 223 Applied Survival Analysis Chapter 7: Time-varying covariates

References

- Cox Proportional-Hazards Regression for Survival Data: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-cox-regression.pdf
- BIO 223 Applied Survival Analysis note chapter 7.
- Time-dependent covariates in the Cox proportional-hazards regression model.: http://www.ncbi.nlm.nih.gov/pubmed/10352854

Recidivism example: Create dataset

"The file Rossi.txt contains data from an experimental study of recidivism of 432 male prisoners, who were observed for a year after being released from prison (Rossi, Berk, and Lenihan, 1980)."

Use of time-varying variables require a counting process (long format) dataset in R. Thus, wide-to-long transformation is done below.

```
## Load survival package
library(survival)

## Load data from online
Rossi <- read.table(file = "http://cran.r-project.org/doc/contrib/Fox-Companion/Rossi.txt", header = TRUE)
head(Rossi)</pre>
```

| V Dmr | week a o12 | arrest | fin | age | race | wexp | mar | paro | pri | o educ | emp1 | emp2 | emp3 | emp4 | emp | os empe | emp7 | emp8 | emp9 e | emp10 | emp11 |
|----------|-----------------------------------|-----------------------------------|--------|-------|------|-------|------|-------|-------|--------|-------|-------|------|------|-----|---------|-------|-------|--------|-------|---------|
| 1 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 | 0 | 0 | 0 |) | 0 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 17 | 1 | 0 | 18 | 1 | 0 | 0 | 1 | 8 | 3 4 | 0 | 0 | 0 | 0 |) | 0 0 | 0 | 0 | 0 | 1 | 1 |
| 1 3 | 25 | 1 | 0 | 19 | 0 | 1 | 0 | 1 | 13 | 3 3 | 0 | 0 | 0 | 0 |) | 0 0 | 0 | 0 | 0 | 0 | 0 |
| 0 4 | 52 | 0 | 1 | 23 | 1 | 1 | 1 | 1 | _ | 1 5 | 0 | 0 | 0 | 0 |) | 1 1 | . 1 | 1 | 1 | 1 | 1 |
| 1 5 | 52 | 0 | 0 | 19 | 0 | 1 | 0 | 1 | 3 | 3 3 | 0 | 0 | 0 | 0 |) | 0 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 52 | 0 | 0 | 24 | 1 | | 0 | 0 | | 2 4 | | 0 | 0 | 0 | | 1 1 | 1 | 1 | 1 | 1 | 0 |
| 0 | | | | | | | | | | | | | | | | emn25 | emn26 | emn27 | emn28 | emn2 | 9 emp30 |
| emp | 031 | 0 | Cilipa | 0 | 0 | 0 | | 0 | 0 | 0 | NA | NA | | | NA | NA | NA | NA | | | |
| NA | 1 | 1 | | | | | | | | | | | | | | | | | | | |
| 2 NA | Τ | 1 | | 0 | 0 | 0 | | | NA | NA | NA | NA | | | NA | NA | NA | NA | | | |
| 3 NA | 0 | 0 | | 0 | 0 | 1 | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | NA | NA | | | |
| 4 0 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 0 | | 0 | 0 | 0 | 0 | 0 | (|) | 0 0 |
| 5 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | . 1 | - | 1 1 |
| 6 | 0 | 0 | | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | | 1 | 1 | 1 | 1 | 1 | . 1 | - | 1 0 |
| O (| emp32 050 | emp33 | emp3 | 84 ei | mp35 | emp36 | emp3 | 7 emp | р38 е | emp39 | emp40 | emp41 | emp4 | 2 em | p43 | emp44 | emp45 | emp46 | emp47 | emp4 | 8 emp49 |
| 1 | NA | NA | N | IA | NA | NA | N | Α | NA | NA | NA | NA | N | A | NA | NA | NA | NA | N.A | N N | A NA |
| NA 2 | NA | NA | N | IA | NA | NA | N | Α | NA | NA | NA | NA | N | Α | NA | NA | NA | NA | . NA | N N | A NA |
| NA 3 | NA | NA | N | IA | NA | NA | N | Α | NA | NA | NA | NA | N | Α | NA | NA | NA | NA | N.A | N N | A NA |
| NA 4 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | . 1 | - | 1 1 |
| 5 | 1 | 1 | | 1 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | C |) | 0 0 |
| 0 | 0 | 0 | | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | (|) | 0 0 |
| 0 | emp51 NA NA NA 1 0 | emp52 NA NA NA 1 0 | | | | | | | | | | | | | | | | | | | |

```
## Change to long format
Rossi.long <- reshape(data
                                                                             = Ross1,
                                                     varying = paste0("emp", 1:52),
                                                     v.names = "employed",
                                                                        = "time",
                                                     timevar
                                                     idvar = "id",
                                                     direction = "long",
                                                                         = "'i`
                                                      sep
## Sort by id and time
library(doBy)
Rossi.long < orderBy(\sim + id + time, data = Rossi.long)
## Drop rows where emp is NA (time after event/censoring)
Rossi.long <- Rossi.long[!is.na(Rossi.long$emp),]</pre>
## Load plyr package
library(plyr)
library(doMC)
registerDoMC()
                                                         # parallel backend to foreach/plyr
                                                         # Turn on multicore processing
## Create time variables and various forms of exposure variables
Rossi.long <- ddply(.data = Rossi.long,</pre>
                                                 .variables = c("id")
                                                 .drop = TRUE, .parallel = TRUE,
                                                 .fun = function(DF) {
                                                          ## Start time is the start of each interval
                                                          DF$start <- c(0, head(DF$time, -1))</pre>
                                                          ## Stop time is the end of each interval
                                                          DF$stop <- DF$time
                                                          ## Event indicator for each interval
                                                          DF$event <- 0
                                                          ## Use arrest value for the last interval
                                                          DF[nrow(DF),"event"] <- DF[nrow(DF),"arrest"]</pre>
                                                          ## Initial employment status
                                                          DF$employed.initial <- DF$employed[1]</pre>
                                                          ## Lagged employment status (employment status from last interval matters)
                                                          DF\temp\left\[overline{\text{op}}\] \[overline{\text{op}}\] \[overline{\text{
                                                          ## Cumulative number of weeks in employment
                                                          DF$employed.cumsum <- cumsum(DF$employed)</pre>
                                                          ## Ever employed status
                                                          DF$employed.ever <- as.numeric(DF$employed.cumsum > 0)
                                                          ## % of time in employment
                                                          DF$employed.percent <- with(DF, employed.cumsum / stop)*100
                                                          ## Return DF
                                                          DF
                                                })
```

Functional forms of exposure

```
## Data for subject 2
Rossi.long[Rossi.long$id == 2,
c("id","start","stop","event","employed","employed.initial","employed.lag1","employed.ever","employed.cumsum","employed
```

| | id | start | stop | event | employed | employed.initial | employed.lag1 | employed.ever | employed.cumsum | employed.percent |
|----|----|-------|------|-------|----------|------------------|---------------|---------------|-----------------|------------------|
| 21 | 2 | 0 | 1 | 0 | 0 | 0 | NA | 0 | 0 | 0.00 |
| 22 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 23 | 2 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 24 | 2 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 25 | 2 | 4 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 26 | 2 | 5 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 27 | 2 | 6 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 28 | 2 | 7 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 29 | 2 | 8 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 |
| 30 | 2 | 9 | 10 | 0 | 1 | 0 | 0 | 1 | 1 | 10.00 |
| 31 | 2 | 10 | 11 | 0 | 1 | 0 | 1 | 1 | 2 | 18.18 |
| 32 | 2 | 11 | 12 | 0 | 1 | 0 | 1 | 1 | 3 | 25.00 |
| 33 | 2 | 12 | 13 | 0 | 1 | 0 | 1 | 1 | 4 | 30.77 |
| 34 | 2 | 13 | 14 | 0 | 1 | 0 | 1 | 1 | 5 | 35.71 |
| 35 | 2 | 14 | 15 | 0 | 0 | 0 | 1 | 1 | 5 | 33.33 |
| 36 | 2 | 15 | 16 | 0 | 0 | 0 | 0 | 1 | 5 | 31.25 |
| 37 | 2 | 16 | 17 | 1 | 0 | 0 | 0 | 1 | 5 | 29.41 |
| | | | | | | | | | | |

The employment status variable has to be modeled in the "right" functional form to make correct inference about the exposure-outcome relationship. There are many different ways to model the exposure, and choosing the "right" functional form requires scientific understanding of the exposure-outcome relationship, not just statistics.

Here we created the current employment status (employed), the initial employment status (employed.initial), the past-week employment status (employed.lag1), ever-employed status (employed.ever), cumulative sum of weeks in employment (employed.cumsum), and % of time after release spent in employment (employed.percent).

Use cluster(id) to indicate the rows with the same id are from the same individual, thus, clustered. This will give the robust SE. However, the same analysis in SAS appear to return the SE with using robust estimator by default.

Using initial employment status (time-invariant)

Employment during the first week is associated with a HR for arrest of 0.7396 (non-significant). This is an equivalent of intention-to-treat (ITT) analysis.

```
res.with.employed.initial <-
    coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed.initial
+ cluster(id),
    data = Rossi.long,
    ties = c("efron","breslow","exact")[1])
summary(res.with.employed.initial)</pre>
```

```
call:
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
    wexp + mar + paro + prio + employed.initial + cluster(id)
    data = Rossi long, ties = c("efron", "breslow", "exact")[1])
 n= 19809, number of events= 114
                    coef exp(coef) se(coef) robust se
                                                           z Pr(>|z|)
                 -0.3801
                                                0.1955 - 1.94
fin
                            0.6838
                                      0.1913
                                                                0.0519
                                                0.0254 -2.23
                                      0.0220
                 -0.0568
                            0.9448
                                                                0.0257 *
age
                            1.3296
                                     0.3095
                  0.2849
                                                0.2949 0.97
                                                               0.3341
race
                                     0.2125
                 -0.1371
                            0.8719
                                                0.2188 - 0.63
                                                                0.5308
wexp
                                     0.3815
                 -0.4161
                            0.6596
                                                0.3785 - 1.10
                                                                0.2717
mar
                                     0.1957
                 -0.0766
                            0.9263
                                                0.1994 - 0.38
                                                                0.7009
paro
                                     0.0286
                  0.0905
                            1.0947
                                                0.0290 3.12
                                                                0.0018 **
prio
employed.initial -0.3017
                            0.7396
                                      0.3211
                                                0.3131 - 0.96
                                                                0.3352
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                 exp(coef) exp(-coef) lower .95 upper .95
fin
                     0.684
                                 1.462
                                           0.466
                                                     1.003
                     0.945
                                1.058
                                           0.899
                                                     0.993
age
                     1.330
                                 0.752
                                           0.746
                                                     2.370
race
                                1.147
                     0.872
                                           0.568
                                                     1.339
wexp
                     0.660
                                           0.314
                                 1.516
                                                     1.385
mar
                     0.926
                                1.080
                                           0.627
                                                     1.369
paro
                     1.095
                                0.913
                                           1.034
                                                     1.159
prio
employed.initial
                     0.740
                                 1.352
                                           0.400
                                                     1.366
Concordance= 0.643 (se = 0.027)
Rsquare= 0.002 (max possible= 0.066)
Likelihood ratio test= 34.2 on 8 df,
                                         p=0.0000371
                     = 31.6 on 8 df,
                                         p=0.000109
Wald test
Score (logrank) test = 34.3 on 8 df,
                                                        Robust = 27.9 p=0.000488
                                         p=0.0000359,
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the Wald and robust score tests do not).
```

Using time-dependent current employment status

Employment during a week is associated with a HR for arrest of 0.2649 during that week. The assumption is that only the current employment status matters, not past history of employment.

```
res.with.employed <-
    coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed +
cluster(id),
    data = Rossi.long,
    ties = c("efron","breslow","exact")[1])
summary(res.with.employed)</pre>
```

```
call:
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
    wexp + mar + paro + prio + employed + cluster(id), data = Rossi.long,
    ties = c("efron", "breslow", "exact")[1])
 n= 19809, number of events= 114
            coef exp(coef) se(coef) robust se
                                                       Pr(>|z|)
fin
                                        0.1961 - 1.82
                    0.7000
                              0.1911
                                                          0.0689
         -0.3567
         -0.0463
                    0.9547
                              0.0217
                                        0.0251 - 1.84
                                                          0.0653 .
age
                    1.4031
                                        0.2978
          0.3387
                              0.3096
                                               1.14
                                                          0.2554
race
                             0.2114
         -0.0256
                    0.9748
                                        0.2194 - 0.12
                                                          0.9073
wexp
                                        0.3975 - 0.74
         -0.2937
                    0.7455
                              0.3830
                                                          0.4599
mar
                    0.9378
                                        0.1992 - 0.32
         -0.0642
                              0.1947
                                                          0.7472
paro
                    1.0889
                             0.0290
                                               2.91
          0.0851
prio
                                        0.0293
                                                          0.0036 **
employed -1.3283
                    0.2649
                              0.2507
                                        0.2515 -5.28 0.00000013
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
         exp(coef) exp(-coef) lower .95 upper .95
                        1.429
fin
                                   0.477
                                             1.028
             0.700
             0.955
                        1.047
                                   0.909
                                             1.003
age
             1.403
                        0.713
                                   0.783
                                             2.515
race
             0.975
                        1.026
                                   0.634
                                             1.498
wexp
             0.745
                        1.341
                                   0.342
                                             1.625
mar
             0.938
                                   0.635
                        1.066
                                             1.386
paro
             1.089
                                   1.028
                        0.918
                                             1.153
prio
             0.265
                         3.775
                                   0.162
                                             0.434
employed
Concordance= 0.708 (se = 0.027)
                 (max possible= 0.066)
Rsquare= 0.003
Likelihood ratio test= 68.7 on 8 df,
                                         p=9.11e-12
Wald test
                     = 57.3 on 8 df,
                                         p=0.0000000159
                                                      Robust = 54.1 p=0.0000000647
Score (logrank) test = 64.5 on 8 df,
                                         p=6.1e-11,
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the wald and robust score tests do not).
```

Using time-dependent past-week employment status (employment status of the previous interval is used)

Employment during a week is associated with a HR for arrest of 0.4553 during the next week. The assumption is that only the past-week employment status matters, not the current employment status or that of two or more weeks before.

```
res.with.employed.lag1 <-
   coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed.lag1 +
cluster(id),
   data = Rossi.long,
   ties = c("efron","breslow","exact")[1])
summary(res.with.employed.lag1)</pre>
```

```
call:
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
    wexp + mar + paro + prio + employed.lag1 + cluster(id), data = Rossi.long,
    ties = c("efron", "breslow", "exact")[1])
 n= 19377, number of events= 113
   (432 observations deleted due to missingness)
                 coef exp(coef) se(coef) robust se
                                                        z Pr(>|z|)
fin
                         0.7038
                                   0.1918
                                             0.1957 - 1.80
                                                           0.07259
              -0.3513
                         0.9514
                                   0.0219
              -0.0498
                                             0.0252 - 1.98
age
                                                           0.04802 *
                         1.3792
                                  0.3091
               0.3215
                                             0.2946
                                                    1.09
                                                           0.27521
race
              -0.0476
                         0.9535
                                  0.2132
                                             0.2185 - 0.22
                                                           0.82741
wexp
                                                           0.37941
                                             0.3922 - 0.88
              -0.3448
                         0.7084
                                  0.3832
mar
                                  0.1963
                         0.9540
                                             0.1987 - 0.24
              -0.0471
                                                           0.81260
paro
               0.0920
                         1.0964
                                  0.0288
                                             0.0286
                                                    3.22
                                                           0.00128 **
prio
                         0.4553
                                   0.2181
                                             0.2173 -3.62 0.00029 ***
employed.lag1 -0.7869
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
              exp(coef) exp(-coef) lower .95 upper .95
fin
                             1.421
                                        0.480
                  0.704
                                                  1.033
                  0.951
                             1.051
                                        0.906
                                                  1.000
age
                  1.379
                             0.725
                                        0.774
                                                  2.457
race
                  0.953
                             1.049
                                        0.621
                                                  1.463
wexp
                  0.708
                             1.412
                                        0.328
                                                  1.528
mar
                  0.954
                             1.048
                                        0.646
                                                  1.408
paro
                                                  1.159
                  1.096
                             0.912
                                        1.037
prio
employed.lag1
                  0.455
                             2.197
                                        0.297
                                                  0.697
Concordance= 0.67 (se = 0.027)
                 (max possible= 0.067)
Rsquare= 0.002
Likelihood ratio test= 47.2 on 8 df.
                                         p=0.00000143
                             on 8 df,
Wald test
                     = 41.6
                                         p=0.0000158
Score (logrank) test = 46.4 on 8 df,
                                                          Robust = 36.8 p=0.0000125
                                         p=0.00000199
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the Wald and robust score tests do not).
```

Using time-dependent ever-employed status

Having experienced any duration of employment after release is associated with a HR for arrest of 0.6464. The assumption is that having experienced employment is equal regardless of duration or continuity.

```
res.with.employed.ever <-
    coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed.ever +
cluster(id),
    data = Rossi.long,
    ties = c("efron","breslow","exact")[1])
summary(res.with.employed.ever)</pre>
```

```
call:
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
    wexp + mar + paro + prio + employed.ever + cluster(id), data = Rossi.long.
    ties = c("efron", "breslow", "exact")[1])
  n= 19809, number of events= 114
                 coef exp(coef) se(coef) robust se
                                                        z Pr(>|z|)
fin
              -0.3773
                         0.6857
                                   0.1913
                                             0.1950 - 1.93
                                                             0.0531
              -0.0583
                         0.9434
                                   0.0220
                                             0.0254 - 2.30
                                                             0.0217 *
age
                         1.3761
                                   0.3078
                                             0.2890 1.10
               0.3192
                                                            0.2693
race
                                   0.2143
              -0.0929
                         0.9113
                                             0.2164 - 0.43
                                                             0.6678
wexp
                                             0.3807 -1.00
              -0.3825
                         0.6821
                                   0.3819
                                                             0.3149
mar
                                                            0.6997
                                   0.1958
                                             0.1990 - 0.39
              -0.0767
                         0.9261
paro
                                   0.0288
                                                            0.0022 **
               0.0878
                                             0.0287 3.06
prio
                         1.0917
employed.ever -0.4364
                         0.6464
                                   0.2168
                                             0.2129 - 2.05
                                                             0.0404 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
              exp(coef) exp(-coef) lower .95 upper .95
fin
                              1.458
                                        0.468
                                                  1.005
                  0.686
                  0.943
                              1.060
                                        0.898
                                                  0.991
age
                  1.376
                             0.727
                                        0.781
                                                  2.425
race
                  0.911
                              1.097
                                        0.596
                                                  1.393
wexp
                                                  1.438
                  0.682
                              1.466
                                        0.323
mar
                  0.926
                              1.080
                                        0.627
                                                  1.368
paro
                  1.092
                             0.916
                                        1.032
                                                  1.155
prio
employed.ever
                  0.646
                              1.547
                                        0.426
                                                  0.981
Concordance = 0.652 (se = 0.027)
                 (max possible= 0.066)
Rsquare= 0.002
Likelihood ratio test= 37.1 on 8 df,
                                         p=0.0000108
                     = 37.7
                             on 8 df,
                                         p=0.00000862
Wald test
Score (logrank) test = 37.5 on 8 df.
                                         p=0.00000926
                                                         Robust = 31.2 p=0.000127
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the wald and robust score tests do not).
```

Using cumulative time in employment (total weeks of employment)

One additional week of employment is associated with a HR for arrest of 0.9825 regardless of when it is from (non-significant). The effect of the cumulative exposure for a given individual can increase (during additional exposure) or stay the same (while unexposed). The assumption is that each additional week of employment is equivalent no matter when it was obtained and that once obtained the benefit of employment will not degrade even if the subject becomes unemployed.

```
res.with.employed.cumsum <-
   coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed.cumsum +
cluster(id),
   data = Rossi.long,
   ties = c("efron","breslow","exact")[1])
summary(res.with.employed.cumsum)</pre>
```

```
call:
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
   wexp + mar + paro + prio + employed cumsum + cluster(id),
   data = Rossi.long, ties = c("efron", "breslow", "exact")[1])
 n= 19809, number of events= 114
                    coef exp(coef) se(coef) robust se
                                                           z Pr(>|z|)
fin
                           0.68412 0.19121
                                              0.19562 - 1.94
                -0.37962
                                                               0.0523
                -0.05294
                           0.94844 0.02212
                                              0.02565 - 2.06
                                                               0.0390 *
age
                           1.36572 0.30848
                                              0.29445 1.06
                 0.31168
                                                               0.2898
race
                -0.08529
                           0.91825
                                   0.21445
                                              0.22215 - 0.38
                                                               0.7010
wexp
                                              0.38253 -0.99
                -0.37944
                           0.68425 0.38307
                                                               0.3212
mar
                           0.93858 0.19571
                                              0.19997 -0.32
                                                               0.7512
                -0.06339
paro
                                                               0.0023 **
                0.08907
                           1.09315 0.02887
                                              0.02920 3.05
prio
employed cumsum -0.01766
                           0.98250 0.00947
                                              0.00964 - 1.83
                                                               0.0671 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                exp(coef) exp(-coef) lower .95 upper .95
fin
                               1.462
                                         0.466
                                                    1.004
                    0.684
                    0.948
                               1.054
                                         0.902
                                                    0.997
age
                    1.366
                               0.732
                                         0.767
                                                    2.432
race
                               1.089
                                         0.594
                                                    1.419
                    0.918
wexp
                               1.461
                    0.684
                                                    1.448
                                         0.323
mar
                    0.939
                                         0.634
                               1.065
                                                   1.389
paro
                               0.915
                    1.093
                                         1.032
                                                   1.158
prio
employed.cumsum
                    0.982
                               1.018
                                         0.964
                                                    1.001
Concordance = 0.649 (se = 0.027)
Rsquare= 0.002
                 (max possible= 0.066)
Likelihood ratio test= 36.9 on 8 df,
                                        p=0.0000121
                     = 34 on 8 df,
                                      p=0.0000414
Wald test
Score (logrank) test = 36.5 on 8 df,
                                                       Robust = 30.2 p=0.000195
                                        p=0.000014
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the wald and robust score tests do not).
```

Using cumulative time in employment as percentage of time after release

One additional percent of time after release in employment is associated with a HR for arrest of 0.9923. The effect of the cumulative exposure for a given individual can increase (during additional exposure) or decrease (while unexposed). The assumption is that more time spent in employment is beneficial, but the benefit will degrade if the subject becomes unemployed.

```
res.with.employed.percent <-
    coxph(formula = Surv(start, stop, event) ~ fin + age + race + wexp + mar + paro + prio + employed.percent
+ cluster(id),
    data = Rossi.long,
    ties = c("efron", "breslow", "exact")[1])
summary(res.with.employed.percent)</pre>
```

```
coxph(formula = Surv(start, stop, event) ~ fin + age + race +
    wexp + mar + paro + prio + employed.percent + cluster(id)
    data = Rossi.long, ties = c("efron", "breslow", "exact")[1])
  n= 19809, number of events= 114
                     coef exp(coef) se(coef) robust se
                                                             z Pr(>|z|)
                                                0.19575 -1.94
fin
                 -0.38011
                                     0.19116
                                                                  0.052
                            0.68379
                                      0.02216
                 -0.05121
                            0.95008
                                                0.02591 - 1.98
                                                                  0.048 *
age
                                     0.30867
                            1.36174
                                                0.29419 1.05
                  0.30876
                                                                  0.294
race
                                     0.21415
                            0.93964
                                                0.22093 -0.28
                                                                  0.778
                 -0.06226
wexp
                 -0.34864
                                      0.38341
                                                0.38707 -0.90
                            0.70565
                                                                  0.368
mar
                                                0.20052 -0.26
                 -0.05265
                            0.94872
                                     0.19576
                                                                  0.793
paro
                                                0.02907 2.97
                                                                  0.003 **
                            1.09002
                  0.08620
prio
                                      0.02887
employed.percent -0.00770
                            0.99233
                                     0.00311
                                                0.00305 - 2.52
                                                                  0.012 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                 exp(coef) exp(-coef) lower .95 upper .95
fin
                                 1.462
                                           0.466
                                                     1.004
                     0.684
                     0.950
                                 1.053
                                           0.903
                                                     1.000
age
                     1.362
                                 0.734
                                           0.765
                                                     2.424
race
                     0.940
                                           0.609
                                 1.064
                                                     1.449
wexp
                                 1.417
                     0.706
                                           0.330
                                                     1.507
mar
                     0.949
                                           0.640
                                 1.054
                                                     1.405
paro
                     1.090
                                0.917
                                           1.030
                                                     1.154
prio
                     0.992
                                 1.008
                                           0.986
                                                     0.998
employed.percent
Concordance = 0.658 (se = 0.027)
Rsquare= 0.002
                 (max possible= 0.066)
Likelihood ratio test= 39.7 on 8 df,
                                         p=0.0000036
                     = 38.6 on 8 df,
                                         p=0.00000583
Wald test
Score (logrank) test = 38.9 on 8 df,
                                         p=0.00000523,
                                                         Robust = 33.7 p=0.0000458
  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the wald and robust score tests do not).
```

Professor Fox's fold function

call:

http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-cox-regression.pdf

Professor Fox's fold function explained in the PDF above is handy in creating the dataset.

```
suffix='.time', cov.times=0:ncov, common.times=TRUE, lag=0){
    Arguments:
##
    data: A data frame or numeric matrix (with column names) to be `unfolded.'
##
        For reasons of efficiency, if there are factors in data these will be
        converted to numeric variables in the output data frame.
##
    time: The quoted name of the event/censoring-time variable in data.
##
    event: The quoted name of the event/censoring-indicator variable in data.
##
    cov: A vector giving the column numbers of the time-dependent covariate
##
##
        in data, or a list of vectors if there is more than one time-varying
##
        covariate.
##
    cov.names: A character string or character vector giving the name or names
##
        to be assigned to the time-dependent covariate(s) in the output data set.
##
    suffix: The suffix to be attached to the name of the time-to-event variable
        in the output data setl defaults to '.time'.
##
##
    cov.times: The observation times for the covariate values, including the start
##
        time. This argument can take several forms:
            The default is integers from 0 to the number of covariate values (i.e.,
##
##
                one more than the length of each vector in cov).
##
            An arbitrary numerical vector with one more entry than the length of each
##
                vector in cov.
##
            The columns in the input data set that give the observations times for each
##
                individual. There should be one more column than the length of each
##
                vector in cov.
##
            common.times: A logical value indicating whether the times of observation
##
                are the same for all individuals; defaults to TRUE.
    lag: Number of observation periods to lag each value of the time-varying
##
                covariate(s); defaults to 0.
##
    vlag <- function(x, lag) c(rep(NA, lag), x[1:(length(x)-lag)])
    xlag <- function(x, lag) apply(as.matrix(x), 2, vlag, lag=lag)</pre>
    all.cov <- unlist(cov)
    if (!is.list(cov)) cov <- list(cov)</pre>
    ncovs <- length(cov)</pre>
    nrow <- nrow(data)</pre>
    ncol <- ncol(data)</pre>
    ncov <- length(cov[[1]])</pre>
    nobs <- nrow*ncov
    if (length(unique(c(sapply(cov, length), length(cov.times)-1))) > 1)
             all elements of cov must be of the same length and \n"
            "cov.times must have one more entry than each element of cov."))
    var.names <- names(data)</pre>
    subjects <- rownames(data)</pre>
    omit.cols <- if (!common.times) c(all.cov, cov.times) else all.cov
    keep.cols <- (1:ncol)[-omit.cols]
    nkeep <- length(keep.cols)</pre>
    if (is.numeric(event)) event <- var.names[event]</pre>
    times <- if (common.times) matrix(cov.times, nrow, ncov+1, byrow=T)
        else data[, cov.times]
    new.data <- matrix(Inf, nobs, 3 + ncovs + nkeep)</pre>
    rownames <- rep("", nobs)</pre>
    colnames(new.data) <- c('start', 'stop', paste(event, suffix, sep=""),
        var.names[-omit.cols], cov.names)
    end.row <- 0
    for (i in 1:nrow) {
        start.row <- end.row + 1
        end.row <- end.row + ncov
        start <- times[i, 1:ncov]</pre>
        stop <- times[i, 2:(ncov+1)]</pre>
        event.time <- ifelse (stop == data[i, time] & data[i, event] == 1, 1, 0)
        keep <- matrix(unlist(data[i, -omit.cols]), ncov, nkeep, byrow=T)</pre>
        select <- apply(matrix(!is.na(data[i, all.cov]), ncol=ncovs), 1, all)</pre>
        rows <- start.row:end.row
        cov.mat <- xlag(matrix(unlist(data[i, all.cov]), nrow=length(rows)), lag)</pre>
        new.data[rows[select], ] <-</pre>
        cbind(start, stop, event.time, keep, cov.mat)[select,]
rownames[rows] <- paste(subjects[i], '.', seq(along=rows), sep="")</pre>
    row.names(new.data) <- rownames</pre>
    as.data.frame(new.data[new.data[, 1] != Inf &
        apply(as.matrix(!is.na(new.data[, cov.names])), 1, all), ])
## Use of lag is easy
Rossi.long2 <- fold(Rossi, time='week', event='arrest', cov=11:62, cov.names='emp', lag = 1)
Rossi.long2
       start stop arrest.time week arrest fin age race wexp mar paro prio educ emp
1 2
```

| 1. Z | Τ | Z | U | 20 | T | U | 21 | Т. | U | U | Т. | 3 | 3 | U |
|-------------|----|----|---|----|---|---|----|----|---|---|----|---|---|---|
| 1.3 | 2 | 3 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.4 | 3 | 4 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.5 | 4 | 5 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.6 | 5 | 6 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.7 | 6 | 7 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.8 | 7 | 8 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.9 | 8 | 9 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.10 | 9 | 10 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.11 | 10 | 11 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |
| 1.12 | 11 | 12 | 0 | 20 | 1 | 0 | 27 | 1 | 0 | 0 | 1 | 3 | 3 | 0 |

| 1.11.1.12.2.3.4.5.67.8.9.0.1.2.3.4.5.0.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2 |
|--|
| 1234567890100000000000000000000000000000000000 |
| $\frac{13}{11}, \frac{13}{11}, \frac{13}{11}, \frac{15}{11}, \frac{15}{11}$ |
| |
| 000000100000000000001000000000000000000 |
| 200000077777777777775555555555555555555 |
| 111111111111111111111111111111111111111 |
| 000000000000000000000000000000000000000 |
| 22222221111111111111111111111111111111 |
| 111111111111111111111111111111111111111 |
| 0000000000000000000000111111111111111 |
| 000000000000000000000000000000000000000 |
| 111111111111111111111111111111111111 |
| 333333388888888888888888888888888888888 |
| 33333333333333333333333333333333333333 |
| 000000000000000011111100000000000000000 |
| |
| |