Example of cure-regression models - on lognormal

abuxton

November 6, 2013

Abstract

The Gamel-Boag model is the mixture model fit with a lognormal failure density and a logistic cure-fraction link. This model seems to be the one used by Schmidt and Witte (1987). The purpose of this report is to show that the form of the lognormal of Sposto (2002) is the same as that of Schmidt and Witte (1987). The scale and shape parameters are the negative of each other.

Contents

List of Tables

List of Figures

1 split-population description

A spilt-population model. Survival, $S(t)$ where $g(t)$ is time, t, itself or $\tilde{t}$. Please note that $\tilde{t}$ is the scaled and shaped time: $\tilde{t} = (\lambda t)^{\gamma}$ where the scale and shape are, respectively, $\lambda = exp(x_{\lambda}\beta_{\lambda})$ and $\gamma = exp(x_{\gamma}\beta_{\gamma})$. $\hat{\mu_t}$ & $\hat{\sigma}$ below, respectively, refer to the mean time to fail in units of time, t, and the standard deviation or shape, (γ). These apply to the two probabilistically equal alternative lognormal failure densities, equations 2 and 3.

1.1 mixture model

$$S(t) = \pi + (1 - \pi)(1 - F(g(t))) \tag{1}$$

1.2 fail density: lognormal, form Sposto

$\hat{\mu_t} = exp(-x_{\lambda}\beta_{\lambda})$ & $\hat{\sigma} = exp(-x_{\gamma}\beta_{\gamma})$

$$F(\tilde{t}) = \int_{-\infty}^{ln\left(\tilde{t}\right)} \frac{1}{\sqrt{2\pi}} e^{\left(-x^2/2\right)} dx \tag{2}$$

This form (Sposto, 2002) is really the same as the standard lognormal.

1.3 fail density: lognormal

$\hat{\mu_t} = exp(x_{\lambda}\beta_{\lambda})$ & $\hat{\sigma} = exp(x_{\gamma}\beta_{\gamma})$

$$F(t) = \int_{-\infty}^{ln(t)} \frac{1}{(\gamma)\sqrt{2\pi}} e^{\left(-(x-(x_{\lambda}\beta_{\lambda}))^2/2\gamma^2\right)} dx \tag{3}$$

The standard lognormal fail density and is what is fit in Schmidt and Witte (1987) and also, for example, in Frankel and Longmate (2002) except that this form replaces the scale invariant σ with shape, (γ).

1.4 fail density: exponential

$$F(t) = 1 - exp\left(-\frac{t}{\lambda}\right) \tag{4}$$

1.5 cure-fraction link function:

logistic

$$\pi = \frac{exp\left(x_{\pi}\beta_{\pi}\right)}{1 + exp\left(x_{\pi}\beta_{\pi}\right)} \tag{5}$$

2 example: Schmidt & Witte model

2.1 sample data description

data description source at Stata Corp. Wooldridge et al. (1998) (does not include the gender (male) factor).

factor	description
tserved1	time served rounded to months (x100) median .12 (range: 0, 2.19)
age1	age in months (x1000) median .307 (range: .198, .933)
priors1	number of previous convictions (x10) median 0 (range: 0, 2.8)
white	=1 if not black (51%) n=744
felon	=1 if felony sentence (31%) n=454
alcohol	=1 if alcohol problems (21%) n=303
drugs	=1 if drug history (24%) n=349
property	=1 if property crime (25%) n=368

Table 1: factors

2.2 output

model featuring the fail density of equation 2 (cureregr Buxton (2013a)).

```
. cureregr tserved1 age1 priors1 white felon alcohol drugs property , dist(logn
> ormal) cl(mix) link(logist) scale() shape() nolog noshow

cf: logistic, kn: lognormal, model: mixture
cf_initial_coef: 0.4700 pi: 0.6154

No. of subjects = 1445                           Number of obs    =      1445
                                                 LR chi2(8)       =    150.71
Log likelihood = -3137.7484                      Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
cure_frac    |
    tserved1 |  -2.880944   .7033178    -4.10   0.000    -4.259422   -1.502467
        age1 |   4.299798   .7287908     5.90   0.000     2.871395    5.728202
     priors1 |  -2.011418   .4729662    -4.25   0.000    -2.938414   -1.084421
       white |   .6758835   .1436679     4.70   0.000     .3942995    .9574675
       felon |    1.00679   .2676913     3.76   0.000     .4821247    1.531455
     alcohol |  -.6741825   .1896951    -3.55   0.000    -1.045978    -.302387
       drugs |  -.4165466   .1662791    -2.51   0.012    -.7424477   -.0906455
    property |  -.5894715   .2427129    -2.43   0.015     -1.06518    -.113763
       _cons |  -.7904601   .2700485    -2.93   0.003    -1.319745   -.2611748
-------------+----------------------------------------------------------------
scale        |
       _cons |  -3.211344   .0949789   -33.81   0.000    -3.397499   -3.025189
-------------+----------------------------------------------------------------
shape        |
       _cons |  -.1818824   .0529575    -3.43   0.001    -.2856771   -.0780877
------------------------------------------------------------------------------
```

model featuring the fail density of equation 3 (cureregr1 Buxton (2013b)).

```
. cureregr1 tserved1 age1 priors1 white felon alcohol drugs property , dist(nor
> mallnt) cl(mix) link(logist) scale() shape() nolog noshow auxuse cformat(%6.4
> f) sformat(%3.2f) pformat(%5.4f)

cf: logistic, kn: lognormal, model: mixture
cf_initial_coef: 0.4700 pi: 0.6154

No. of subjects = 1445                          Number of obs   =       1445
                                                LR chi2(8)      =     150.71
Log likelihood = -3137.7484                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
cure_frac    |
    tserved1 |    -2.8809     0.7033    -4.10   0.0000      -4.2594     -1.5025
        age1 |     4.2998     0.7288     5.90   0.0000       2.8714      5.7282
     priors1 |    -2.0114     0.4730    -4.25   0.0000      -2.9384     -1.0844
       white |     0.6759     0.1437     4.70   0.0000       0.3943      0.9575
       felon |     1.0068     0.2677     3.76   0.0002       0.4821      1.5315
     alcohol |    -0.6742     0.1897    -3.55   0.0004      -1.0460     -0.3024
       drugs |    -0.4165     0.1663    -2.51   0.0122      -0.7424     -0.0906
    property |    -0.5895     0.2427    -2.43   0.0152      -1.0652     -0.1138
       _cons |    -0.7905     0.2700    -2.93   0.0034      -1.3197     -0.2612
-------------+----------------------------------------------------------------
mean_ln      |
       _cons |     3.2113     0.0950    33.81   0.0000       3.0252      3.3975
-------------+----------------------------------------------------------------
shape        |
       _cons |     0.1819     0.0530     3.43   0.0006       0.0781      0.2857
-------------+----------------------------------------------------------------
      mean_t |    24.8124     2.3567                        20.5979     29.8893
       stdev |     1.1995     0.0635                         1.0812      1.3307
------------------------------------------------------------------------------
```

Alas, the ancillary parameter estimates *mean_t* and *stdev* only are displayed for the constant parameter even if more than one factor is included the respective *mean_ln* and *shape* models.

2.3 model results

These results in tables 2 and 3 can be compared to those in Schmidt and Witte (1987). The results compare closely even though the factor, male, is not available as it is in Schmidt & Witte.

	cox PH[a]		
	b	sig[b]	z
main			
tserved1	1.3650	***	8.10
age1	-3.4386	***	-6.88
priors1	0.8882	***	6.63
white	-0.4415	***	-5.01
felon	-0.5735	***	-3.97
alcohol	0.4280	***	4.08
drugs	0.2997	**	3.06
property	0.4190	**	3.03
lnL	-3813.672		

[a] re: proportional hazards model, *final specification*, in table 1 of Schmidt & Witte.
[b] $* < 0.05$, $** < 0.01$, $*** < 0.001$

Table 2: cox PH

factor	split lognormal model t2a[a]			logit lognormal model t2b[b]			logit/individual lognormal t3[c]			logit/individual exponential t4[d]		
	b	sig[e]	z	b	sig	z	b	sig	z	b	sig	z
cure_frac												
tserved1				-2.8809	***	-4.10	-1.5831	***	-3.80	-1.2902	**	-3.23
age1				4.2998	***	5.90	3.7469	***	4.99	3.8249	***	4.58
priors1				-2.0114	***	-4.25	-1.1582	***	-3.65	-2.0488	***	-3.88
white				0.6759	***	4.70	0.6695	***	4.32	0.7796	***	4.70
felon				1.0068	***	3.76	0.4735		1.67	0.3414		1.16
alcohol				-0.6742	***	-3.55	-0.4832	**	-2.59	-0.5197	**	-2.61
drugs				-0.4165	*	-2.51	-0.4226	*	-2.40	-0.3479	*	-2.02
property				-0.5895	*	-2.43	-0.2338		-0.92	-0.1196		-0.48
_cons	-0.8723	**	-3.09	-0.7905	**	-2.93	-0.8773	**	-3.14	-0.8387	**	-2.88
mean_ln												
tserved1	-1.9602	***	-5.98				-1.1552	***	-3.85			
age1	3.4541	***	6.07				1.1795		1.95			
priors1	-1.4350	***	-6.07				-0.6491	**	-2.87			
white	0.4729	***	3.92				-0.0072		-0.06			
felon	0.9364	***	4.75				0.6707	**	2.88			
alcohol	-0.6339	***	-4.31				-0.3029	*	-2.04			
drugs	-0.2957	*	-2.20				-0.0070		-0.05			
property	-0.6869	***	-3.69				-0.5596	**	-2.86			
_cons	3.3187	***	14.08	3.2113	***	33.81	3.1750	***	14.48			
mean	27.62			24.81			23.93					
shape												
_cons	0.3853	***	6.38	0.1819	***	3.43	0.1118	*	2.21			
stdev	1.47			1.20			1.12					
scale												
tserved1										-1.4549	***	-5.69
age1										1.2943		1.63
priors1										0.0662		0.28
white										-0.1685		-1.08
felon										0.7858	**	3.21
alcohol										-0.2227		-1.26
drugs										-0.1239		-0.80
property										-0.6842	***	-3.62
_cons										3.5309	***	12.74
lnL	-3145.759			-3137.748			-3121.393			-3138.226		

[a] re: split lognormal model in table 2 of Schmidt & Witte.
[b] re: logit lognormal model in table 2.
[c] re: logit/individual lognormal model in table 3 of Schmidt & Witte.
[d] re: logit/individual exponential model in table 3.
[e] $* < 0.05$, $** < 0.01$, $*** < 0.001$

Table 3: split population models

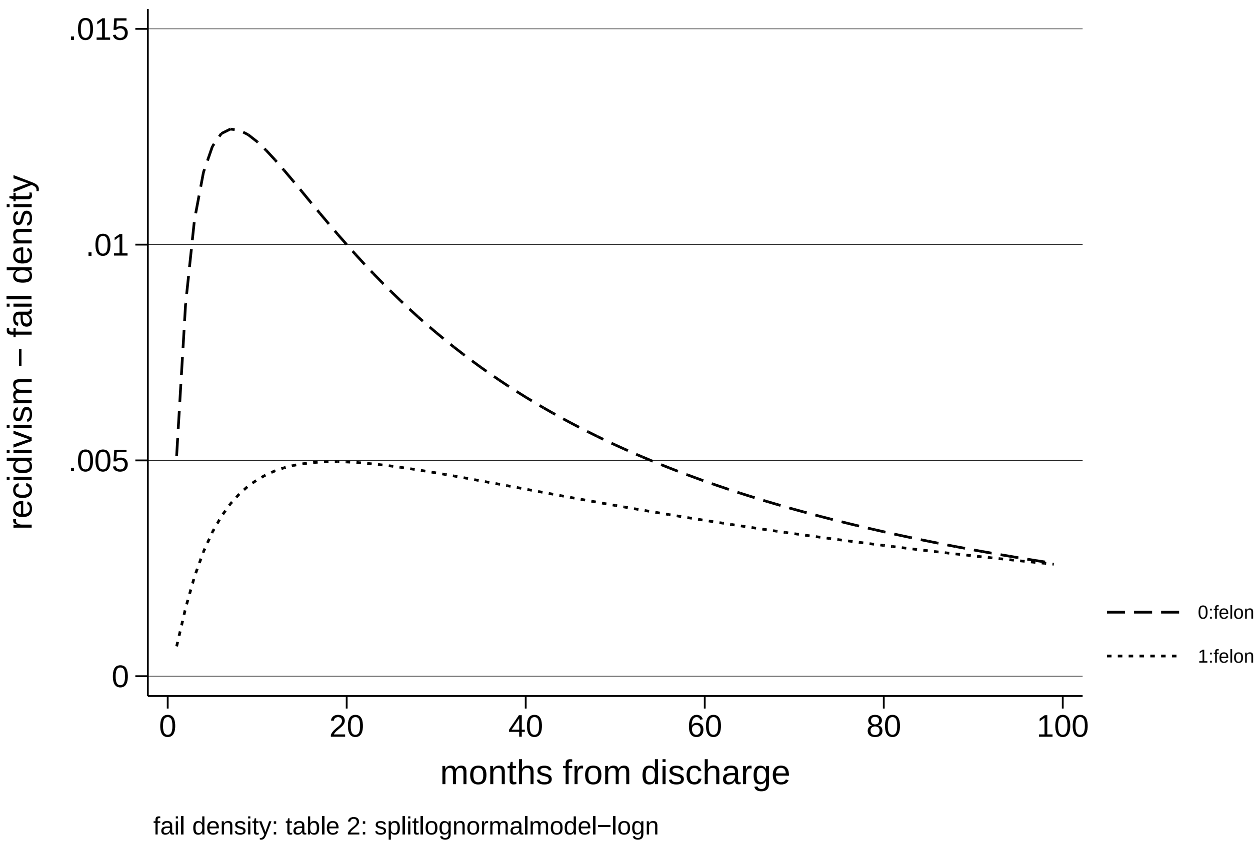

Figure 1: fail density split lognormal model
at median tserved1, age1, & priors1; alcohol, drugs, property, & white are ==0 except felon as noted. The average time to recidivism, among those reincarerated, is about 25 months.

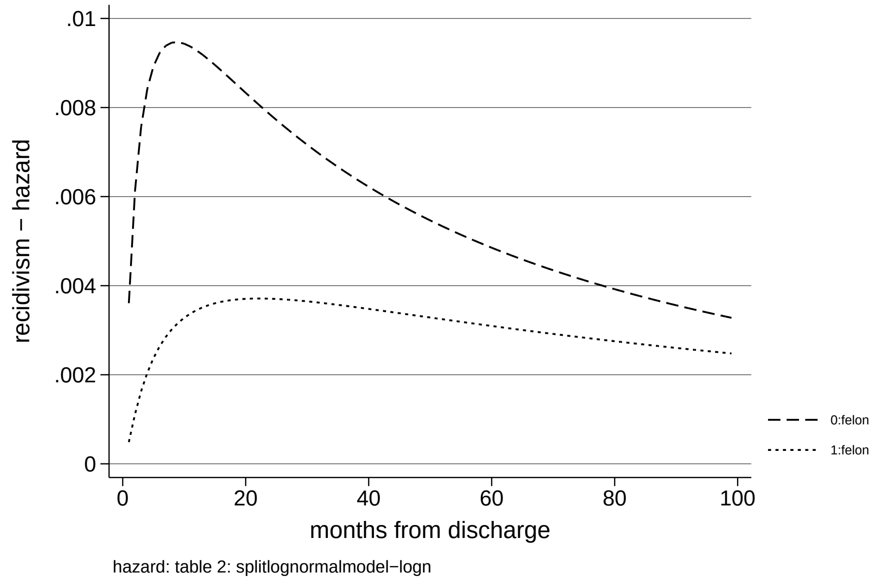

Figure 2: hazard split lognormal model
at median tserved1, age1, & priors1; alcohol, drugs, property, & white are ==0 except felon as noted.

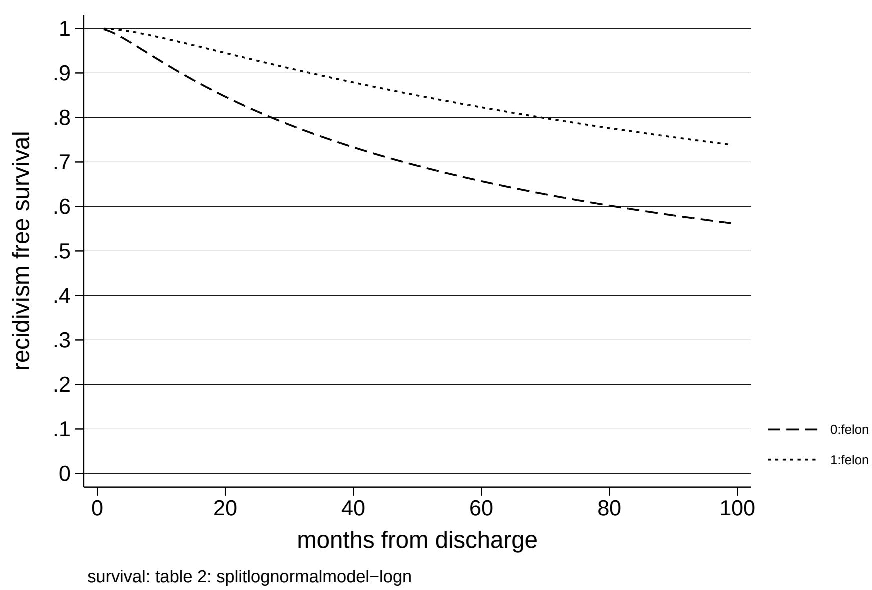

Figure 3: survival split lognormal model
at median tserved1, age1, & priors1; alcohol, drugs, property, & white are ==0 except felon as noted.

References

Buxton, A. (2013a). cureregr (v2.2): module to estimate parametric cure regression. Statistical Software Components (SSC) archive, which is often called the Boston College Archive and is provided by http://www.repec.org. Distribution-Date: 20131025.

Buxton, A. (2013b). cureregr1 herein is cureregr (v2.3): module to estimate parametric cure regression. Statistical Software Components (SSC) archive, which is often called the Boston College Archive and is provided by http://www.repec.org. Distribution-Date: 201311dd. 03Nov2013; add alternative lognormal.

Frankel, P. and Longmate, J. (2002). Parametric models for accelerated and long-term survival: a comment on proportional hazards. *Statistics in Medicine*, 21:3279–3289.

Schmidt, P. and Witte, A. D. (1987). Predicting criminal recidivism using *spilt population* survival time models. Working Paper 2445, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138.

Sposto, R. (2002). Cure model analysis in cancer: an application to data from the children's oncology group. *Statistics in Medicine*, 21:293–312.

Wooldridge, Chung, Schmidt, P., and Witte, A. D. (1998). Recidivism - a random sample of convicts. internet, Stata dataset. The dataset considered here is analyzed in Wooldridge (2002) and credited to Chung, Schmidt and Witte (1991). The data pertain to a random sample of convicts released from prison between July 1, 1977 and June 30, 1978. Of interest is the time until they return to prison. The information was collected retrospectively by looking at records in April 1984, so the maximum possible length of observation is 81 months. The data are available from the Stata website and can be accessed using the command: 'use http://www.stata.com/data/jwooldridge/eacsap/recid.dta'. Also see: http://data.princeton.edu/pop509a/recid1.html.