

# ICPSR 2017 “Advanced Maximum Likelihood”: Survival Analysis

## Day Three

August 9, 2017

Basic idea:

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

Note:

- $h_0(t) \equiv h(t|\mathbf{X} = 0)$
- Changes in  $\mathbf{X}$  shift  $h(t)$  *proportionally*

## Cox (1972) (continued)

$$\begin{aligned}\text{HR} &= \frac{h_0(t)\exp(X_1\hat{\beta})}{h_0(t)\exp(X_0\hat{\beta})} \\ &= \exp[(1 - 0)\hat{\beta}] \\ &= \exp(\hat{\beta})\end{aligned}$$

# Cox (1972) (continued)

Also, because

$$S(t) = \exp[-H(t)]$$

then

$$\begin{aligned} S(t) &= \exp \left[ - \int_0^t h(t) dt \right] \\ &= \exp \left[ - \exp(\mathbf{X}_i \beta) \int_0^t h_0(t) dt \right] \\ &= \left[ \exp \left( - \int_0^t h_0(t) dt \right) \right]^{\exp(\mathbf{X}_i \beta)} \\ &= [S_0(t)]^{\exp(\mathbf{X}_i \beta)} \end{aligned}$$

# Partial Likelihood

Assume  $N_C$  distinct event times  $t_j$ , with no “ties.”

Then:

$$\begin{aligned} & \Pr(\text{Individual } k \text{ experienced the event at } t_j \mid \text{One observation experienced the event at } t_j) \\ &= \frac{\Pr(\text{At-risk observation } k \text{ experiences the event of interest at } t_j)}{\Pr(\text{One at-risk observation experiences the event of interest at } t_j)} \\ &= \frac{h_k(t_j)}{\sum_{\ell \in R_j} h_\ell(t_j)} \end{aligned}$$

# Partial Likelihood (continued)

$$\begin{aligned} L_i &= \frac{h_0(t_j) \exp(\mathbf{X}_i \beta)}{\sum_{\ell \in R_j} h_0(t_j) \exp(\mathbf{X}_\ell \beta)} \\ &= \frac{h_0(t_j) \exp(\mathbf{X}_i \beta)}{h_0(t_j) \sum_{\ell \in R_j} \exp(\mathbf{X}_\ell \beta)} \\ &= \frac{\exp(\mathbf{X}_i \beta)}{\sum_{\ell \in R_j} \exp(\mathbf{X}_\ell \beta)} \end{aligned}$$

$$L = \prod_{i=1}^N \left[ \frac{\exp(\mathbf{X}_i \beta)}{\sum_{\ell \in R_j} \exp(\mathbf{X}_\ell \beta)} \right]^{C_i}$$

$$\ln L = \sum_{i=1}^N C_i \left\{ \mathbf{X}_i \beta - \ln \left[ \sum_{\ell \in R_j} \exp(\mathbf{X}_\ell \beta) \right] \right\}$$

# Notes on Partial Likelihood

- PL is
  - Consistent
  - Asymptotically normal
  - Slightly inefficient (but asymptotically efficient)
- Considers order of events, but not actual duration
- Censored events: Modify  $R_j$
- No ties

# Example: Interstate War, 1950-1985

- Dyad-years for “politically-relevant” dyads
- $N = 827$ ,  $NT = 20448$ .
- Covariates:
  - Whether (=1) or not the two countries are *allies*,
  - Whether (=1) or not the two countries are *contiguous*,
  - The *capability ratio* of the two countries,
  - The lower of the two countries' (GDP) *growth* (rescaled),
  - The lower of the two countries' *democracy* (POLITY IV) scores (rescaled to [-1,1]), and
  - The amount of *trade* between the two countries, as a fraction of joint GDP.

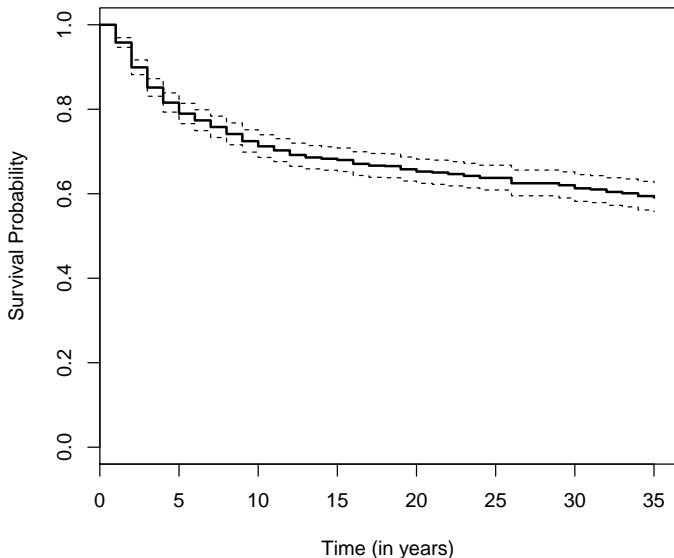


# The Data

```
> summary(OR)
```

dyadid	year	start	stop	futime
Min. : 2020	Min. :1951	Min. : 0.00	Min. : 1.00	Min. : 5.00
1st Qu.:100365	1st Qu.:1965	1st Qu.: 5.00	1st Qu.: 6.00	1st Qu.:23.00
Median :220235	Median :1972	Median :11.00	Median :12.00	Median :31.00
Mean :253305	Mean :1971	Mean :12.32	Mean :13.32	Mean :28.97
3rd Qu.:365600	3rd Qu.:1979	3rd Qu.:19.00	3rd Qu.:20.00	3rd Qu.:35.00
Max. :900920	Max. :1985	Max. :34.00	Max. :35.00	Max. :35.00
dispute	allies	contig	trade	
Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.00000	
1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000	
Median :0.00000	Median :0.0000	Median :0.0000	Median :0.00020	
Mean :0.01981	Mean :0.3563	Mean :0.3099	Mean :0.00231	
3rd Qu.:0.00000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.00120	
Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :0.17680	
growth	democracy	capratio		
Min. : -0.264900	Min. : -1.0000	Min. : 0.0100		
1st Qu.: -0.004800	1st Qu.: -0.8000	1st Qu.: 0.0462		
Median : 0.014700	Median : -0.7000	Median : 0.2220		
Mean : 0.007823	Mean : -0.3438	Mean : 1.6677		
3rd Qu.: 0.027800	3rd Qu.: 0.2000	3rd Qu.: 1.1560		
Max. : 0.164700	Max. : 1.0000	Max. :78.9296		

# The Data (Kaplan-Meier plot)



## R:

- `coxph` in `survival` (preferred)
- `cph` in `design`
- Plots: `plot(survfit(PHobject))`

## Stata:

- Basic command = `stcox`
- `stset` first
- Options: `robust`, various methods for ties, postestimation commands

# Model Fitting

```
> ORCox.br<-coxph(OR.S~allies+contig+capratio+growth+democracy+trade,  
                  data=OR,na.action=na.exclude,method="breslow")
```

```
> summary(ORCox.br)
```

```
n= 20448, number of events= 405
```

	coef	exp(coef)	se(coef)	z	Pr(> z )	
allies	-0.34849	0.70576	0.11096	-3.141	0.001686	**
contig	0.94861	2.58213	0.12173	7.793	6.55e-15	***
capratio	-0.22303	0.80009	0.05164	-4.319	1.57e-05	***
growth	-3.69487	0.02485	1.19950	-3.080	0.002068	**
democracy	-0.38194	0.68254	0.09915	-3.852	0.000117	***
trade	-3.22857	0.03961	9.45588	-0.341	0.732776	

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
.  
.  
.
```

# Model Fitting (continued)

```
.  
.   
.   
exp(coef) exp(-coef) lower .95 upper .95  
allies      0.70576      1.4169 5.678e-01 8.772e-01  
contig      2.58213      0.3873 2.034e+00 3.278e+00  
capratio    0.80009      1.2499 7.231e-01 8.853e-01  
growth      0.02485     40.2402 2.368e-03 2.608e-01  
democracy   0.68254      1.4651 5.620e-01 8.289e-01  
trade       0.03961     25.2436 3.540e-10 4.433e+06
```

Concordance= 0.714 (se = 0.015 )

Rsquare= 0.01 (max possible= 0.234 )

Likelihood ratio test= 210.3 on 6 df, p=0

Wald test = 159.8 on 6 df, p=0

Score (logrank) test = 185.8 on 6 df, p=0

# Interpretation: Hazard Ratios

$$HR = \exp[(\mathbf{X}_j - \mathbf{X}_k)\hat{\beta}]$$

Means:

- $HR = 1 \Leftrightarrow \hat{\beta} = 0$
- $HR > 1 \Leftrightarrow \hat{\beta} > 0$
- $HR < 1 \Leftrightarrow \hat{\beta} < 0$

$$\text{Percentage difference} = 100 \times \{\exp[(\mathbf{X}_j - \mathbf{X}_k)\hat{\beta}] - 1\}.$$

# Example: Hazard Ratios

From above:

	exp(coef)	exp(-coef)	lower .95	upper .95
allies	0.70576	1.4169	5.678e-01	8.772e-01
contig	2.58213	0.3873	2.034e+00	3.278e+00
capratio	0.80009	1.2499	7.231e-01	8.853e-01
growth	0.02485	40.2402	2.368e-03	2.608e-01
democracy	0.68254	1.4651	5.620e-01	8.289e-01
trade	0.03961	25.2436	3.540e-10	4.433e+06

Interpretation:

- Countries which are *allies* have an expected  $(0.706 - 1) \times 100 = 29.4$  percent lower hazard of conflict than those that are not.
- *Contiguous* countries have  $(2.582 - 1) \times 100 = 158$  percent higher hazards of conflict than non-contiguous ones.
- A one-unit increase in *democracy* corresponds to a  $(0.683 - 1) \times 100 = 31.7$  percent decrease in the expected hazard of conflict.

# Hazard Ratios: Scaling Covariates

It is good for one-unit changes to be meaningful / realistic...

```
> OR$growthPct<-OR$growth*100  
> summary(coxph(OR.S~allies+contig+capratio+growthPct+democracy+trade,  
               data=OR,na.action=na.exclude, method="breslow"))
```

```
.  
.   
.   
exp(coef) exp(-coef) lower .95 upper .95  
allies      0.70576      1.4169 5.678e-01 8.772e-01  
contig      2.58213      0.3873 2.034e+00 3.278e+00  
capratio    0.80009      1.2499 7.231e-01 8.853e-01  
growthPct   0.96373      1.0376 9.413e-01 9.867e-01  
democracy   0.68254      1.4651 5.620e-01 8.289e-01  
trade       0.03961     25.2436 3.540e-10 4.433e+06
```

Note:

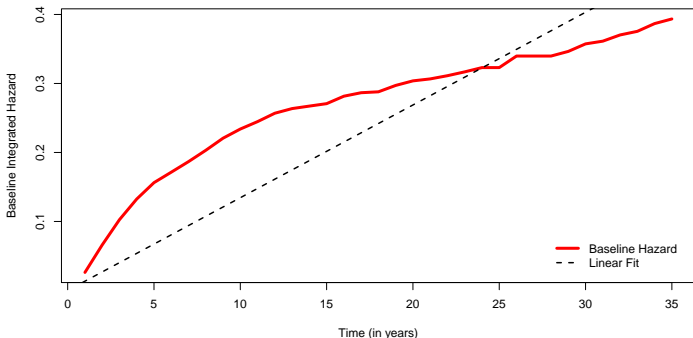
- Previous HR for growth = 0.02485  $\rightarrow$  97.5 percent decrease in  $\hat{h}(t)$
- HR for growthPct is now 0.964; 1 unit increase  $\rightarrow$  4% decrease in  $\hat{h}(t)$
- Same result, proportionally:  $0.96373^{100} = 0.02485$



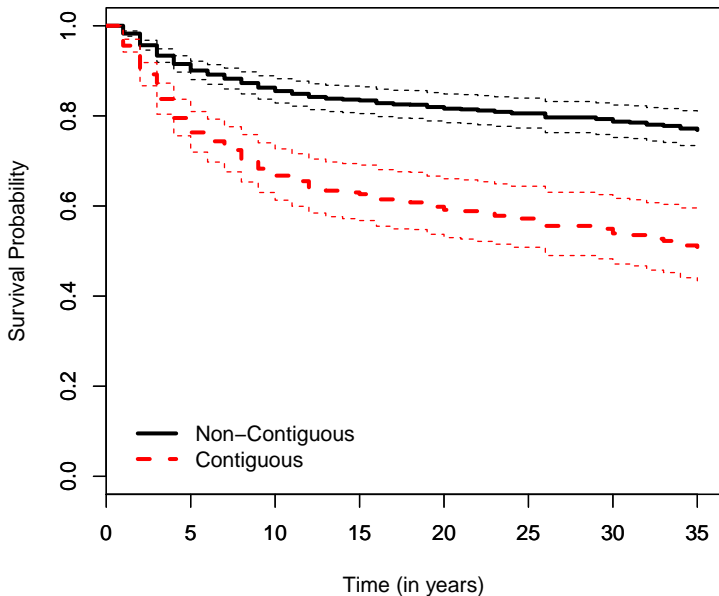
# Baseline Hazards

Because the Cox model is semiparametric, it uses a conventional / univariate (Nelson-Aalen) estimate of the “baseline” hazard:

```
OR.BH<-basehaz(ORCox.br,centered=FALSE)
```



# Comparing Survival Curves



$\exists$  ties...

Their presence biases Cox  $\hat{\beta}$ s toward zero.

- Call  $d_j > 0$  the number of events occurring at  $t_j$ , and
- $D_j$  the set of  $d_j$  observations that have the event at  $t_j$ .

# Ties (continued)

Means of handling ties:

- Breslow:

$$L_{\text{Breslow}}(\beta) = \prod_{i=1}^N \frac{\exp \left[ \left( \sum_{q \in D_j} \mathbf{x}_q \right) \beta \right]}{\left[ \sum_{\ell \in R_j} \exp(\mathbf{x}_\ell \beta) \right]^{d_j}}$$

- Efron

$$\ln L_{\text{Efron}}(\beta) = \sum_{j=1}^J \sum_{i \in D_j} \left\{ \mathbf{x}_i \beta - \frac{1}{d_j} \sum_{k=1}^{d_j-1} \ln \left[ \sum_{\ell \in R_j} \exp(\mathbf{x}_\ell \beta) \right] - k \left( \frac{1}{d_j} \sum_{\ell \in D_j} \exp(\mathbf{x}_\ell \beta) \right) \right\}$$

# Ties (continued)

- “Exact” (partial likelihood)

$$\ln L_{\text{Exact}}(\beta) = \sum_{j=1}^J \left\{ \sum_{i \in R_j} \delta_{ij}(\mathbf{X}_i; \beta) - \ln[f(r_j, d_j)] \right\}$$

where

$$f(r, d) = g(r-1, d) + g(r-1, d-1) \exp(\mathbf{X}_k \beta),$$

$k = r$ th observation in  $R_j$ ,

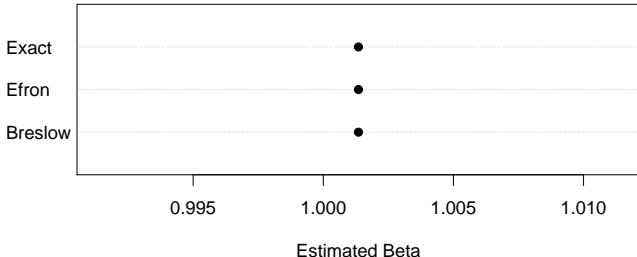
$r_j =$  cardinality of  $R_j$ , and

$$g(r, d) = \begin{cases} 0 & \text{if } r < d, \\ 1 & \text{if } d = 0 \end{cases}$$

# Ties: Example

```
set.seed(7222009)
Data<-as.data.frame(cbind(c(rep(1,times=400)),
                           c(rep(c(0,1),times=200))))
colnames(Data)<-c("C", "X")
Data$T<-rexp(400,exp(0+1*Data$X)) # B = 1.0
Data.S<-Surv(Data$T,Data$C)

D.br<-coxph(Data.S~X,data=Data,method="breslow")
D.ef<-coxph(Data.S~X,data=Data,method="efron")
D.ex<-coxph(Data.S~X,data=Data,method="exact")
```



# Ties: Example (continued)

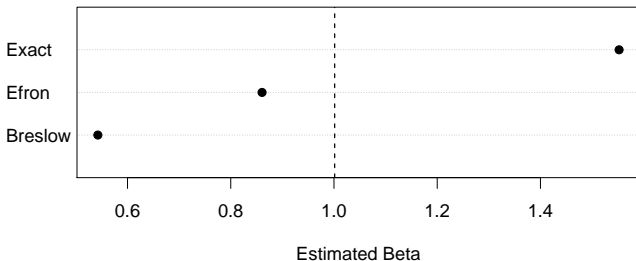
```
Data$Tied<-round(Data$T,0)
```

```
DataT.S<-Surv(Data$Tied,Data$C)
```

```
DT.br<-coxph(DataT.S~X,data=Data,method="breslow")
```

```
DT.ef<-coxph(DataT.S~X,data=Data,method="efron")
```

```
DT.ex<-coxph(DataT.S~X,data=Data,method="exact")
```



# Ties: Practical Advice

- All approx. are identical if  $\nexists$  ties
- Few ties = similar results
- When ties are present, Breslow  $<$  Efron  $<$  “Exact” methods
- If you want to learn more about ties in the Cox model, [read this](#).



# Cox vs. Parametric Models

Conceptual considerations:

- Theory
- Nature of  $h(t)$
- Relative importance: Bias vs. efficiency
- Need / willingness for out-of-sample predictions / forecasting

# Cox, On His Model

Reid: “What do you think of the cottage industry that’s grown up around [the Cox model]?”

Cox: “In the light of further results one knows since, I think I would normally want to tackle the problem parametrically... I’m not keen on non-parametric formulations normally.”

Reid: “So if you had a set of censored survival data today, you might rather fit a parametric model, even though there was a feeling among the medical statisticians that that wasn’t quite right.”

Cox: “That’s right, but since then various people have shown that the answers are very insensitive to the parametric formulation of the underlying distribution. And if you want to do things like predict the outcome for a particular patient, it’s much more convenient to do that parametrically.”

– From [Reid \(1994\)](#).

# Extensions (future classes)

- Stratification
- Cox Models for *repeated events*
- Models with “frailties”
- Competing risks / “cured”  
subpopulations
- etc.