ICPSR 2017 "Advanced Maximum Likelihood": Survival Analysis

Day Eight

August 16, 2017

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.

Mixture Cure Model

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 $C_i = 0$:

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

Mixture Cure Model: Likelihood

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:

$$\pi_i = \text{Pr}(N_i = 0).$$

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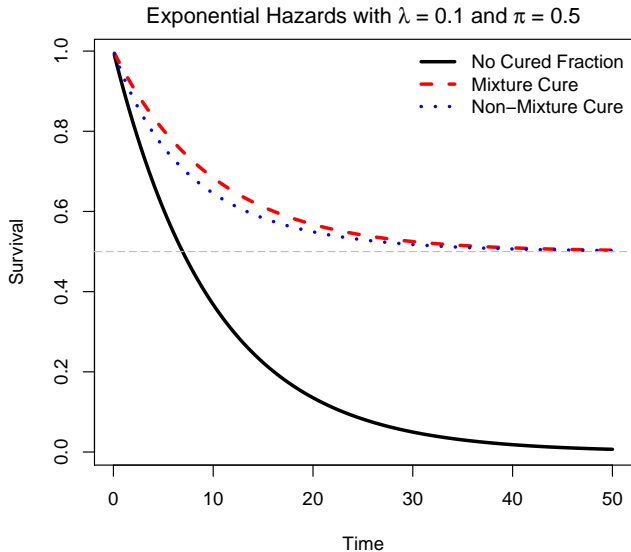
$$S(t) = \exp[\ln(\pi)F(t)].$$

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

Exponential Hazards with $\lambda = 0.1$ and $\pi = 0.5$

Legend:
- No Cured Fraction
- Mixture Cure
- Non−Mixture Cure

Survival (y-axis) vs Time (x-axis)

Discrete-Time Cure 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)
```
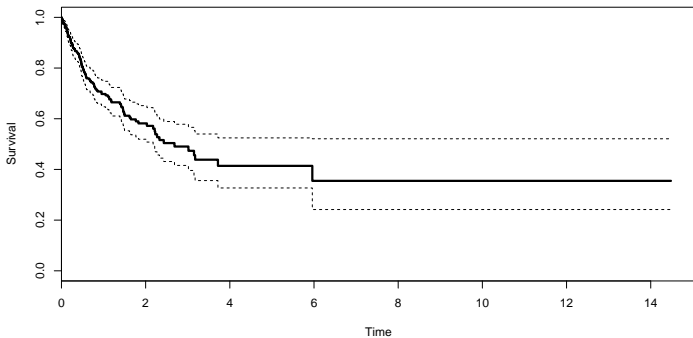
Cox Models

```
> 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 Model

```
> 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
```
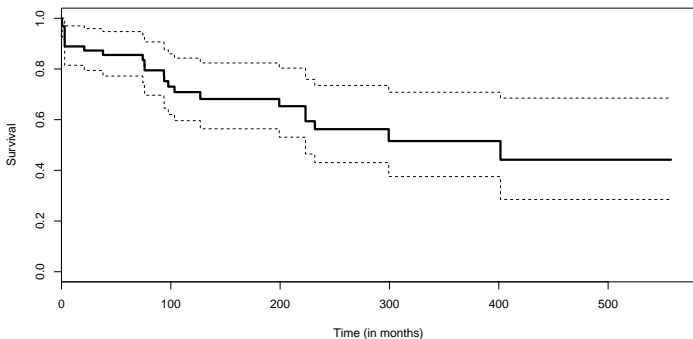
An Interesting Plot

```
> 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")
```
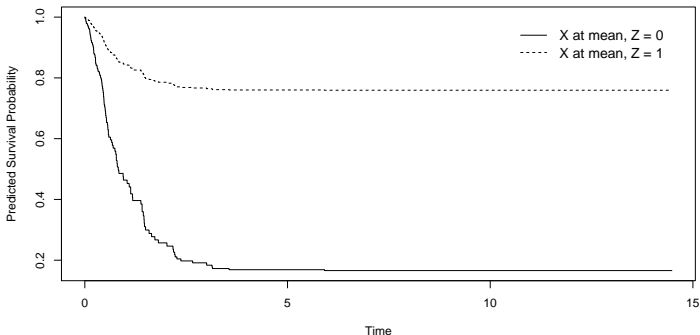
An Example: Ceasefire Durability

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 <u>among allies</u>

- Time-varying dyadic data, 1816-2001 ($NT = 57,819$)

Kaplan–Meier Survival
Estimate, Long et al. (2007)
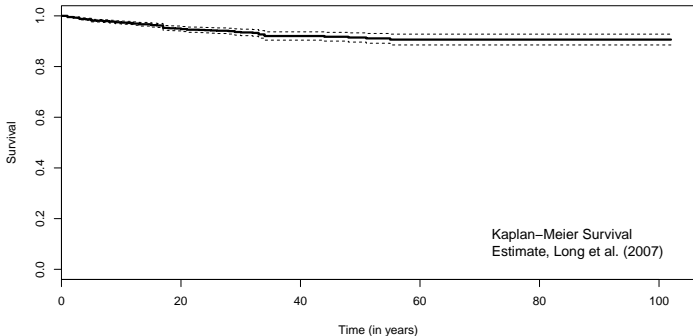
Cox Model (replicating LNB)

```
> 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

Cure Models

<div align="center">

(hours of fiddling...)

</div>

```
> 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...
```

Cure Models (Stata Remix)

```
. 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	.9298305	.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

Some Lessons

Cure models...

- ...Powerful

- ...Intuitive

- ...Temperamental

- ...Ask a lot of your data