GSERM - 2018
Survival Model Extensions

June 8, 2018 (morning session)

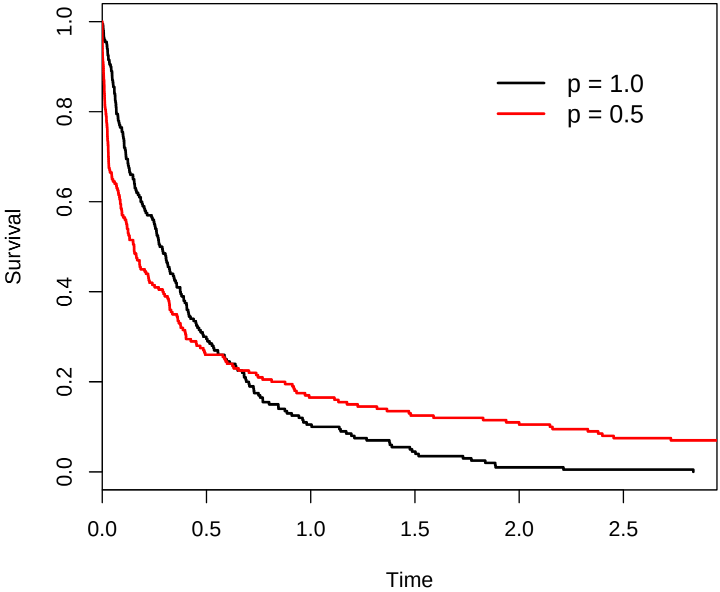
Stratification

Stratification

- Allow different groups to have <u>different baseline hazards</u>
- Akin to different intercepts, but more flexible.
- *Assumes covariate effects are otherwise identical*
- Uses:
  - Unit/group heterogeneity
  - Nonproportional hazards
  - Simple models for duration dependence

```
> set.seed=7222009
> Z<-rnorm(200)
> X0<-rep(0,times=200)
> X1<-rep(1,times=200)
> T0<-rweibull(200,shape=1,scale=1/exp(2+0.5*Z))
> T1<-rweibull(200,shape=0.5,scale=1/exp(2+0.5*Z))
> C<-rep(1,times=400)
> X<-append(X0,X1)
> T<-append(T0,T1)
> data<-as.data.frame(cbind(T,C,X,rep(Z,times=2)))
> colnames(data)<-c("T","C","X","Z")
```

Stratified Weibull Hazards



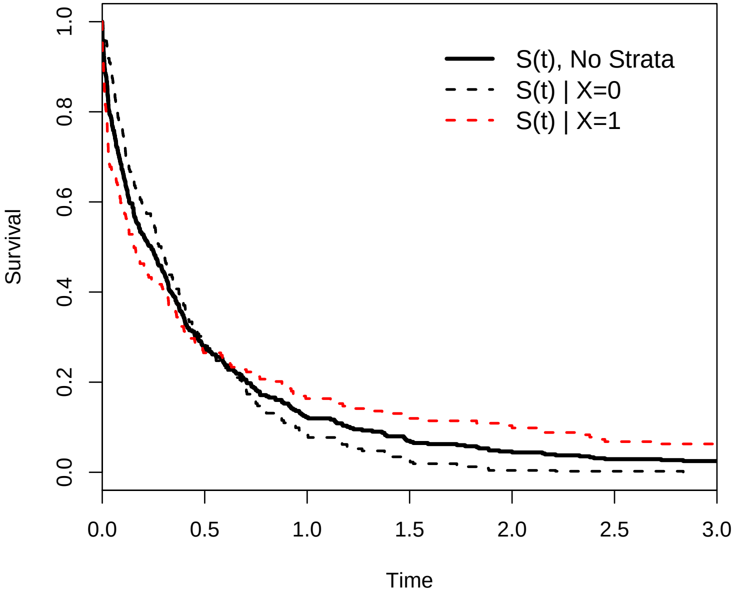
Stratification, Simulated

```
> cox<-coxph(S~Z+X,data=data)
> summary(cox)
Call:
coxph(formula = S ~ Z + X, data = data)

  n= 400, number of events= 400

      coef exp(coef) se(coef)      z Pr(>|z|)
Z  0.28286   1.32692  0.05133  5.510 3.58e-08 ***
X -0.22866   0.79560  0.10639 -2.149   0.0316 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
Z    1.3269     0.7536    1.1999    1.4674
X    0.7956     1.2569    0.6459    0.9801

Concordance= 0.571  (se = 0.017 )
Rsquare= 0.08   (max possible= 1 )
Likelihood ratio test= 33.25  on 2 df,   p=6.022e-08
Wald test            = 33.02  on 2 df,   p=6.749e-08
Score (logrank) test = 33.07  on 2 df,   p=6.601e-08
```

Stratification, Simulated

```
> cox.strata<-coxph(S~Z+strata(X),data=data)
> summary(cox.strata)
Call:
coxph(formula = S ~ Z + strata(X), data = data)

  n= 400, number of events= 400

     coef exp(coef) se(coef)    z Pr(>|z|)
Z 0.32140   1.37906  0.05176 6.21  5.3e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
Z     1.379     0.7251     1.246     1.526

Concordance= 0.597  (se = 0.024 )
Rsquare= 0.092   (max possible= 1 )
Likelihood ratio test= 38.69  on 1 df,   p=4.955e-10
Wald test            = 38.56  on 1 df,   p=5.303e-10
Score (logrank) test = 38.62  on 1 df,   p=5.151e-10
```

Cox $\widehat{S(t)}$s



Stratified Weibull Model

```
> summary(survreg(S~Z+strata(X),data=data,dist="weibull"))

Call:
survreg(formula = S ~ Z + strata(X), data = data, dist = "weibull")
              Value Std. Error       z        p
(Intercept) -0.9976     0.0675 -14.781 1.93e-49
Z           -0.4140     0.0577  -7.178 7.06e-13
X=0          0.0152     0.0555   0.274 7.84e-01
X=1          0.6864     0.0543  12.650 1.11e-36

Scale:
 X=0  X=1   # Recall: scale = 1 / p
1.02 1.99

Weibull distribution
Loglik(model)= -7.8   Loglik(intercept only)= -31.4
Chisq= 47.36 on 1 degrees of freedom, p= 5.9e-12
Number of Newton-Raphson Iterations: 6
n= 400
```

Duration Dependence

Duration Dependence

1. *State Dependence*
   - E.g., Institutionalization / Degradation

Positive State Dependence $\longrightarrow$ Negative Duration Dependence

Negative State Dependence $\longrightarrow$ Positive Duration Dependence

2. *Unobserved / Unmodeled Heterogeneity*
   - $h(t|\mathbf{X}_i) \neq h(t|\mathbf{X}_j)$ for $\mathbf{X}_i = \mathbf{X}_j$
   - Adverse selection in the sample / data
   - Result: "Spurious" duration dependence

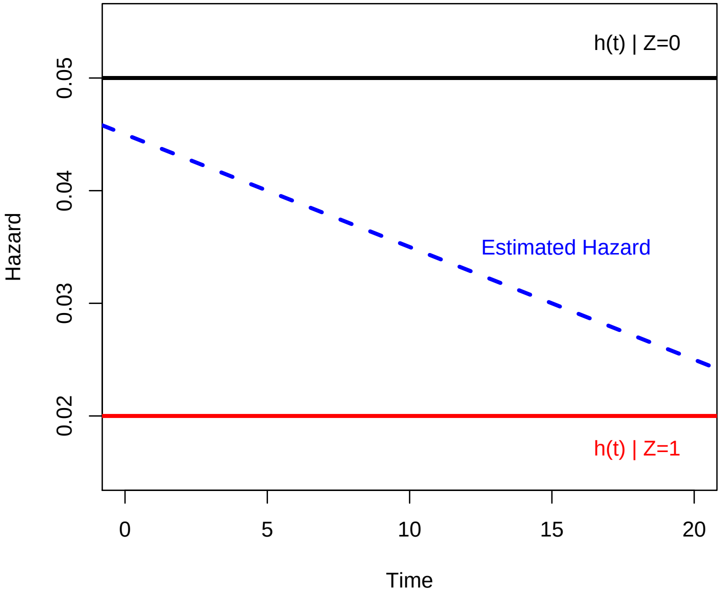
Suppose we have an unobserved Z, with

$$h_i(t|\mathbf{X}_i, Z_i = 0) = 0.05$$

and

$$h_i(t|\mathbf{X}_i, Z_i = 1) = 0.02.$$

Unobserved Heterogeneity Illustrated



Unobserved Heterogeneity: A Simulation

```
> set.seed(7222009)
> W<-rnorm(500)
> X<-rnorm(500)
> Z<-rnorm(500)
> T<-rexp(500,rate=(exp(0+0.5*W+0.5*X-0.6*Z))) # exponential hazard
> C<-rep(1,times=500)
> S<-Surv(T,C)
> summary(survreg(S~W,dist="weibull"))

Call:
survreg(formula = S ~ W, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0101     0.0629  -0.16 8.73e-01
W           -0.6339     0.0610 -10.40 2.47e-25
Log(scale)   0.2833     0.0333   8.52 1.62e-17

Scale= 1.33 # implies p = 1/Scale = 0.753

Weibull distribution
Loglik(model)= -568.1   Loglik(intercept only)= -615.3
Chisq= 94.47 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

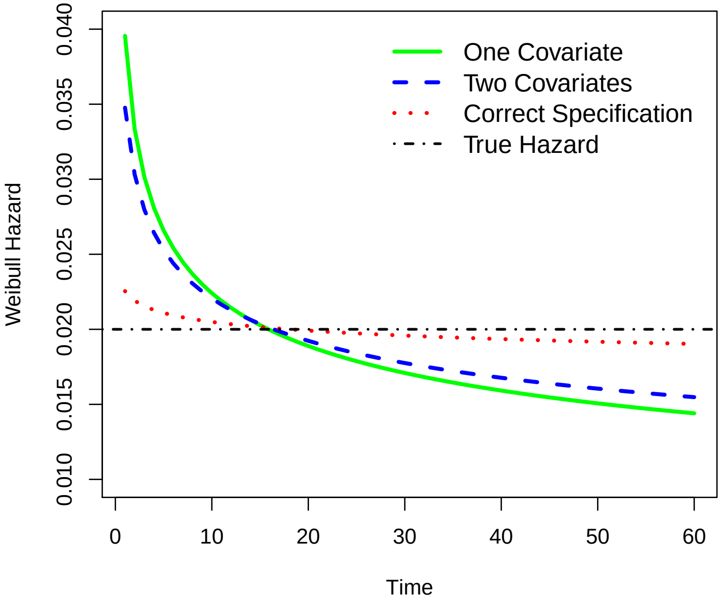
Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X,dist="weibull"))

Call:
survreg(formula = S ~ W + X, dist = "weibull")
              Value Std. Error       z        p
(Intercept) -0.0511     0.0591  -0.865 3.87e-01
W           -0.5907     0.0581 -10.160 2.98e-24
X           -0.4750     0.0556  -8.549 1.24e-17
Log(scale)   0.2202     0.0329   6.689 2.24e-11

Scale= 1.25 # implies p = 1/Scale = 0.802

Weibull distribution
Loglik(model)= -534.5   Loglik(intercept only)= -615.3
Chisq= 161.6 on 2 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X+Z,dist="weibull"))

Call:
survreg(formula = S ~ W + X + Z, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0777     0.0494  -1.57 1.16e-01
W           -0.5665     0.0468 -12.11 9.17e-34
X           -0.5041     0.0473 -10.66 1.58e-26
Z            0.5923     0.0446  13.29 2.73e-40
Log(scale)   0.0423     0.0345   1.22 2.21e-01

Scale= 1.04 # implies p = 1/Scale = 0.959

Weibull distribution
Loglik(model)= -464.3   Loglik(intercept only)= -615.3
Chisq= 302.01 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation



Duration Dependence: What To Do?

(At least) Three Options:

1. Model Specification
2. Unit-Level Effects
3. <u>Model</u> the Duration Dependence

Weibull with:

$$p = \exp(\mathbf{Z}_i\gamma)$$

Gives:

$$h_i(t) = \exp(\mathbf{X}_i\beta)\exp(\mathbf{Z}_i\gamma)[\exp(\mathbf{X}_i\beta)t]^{[\exp(\mathbf{Z}_i\gamma)]-1}$$

and (more usefully):

$$S(t) = \exp(-\exp(\mathbf{X}_i\beta)t)^{\exp(\mathbf{Z}_i\gamma)}$$

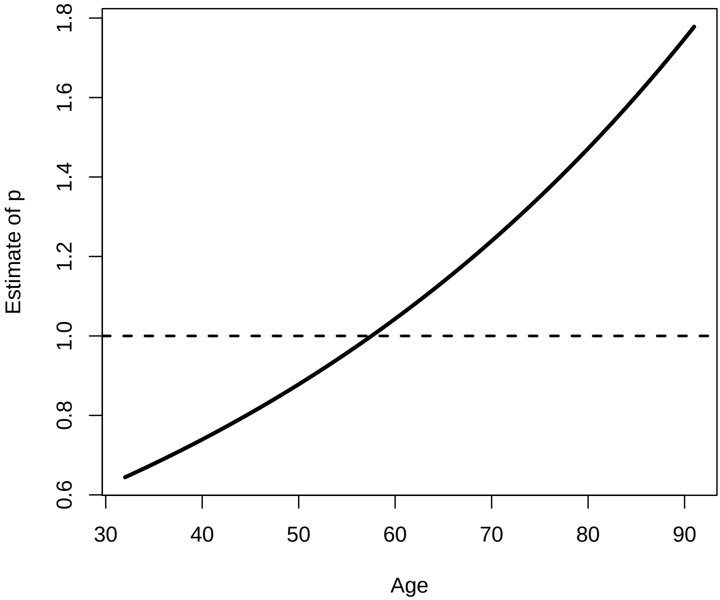
Example: SCOTUS Departures

```
> library(flexsurv)
> ct.weib<-flexsurvreg(scotus.S~age+pension+pagree,
                      data=scotus,dist="weibull")
> ct.weib

Estimates:
        data mean  est        L95%       U95%       exp(est)
shape          NA      0.999      0.637      1.570          NA
scale          NA    942.000     13.700  64800.000          NA
age        62.100     -0.041     -0.102      0.019       0.959
pension     0.199     -1.310     -2.360     -0.265       0.269
pagree      0.616     -0.113     -0.673      0.447       0.893
        L95%       U95%
shape          NA         NA
scale          NA         NA
age         0.903      1.020
pension     0.095      0.767
pagree      0.510      1.560

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -209, df = 5
AIC = 429
```

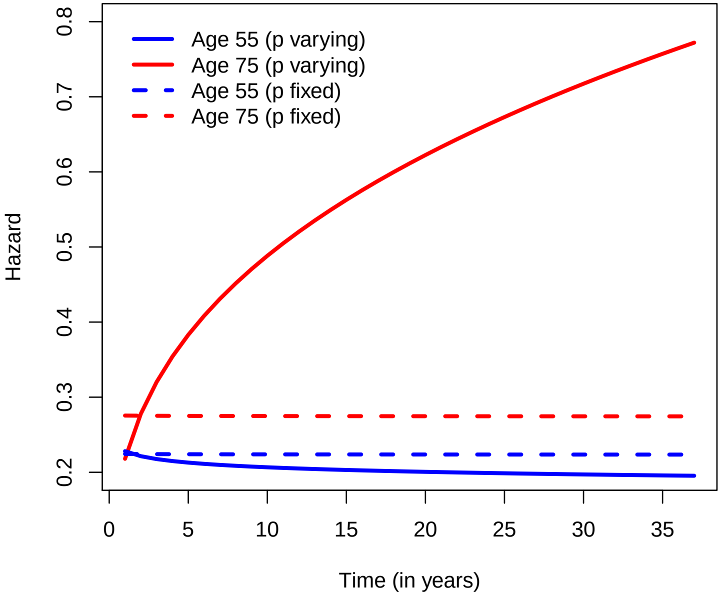
Example: SCOTUS Departures

```
> ct.weib.DD<-flexsurvreg(scotus.S~age+pension+pagree+shape(age),
                         data=scotus,dist="weibull")
> ct.weib.DD

Estimates:
            data mean   est         L95%        U95%
shape              NA      0.3710      0.1260      1.0900
scale              NA    491.0000     16.7000  14500.0000
age           62.1000     -0.0307     -0.0779      0.0164
pension        0.1990     -1.0900     -1.9700     -0.2190
pagree         0.6160     -0.0328     -0.4840      0.4180
shape(age)    62.1000      0.0172     -0.0011      0.0356
            exp(est)    L95%        U95%
shape              NA          NA          NA
scale              NA          NA          NA
age            0.9700      0.9250      1.0200
pension        0.3350      0.1400      0.8030
pagree         0.9680      0.6160      1.5200
shape(age)     1.0200      0.9990      1.0400

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -208, df = 6
AIC = 427
```

$\hat{p}$ by Age



$\widehat{h(t)}$s by Age and Model



Separation

Separation

"Separation" = "perfect prediction"

Intuition:

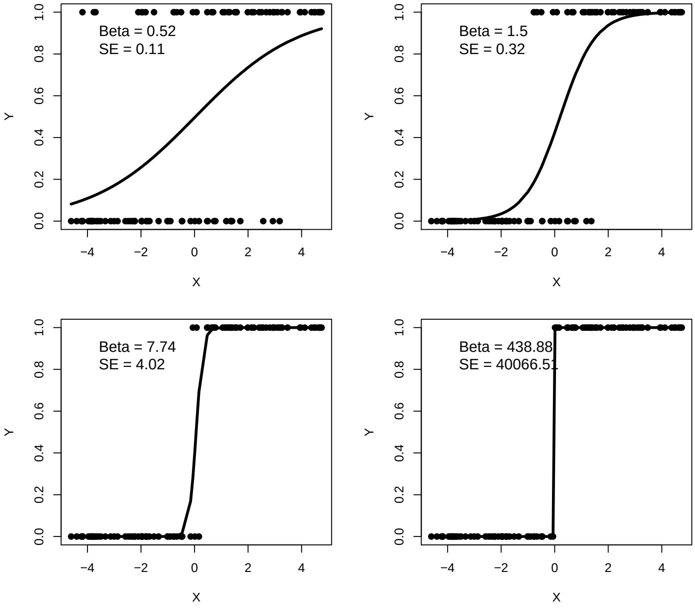
```
      Dems
Yeas    0   1
   0 178  34
   1   0 219
```

$\Pr(Y = 1|X = 0) =?$

Separation: Effects

- $\hat{\beta}_X = \pm\infty$
- $\widehat{\text{s.e.}}_\beta = \infty$
- $\left.\frac{\partial^2 \ln L}{\partial X^2}\right|_{\hat{\beta}} = 0$

Separation Illustrated



Separation: What Happens

```
> summary(glm(Y~W+Z+X,family="binomial"))

Call:
glm(formula = Y ~ W + Z + X, family = "binomial")

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.363      0.111   -3.26  0.00111 **
W              0.424      0.119    3.57  0.00036 ***
Z             -0.412      0.112   -3.67  0.00024 ***
X             18.746    541.835    0.03  0.97240
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 684.41  on 499  degrees of freedom
Residual deviance: 464.69  on 496  degrees of freedom
AIC: 472.7

Number of Fisher Scoring iterations: 17
```

Separation: What Happens (Stata Remix)

```
. logit Y W Z X

note: X != 0 predicts success perfectly
      X dropped and 136 obs not used

Iteration 0:   log likelihood = -245.53269
Iteration 1:   log likelihood = -232.41173
Iteration 2:   log likelihood = -232.34436
Iteration 3:   log likelihood = -232.34436

Logistic regression                               Number of obs   =        364
                                                  LR chi2(2)      =      26.38
                                                  Prob > chi2     =     0.0000
Log likelihood = -232.34436                       Pseudo R2       =     0.0537

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |   .4242117   .1187869     3.57   0.000     .1913936    .6570298
           Z |  -.4120285   .1123154    -3.67   0.000    -.6321628   -.1918943
           X |          0  (omitted)
       _cons |  -.3626348   .1112033    -3.26   0.001    -.5805892   -.1446803
------------------------------------------------------------------------------
```

Solution (?): Exact Logistic Regression

- Cox (1970, Ch. 4); Hirji et al. (1987 *JASA*); Mehta & Patel (1995 *Stat. Med.*)
- Conditions on permutations of covariate patterns
- Always has finite solutions;
- Computational issues...

Firth proposed:

$$L(\beta|Y)^* = L(\beta|Y)\,|\mathbf{I}(\beta)|^{\frac{1}{2}}$$

$$\ln L(\beta|Y)^* = \ln L(\beta|Y) + 0.5 \ln |\mathbf{I}(\beta)|$$

"Penalized likelihood":

- Consistent
- Eliminates small-sample bias
- Exist given separation
- Bayesians: "Jeffreys' prior"

- "Profile" (= "concentrated") likelihood
- $L(\hat{\beta})$ can be asymmetrical...
- $\rightarrow$ inference...

Software

- R
  - `elrm` (exact logistic regression via MCMC)
  - `brlr` ("bias-reduced logistic regression")
  - `logistf` ("Firth's logistic regression")

- Stata
  - `exlogistic` (exact logistic regression)
  - `firthlogit` (Firth corrected logit)

Example: Pets as Family

- CBS/NYT Poll, April 1997
- Standard political/demographics, plus
- "Do you consider your pet to be a member of your family, or not?"
- Yes $=$ 84.4%, No $=$ 15.6%

Pets as Family: Data

```
> summary(Pets)
   petfamily        female            married
 Min.   :0.00   Female:403   Divorced/Sep:118
 1st Qu.:1.00   Male  :321   Married     :442
 Median :1.00                NBM         :118
 Mean   :0.84                Widowed     : 46
 3rd Qu.:1.00
 Max.   :1.00
       partyid            education
            : 58   < HS         : 71
 Democrat   :224   College Grad:131
 GOP        :228   HS diploma  :244
 Independent:214   Post-Grad   : 96
                   Some college:182
```

Pets as Family: Basic Model

```
> Pets.1<-glm(petfamily~female+as.factor(married)+as.factor(partyid)
             +as.factor(education),data=Pets,family=binomial)

> summary(Pets.1)

Call:
glm(formula = petfamily ~ female + as.factor(married) + as.factor(partyid) +
    as.factor(education), family = binomial, data = Pets)

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                        2.0133     0.5388    3.74  0.00019 ***
femaleMale                        -0.6959     0.2142   -3.25  0.00116 **
as.factor(married)Married         -0.0657     0.2911   -0.23  0.82147
as.factor(married)NBM              0.4599     0.3957    1.16  0.24504
as.factor(married)Widowed         -0.1568     0.4921   -0.32  0.75007
as.factor(partyid)Democrat        -0.1241     0.4286   -0.29  0.77213
as.factor(partyid)GOP             -0.0350     0.4321   -0.08  0.93537
as.factor(partyid)Independent     -0.1521     0.4299   -0.35  0.72338
as.factor(education)College Grad   0.2511     0.4121    0.61  0.54228
as.factor(education)HS diploma     0.0595     0.3685    0.16  0.87182
as.factor(education)Post-Grad      0.1946     0.4331    0.45  0.65321
as.factor(education)Some college   0.0587     0.3867    0.15  0.87928
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 627.14  on 723  degrees of freedom
Residual deviance: 612.76  on 712  degrees of freedom
AIC: 636.8

Number of Fisher Scoring iterations: 4
```

Pets as Family: More Complicated Model

```
> Pets.2<-glm(petfamily~female+as.factor(married)*female+as.factor(partyid)+
             as.factor(education),data=Pets,family=binomial)

> summary(Pets.2)

Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                            2.2971     0.6166    3.73   0.0002 ***
femaleMale                            -1.1833     0.5305   -2.23   0.0257 *
as.factor(married)Married             -0.3218     0.4470   -0.72   0.4716
as.factor(married)NBM                  0.1854     0.6140    0.30   0.7628
as.factor(married)Widowed             -0.7415     0.5780   -1.28   0.1995
as.factor(partyid)Democrat            -0.1575     0.4297   -0.37   0.7140
as.factor(partyid)GOP                 -0.0445     0.4334   -0.10   0.9182
as.factor(partyid)Independent         -0.1757     0.4312   -0.41   0.6837
as.factor(education)College Grad       0.2332     0.4137    0.56   0.5730
as.factor(education)HS diploma         0.0558     0.3703    0.15   0.8801
as.factor(education)Post-Grad          0.2171     0.4342    0.50   0.6171
as.factor(education)Some college       0.0358     0.3890    0.09   0.9266
femaleMale:as.factor(married)Married   0.4853     0.5908    0.82   0.4114
femaleMale:as.factor(married)NBM       0.5260     0.8051    0.65   0.5136
femaleMale:as.factor(married)Widowed  15.2516   549.3719    0.03   0.9779
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 627.14  on 723  degrees of freedom
Residual deviance: 607.42  on 709  degrees of freedom
AIC: 637.4

Number of Fisher Scoring iterations: 14
```

What's Going On?

```
> with(Pets, xtabs(~petfamily+as.factor(married)+female))

, , female = Female

         as.factor(married)
petfamily Divorced/Sep Married NBM Widowed
        0            7      28   5       7
        1           67     199  58      32

, , female = Male

         as.factor(married)
petfamily Divorced/Sep Married NBM Widowed
        0           11      47   8       0
        1           33     168  47       7
```

Pets as Family: Firth Model

```
> Pets.Firth<-logistf(petfamily~female+as.factor(married)*female+as.factor(partyid)+
                   as.factor(education),data=Pets)

> Pets.Firth

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

                                       coef se(coef) lower 0.95 upper 0.95    Chisq        p
(Intercept)                          2.1589     0.60       1.05       3.40 16.17636 0.000058
femaleMale                          -1.1387     0.52      -2.19      -0.14  5.04186 0.024742
as.factor(married)Married           -0.2739     0.43      -1.19       0.53  0.41518 0.519353
as.factor(married)NBM                0.1589     0.59      -0.99       1.37  0.07322 0.786705
as.factor(married)Widowed           -0.7263     0.56      -1.84       0.38  1.67233 0.195947
as.factor(partyid)Democrat          -0.1182     0.42      -0.99       0.66  0.08159 0.775159
as.factor(partyid)GOP               -0.0078     0.42      -0.89       0.78  0.00034 0.985289
as.factor(partyid)Independent       -0.1364     0.42      -1.01       0.65  0.10813 0.742278
as.factor(education)College Grad     0.2390     0.40      -0.57       1.02  0.34480 0.557069
as.factor(education)HS diploma       0.0753     0.36      -0.67       0.76  0.04289 0.835933
as.factor(education)Post-Grad        0.2184     0.43      -0.63       1.05  0.26307 0.608019
as.factor(education)Some college     0.0524     0.38      -0.72       0.78  0.01888 0.890698
femaleMale:as.factor(married)Married 0.4558     0.58      -0.66       1.61  0.63550 0.425347
femaleMale:as.factor(married)NBM     0.5233     0.78      -1.02       2.05  0.45133 0.501702
femaleMale:as.factor(married)Widowed 2.4017     1.68      -0.14       7.37  3.37453 0.066212

Likelihood ratio test=17 on 14 df, p=0.24, n=724
```

- Separation $\nrightarrow$ dropping covariates!
- Firth's approach $>$ ELR
- Can also be applied to other sparse-data situations...

From the `coxph` Documentation

Convergence

In certain data cases the actual MLE estimate of a coefficient is infinity, e.g., a dichotomous variable where one of the groups has no events. When this happens the associated coefficient grows at a steady pace and a race condition will exist in the fitting routine: either the log likelihood converges, the information matrix becomes effectively singular, an argument to exp becomes too large for the computer hardware, or the maximum number of interactions is exceeded. (Nearly always the first occurs.) The routine attempts to detect when this has happened, not always successfully. The primary consequence for he user is that the Wald statistic = coefficient/se(coefficient) is not valid in this case and should be ignored; the likelihood ratio and score tests remain valid however.

```
> set.seed(7222009)
> X<-rep(0:1,times=100)
> T<-abs(rweibull(200,shape=1.2,
      scale=1/(exp(0+0.2*X))))
> C<-rbinom(200,1,0.2)
> C<-ifelse(X==0,0,C)

> table(C,X)
   X
C     0   1
  0 100  81
  1   0  19
```

Cox Results

```
> cox.fit<-coxph(Surv(T,C)~X,method="efron")
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  1 ; beta may be infinite.

> summary(cox.fit)

Call:
coxph(formula = Surv(T, C) ~ X, method = "efron")

  n= 200, number of events= 19

       coef exp(coef)  se(coef)     z Pr(>|z|)
X 2.038e+01 7.112e+08 5.630e+03 0.004    0.997

  exp(coef) exp(-coef) lower .95 upper .95
X 711225014  1.406e-09         0       Inf

Concordance= 0.761  (se = 0.064 )
Rsquare= 0.137   (max possible= 0.583 )
Likelihood ratio test= 29.37  on 1 df,   p=5.994e-08
Wald test            = 0  on 1 df,   p=0.9971
Score (logrank) test = 22.11  on 1 df,   p=2.58e-06
```

Parametric Model = No Help

```
> weib.fit<-survreg(Surv(T,C)~X,dist="weibull")
> summary(weib.fit)

Call:
survreg(formula = Surv(T, C) ~ X, dist = "weibull")
               Value Std. Error       z     p
(Intercept)  61.5536      0.337 182.851 0.000
X           -60.1634      0.000    -Inf 0.000
Log(scale)   -0.0184      0.190  -0.097 0.923

Scale= 0.982

Weibull distribution
Loglik(model)= -45.9   Loglik(intercept only)= -60.8
        Chisq= 29.94 on 1 degrees of freedom, p= 4.5e-08
Number of Newton-Raphson Iterations: 9
n= 200
```

Heinze and Schemper (2001):

$$\ln PL(\boldsymbol{\beta}|Y)^* = \ln PL(\boldsymbol{\beta}|Y) + 0.5 \ln |\mathbf{I}(\boldsymbol{\beta})|$$

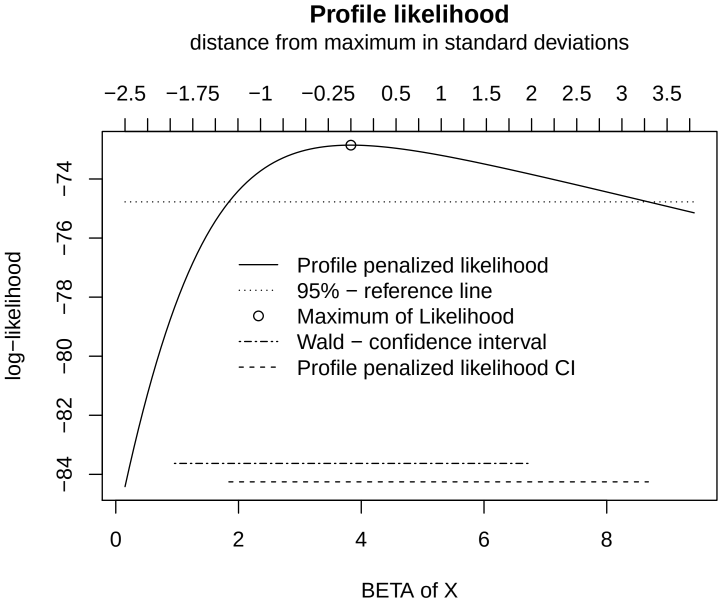
with $\mathbf{I}(\boldsymbol{\beta}) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln PL(X, \boldsymbol{\beta}) \middle| \boldsymbol{\beta} \right]$.

Also, software: `coxphf`...

Firth-Corrected Cox

```
> SIM<-cbind(T,C,X)
> SIM<-data.frame(SIM)
> firth.fit<-coxphf(SIM,formula=Surv(T,C)~X)

> firth.fit
coxphf(formula = Surv(T, C) ~ X, data = SIM)
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

      coef se(coef) exp(coef) lower 0.95 upper 0.95    Chisq            p
X 3.830922 1.472103  46.10502   6.316689   5871.037 26.08802 3.262011e-07

Likelihood ratio test=26.08802 on 1 df, p=3.262011e-07, n=200
```

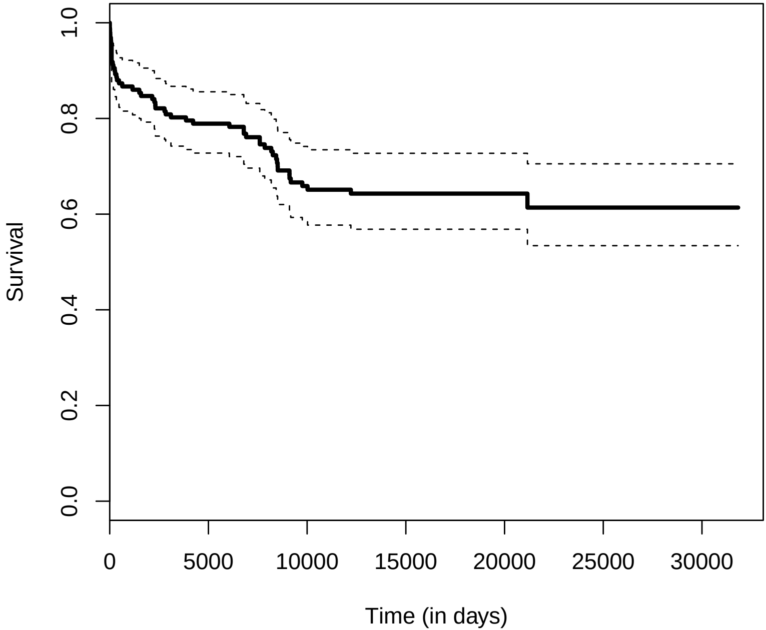
Examining the Profile Likelihood



Example: Lo et al. (2008)

- Outcome: "Cease-Fire Duration"
- Key covariate: *Foreign-Imposed Regime Change* ("FIRC")
- Data: Annual data on cease-fires, 1914-2001 (expanding Fortna 1998)
- Hypotheses:
  - FIRCs $\rightarrow$ more durable cease-fires
  - Pacifying influence of FIRCs declines over time

Lo et al.: Kaplan-Meier



From Lo et al. (2008)

Variables	*Model 1 (ARCHIGOS data)*	*Model 2 (ARCHIGOS data)*
FOREIGN-IMPOSED REGIME CHANGE	−161***	—
	(29.3)	
FIRC*ln(t)	16.8***	—
	(3.03)	
PUPPET-FIRC	—	−161***
		(29.4)
PUPPET-FIRC*ln(t)	—	16.8***
		(3.04)
CHANGE IN CAPABILITIES	.272	.274
	(.376)	(.376)
BATTLE CONSISTENCY	−.796**	−.809**
	(.336)	(.342)

Lo et al. Odds Ratio

$\widehat{\mathsf{OR}}_{\mathsf{FIRC}} =$

0.0000000000000000000000000000000000000000000000000000000000000000000001198

Simplified Cox Model

```
> LHR.Cox<-coxph(LHR.S~archigosFIRC+archigosFIRClnt,data=LHR,method="efron",
                iter.max=10000)
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Ran out of iterations and did not converge

> LHR.Cox

Call:
coxph(formula = LHR.S ~ archigosFIRC + archigosFIRClnt, data = LHR,
    method = "efron", iter.max = 10000)

                  coef exp(coef) se(coef)     z     p
archigosFIRC    -44.11  6.99e-20    20.92 -2.11 0.035
archigosFIRClnt   4.69  1.09e+02     2.24  2.09 0.036

Likelihood ratio test=20.8  on 2 df, p=3.04e-05  n= 6368, number of events= 54
```

Firth-Corrected Cox Model

```
> LHR.CoxF<-coxphf(LHR.S~archigosFIRC+archigosFIRClnt,data=LHR,maxit=1000)

> LHR.CoxF

coxphf(formula = LHR.S ~ archigosFIRC + archigosFIRClnt, data = LHR,
    maxit = 1000)
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

                      coef  se(coef)   exp(coef)   lower 0.95
archigosFIRC    -55.591223 26.330771 7.19513e-25 4.808432e-65
archigosFIRClnt   5.848163  2.775927 3.46597e+02          NaN

                upper 0.95    Chisq           p
archigosFIRC           NaN 9.738246 0.001804731
archigosFIRClnt    7320112 8.647141 0.003275750

Likelihood ratio test=15.09131 on 2 df, p=0.0005284014, n=6368
```

What Is Going On?

```
> table(LHR$archigosFIRC,LHR$X_d)

       0    1
  0 5265   52
  1 1049    2
```

Days $\rightarrow$ Years = Little Help

	Cox	Firth-Corrected
FIRC	-53.83	-21.08
	(51.30)	(9.99)
FIRC $\times$ $\ln(T)$	14.57	5.85
	(13.54)	(2.78)
AIC	494.07	496
Num. events	54	54

Proportional Hazards

Proportional Hazards

For two individuals A and B, their relative hazards will be:

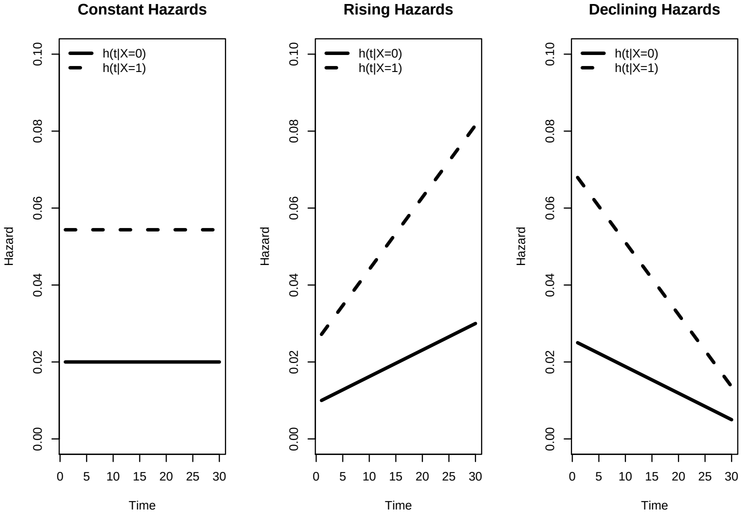
$$h_A(t) = Ch_B(t)$$

where C is the hazard ratio between A and B.

Proportionality:

- "Flat" hazards $\rightarrow$ parallel
- Rising hazards $\rightarrow$ diverging
- Falling hazards $\rightarrow$ converging

Proportional Hazards, Illustrated



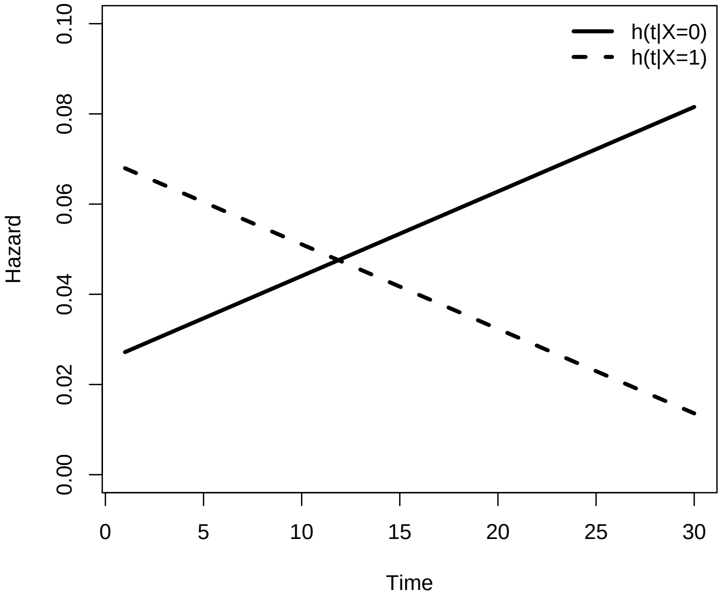
Proportional Hazards, continued

Why might hazards not be proportional?

- Resistance ($\rightarrow$ converging hazards)
- Learning ($\rightarrow$ converging hazards)
- Reinforcement ($\rightarrow$ diverging hazards)

Also, crossing hazards (always non-proportional)

Crossing Hazards



What Proportional Hazards Mean

Covariate influence over time

- PH assumes that the (proportional) influence of covariates **X** on the hazard will be the same at any point in the duration.
- Suggests how to think about it:

Conventional model:

$$h(t|\mathbf{X}_i) = h_0(t)\exp(\mathbf{X}_i\beta)$$

Generalized model:

$$h(t|\mathbf{X}_i) = h_0(t)\exp[\mathbf{X}_i\beta + \mathbf{X}_i g(t)\gamma]$$

Three kinds of tests for nonproportionality:

1. Tests for *changes in parameter values for coefficients estimated on a subsample of the data* defined by t,
2. Tests based on *plots of survival estimates and regression residuals against time*, and
3. Explicit tests of *interactions of covariates and time*.

Step function:

$$\begin{aligned} g(t) &= 0 \;\forall\; t \leq \tau \\ &= 1 \;\forall\; t > \tau \end{aligned}$$

Implies:

$$h_i(t) = f\{X_i\beta_1 + [g(t)_i]\beta_2 + X_i[g(t)_i]\beta_3\}$$

Things to think about:

- Abrupt change?
- Choice of t in $g(t)$
- Multiple "steps"?

log-log-Survival Plots

Kalbfleisch and Prentice (1980) note that in the Cox model:

$$S(t) = \exp\left[-\exp(\mathbf{X}_i\beta)\int_0^t h_0(t)\,dt\right]$$

which means

$$\ln\{-\ln[S(t)]\} = H_0(t) \times \mathbf{X}_i\beta.$$

Implies that plots of $\ln\{\widehat{-\ln[S(t)}]\}$ vs. $\ln(T)$ for different values of $\mathbf{X}$ should be parallel to one another.

Recall:

$$\hat{M}_i(t) = C_i(t) - \hat{H}_i(t)$$

where $C_i(t) \equiv N_i(t)$ is the censoring indicator at t and $\hat{H}_i(t)$ is the integrated hazard.

Proportional hazards implies:

$$\hat{M}_i(t) = C_i(t) - \underbrace{\exp(\mathbf{X}_{it}\hat{\beta})\hat{H}_0(t)}_{\text{("Cox-Snell" residual)}}$$

Martingale Residuals

Under the usual assumptions:

- $\mathrm{E}(M_i) = 0$ and
- $\mathrm{Cov}(M_i, M_j) = 0$ asymptotically.

If data are time-varying, then $M_i(t)$ is the "partial" martingale residual, and

$$M_i = M_i(\infty) = \sum_{t=1}^{t_i} M_i(t)$$

$$\begin{aligned}\frac{\partial \ln L(\beta)}{\partial \beta_k} &= \sum_{i=1}^{N} C_i \left\{ X_{ik} - \frac{\sum_{j \in R(t)} X_{jk} \exp(X_j \beta)}{\sum_{j \in R(t)} \exp(X_j \beta)} \right\} \\ &= \sum_{i=1}^{N} C_i (X_{ik} - \bar{X}_{w_i k}).\end{aligned}$$

$$\hat{r}_{ik} = C_i \left[X_{ik} - \frac{\sum_{j \in R(t)} X_{jk} \exp(X_j \hat{\beta})}{\sum_{j \in R(t)} \exp(X_j \hat{\beta})} \right]$$

Schoenfeld Residuals

Intuition:

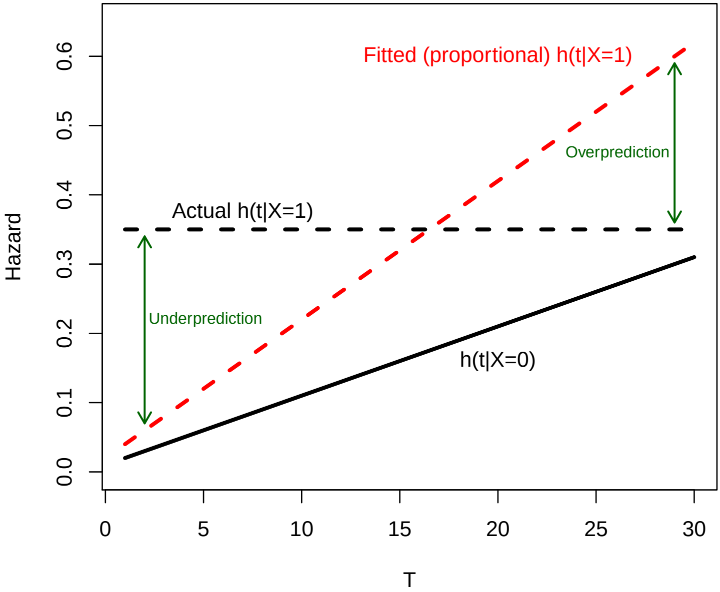
> *"(Schoenfeld residuals) ...can essentially be thought of as the observed minus the expected values of the covariate at each failure time."*

– Box-Steffensmeier and Jones (2004, 121)

Properties:

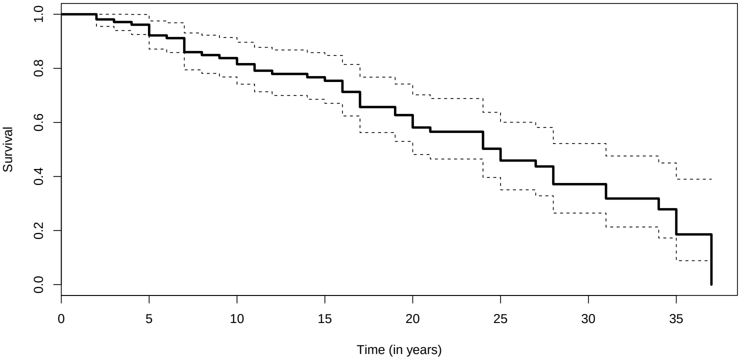
- Are defined only at event times, for non-censored observations,
- $\sum_{i=1}^{N} \hat{r}_{ik} = 0$
- $\text{Cov}(\hat{r}_{ik}, T) = 0$ if X_k's effect is proportional
- Tend to be skewed; in practice, *scaled* Schoenfeld residuals are used (see Grambsch and Therneau 1994).

Schoenfeld Residuals: Intuition



Example: Supreme Court Departures

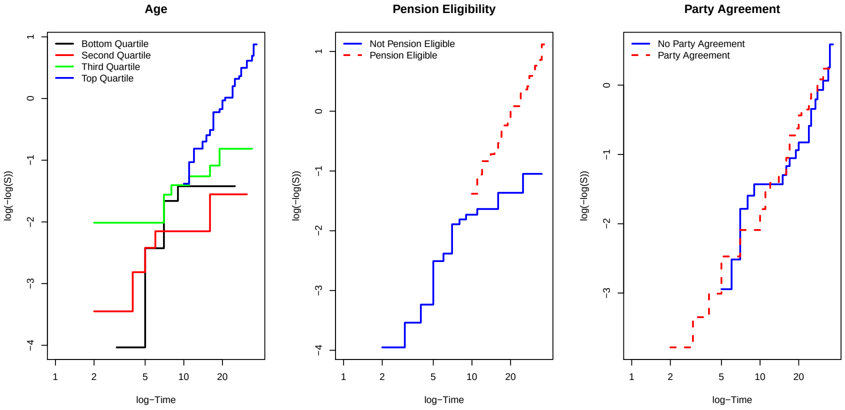
```
> summary(scotus)
   justice          service          retire            age            pension          pagree
 Min.   :  1.00   Min.   : 1.00   Min.   :0.0000   Min.   :32.0   Min.   :0.0000   Min.   :0.0000
 1st Qu.: 26.00   1st Qu.: 5.00   1st Qu.:0.0000   1st Qu.:56.0   1st Qu.:0.0000   1st Qu.:0.0000
 Median : 51.00   Median :10.00   Median :0.0000   Median :62.0   Median :0.0000   Median :1.0000
 Mean   : 52.13   Mean   :11.74   Mean   :0.0289   Mean   :62.1   Mean   :0.1989   Mean   :0.6164
 3rd Qu.: 78.00   3rd Qu.:17.00   3rd Qu.:0.0000   3rd Qu.:69.0   3rd Qu.:0.0000   3rd Qu.:1.0000
 Max.   :107.00   Max.   :37.00   Max.   :1.0000   Max.   :91.0   Max.   :1.0000   Max.   :1.0000
```



SCOTUS Departures: Cox Regression

```
> scotus.Cox<-coxph(scotus.S~age+pension+pagree,data=scotus,ties="efron")

> summary(scotus.Cox)
Call:
coxph(formula = scotus.S ~ age + pension + pagree, data = scotus,
    ties = "efron")

  n= 1765, number of events= 51

           coef exp(coef) se(coef)     z Pr(>|z|)
age     0.06395   1.06604  0.02731 2.341 0.019216 *
pension 2.05136   7.77847  0.55040 3.727 0.000194 ***
pagree  0.13748   1.14738  0.29831 0.461 0.644898
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

        exp(coef) exp(-coef) lower .95 upper .95
age         1.066     0.9381    1.0105     1.125
pension     7.778     0.1286    2.6448    22.877
pagree      1.147     0.8716    0.6394     2.059

Concordance= 0.647  (se = 0.049 )
Rsquare= 0.022   (max possible= 0.194 )
Likelihood ratio test= 38.82  on 3 df,   p=1.898e-08
Wald test            = 26.82  on 3 df,   p=6.426e-06
Score (logrank) test = 35.27  on 3 df,   p=1.068e-07
```

log-log-Survival Plots

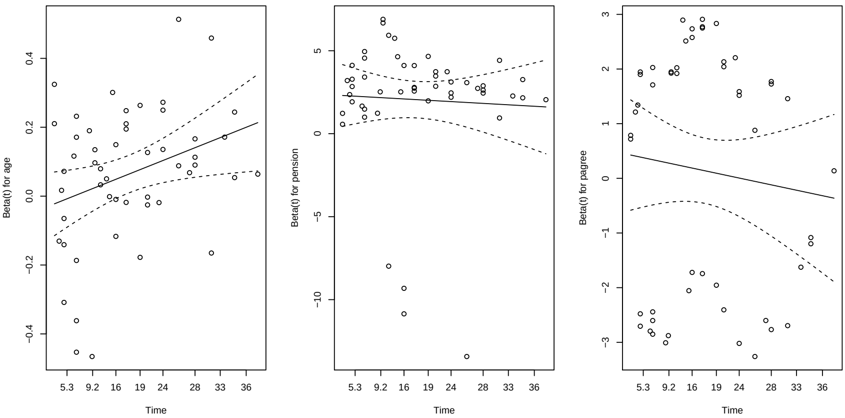


Martingale Residuals

```
>scotus$mgres<-residuals(scotus.Cox,type="martingale")

> # William Howard Taft...
> print(scotus[scotus$justice==69,])
     justice service retire age pension pagree       mgres
1173      69       1      0  63       0      1  0.00000000
1174      69       2      0  64       0      1 -0.03510077
1175      69       3      0  65       0      1 -0.01816026
1176      69       4      0  66       0      1 -0.01899776
1177      69       5      0  67       0      1 -0.07903096
1178      69       6      0  68       0      1 -0.02063125
1179      69       7      0  69       0      1 -0.11090925
1180      69       8      0  70       0      1 -0.02384340
1181      69       9      0  71       0      1 -0.02117129
1182      69      10      1  72       1      1  0.87052892

> L.Q.C. Lamar:
> print(scotus[scotus$justice==49,])
    justice service retire age pension pagree       mgres
851      49       1      0  62       0      1  0.00000000
852      49       2      0  63       0      0 -0.02869710
853      49       3      0  64       0      0 -0.01484716
854      49       4      0  65       0      0 -0.01553187
855      49       5      0  66       0      0 -0.06461280
856      49       6      0  67       0      1 -0.01935322
```

```
> # scotus.schres<-residuals(scotus.Cox,type="schoenfeld")
> # scotus.scares<-residuals(scotus.Cox,type="scaledsch")

> PHtest<-cox.zph(scotus.Cox)
> PHtest
            rho chisq      p
age      0.3160 5.359 0.0206
pension -0.0471 0.113 0.7370
pagree  -0.0962 0.504 0.4779
GLOBAL       NA 5.824 0.1205
```

Plots of Schoenfeld Residuals



log-Time Interactions

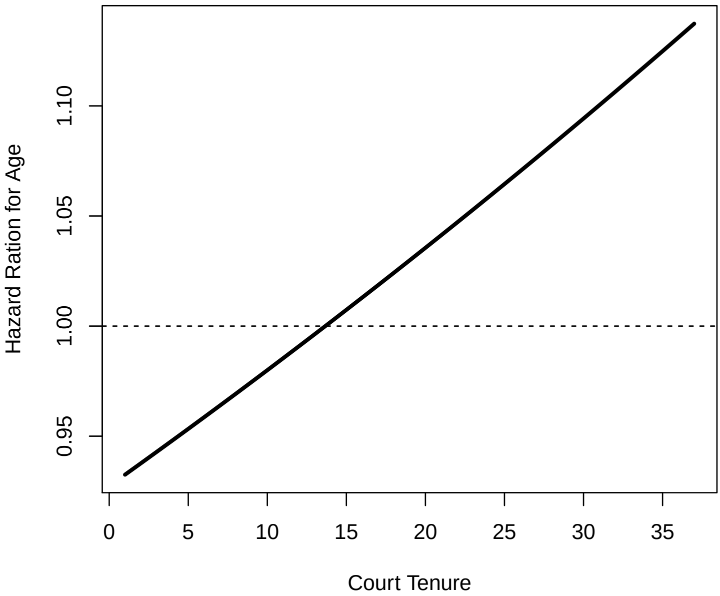
Model becomes:

$$h_i(t) = h_0(t) \exp[X_i \beta + X_i \ln(T_i) \gamma + ...]$$

- Implies that the effect of the covariate on $h(t)$ varies linearly in T
- No T term is included
- Interpretation is standard

```
> scotus$lnT<-log(scotus$service)
> scotus$ageLnT<-scotus$age*(scotus$lnT)
> scotus.NPH<-coxph(scotus.S~age+pension+pagree+ageLnT,
          data=scotus,ties="efron")
> summary(scotus.NPH)

  n= 1765, number of events= 51

            coef exp(coef) se(coef)      z Pr(>|z|)
age     -0.06988   0.93251  0.07729 -0.904 0.365933
pension  1.99866   7.37915  0.55167  3.623 0.000291 ***
pagree   0.09501   1.09966  0.30298  0.314 0.753849
ageLnT   0.05499   1.05653  0.03062  1.796 0.072552 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Concordance= 0.605  (se = 0.049 )
Rsquare= 0.023   (max possible= 0.194 )
Likelihood ratio test= 41.88  on 4 df,   p=1.768e-08
Wald test            = 28.84  on 4 df,   p=8.429e-06
Score (logrank) test = 36.55  on 4 df,   p=2.232e-07
```

Hazard Ratio Changes Over Time



More Proportional Hazards Tests

```
> PHtest2<-cox.zph(scotus.NPH)

> PHtest2
            rho   chisq     p
age     -0.1388 1.02621 0.311
pension  0.0126 0.00814 0.928
pagree  -0.1086 0.66902 0.413
ageLnT   0.1878 1.66856 0.196
GLOBAL       NA 2.58245 0.630
```

Additional Considerations

- Keele (2010): Residual-based tests for nonproportionality can also be detecting model misspecification (specifically, unmodeled nonlinearity).
- Licht (2012): Inclusion of $\ln(T)$ interactions alters the substantive interpretation of the regression results.
- Park and Hendry (2015): Residual-based tests for non-proportionality require careful attention to the transformation of the time scale T.
- Jin and Boehmke (2017): Modeling time-varying effects requires specification of time-varying covariates, even if the covariate in question does not vary over time.

Competing Risks

R multiple kinds of events:

$$T_i \in T_{i1}, ..., T_{iR}$$

Observed duration:

$$T_i = \min(T_{i1}, ... T_{iR})$$

Event indicator:

$$D_i = r \text{ iff } T_i = T_{ri}$$

R censoring indicators:

$$C_{ir} = \begin{cases} 1 \text{ if observation } i \text{ experienced event } r \\ 0 \text{ otherwise} \end{cases}$$

Likelihoods

$$L_i = f_r(T_i|\mathbf{X}_{ir}, \beta_r) \prod_{r \neq D_i} S_r(T_i|\mathbf{X}_{ir}, \beta_r)$$

$$\begin{aligned}
L &= \prod_{i=1}^{N} \left\{ f_r(T_i|\mathbf{X}_{ir}, \beta_r) \prod_{r \neq D_i} S_r(T_i|\mathbf{X}_{ir}, \beta_r) \right\} \\
&= \prod_{r=1}^{R} \prod_{i=1}^{N_r} \{ f_r(T_i|\mathbf{X}_{ir}, \beta_r)\, S_r(T_i|\mathbf{X}_{ir}, \beta_r) \} \\
&= \prod_{r=1}^{R} \prod_{i=1}^{N} [f_r(T_i|\mathbf{X}_{ir}, \beta_r)]^{C_{ir}} \, [S_r(T_i|\mathbf{X}_{ri}, \beta_r)]^{1-C_{ir}}
\end{aligned}$$

Practical Estimation

- Independent risks = separate models
- Otherwise identical estimation, interpretation, etc.
- No identification problem
- Discrete-Time $\rightarrow$ MNL
- See (e.g.) Diermeier and Stevenson 1999; Zorn and Van Winkle 2000; Goemans 2008

Independent Risks

- Key: <u>Conditional</u> independence
- $\rightsquigarrow$ Model specification
- Dependent risks:
  - Using frailties (Gordon 2002)
  - Discrete-time: strategic (Fukumoto 2009)
  - Discrete-time: bivariate probit (Quiros Flores 2012)
  - SUR?

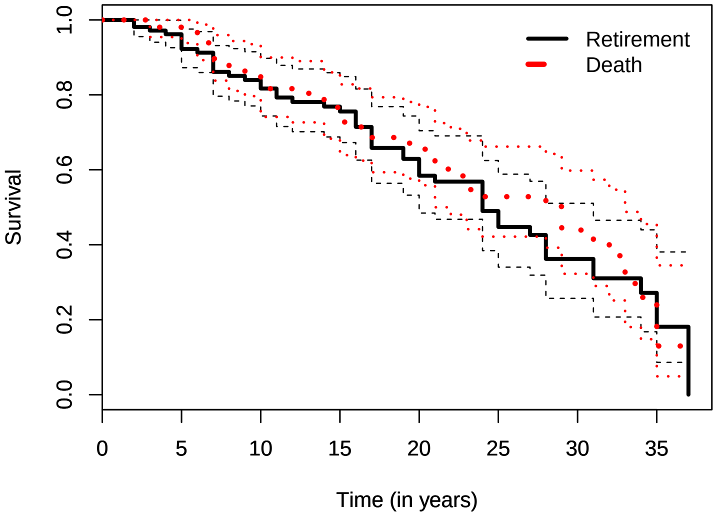
Example: SCOTUS Vacancies

- Supreme Court Vacancies, 1789-1992 ($NT = 1783$)
- Departures $\in$ {Retirement, Mortality}
- Independent competing risks models: Cox + MNL

SCOTUS Data

```
> summary(scotus)
    justice        svcstart       service         retire    
 Min.   :  1   Min.   : 0   Min.   : 1   Min.   :0.00  
 1st Qu.: 26   1st Qu.: 4   1st Qu.: 5   1st Qu.:0.00  
 Median : 51   Median : 9   Median :10   Median :0.00  
 Mean   : 53   Mean   :11   Mean   :12   Mean   :0.03  
 3rd Qu.: 79   3rd Qu.:16   3rd Qu.:17   3rd Qu.:0.00  
 Max.   :109   Max.   :36   Max.   :37   Max.   :1.00  
     death           chief           south            age    
 Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :32  
 1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:56  
 Median :0.00   Median :0.00   Median :0.00   Median :62  
 Mean   :0.03   Mean   :0.12   Mean   :0.31   Mean   :62  
 3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:69  
 Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :91  
    pension         pagree        threecat   
 Min.   :0.0   Min.   :0.00   Min.   :0.00  
 1st Qu.:0.0   1st Qu.:0.00   1st Qu.:0.00  
 Median :0.0   Median :1.00   Median :0.00  
 Mean   :0.2   Mean   :0.61   Mean   :0.08  
 3rd Qu.:0.0   3rd Qu.:1.00   3rd Qu.:0.00  
 Max.   :1.0   Max.   :1.00   Max.   :2.00  
```

SCOTUS: Death and Retirement



Independent Risks (Cox) Models

	Combined	Retirement	Death
Age	0.06	0.07	0.04
	(0.02)	(0.03)	(0.02)
Chief	−0.03	−0.23	0.09
	(0.30)	(0.44)	(0.40)
South	0.29	0.06	0.45
	(0.23)	(0.34)	(0.33)
Pension Eligibility	0.59	2.04	−0.48
	(0.28)	(0.55)	(0.41)
Party Agreement	−0.01	0.10	−0.10
	(0.21)	(0.29)	(0.31)
AIC	713.26	356.70	348.83
Num. events	99	52	47

Multinomial Logit

	Retirement	Death
Intercept	−7.77	−8.28
	(1.45)	(1.28)
Age	−0.29	0.00
	(0.45)	(0.42)
Chief	0.06	0.48
	(0.34)	(0.32)
South	0.07	0.06
	(0.03)	(0.02)
Pension Eligibility	1.40	−0.56
	(0.42)	(0.41)
Party Agreement	0.03	−0.26
	(0.30)	(0.31)
log(Time)	−0.30	0.51
	(0.27)	(0.29)
AIC	847.51	847.51
BIC	924.31	924.31
Log Likelihood	-409.75	-409.75

Cure Models

Standard models (e.g.):

$$h(T_i|\mathbf{X}_i, \beta) = \frac{f(T_i|\mathbf{X}_i, \beta)}{S(T_i|\mathbf{X}_i, \beta)}$$

assume:

$$\int_0^{\infty} f(t)\, dt = 1 \; \forall \; i.$$

All observations will (eventually) experience the event of interest.

Assume (unobserved):

$$Y_i = \begin{cases} 1 \text{ for observations that will eventually fail,} \\ 0 \text{ for those that will not.} \end{cases}$$

For observations with $Y = 1$:

$$\begin{aligned} f(T_i|\mathbf{X}_i, \beta, Y_i = 1) &= g(T|\mathbf{X}_i, \beta) \\ F(T_i|\mathbf{X}_i, \beta, Y_i = 1) &= G(T|\mathbf{X}_i, \beta) \end{aligned}$$

For observations with $Y = 0$, $f(T)$ and $F(T)$ are undefined.

Mixture Cure Model (continued)

Define:

$$\Pr(Y_i = 1) = \delta_i.$$

Overall survival is then just:

$$S_i(T) = (1 - \delta_i) + \delta_i[1 - G_i(t)]$$

Mixture Cure Model: Likelihood

Then for $C_i = 1$:

$$\begin{aligned} L_i | C_i = 1 &= \Pr(Y_i = 1) \Pr(T_i = t | Y_i = 1, \mathbf{X}_i, \beta) \\ &= \delta_i g(T_i | \mathbf{X}_i, \beta) \end{aligned}$$

For $C_i = 0$:

$$\begin{aligned} L_i | C_i = 0 &= \Pr(Y_i = 0) + \Pr(Y_i = 1) \Pr(T_i > t_i | Y_i = 1, \mathbf{X}_i, \beta) \\ &= (1 - \delta_i) + \delta_i [1 - G(T_i | \mathbf{X}_i, \beta)] \end{aligned}$$

Implies:

$$\mathbf{L} = \prod_{i=1}^{N} \left[\delta_i g(T_i|\mathbf{X}_i, \beta)\right]^{C_i} \left\{(1-\delta_i) + \delta_i \left[1 - G(T_i|\mathbf{X}_i, \beta)\right]\right\}^{(1-C_i)}$$

and:

$$\begin{aligned} ln\mathbf{L} &= \sum_{i=1}^{N} C_i \left\{\ln(\delta_i) + \ln\left[g(T_i|\mathbf{X}_i, \beta)\right]\right\} \\ &\quad +(1-C_i)\ln\left\{(1-\delta_i) + \delta_i\left[1 - G(T_i|\mathbf{X}_i, \beta)\right]\right\} \end{aligned}$$

Mixture Cure Model: Specification

Typically:

$$\delta_i = \frac{\exp(\mathbf{Z}_i\gamma)}{1 + \exp(\mathbf{Z}_i\gamma)}$$

or:

$$\delta_i = \Phi(\mathbf{Z}_i\gamma).$$

Identified even if $\mathbf{Z} \equiv \mathbf{X}$.

Non-Mixture Cure Model (e.g. Sposto 2002)

N_i = number of pre-cancerous cell clusters, with:

$$N_i \sim \text{Poisson}(\lambda).$$

Pr(Cure) is:

$$\pi_i = \Pr(N_i = 0).$$

Time to cancer onset for cluster j of observation i is:

$$Z_{ij} \sim F(t),\ j = \{1, 2, ...N_i\}.$$

Non-Mixture Cure Model (continued)

Survival to first onset:

$$S(t) = \pi^{F(t)}$$

with hazard function:

$$h(t) = -\ln(\pi) f(t)$$

which reflects the fact that $\int_0^\infty h(t)dt = -\ln(\pi)$.

Non-Mixture Cure Model (continued)

Rewritten $S(t)$:

$$S(t) = \exp[\ln(\pi)F(t)].$$

Assuming:

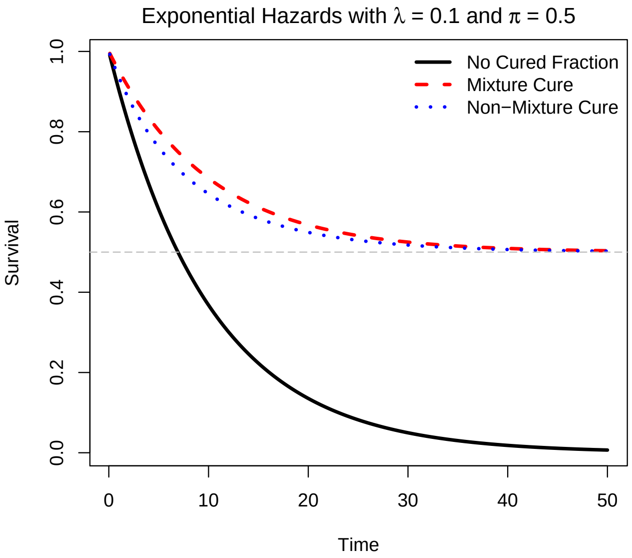
$$\pi_i = \exp[-\exp(\mathbf{X}_i\beta)]$$

we get:

$$S(t) = \exp\{[-\exp(\mathbf{X}_i\beta)]F(t)\}.$$

which is the Cox.

Mixture vs. Non-Mixture Models



- Parametric / Cox $\longrightarrow$ Poisson
- Mixture Cure Model $\longrightarrow$ Zero-Inflated Poisson
- Non-Mixture Cure Model $\longrightarrow$ "Hurdle" Poisson

Software

R

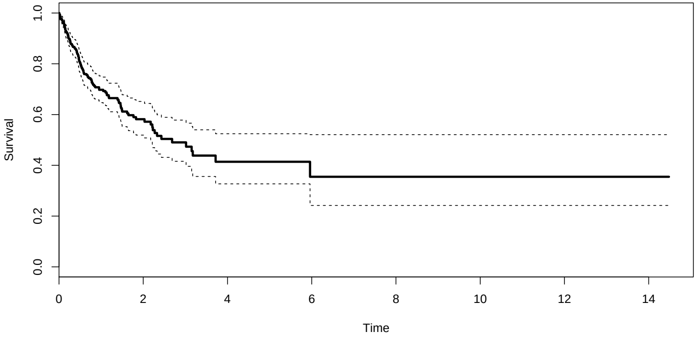
- `smcure` (semiparametric mixture models via EM)
- `semicure` (same; old)
- `nltm` (various; see Tsodikov 2003)
- `CR`, `NPHMC` (power analysis for cure models)

Stata

- `strsmix` and `strsnmix` (general parametric mixture & non-mixture cure models)
- `cureregr` (an old version)
- `lncure` (log-normal cure model)
- `spsurv` (discrete-time cure model)
- `zip / zinb` (discrete-time kludge)

A Simulated Example

```
> set.seed=7222009
> X<-rnorm(500)
> Z<-rbinom(500,1,0.5)
> T<-rweibull(500,shape=1.2,scale=1/(exp(0.5+1*X)))
> C<-rbinom(500,1,(0.4-0.3*Z))
> S<-Surv(T,C)
```

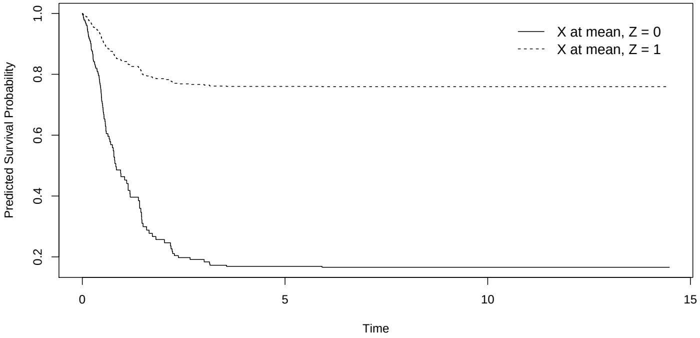


```
> coxph(S~X)
Call:
coxph(formula = S ~ X)

  coef exp(coef) se(coef)    z p
X 1.05      2.85    0.124 8.44 0

Likelihood ratio test=77.7  on 1 df, p=0  n= 500, number of events= 130


> coxph(S~X+Z)
Call:
coxph(formula = S ~ X + Z)

   coef exp(coef) se(coef)    z       p
X  1.08     2.956    0.122  8.9 0.0e+00
Z -1.59     0.204    0.230 -6.9 5.4e-12

Likelihood ratio test=140  on 2 df, p=0  n= 500, number of events= 130
```

```
> cure.fit<-smcure(S~X,cureform=~Z,data=data.cure,model="ph")

Program is running..be patient... done.

Call:
smcure(formula = S ~ X, cureform = ~Z, data = data.cure, model = "ph")

Cure probability model:
            Estimate Std.Error Z value Pr(>|Z|)
(Intercept)      1.6      0.39     4.1  3.4e-05
Z               -2.8      0.41    -6.7  2.5e-11


Failure time distribution model:
  Estimate Std.Error Z value Pr(>|Z|)
X      1.1      0.14     8.1  6.7e-16
```

An Interesting Plot

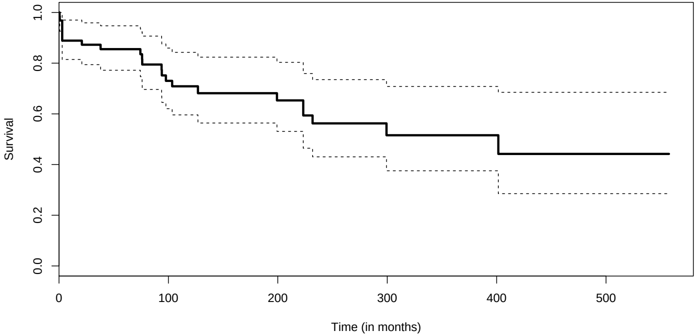
```
> cure.hat<-predictsmcure(cure.fit,c(rep(mean(X),times=2)),
                          c(0,1),model="ph")

> cure.pic<-plotpredictsmcure(cure.hat,type="S",model="ph")
```



An Example: Ceasefire Durability

Data are a subset used in Fortna (2004) (full data are here).

- $N = 63$
- Non-time-varying



Ceasefires: Cox Model

```
> CF.cox<-coxph(CF.S~tie+imposed+lndeaths+contig+onedem+twodem,
             data=CF,method="efron")
> CF.cox

Call:
coxph(formula = CF.S ~ tie + imposed + lndeaths + contig + onedem +
    twodem, data = CF, method = "efron")

            coef exp(coef) se(coef)      z       p
tie        1.845     6.327    0.557  3.314 0.00092
imposed    0.210     1.233    0.594  0.353 0.72000
lndeaths  -0.135     0.874    0.193 -0.699 0.48000
contigyes  2.898    18.143    0.948  3.058 0.00220
onedem     3.423    30.648    1.144  2.991 0.00280
twodem    -0.723     0.485    1.209 -0.598 0.55000

Likelihood ratio test=36.8  on 6 df, p=0.00000197  n= 63, number of events= 23
```

(hours of fiddling...)

A Typical Result

```
> CF.cure1.fit<-smcure(CF.S~tie+lndeaths+imposed,
                     cureform=~contig,data=CF,model="ph",
                     link="logit",emmax=500)

Program is running..be patient... done.

Call:
smcure(formula = CF.S ~ tie + lndeaths + imposed, cureform = ~contig,
    data = CF, model = "ph", link = "logit", emmax = 500)

Cure probability model:
            Estimate Std.Error Z value Pr(>|Z|)
(Intercept)     -3.4      12.4   -0.27     0.79
contig           2.1       7.4    0.28     0.78

Failure time distribution model:
         Estimate Std.Error Z value Pr(>|Z|)
tie          2.05      4.06    0.50     0.61
lndeaths    -0.37      0.34   -1.10     0.27
imposed      0.97      2.40    0.41     0.68
There were 50 or more warnings (use warnings() to see the first 50)
```

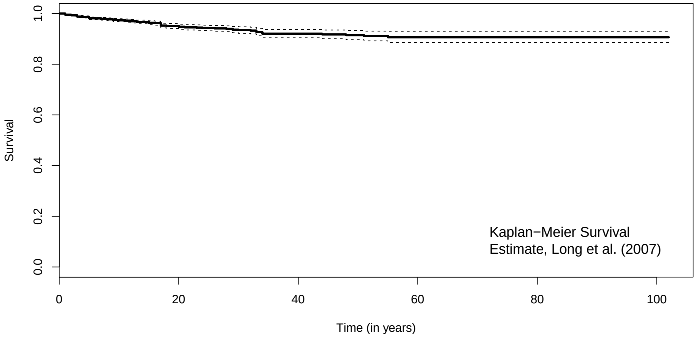
From Svolik (2008)

Consolidation status model[b]				
GDP per capita	2.121***	—	2.045***	2.121***
	(0.586)	—	(0.555)	(0.586)
GDP growth	−0.014	—	−0.048	−0.014
	(0.227)	—	(0.246)	(0.227)
Military (vs. *Not independent*)	−4.061**	—	−3.985**	−4.061**
	(1.895)	—	(1.857)	(1.895)
Civilian (vs. *Not independent*)	−0.421	—	−0.549	−0.421
	(1.097)	—	(1.067)	(1.097)
Monarchy (vs. *Not independent*)	−20.158	—	−15.844	−13.965
	(2888.609)	—	(680.185)	(891.870)
Parliamentary (vs. *Mixed*)	2.231	—	2.290	2.231
	(2.230)	—	(2.326)	(2.230)
Presidential (vs. *Mixed*)	−8.310**	—	−8.186**	−8.310**
	(3.958)	—	(4.035)	(3.958)
Intercept	−6.144**	—	−5.920**	−6.145**
	(2.646)	—	(2.644)	(2.647)

Another Example: Peace Duration

Long, Nordstrom and Baek (2007 *JOP*)

- Peace duration among allies
- Time-varying dyadic data, 1816-2001 ($NT = 57,819$)



Cox Model (replicating LNB)

```
> LNB.cox<-coxph(LNB.S~relcap+major+jdem+border+wartime+s_wt_glo+
               medarb+noagg+arbcom+organ+milinst+cluster(dyad),
               data=LNB,method="breslow")
> LNB.cox
Call:
coxph(formula = LNB.S ~ relcap + major + jdem + border + wartime +
    s_wt_glo + medarb + noagg + arbcom + organ + milinst + cluster(dyad),
    data = LNB, method = "breslow")


           coef exp(coef) se(coef) robust se      z        p
relcap   -1.431     0.239    0.614     0.683 -2.096 0.036000
major     1.137     3.118    0.241     0.280  4.064 0.000048
jdem     -0.987     0.373    0.367     0.380 -2.600 0.009300
border    1.931     6.897    0.190     0.206  9.378 0.000000
wartime  -0.359     0.699    0.367     0.467 -0.768 0.440000
s_wt_glo -0.284     0.752    0.332     0.355 -0.802 0.420000
medarb   -0.367     0.693    0.285     0.306 -1.202 0.230000
noagg    -0.463     0.630    0.126     0.152 -3.051 0.002300
arbcom    1.306     3.690    0.325     0.316  4.133 0.000036
organ     0.353     1.423    0.280     0.285  1.236 0.220000
milinst  -0.373     0.689    0.187     0.177 -2.101 0.036000
```

(hours of fiddling...)

```
> LNB.cure<-smcure(LNB.altS~relcap+major+jdem+border+wartime+s_wt_glo+
                    medarb+noagg+arbcom+organ+milinst,
                    cureform=~border,model="ph",data=LNB)

Program is running..be patient...
```

```
. stset count1, id(episode) f(buofmzmid==1)
. gen h0=0
. strsmix major jdem border wartime, bhazard(h0) distribution(weibull) link(logistic) k1
> (relcap major jdem border wartime s_wt_glo medarb noagg arbcom organ milinst)

                                                  Number of obs   =      57819
                                                  Wald chi2(4)    =      36.82
Log likelihood = -793.21263                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pi           |
       major |  -7.921296   3.764002    -2.10   0.035     -15.2986   -.5439877
        jdem |  -.6177566   .7656096    -0.81   0.420    -2.118324    .8828107
      border |  -1.943181   .3786093    -5.13   0.000    -2.685241    -1.20112
     wartime |   2.583909   1.051959     2.46   0.014     .5221065    4.645711
       _cons |   2.659179   .3980719     6.68   0.000     1.878972    3.439385
-------------+----------------------------------------------------------------
ln_lambda    |
      relcap |  -1.408332   .7129111    -1.98   0.048    -2.805613   -.0110523
       major |  -1.232928    .395653    -3.12   0.002    -2.008394   -.4574626
        jdem |   -1.69796   .4596442    -3.69   0.000    -2.598846   -.7970736
      border |   1.224114   .2622007     4.67   0.000     .7102103    1.738018
     wartime |     .42086   .4072876     1.03   0.301     -.377409    1.219129
    s_wt_glo |   -.274703   .3579769    -0.77   0.443    -.9763249    .4269188
      medarb |  -.8221547   .3503126    -2.35   0.019    -1.508755   -.1355545
       noagg |    -.68365   .1465971    -4.66   0.000     -.970975   -.3963251
      arbcom |   1.667284   .4562532     3.65   0.000     .7730438    2.561524
       organ |   .9298395   .3595899     2.59   0.010     .2250563    1.634623
     milinst |  -.4428979   .2251323    -1.97   0.049    -.8841491   -.0016468
       _cons |  -2.060399   .7260061    -2.84   0.005    -3.483344   -.6374528
-------------+----------------------------------------------------------------
ln_gamma     |
       _cons |   .0969349   .0733007     1.32   0.186    -.0467319    .2406018
------------------------------------------------------------------------------
```

Cure models...

- ...Powerful
- ...Intuitive
- ...Temperamental
- ...Ask a lot of your data

Repeated Events

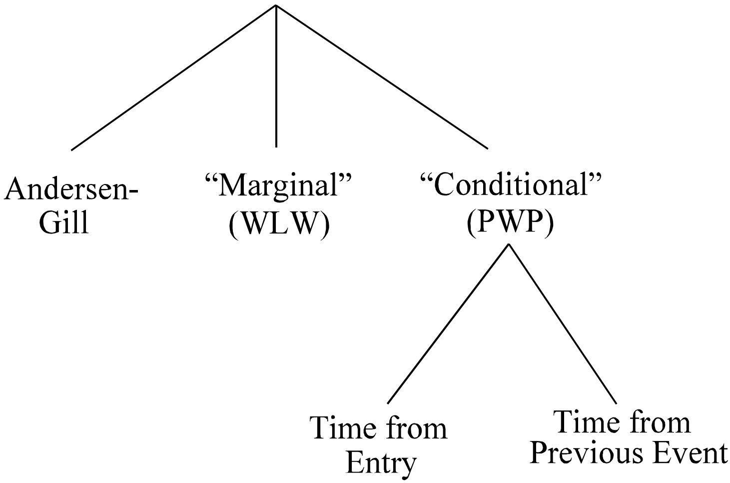
Multiple / Repeated Events

Events are not "absorbing" $\rightarrow$ capable of repetition

Raises (at least) two issues:

- Dependence across events
- Parameter variability

Variance-Correction Models



Variance Correction Model Properties

Model Property	Andersen-Gill (AG)	Marginal (WLW)	Conditional (PWP), Elapsed Time	Conditional (PWP), Gap Time
Risk Set for Event k at Time t	Independent Events	All Subjects that Haven't Experienced Event k at Time t	All Subjects that Have Experienced Event k - 1, and Haven't Experienced Event k, at Time t	
Time Scale	Duration Since Starting Observation	Duration Since Starting Observation	Duration Since Starting Observation	Duration Since Previous Event
Robust standard errors?	Yes	Yes	Yes	
Stratification by Event?	No	Yes	Yes	

Data Organization

```
> OR<-OR[order(OR$dyadid,OR$year),]
> OR$one<-rep(1,times=nrow(OR))
> OR<-ddply(OR,"dyadid",mutate,eventno=cumsum(dispute)+1,
          altstart=cumsum(one)-1,altstop=cumsum(one))

    dyadid year start stop altstart altstop dispute eventno
461   2130 1951     0    1        0       1       0       1
462   2130 1952     1    2        1       2       1       1
463   2130 1953     0    1        2       3       0       2
464   2130 1954     1    2        3       4       1       2
465   2130 1956     0    1        4       5       0       3
466   2130 1957     1    2        5       6       0       3
467   2130 1958     2    3        6       7       0       3
468   2130 1959     3    4        7       8       0       3
469   2130 1960     4    5        8       9       0       3
470   2130 1961     5    6        9      10       0       3
471   2130 1962     6    7       10      11       0       3
472   2130 1963     7    8       11      12       1       3
473   2130 1964     0    1       12      13       0       4
474   2130 1965     1    2       13      14       0       4
.
.
.
```

First Events

```
> OR1st<-OR[OR$eventno==1,]
> OR.1st<-Surv(OR1st$altstart,OR1st$altstop,OR1st$dispute)
> OR.Cox.1st<-coxph(OR.1st~allies+contig+capratio+growth+democracy+
                  trade+cluster(dyadid),data=OR1st,method="efron")

> OR.Cox.1st
Call:
coxph(formula = OR.1st ~ allies + contig + capratio + growth +
    democracy + trade + cluster(dyadid), data = OR1st, method = "efron")


            coef exp(coef) se(coef) robust se      z            p
allies    -0.448    0.6389   0.1585    0.1640 -2.732 0.0063000000
contig     1.070    2.9167   0.1681    0.1767  6.059 0.0000000014
capratio  -0.196    0.8223   0.0603    0.0779 -2.510 0.0120000000
growth    -2.198    0.1110   1.7195    1.9005 -1.157 0.2500000000
democracy -0.424    0.6547   0.1298    0.1259 -3.365 0.0007600000
trade     -6.728    0.0012  12.3255   13.9025 -0.484 0.6300000000

Likelihood ratio test=121  on 6 df, p=0  n= 17158, number of events= 205
```

```
> OR.AGS<-Surv(OR$altstart,OR$altstop,OR$dispute)
> OR.Cox.AG<-coxph(OR.AGS~allies+contig+capratio+growth+democracy+
                  trade+cluster(dyadid),data=OR,method="efron")

> OR.Cox.AG
Call:
coxph(formula = OR.AGS ~ allies + contig + capratio + growth +
    democracy + trade + cluster(dyadid), data = OR, method = "efron")


             coef  exp(coef) se(coef) robust se      z       p
allies     -0.414 0.66090755   0.1107    0.1703 -2.431 1.5e-02
contig      1.213 3.36515975   0.1209    0.1782  6.811 9.7e-12
capratio   -0.214 0.80717357   0.0514    0.0817 -2.620 8.8e-03
growth     -3.227 0.03967003   1.2279    1.3169 -2.451 1.4e-02
democracy  -0.439 0.64437744   0.0998    0.1231 -3.571 3.6e-04
trade     -13.162 0.00000192  10.3266   13.8188 -0.953 3.4e-01

Likelihood ratio test=272  on 6 df, p=0  n= 20448, number of events= 405
```

```
> OR.PWPES<-Surv(OR$altstart,OR$altstop,OR$dispute)
> OR.Cox.PWPE<-coxph(OR.PWPES~allies+contig+capratio+growth+democracy+
                     trade+strata(eventno)+cluster(dyadid),data=OR,
                     method="efron")

> OR.Cox.PWPE
Call:
coxph(formula = OR.PWPES ~ allies + contig + capratio + growth +
    democracy + trade + strata(eventno) + cluster(dyadid), data = OR,
    method = "efron")


            coef exp(coef) se(coef) robust se      z       p
allies    -0.240    0.7865   0.1122    0.1283 -1.872 6.1e-02
contig     0.868    2.3811   0.1223    0.1329  6.526 6.8e-11
capratio  -0.162    0.8506   0.0472    0.0618 -2.618 8.8e-03
growth    -3.625    0.0266   1.2371    1.2032 -3.013 2.6e-03
democracy -0.273    0.7612   0.1036    0.1074 -2.541 1.1e-02
trade     -2.514    0.0810   9.2934    9.9432 -0.253 8.0e-01

Likelihood ratio test=133  on 6 df, p=0  n= 20448, number of events= 405
```

```
> OR.PWPGS<-Surv(OR$start,OR$stop,OR$dispute)
> OR.Cox.PWPG<-coxph(OR.PWPGS~allies+contig+capratio+growth+democracy+
                      trade+strata(eventno)+cluster(dyadid),data=OR,
                      method="efron")

> OR.Cox.PWPG
Call:
coxph(formula = OR.PWPGS ~ allies + contig + capratio + growth +
    democracy + trade + strata(eventno) + cluster(dyadid), data = OR,
    method = "efron")


            coef exp(coef) se(coef) robust se     z       p
allies    -0.329    0.7193   0.1119    0.1229 -2.68 7.3e-03
contig     0.885    2.4232   0.1222    0.1285  6.89 5.6e-12
capratio  -0.171    0.8431   0.0481    0.0636 -2.68 7.3e-03
growth    -3.459    0.0315   1.2189    1.2102 -2.86 4.3e-03
democracy -0.284    0.7530   0.1028    0.1016 -2.79 5.2e-03
trade     -4.287    0.0137   9.9352   10.4592 -0.41 6.8e-01

Likelihood ratio test=139  on 6 df, p=0  n= 20448, number of events= 405
```

WLW: Data Organization

```
> OR.expand<-OR[rep(1:nrow(OR),each=max(OR$eventno)),]
> OR.expand<-ddply(OR.expand,c("dyadid","year"),mutate,
               eventrisk=cumsum(one))
> OR.expand$dispute<-ifelse(OR.expand$eventno==OR.expand$eventrisk
                        & OR.expand$dispute==1,1,0)
> dim(OR.expand)
[1] 163584     17

> head(OR.expand,9)
  dyadid year start stop futime dispute allies contig trade growth
1   2020 1951     0    1     35       0      1      1 0.014 0.0085
2   2020 1951     0    1     35       0      1      1 0.014 0.0085
3   2020 1951     0    1     35       0      1      1 0.014 0.0085
4   2020 1951     0    1     35       0      1      1 0.014 0.0085
5   2020 1951     0    1     35       0      1      1 0.014 0.0085
6   2020 1951     0    1     35       0      1      1 0.014 0.0085
7   2020 1951     0    1     35       0      1      1 0.014 0.0085
8   2020 1951     0    1     35       0      1      1 0.014 0.0085
9   2020 1952     1    2     35       0      1      1 0.015 0.0259
  democracy capratio one eventno altstart altstop eventrisk
1         1     0.20   1       1        0       1         1
2         1     0.20   1       1        0       1         2
3         1     0.20   1       1        0       1         3
4         1     0.20   1       1        0       1         4
5         1     0.20   1       1        0       1         5
6         1     0.20   1       1        0       1         6
7         1     0.20   1       1        0       1         7
8         1     0.20   1       1        0       1         8
9         1     0.19   1       1        1       2         1
```

```
> OR.expand.S<-Surv(OR.expand$altstart,OR.expand$altstop,
                   OR.expand$dispute)
> OR.Cox.WLW<-coxph(OR.expand.S~allies+contig+capratio+growth+
                     democracy+trade+strata(eventno)+
                     cluster(dyadid),data=OR.expand,
                     method="efron")

> OR.Cox.WLW
Call:
coxph(formula = OR.expand.S ~ allies + contig + capratio + growth +
    democracy + trade + strata(eventno) + cluster(dyadid), data = OR.expand,
    method = "efron")


            coef exp(coef) se(coef) robust se      z       p
allies    -0.230    0.7947   0.1122    0.1248 -1.841 6.6e-02
contig     0.852    2.3435   0.1223    0.1297  6.568 5.1e-11
capratio  -0.160    0.8524   0.0471    0.0609 -2.621 8.8e-03
growth    -3.508    0.0300   1.2370    1.1671 -3.005 2.7e-03
democracy -0.271    0.7625   0.1037    0.1055 -2.570 1.0e-02
trade     -2.656    0.0702   9.2807    9.6144 -0.276 7.8e-01

Likelihood ratio test=129  on 6 df, p=0  n= 163584, number of events= 405
```

	First	AG	PWP-E	PWP-G	WLW
Allies	−0.45	−0.41	−0.24	−0.33	−0.23
	(0.16)	(0.17)	(0.13)	(0.12)	(0.12)
Contiguity	1.07	1.21	0.87	0.89	0.85
	(0.18)	(0.18)	(0.13)	(0.13)	(0.13)
Capability Ratio	−0.20	−0.21	−0.16	−0.17	−0.16
	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)
Growth	−2.20	−3.23	−3.63	−3.46	−3.51
	(1.90)	(1.32)	(1.20)	(1.21)	(1.17)
Democracy	−0.42	−0.44	−0.27	−0.28	−0.27
	(0.13)	(0.12)	(0.11)	(0.10)	(0.11)
Trade	−6.73	−13.16	−2.51	−4.29	−2.66
	(13.90)	(13.82)	(9.94)	(10.46)	(9.61)
AIC	2538.02	5015.77	3892.77	4103.47	5597.54
Num. events	205	405	405	405	405

Parameter Change Across Events

- Values of β differ from k to $k + 1$
- Again: Institutionalization, learning, etc.
- Addressed using *strata by covariate interactions*

Parameter Change Example

```
> OR$capXevent<-OR$capratio*OR$eventno
> OR.Cox.BVary<-coxph(OR.PWPGS~allies+contig+growth+democracy+
                      trade+capratio+capXevent+strata(eventno)+
                      cluster(dyadid),data=OR,
                      method="efron")

> OR.Cox.BVary
Call:
coxph(formula = OR.PWPGS ~ allies + contig + growth + democracy +
    trade + capratio + capXevent + strata(eventno) + cluster(dyadid),
    data = OR, method = "efron")


            coef exp(coef) se(coef) robust se      z       p
allies    -0.349    0.7053   0.1120    0.1177 -2.967 3.0e-03
contig     0.897    2.4517   0.1221    0.1254  7.150 8.7e-13
growth    -3.519    0.0296   1.2196    1.2129 -2.901 3.7e-03
democracy -0.305    0.7374   0.1037    0.0972 -3.135 1.7e-03
trade     -3.297    0.0370   9.7624   10.1869 -0.324 7.5e-01
capratio  -0.340    0.7117   0.0997    0.1054 -3.227 1.2e-03
capXevent  0.135    1.1443   0.0631    0.0581  2.321 2.0e-02

Likelihood ratio test=143  on 7 df, p=0  n= 20448, number of events= 405
```

Conclusions / Recommendations

As a practical matter, estimating these models is simply a function of:

- Setting up the data correctly (so as to define the right risk sets),
- Stratifying when appropriate, and
- Calculating / using robust standard errors...

Frailty Models

"Frailty" Models

$$h_i(t) = \lambda_i(t)\nu_i$$

- $\nu_i = 1 \approx$ "baseline,"
- $\nu_i > 1 \rightarrow i$ has a greater-than-average hazard,
- $\nu_i < 1 \rightarrow$ the opposite.

More Frailty

Implies:

$$
\begin{aligned}
S(t|\nu_i) &= \exp\left[-\int_0^t h(t|\nu_i)dt\right] \\
&= \exp\left[-\int_0^t \nu_i h(t)dt\right] \\
&= \exp\left[-\int_0^t h(t)dt\right]^{\nu_i} \\
&= S(t)^{\nu_i}
\end{aligned}
$$

Typically:

- Assume $\nu_i \sim g(\nu)$, with
- $E(\nu) = 1$ and
- $\text{Var}(\nu) = \theta$

Example: Cox with Frailty

$$
\begin{aligned}
h_i(t) &= h_0(t)\nu_i \exp(\mathbf{X}_i\beta) \\
&= h_0(t)\exp(\mathbf{X}_i\beta + \alpha_i)
\end{aligned}
$$

where $\alpha_i = \ln(\nu_i)$.

(Also weibull, log-normal, etc.)

Frailty Distributions: Gamma

$$\begin{aligned} g(\nu) &= \mathcal{G}(\theta, 1/\theta) \\ &= \frac{\nu^{1/\theta - 1} \mathsf{exp}\left(\frac{-\nu}{\theta}\right)}{\theta^{(1/\theta)} \Gamma(1/\theta)} \end{aligned}$$

with

$$S_\theta(t) = \{1 - \theta \ln[S(t)]\}^{-1/\theta}$$

Frailty Distributions: Inverse-Gaussian

$$
\begin{aligned}
g(\nu) &= \mathcal{IG}(\theta, 1/\theta) \\
&= (2\pi\theta\nu^3)^{-1/2} \exp\left[-\frac{1}{2\theta}\left(\alpha - 2 + \frac{1}{\nu}\right)\right]
\end{aligned}
$$

with

$$
S_\theta(t) = \exp\left\{\frac{1}{\theta}\left[1 - \left(1 - 2\theta \ln\{S(t)\}\right)^{1/2}\right]\right\}
$$

Individual- (or *Unit*-) Specific Survival Function:

$$S(t|\nu_i) = S(t)^{\nu_i}$$

Population Average Survival Function:

$$\overline{S(t)} = \int_0^\infty S(t|\nu_i)g(\nu)d\nu$$

- Originally: E-M algorithm (e.g. Klein 1992)
- Later: Penalized Likelihood
  - Two-level iterative procedure
  - Intuition: Iterate between fitting $\hat{\beta}|\theta$ for a range of θs, and searching over the (univariate) marginal likelihood for θ to obtain $\hat{\theta}$
  - Details: Therneau and Grambsch (2000, §9.6)

Practical Matters

- Computation...

  "...if there are 300 families, each with their own frailty, and four other variables, then the full information matrix has $304^2 = 92,416$ *elements. The Cholesky decomposition must be applied to this matrix with each Newton-Raphson iteration."*

  – Therneau and Grambsch (2000, p. 258)

- Fitting choices (fix θ vs. estimation, etc.)

- Predictions / interpretation (typically assume $\hat{\nu}_i = 1$).

R

- survival: Fits a single frailty term via frailty.gamma, frailty.gaussian, or frailty.t to either Cox or parametric models.
- coxme (Cox w/Gaussian random effects; see below)
- frailtypack (parallel to frailty and coxme)
- Others (see the task view)

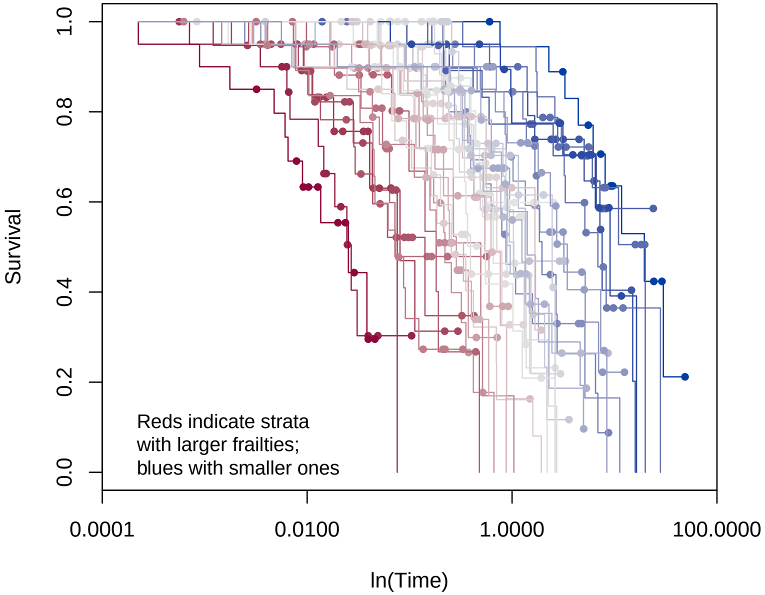
Stata

- The option shared() introduces one-level gamma-distributed frailties into stcox
- streg allows unshared or shared frailties (via frailty() and shared(), respectively) in both gamma and inverse-gaussian flavors in its parametric survival models; see Guiterrez (2002) for a good starting point.

Simulated Example

```
> set.seed(7222009)
> G<-1:40        # "groups"
> F<-rnorm(40)   # frailties
> data<-data.frame(cbind(G,F))
> data<-data[rep(1:nrow(data),each=20),]
> data$X<-rbinom(nrow(data),1,0.5)
> data$T<-rexp(nrow(data),rate=exp(0+1*data$X+(2*data$F)))
> data$C<-rbinom(nrow(data),1,0.5)
> data<-data[order(data$F),]

> S<-Surv(data$T,data$C)
```

K-M Plots By Strata



Cox Fit (No Frailty)

```
> cox.noF<-coxph(S~X,data=data)
> summary(cox.noF)
Call:
coxph(formula = S ~ X, data = data)

  n= 800, number of events= 381

   coef exp(coef) se(coef)    z   Pr(>|z|)
X 0.522     1.685    0.104 5.02 0.00000051 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
X      1.69      0.593      1.37      2.07

Concordance= 0.577  (se = 0.015 )
Rsquare= 0.031   (max possible= 0.996 )
Likelihood ratio test= 25.2  on 1 df,   p=0.000000521
Wald test            = 25.2  on 1 df,   p=0.000000508
Score (logrank) test = 25.8  on 1 df,   p=0.000000382
```

Weibull Fit (No Frailty)

```
> weib.noF<-survreg(S~X,data=data,dist="weib")
> summary(weib.noF)

Call:
survreg(formula = S ~ X, data = data, dist = "weib")
             Value Std. Error     z        p
(Intercept)  1.595     0.1450 11.00 3.92e-28
X           -1.031     0.1974 -5.22 1.76e-07
Log(scale)   0.653     0.0383 17.04 3.98e-65

Scale= 1.92

Weibull distribution
Loglik(model)= -581   Loglik(intercept only)= -594
Chisq= 27 on 1 degrees of freedom, p= 0.00000023
Number of Newton-Raphson Iterations: 5
n= 800
```

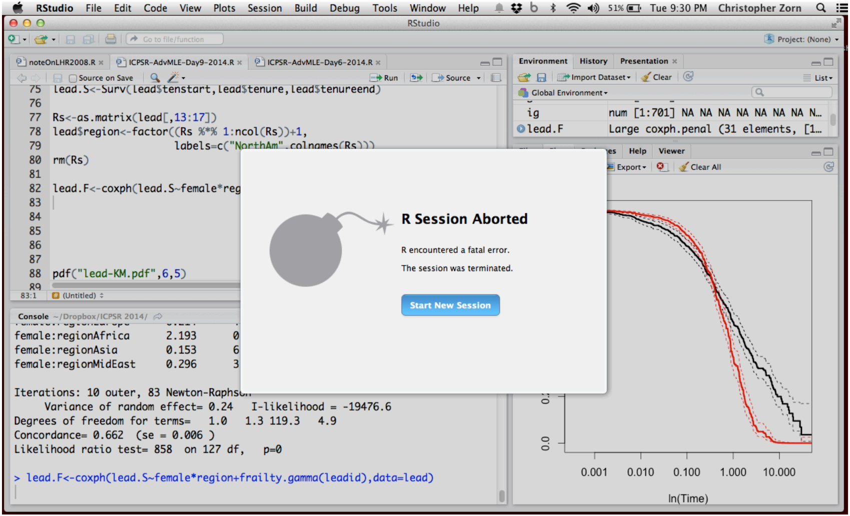
```
> cox.F<-coxph(S~X+frailty.gaussian(F),data=data)
> summary(cox.F)
Call:
coxph(formula = S ~ X + frailty.gaussian(F), data = data)

  n= 800, number of events= 381

                    coef se(coef) se2   Chisq DF   p
X                   1.01 0.112    0.112  81.9  1.0 0
frailty.gaussian(F)                     609.0 37.6 0

  exp(coef) exp(-coef) lower .95 upper .95
X      2.76      0.363      2.21      3.43

Iterations: 7 outer, 47 Newton-Raphson
     Variance of random effect= 1.8
Degrees of freedom for terms=  1.0 37.6
Concordance= 0.791  (se = 0.017 )
Likelihood ratio test= 414  on 38.5 df,   p=0
```

Weibull Fit With Frailty

```
> weib.F<-survreg(S~X+frailty.gaussian(F),data=data,dist="weib")

> summary(weib.F)

Call:
survreg(formula = S ~ X + frailty.gaussian(F), data = data, dist = "weib")
              Value Std. Error      z        p
(Intercept)  0.6188     0.2622   2.36 1.83e-02
X           -1.1386     0.1121 -10.16 3.12e-24
Log(scale)   0.0546     0.0417   1.31 1.91e-01

Scale= 1.06

Weibull distribution
Loglik(model)= -372   Loglik(intercept only)= -594
Chisq= 443 on 37 degrees of freedom, p= 0
Number of Newton-Raphson Iterations:  5 18
n= 800
```

Example: Leader Tenure

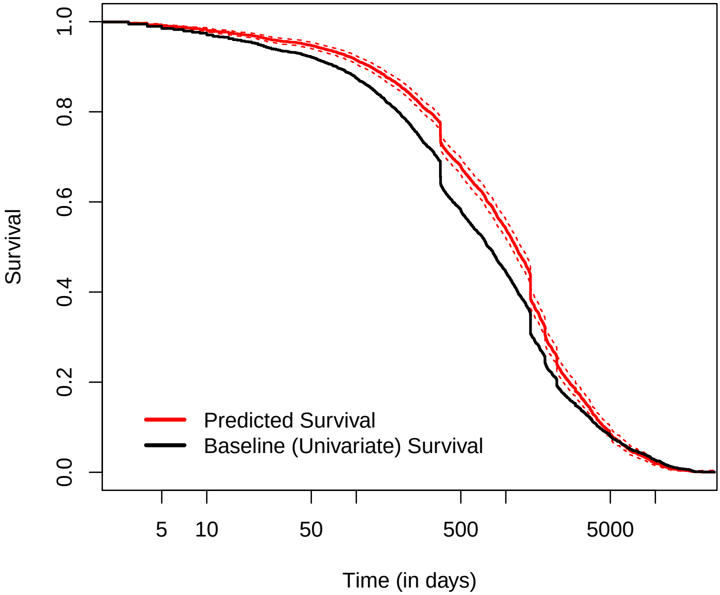
```
> lead.S<-Surv(lead$tenstart,lead$tenure,lead$tenureend)

> Rs<-as.matrix(lead[,13:17])
> lead$region<-factor((Rs %*% 1:ncol(Rs))+1,
                  labels=c("NorthAm",colnames(Rs)))
> rm(Rs)

> lead.F<-coxph(lead.S~female*region+frailty.gamma(leadid),data=lead)
```


R Session Aborted
R encountered a fatal error.
The session was terminated.
Start New Session
lead.F<-coxph(lead.S~female*region+frailty.gamma(leadid),data=lead)

Let's Try That Again

```
> lead.F<-coxph(lead.S~female*region+frailty.gamma(ccode),data=lead)
Warning message:
In coxpenal.fit(X, Y, strats, offset, init = init, control, weights = weights,  :
  Inner loop failed to coverge for iterations 2 3

> summary(lead.F)
Call:
coxph(formula = lead.S ~ female * region + frailty.gamma(ccode),
    data = lead)

  n= 15222, number of events= 2806
   (22 observations deleted due to missingness)

                     coef    se(coef) se2    Chisq  DF  p
female                1.2427 0.462    0.4594   7.24   1 0.007100
regionLatinAm        -0.1259 0.208    0.0333   0.37   1 0.540000
regionEurope          0.0414 0.160    0.0545   0.07   1 0.800000
regionAfrica         -0.7047 0.160    0.0840  19.45   1 0.000010
regionAsia           -0.3896 0.164    0.0742   5.65   1 0.017000
regionMidEast        -0.7478 0.186    0.0986  16.13   1 0.000059
frailty.gamma(ccode)                         523.81 119 0.000000
female:regionLatinAm -1.8826 0.851    0.8495   4.89   1 0.027000
female:regionEurope  -1.5424 0.624    0.6212   6.11   1 0.013000
female:regionAfrica   0.7854 0.861    0.8556   0.83   1 0.360000
female:regionAsia    -1.8765 0.572    0.5666  10.76   1 0.001000
female:regionMidEast -1.2175 0.861    0.8551   2.00   1 0.160000

Iterations: 10 outer, 83 Newton-Raphson
     Variance of random effect= 0.24   I-likelihood = -19476.6
Degrees of freedom for terms=   1.0   1.3 119.3   4.9
Concordance= 0.662  (se = 0.006 )
Likelihood ratio test= 858  on 127 df,   p=0
```

Predicted vs. Actual



Extensions: Mixed-Effects Survival Models

- HLMs for survival data / outcomes
- Combined fixed, random, and mixed effects (random-coefficient) models
- R: Implemented in `coxme`
- Stata: `stmixed` (parametric models)
- Terry Therneau has a nice vignette

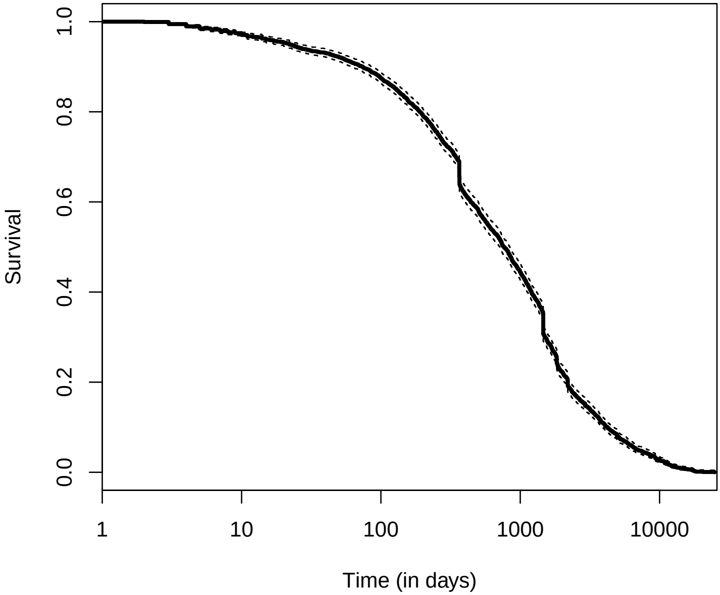
Mixed Effects Example

```
> lead.coxME<-coxme(lead.S~female + (1 | ccode/female),data=lead)
> lead.coxME
Cox mixed-effects model fit by maximum likelihood
  Data: lead
  events, n = 2806, 15222 (22 observations deleted due to missingness)
  Iterations= 38 160
                 NULL Integrated Fitted
Log-likelihood -19738     -19505 -19314

                  Chisq  df p AIC  BIC
Integrated loglik   465   3 0 459  441
 Penalized loglik   849 129 0 590 -177

Model:  lead.S ~ female + (1 | ccode/female)
Fixed coefficients
        coef exp(coef) se(coef)     z    p
female -0.07      0.93     0.22 -0.31 0.75

Random effects
 Group        Variable    Std Dev Variance
 ccode/female (Intercept) 0.279   0.078
 ccode        (Intercept) 0.487   0.237
```

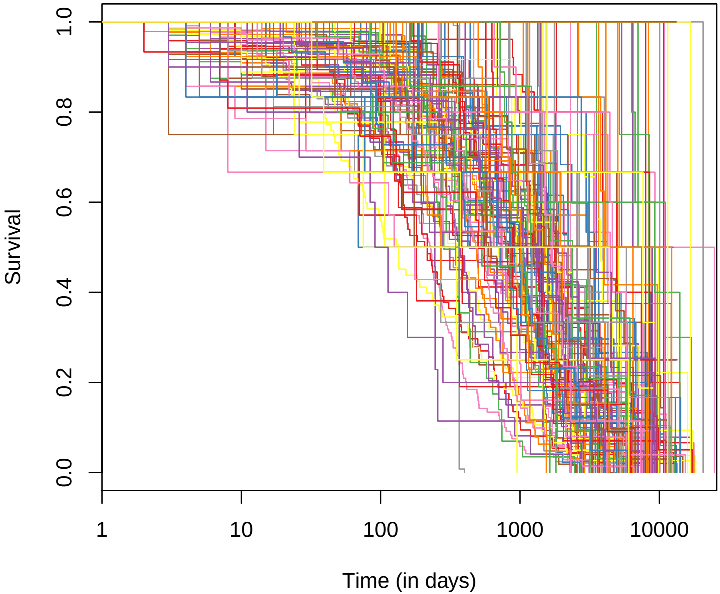
Stratify? Frailties? Clustering?

- Stratification $\approx$ "fixed effects"
- Frailties $\approx$ "random effects"
- "Robust" / cluster $\approx$ GEE / PCSEs, etc.
- Not all combinations are possible, or make sense

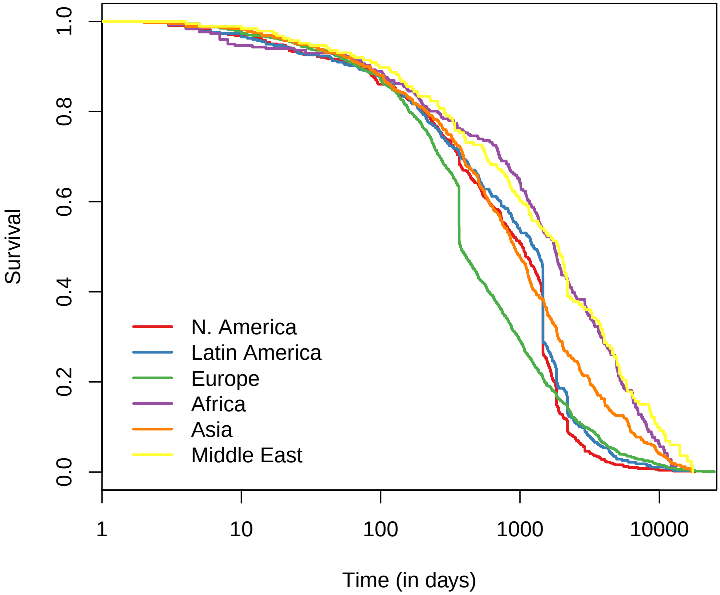
K-M Plot: Leaders



K-M Plot: Leaders (by country)



K-M Plot: Leaders (by region)



Strata + Frailty

```
> lead.Fstrat<-coxph(lead.S~female*strata(region)+
                      frailty.gamma(ccode),data=lead)
Warning message:
In coxpenal.fit(X, Y, strats, offset, init = init, control, weights = weights,  :
  Inner loop failed to coverge for iterations 2 3 4

> summary(lead.Fstrat)
Call:
coxph(formula = lead.S ~ female * strata(region) + frailty.gamma(ccode),
    data = lead)

  n= 15222, number of events= 2806
   (22 observations deleted due to missingness)

                          coef  se(coef) se2    Chisq  DF  p
female                     1.46 0.463     0.461   9.88   1 0.00170
frailty.gamma(ccode)                             594.82 121 0.00000
female:strata(region)regi -2.20 0.853     0.851   6.63   1 0.01000
female:strata(region)regi -1.75 0.625     0.623   7.81   1 0.00520
female:strata(region)regi  0.13 0.869     0.864   0.02   1 0.88000
female:strata(region)regi -2.07 0.573     0.568  13.04   1 0.00031
female:strata(region)regi -1.31 0.862     0.857   2.32   1 0.13000
```

Strata + Clustering

```
> lead.stratCl<-coxph(lead.S~female*strata(region)+
                      cluster(ccode),data=lead)

> summary(lead.stratCl)
Call:
coxph(formula = lead.S ~ female * strata(region) + cluster(ccode),
    data = lead)

  n= 15222, number of events= 2806
   (22 observations deleted due to missingness)

                                      coef exp(coef) se(coef) robust se     z
female                               1.234     3.436    0.453     0.288  4.28
female:strata(region)region=LatinAm -1.881     0.152    0.842     0.627 -3.00
female:strata(region)region=Europe  -1.618     0.198    0.610     0.415 -3.90
female:strata(region)region=Africa   0.473     1.605    0.849     0.382  1.24
female:strata(region)region=Asia    -1.711     0.181    0.555     0.342 -5.00
female:strata(region)region=MidEast -0.709     0.492    0.846     0.349 -2.03

Concordance= 0.503  (se = 0.002 )
Rsquare= 0.001   (max possible= 0.864 )
Likelihood ratio test= 13.8  on 6 df,   p=0.0323
Wald test            = 81.6  on 6 df,   p=1.67e-15
Score (logrank) test = 20.1  on 6 df,   p=0.00263,   Robust = 14.4  p=0.0255

  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the Wald and robust score tests do not).
```

Choices...

From the `frailty` documentation:

"Note that use of a frailty term implies a mixed effects model and use of a cluster term implies a GEE approach; these cannot be mixed."

```
> lead.FstratCl<-coxph(lead.S~female*strata(region)+frailty.gamma(ccode)+
                     cluster(ccode),data=lead)
Error in residuals.coxph(fit2, type = "dfbeta", collapse = cluster,
weighted = TRUE) :
  length of 'dimnames' [2] not equal to array extent
In addition: Warning message:
In coxpenal.fit(X, Y, strats, offset, init = init, control, weights = weights,
  Inner loop failed to coverge for iterations 2 3 4
```

Therneau, Terry M., Jun 27, 2011; 8:02am **Re: cluster() or frailty() in coxph**

693 posts

In reply to this post by Ehsan Karim

Addition of a cluster() term fits a Generalized Estimating Equations (GEE) type of model, addition of frailty() fits a random effects model (Mixed Effect or ME). In glm analysis (linear regression, logistic regression, etc) the arguments about the advantages/disadvantages of GEE ve ME would easily fill a volume. Most of this argument carries over to the coxph case; I find both approaches useful.

Caveats:

1. Coxph with cluster() only allows the "working independence" variance structure. The details for other variance structures were worked out by Alicia Z in her Iowa State PhD thesis, but I've never gotton around to implementing it.
2. For random effects, the coxme function is preferred.
3. In comparing GEE and ME one part of the arguement is that the former model is "marginal" and the second "conditional", and thus the coefficients from the models mean different things. I take this with a grain of salt. Remember that ALL models are wrong.

Terry Therneau

____

[hidden email] mailing list
https://stat.ethz.ch/mailman/listinfo/r-help
PLEASE do read the posting guide http://www.R-project.org/posting-guide.html
and provide commented, minimal, self-contained, reproducible code.

Topics We Didn't Cover

- ⋆ Joint Models for Survival and Longitudinal Outcomes
  - e.g., survival + binary / multinomial / continuous variables
  - *inter alia* R package `JM` (Rizopolous 2010)
  - Recent reference is Viviani et al. (2014)
- ⋆ Causal Inference (IVs, RDDs, matching, etc.)
- ⋆ Variable Selection: regularization, bagging, boosting, stacking, lasso, etc.
- ⋆ Bayesian approaches (esp. for high-dimensional competing risks & hierarchical models); see Ibrahim et al. (2005)
- ⋆ New / better tools for interpretation and graphics (e.g. `simPH`)

General Tips

Journals:

- *Biometrics* / *Biometrika*
- *Statistics in Medicine*
- *Statistical Methods in Medical Research*
- *Lifetime Data Analysis*

Places:

- Biostatistics / Epidemiology / Public Health
- Statistics departments
- *Not* economics, psychology, etc.