Duration Models: Modeling Strategies

Brad Jones[1]

[1]Department of Political Science
University of California, Davis

February 28, 2007

Today: Some Applications using R and Stata

Parametrics

- Let's consider implementation of these models in R and Stata

Parametrics

- Let's consider implementation of these models in R and Stata
- Both environments are tremendous with survival data.

Parametrics

- Let's consider implementation of these models in R and Stata
- Both environments are tremendous with survival data.
- R is a descendent of S, which has a strong bio-stats history.

Parametrics

- Let's consider implementation of these models in R and Stata
- Both environments are tremendous with survival data.
- R is a descendent of S, which has a strong bio-stats history.
- Some applications first using UN Peacekeeping Mission Data

Exponential: Stata streg

```
. streg civil interst, dist(exp) nohr

        failure _d:  failed
  analysis time _t:  duration

Iteration 5:   log likelihood = -86.354481

Exponential regression -- log relative-hazard form

No. of subjects =           54                 Number of obs   =          54
No. of failures =           39
Time at risk    =         3994
                                                LR chi2(2)      =       33.36
Log likelihood  =   -86.354481                 Prob > chi2     =      0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       civil |   1.169344   .3588703     3.26   0.001     .4659714    1.872717
     interst |    -1.6401   .4954337    -3.31   0.001    -2.611132   -.6690679
       _cons |  -4.350864   .2132007   -20.41   0.000     -4.76873   -3.932999
```

Exponential: R survreg

```
   > UN.exp<-survreg(Surv(duration, failed)~ civil + interst, data=UN,
+ dist='weibull',scale=1)
>
> summary(UN.exp)
> UNexp<-cbind(UN.exp$coef)

Call:
survreg(formula = Surv(duration, failed) ~ civil + interst, data = UN,
    dist = "weibull", scale = 1)
            Value Std. Error     z        p
(Intercept)  4.35      0.213 20.41 1.44e-92
civil       -1.17      0.359 -3.26 1.12e-03
interst      1.64      0.495  3.31 9.32e-04

Scale fixed at 1

Weibull distribution
Loglik(model)= -202.9   Loglik(intercept only)= -219.5
        Chisq= 33.36 on 2 degrees of freedom, p= 5.7e-08
Number of Newton-Raphson Iterations: 5
n=54 (4 observations deleted due to missingness)
```

Notes

- Odd ball difference in log-likelihoods between R and Stata.
- As to difference, I do not yet know. If someone knows, please let me know.

Notes

- Odd ball difference in log-likelihoods between R and Stata.
- As to difference, I do not yet know. If someone knows, please let me know.
- Note sign differences: Stata is in hazard rates; R is AFT.

Notes

- Odd ball difference in log-likelihoods between R and Stata.
- As to difference, I do not yet know. If someone knows, please let me know.
- Note sign differences: Stata is in hazard rates; R is AFT.
- Might be useful to compute hazard ratio for a covariate profile:

Case where `civil=1`

```
Stata first:

        display exp(_b[civil])
        Returns 3.2198805

 R:

 UNexp<-cbind(UN.exp$coef)

 hr.civil.exp<-exp(-UNexp[2,1]); hr.civil.exp
 Returns: 3.219880
```

Same number but note difference between HR and AFT parameterizations. R uses AFT by default; therefore, I must take negative of β in computing the hazard.

Interpretation? Interventions prompted by civil wars are about 3.2 times more likely to fail than when compared to the baseline category of internationalized civil wars.

Proportional Hazards Property

- Exponential, Weibull, and Cox are PH Models.

Proportional Hazards Property

- Exponential, Weibull, and Cox are PH Models.
- PH Property: the increase (or decrease) in the hazard rate is a multiple of the baseline hazard rate.

Proportional Hazards Property

- Exponential, Weibull, and Cox are PH Models.
- PH Property: the increase (or decrease) in the hazard rate is a multiple of the baseline hazard rate.
- Therefore, the change in the hazard rate is proportional to the baseline hazard.
- Property:

$$\frac{h_i(t)}{h_0(t)} = \exp[\beta'(\mathbf{x}_i - \mathbf{x}_j)], \tag{1}$$

Illustration

```
Stata: First, compute the estimated hazard rates for each covariate (lambda):

Civil Wars:
. display exp(-(_b[_cons]+_b[civil]*1))
.04152249

Interstate Conflicts:
. display exp(-(_b[_cons]+_b[interst]*1))
.00250125

ICWs:
. display exp(-(_b[_cons]))
.01289566

Second, compute hazard ratios (computed in Stata):

Civil Wars:
. display .04152249/.01289566
3.219881

Interstate Conflicts:
. display .00250125/.01289566
.1939606

ICWs:
. display .01289566/.01289566
1
```

Illustration

```
R: Computing lambda
> ##Civil Wars
>
> exp(-(UNexp[1,1]+UNexp[2,1]))
[1] 0.04152249
>
> ##Interstate Conflicts
>
> exp(-(UNexp[1,1]+UNexp[3,1]))
[1] 0.002501251
>
> ## ICW
>
> exp(-(UNexp[1,1]))
[1] 0.01289566

Second, computing ratios:

> exp(-(UNexp[1,1]+UNexp[2,1]))/exp(-(UNexp[1,1]))
[1] 3.219880
>
>
> exp(-(UNexp[1,1]+UNexp[3,1]))/exp(-(UNexp[1,1]))
[1] 0.1939606
>
>
> exp(-(UNexp[1,1]))/exp(-(UNexp[1,1]))
[1] 1
```

Weibull

- Note that if we had plotted λ, the plot would be flat.
- Let's consider the Weibull.
- Illustrations in Stata and in R.

Weibull

- Note that if we had plotted λ, the plot would be flat.
- Let's consider the Weibull.
- Illustrations in Stata and in R.
- Useful to recall the hazard function:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{2}$$

λ is positive scale parameter; p is shape parameter.

Weibull

- Note that if we had plotted λ, the plot would be flat.
- Let's consider the Weibull.
- Illustrations in `Stata` and in `R`.
- Useful to recall the hazard function:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{2}$$

  λ is positive scale parameter; p is shape parameter.
- $p > 1$, the hazard rate is *monotonically* increasing with time.
- $p < 1$, the hazard rate is *monotonically* decreasing with time.
- $p = 1$, the hazard is flat, i.e. *exponential*.

Weibull

- Note that if we had plotted λ, the plot would be flat.
- Let's consider the Weibull.
- Illustrations in `Stata` and in `R`.
- Useful to recall the hazard function:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{2}$$

λ is positive scale parameter; p is shape parameter.

- $p > 1$, the hazard rate is *monotonically* increasing with time.
- $p < 1$, the hazard rate is *monotonically* decreasing with time.
- $p = 1$, the hazard is flat, i.e. *exponential*.
- Note that λ corresponds to covariates $(\exp -\beta_k x_i)$

Weibull

- Note that if we had plotted λ, the plot would be flat.
- Let's consider the Weibull.
- Illustrations in Stata and in R.
- Useful to recall the hazard function:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{2}$$

  λ is positive scale parameter; p is shape parameter.
- $p > 1$, the hazard rate is *monotonically* increasing with time.
- $p < 1$, the hazard rate is *monotonically* decreasing with time.
- $p = 1$, the hazard is flat, i.e. *exponential*.
- Note that λ corresponds to covariates $(\exp -\beta_k x_i)$
- But BE AWARE of your parameterization!

Stata streg (AFT formulation):

```
. streg civil interst, dist(weib) time

         failure _d:  failed
   analysis time _t:  duration

Iteration 4:   log likelihood = -84.655157

Weibull regression -- accelerated failure-time form

No. of subjects =           54                 Number of obs   =         54
No. of failures =           39
Time at risk    =         3994
                                               LR chi2(2)      =      17.67
Log likelihood  =   -84.655157                 Prob > chi2     =     0.0001
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       civil |  -1.100421   .4457861    -2.47   0.014    -1.974146   -.2266966
     interst |   1.736832   .6165459     2.82   0.005     .5284242     2.94524
       _cons |    4.28793   .2652436    16.17   0.000     3.768062    4.807798
-------------+----------------------------------------------------------------
       /ln_p |  -.2145617   .1237889    -1.73   0.083    -.4571834      .02806
-------------+----------------------------------------------------------------
           p |    .806895   .0998846                      .6330642    1.028457
         1/p |   1.239319   .1534138                        .97233    1.579619
```

R using survreg:

```
> ##Weibull Model for UN Data:
>
> UN.weib<-survreg(Surv(duration, failed)~ civil + interst, data=UN,
+ dist='weibull')
>
> summary(UN.weib)

Call:
survreg(formula = Surv(duration, failed) ~ civil + interst, data = UN,
    dist = "weibull")
             Value Std. Error     z        p
(Intercept)  4.288      0.265 16.17 8.76e-59
civil       -1.100      0.446 -2.47 1.36e-02
interst      1.737      0.617  2.82 4.85e-03
Log(scale)   0.215      0.124  1.73 8.30e-02

Scale= 1.24

Weibull distribution
Loglik(model)= -201.2   Loglik(intercept only)= -210
        Chisq= 17.67 on 2 degrees of freedom, p= 0.00015
Number of Newton-Raphson Iterations: 5
n=54 (4 observations deleted due to missingness)
```

R using eha weibreg:

```
> UN.weib2<-weibreg(Surv(duration, failed)~ civil + interst, data=UN, shape=0)
>
>
> summary(UN.weib2)
Call:
weibreg(formula = Surv(duration, failed) ~ civil + interst, data = UN,
    shape = 0)

Covariate           Mean       Coef  Exp(Coef)  se(Coef)     Wald p
civil               0.072     0.888     2.430     0.383      0.020
interst             0.501    -1.401     0.246     0.512      0.006

log(scale)                    4.288    72.816     0.265      0.000
log(shape)                   -0.215     0.807     0.124      0.083

Events                    39
Total time at risk          3994
Max. log. likelihood      -201.15
LR test statistic         17.7
Degrees of freedom        2
Overall p-value           0.00014576
```

(This is a bit of odd programming. EHA reports log(scale) which is equivalent to intercept for AFT formulation; note, however, that the coefficients are in log relative hazard (PH) form. To retreive AFT parameters, do -b/p. For civil war covariate, -.888/.807=-1.10.)

Reminder of Translation

- There are a couple of ways to express the Weibull (exponential)
- (1): Model $h(t)$; (2): Model $\log(T)$
- In (1), coefficients relate to the hazard function.
- In (2), coefficients relate to log of the failure time.
- Signs will differ depending on choice.
- `Stata` defaults to (1); `R` (`survreg`) defaults to (2).
- (2) is sometimes called "accelerated failure time"

The Two "Different" Models

- Proportional Hazards:

$$h(t \mid \mathbf{x}) = h_{0t} \exp(\alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_j x_{ij}), \qquad (3)$$

- Accelerated Failure Time:

$$\log(T) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_j x_{ij} + \sigma\epsilon, \qquad (4)$$

  ϵ is a stochastic disturbance term with *type-1 extreme-value* distribution scaled by σ.

- $\sigma = 1/p$.
- $F(\epsilon)$ is a type-1 extreme value distribution.
- Close connection to Weibull: the distribution of the log of a Weibull distributed random variable yields a type-1 extreme value distribution.
- Sometimes this parameterization is referred to as a log-Weibull distribution.

Connection between Parameterizations

P.H. Parm.	A.F.T. Parm.	Relationship Between Parameters	Interp. of P.H. Parm.	Interp. of A.F.T. Parm.
α	β	$\beta = \frac{-\alpha}{p}$	$+\alpha \equiv \uparrow h(t \mid x_{ij})$	$+\beta \equiv \uparrow \log(T)$
		$\alpha = -\beta p$	$-\alpha \equiv \downarrow h(t \mid x_{ij})$	$-\beta \equiv \downarrow \log(T)$
p	σ	$\sigma = \frac{1}{p}$	$p > 1 \equiv \uparrow h(t \mid x_{ij})$	$\sigma > 1 \equiv \downarrow h(t \mid x_{ij})$
		$p = \frac{1}{\sigma}$	$p < 1 \equiv \downarrow h(t \mid x_{ij})$	$\sigma < 1 \equiv \uparrow h(t \mid x_{ij})$

Weibull hazards

- Hazard rates are useful to examine:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{5}$$

Weibull hazards

- Hazard rates are useful to examine:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{5}$$

- You may want to compute them and plot them.

Weibull hazards

- Hazard rates are useful to examine:

$$h(t) = \lambda p(\lambda t)^{p-1} \qquad t > 0 \lambda > 0, p > 0 \tag{5}$$

- You may want to compute them and plot them.
- Examples

```
    Stata: Generating the Hazard Rates "the hard way."

. gen lambda_civil=exp(-(_b[_cons]+_b[civil]))
THIS CORRESPONDS TO LAMBDA in EQUATION 3

. gen haz_civil=lambda_civil*e(aux_p)*(lambda_civil*duration)^(e(aux_p)-1)
THIS IS EQUATION 3 COME TO LIFE

Stata makes life (too?) easy:

.  predict hazard_civil, hazard, if civil==1

(I COULD DO THIS FOR ALL THREE MISSION TYPES)
 Then I could plot them:

twoway (scatter hazard_civil _t, connect(s) msymbol(O))
 (scatter hazard_interst_t, connect(s) msymbol(D))
 (scatter hazard_icw _t, connect(s) msymbol(S)),
 xtitle(Duration Time of Peacekeeping Mission)
title(Estimated Hazard Rates ) subtitle((By Mission-Type))
saving(c:\ehbook\icpsr_unhazrates, replace)

which returns:
```

Hazard Rates: Weibull

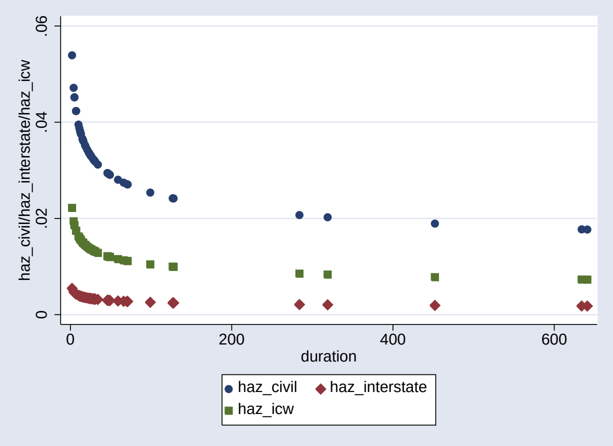
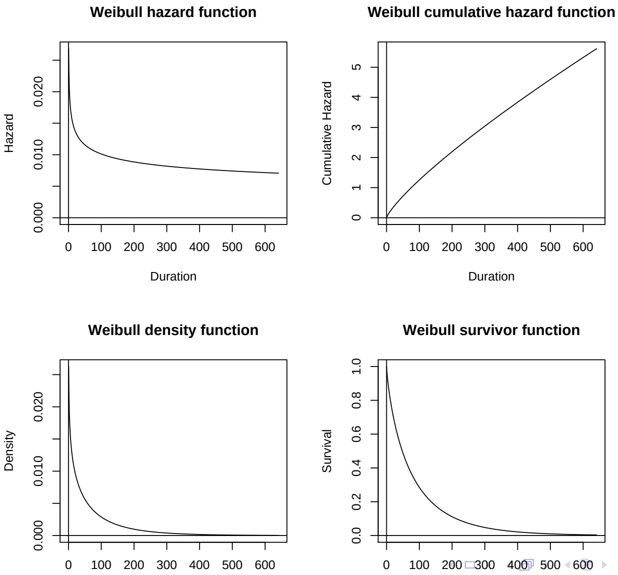


Figure: *This figures graphs the hazard rates from the Weibull.*

```
In R, I could write out the statement for lambda
as is done above.  I would simply need to
retrieve the coefficients from the column matrix
(after cbind-ing it) and write the function
(equation 3).  I could then plot these.

In eha, I can use weibreg.plot.  This
returns several plots, including the hazard
(setting covariates to mean [it is essentially
the "average" hazard]).

Code looks like:


UN.weib2<-weibreg(Surv(duration, failed)~ civil + interst, data=UN, shape=0)

summary(UN.weib2)

UNweib2<-cbind(UN.weib2$coef); UNweib2

plot.weibreg(UN.weib2)
```

Hazard Rates: Weibull



GENERATING HAZARD RATIOS:

```
Stata:
. display exp(-(_b[interst]))^(e(aux_p))
.24624185

. display exp(-(_b[civil]))^(e(aux_p))
2.4300808

. display exp(-(0))^(e(aux_p))
1

I could use predict in Stata:

. predict hr_interst, hr, if interst==1
(48 missing values generated)

. predict hr_civil, hr, if civil==1
(44 missing values generated)

. predict hr_icw, hr, if civil==0 & interst==0
(28 missing values generated)

R (survreg):

> hr.civil.weib<-exp(-UNweib[2,1])^(1/UN.weib$scale); hr.civil.weib
[1] 2.430080
> hr.inter.weib<-exp(-UNweib[3,1])^(1/UN.weib$scale); hr.inter.weib
[1] 0.2462418
> hr.icw.weib<-exp(0)^(1/UN.weib$scale); hr.icw.weib
[1] 1
```

Let's include semi-continuous covariate. Stata:

```
. streg civil interst borders, dist(weib) time nolog

         failure _d:  failed
   analysis time _t:  duration

Weibull regression -- accelerated failure-time form

No. of subjects =           46                     Number of obs   =        46
No. of failures =           36
Time at risk    =         3840
                                                   LR chi2(3)      =     18.45
Log likelihood  =   -76.493097                     Prob > chi2     =    0.0004

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       civil |  -1.380352   .4921063    -2.80   0.005    -2.344862   -.4158411
     interst |   1.806995   .6347777     2.85   0.004     .5628534    3.051136
     borders |  -.1368689   .0972727    -1.41   0.159    -.3275199     .053782
       _cons |   4.800974   .4777848    10.05   0.000     3.864533    5.737415
-------------+----------------------------------------------------------------
       /ln_p |  -.2278767   .1328443    -1.72   0.086    -.4882467    .0324932
-------------+----------------------------------------------------------------
           p |   .7962224   .1057736                      .6137014    1.033027
         1/p |   1.255931   .1668432                       .968029    1.629457
```

```
R (survreg):

Call:
survreg(formula = Surv(duration, failed) ~ civil + interst +
    borders, data = UN, dist = "weibull")
             Value Std. Error     z        p
(Intercept)  4.801     0.4778 10.05 9.34e-24
civil       -1.380     0.4921 -2.80 5.03e-03
interst      1.807     0.6348  2.85 4.42e-03
borders     -0.137     0.0973 -1.41 1.59e-01
Log(scale)   0.228     0.1328  1.72 8.63e-02

Scale= 1.26

Weibull distribution
Loglik(model)= -184.8   Loglik(intercept only)= -194.1
        Chisq= 18.45 on 3 degrees of freedom, p= 0.00036
Number of Newton-Raphson Iterations: 5
n=46 (12 observations deleted due to missingness)
```

```
Proportional Hazards Property again:

This is the hazard ratio for each value the covariate takes (done in Stata):

. gen hazratio_borders=exp(-_b[borders]*borders)^e(aux_p)

Done in R:
hr.borders.weib<-exp(-UNweibc[4,1]*borders)^(1/UN.weibc$scale)

They look like this:
. table hazratio_borders borders

----------------------------------------------------------------
hazratio_ |                       borders
borders   |    1     2     3     4     5     6     8     9    13
----------+-----------------------------------------------------
 1.115138 |   10
 1.243533 |          7
  1.38671 |                6
 1.546373 |                     12
  1.72442 |                            8
 1.922966 |                                  3
 2.391271 |                                        2
 2.666597 |                                              1
 4.123554 |                                                    1
----------------------------------------------------------------

The PH property must hold.  Take the ratio of any adjacent pair:
. display 1.546373/1.38671
1.115138
Note that this is equivalent to:
. display exp(-_b[borders])^e(aux_p)
1.1151379
which is the hazard ratio for the "baseline case".
```

Many Applications

- These are "plug and play" estimators.

Many Applications

- These are “plug and play” estimators.
- They are easy to do.

Many Applications

- These are "plug and play" estimators.
- They are easy to do.
- Let's run through some illustrations, first in Stata and then in R

Many Applications

- These are "plug and play" estimators.
- They are easy to do.
- Let's run through some illustrations, first in Stata and then in R
- I use the cabinet duration data.

Weibull

```
. streg invest polar numst format postelec caretakr, dist(weib) time nolog

         failure _d:  censor
   analysis time _t:  durat

Weibull regression -- accelerated failure-time form

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =    171.94
Log likelihood  =   -414.07496                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |  -.2958188   .1059024    -2.79   0.005    -.5033838   -.0882538
       polar |   -.017943   .0042784    -4.19   0.000    -.0263285   -.0095575
       numst |   .4648894   .1005815     4.62   0.000     .2677533    .6620255
      format |  -.1023747   .0335853    -3.05   0.002    -.1682006   -.0365487
    postelec |   .6796125    .104382     6.51   0.000     .4750276    .8841974
    caretakr |   -1.33401   .2017528    -6.61   0.000    -1.729438   -.9385818
       _cons |   2.985428   .1281146    23.30   0.000     2.734328    3.236528
-------------+----------------------------------------------------------------
       /ln_p |    .257624   .0500578     5.15   0.000     .1595126    .3557353
-------------+----------------------------------------------------------------
           p |   1.293852   .0647673                      1.172939     1.42723
         1/p |   .7728858   .0386889                       .700658    .8525593
```

Exponential

```
. streg invest polar numst format postelec caretakr, dist(exp) time nolog

        failure _d:  censor
  analysis time _t:  durat

Exponential regression -- accelerated failure-time form

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =    148.53
Log likelihood  =   -425.90641                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |  -.3322088   .1376729    -2.41   0.016    -.6020426   -.0623749
       polar |  -.0193017   .0055465    -3.48   0.001    -.0301725   -.0084308
       numst |    .515435   .1291486     3.99   0.000     .2623084    .7685616
      format |  -.1079432   .0435233    -2.48   0.013    -.1932474    -.022639
    postelec |   .7403427    .134558     5.50   0.000     .4766138    1.004072
    caretakr |  -1.319272   .2595422    -5.08   0.000    -1.827965   -.8105783
       _cons |   2.944518   .1663401    17.70   0.000     2.618498    3.270539
------------------------------------------------------------------------------
```

```
Log-logistic

. streg invest polar numst format postelec caretakr, dist(loglog) time nolog

         failure _d:  censor
   analysis time _t:  durat

Log-logistic regression -- accelerated failure-time form

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =    148.72
Log likelihood  =   -424.10921                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |  -.3367541   .1278083    -2.63   0.008    -.5872538   -.0862544
       polar |  -.0221958   .0052638    -4.22   0.000    -.0325127   -.0118789
       numst |   .4830709   .1212506     3.98   0.000     .2454241    .7207177
      format |  -.1093453   .0419715    -2.61   0.009    -.1916078   -.0270827
    postelec |   .6408808   .1240329     5.17   0.000     .3977807    .8839808
    caretakr |   -1.26921   .2310272    -5.49   0.000    -1.722015   -.8164046
       _cons |   2.728818   .1595866    17.10   0.000     2.416034    3.041602
-------------+----------------------------------------------------------------
     /ln_gam |  -.5657686   .0511353   -11.06   0.000     -.665992   -.4655451
-------------+----------------------------------------------------------------
       gamma |   .5679235    .029041                       .5137636    .6277928
------------------------------------------------------------------------------
```

```
  Log-normal

. streg invest polar numst format postelec caretakr, dist(lognorm) time nolog

         failure _d:  censor
   analysis time _t:  durat

Log-normal regression -- accelerated failure-time form

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =    150.66
Log likelihood  =   -425.30621                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |  -.3738013   .1327055    -2.82   0.005    -.6338993   -.1137032
       polar |   -.021988   .0054825    -4.01   0.000    -.0327336   -.0112424
       numst |   .5717579   .1232281     4.64   0.000     .3302353    .8132805
      format |  -.1194982   .0432516    -2.76   0.006    -.2042698   -.0347266
    postelec |   .6668079   .1292366     5.16   0.000     .4135088     .920107
    caretakr |  -1.126047   .2576962    -4.37   0.000    -1.631122   -.6209713
       _cons |   2.632497    .164494    16.00   0.000     2.310095    2.954899
-------------+----------------------------------------------------------------
     /ln_sig |   .0078719   .0439881     0.18   0.858    -.0783432    .0940871
-------------+----------------------------------------------------------------
       sigma |   1.007903   .0443358                       .924647    1.098655
```

```
  Weibull
  > cab.weib<-survreg(Surv(durat,censor)~invest + polar + numst +
+ format + postelec + caretakr,data=cabinet,
+ dist='weibull')
>
> summary(cab.weib)

Call:
survreg(formula = Surv(durat, censor) ~ invest + polar + numst +
    format + postelec + caretakr, data = cabinet, dist = "weibull")
              Value Std. Error     z          p
(Intercept)  2.9854    0.12811 23.30 4.15e-120
invest      -0.2958    0.10590 -2.79  5.22e-03
polar       -0.0179    0.00428 -4.19  2.74e-05
numst        0.4649    0.10058  4.62  3.80e-06
format      -0.1024    0.03359 -3.05  2.30e-03
postelec     0.6796    0.10438  6.51  7.47e-11
caretakr    -1.3340    0.20175 -6.61  3.79e-11
Log(scale)  -0.2576    0.05006 -5.15  2.65e-07

Scale= 0.773

Weibull distribution
Loglik(model)= -1014.6   Loglik(intercept only)= -1100.6
        Chisq= 171.94 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 314
```

```
  Log-Logistic
> cab.ll<-survreg(Surv(durat,censor)~invest + polar + numst +
+ format + postelec + caretakr,data=cabinet,
+ dist='loglogistic')
>
> summary(cab.ll)

Call:
survreg(formula = Surv(durat, censor) ~ invest + polar + numst +
    format + postelec + caretakr, data = cabinet, dist = "loglogistic")
              Value Std. Error      z        p
(Intercept)  2.7288    0.15959  17.10 1.50e-65
invest      -0.3368    0.12781  -2.63 8.42e-03
polar       -0.0222    0.00526  -4.22 2.48e-05
numst        0.4831    0.12125   3.98 6.77e-05
format      -0.1093    0.04197  -2.61 9.18e-03
postelec     0.6409    0.12403   5.17 2.38e-07
caretakr    -1.2692    0.23103  -5.49 3.93e-08
Log(scale)  -0.5658    0.05114 -11.06 1.87e-28

Scale= 0.568

Log logistic distribution
Loglik(model)= -1024.7   Loglik(intercept only)= -1099
        Chisq= 148.72 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 4
n= 314
```

```
  > ##Log-Normal can be fit using survreg:
>
> cab.ln<-survreg(Surv(durat,censor)~invest + polar + numst +
+ format + postelec + caretakr,data=cabinet,
+ dist='lognormal')
>
> summary(cab.ln)

Call:
survreg(formula = Surv(durat, censor) ~ invest + polar + numst +
    format + postelec + caretakr, data = cabinet, dist = "lognormal")
               Value Std. Error      z        p
(Intercept)  2.63250    0.16449 16.004 1.21e-57
invest      -0.37380    0.13271 -2.817 4.85e-03
polar       -0.02199    0.00548 -4.011 6.06e-05
numst        0.57176    0.12323  4.640 3.49e-06
format      -0.11950    0.04325 -2.763 5.73e-03
postelec     0.66681    0.12924  5.160 2.47e-07
caretakr    -1.12605    0.25770 -4.370 1.24e-05
Log(scale)   0.00787    0.04399  0.179 8.58e-01

Scale= 1.01

Log Normal distribution
Loglik(model)= -1025.9   Loglik(intercept only)= -1101.2
        Chisq= 150.66 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 4
n= 314
```

Comparing Log-Likelihoods (note: non-nested models). I did this in R:

```
 anova(cab.weib, cab.ln, cab.ll)
1 invest + polar + numst + format + postelec + caretakr
2 invest + polar + numst + format + postelec + caretakr
3 invest + polar + numst + format + postelec + caretakr
  Resid. Df    -2*LL Test Df   Deviance P(>|Chi|)
1       306 2029.238      NA         NA        NA
2       306 2051.701    =  0 -22.462507        NA
3       306 2049.307    =  0   2.394004        NA
```

```
Back to Stata: Generalized Gamma
. streg invest polar numst format postelec caretakr, dist(gamma) nolog

         failure _d:  censor
   analysis time _t:  durat

Gamma regression -- accelerated failure-time form

No. of subjects =          314                     Number of obs   =        314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =     165.78
Log likelihood  =   -414.00944                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |  -.3005269    .108745    -2.76   0.006    -.5136633   -.0873906
       polar |  -.0182998   .0044674    -4.10   0.000    -.0270559   -.0095438
       numst |   .4692142   .1030895     4.55   0.000     .2671626    .6712659
      format |  -.1031368   .0342637    -3.01   0.003    -.1702925   -.0359811
    postelec |   .6807161   .1061356     6.41   0.000     .4726942     .888738
    caretakr |  -1.328476   .2066422    -6.43   0.000    -1.733487   -.9234647
       _cons |   2.963114   .1447075    20.48   0.000     2.679492    3.246735
-------------+----------------------------------------------------------------
     /ln_sig |   -.234325   .0802121    -2.92   0.003    -.3915378   -.0771122
      /kappa |   .9241712   .2065399     4.47   0.000     .5193605    1.328982
-------------+----------------------------------------------------------------
       sigma |   .7911047   .0634561                      .6760165    .9257859
------------------------------------------------------------------------------
```

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa - 1} \exp[-(\lambda t)^{p}]}{\Gamma(\kappa)} \tag{6}$$

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa-1}\exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa - 1} \exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa-1} \exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa - 1}\exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

- When $\kappa = 1$, the Weibull is implied; when $\kappa = p = 1$, the exponential distribution is implied; when $\kappa = 0$, the log-normal distribution is implied; and when $p = 1$, the gamma distribution is implied.

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa - 1} \exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

- When $\kappa = 1$, the Weibull is implied; when $\kappa = p = 1$, the exponential distribution is implied; when $\kappa = 0$, the log-normal distribution is implied; and when $p = 1$, the gamma distribution is implied.
- In illustrations above, verify that Weibull would be preferred model among the choices.

Adjudication

- Lots of Choices
- Selection can be arbitrary
- If parametrically nested, standard LR tests apply.
- Encompassing Distribution: generalized gamma:

$$f(t) = \frac{\lambda p(\lambda t)^{p\kappa-1} \exp[-(\lambda t)^p]}{\Gamma(\kappa)} \tag{6}$$

- When $\kappa = 1$, the Weibull is implied; when $\kappa = p = 1$, the exponential distribution is implied; when $\kappa = 0$, the log-normal distribution is implied; and when $p = 1$, the gamma distribution is implied.
- In illustrations above, verify that Weibull would be preferred model among the choices.
- AIC $(-2(\log L) + 2(c + p + 1))$ also confirms Weibull is preferred model among choices.