DURATION MODELS

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Introduction
2. Hazard Functions
3. Estimation with Time-Invariant Covariates
4. Grouped Duration Data

1. Introduction

- Some response variables come in the form of a duration, or the time elapsed until a certain event occurs. Unemployment duration (measured in, say, weeks) is an important example.
- Duration methods have their history in survival analysis, where the duration is the survival time of a subject.
- Sometimes interested is in how observed covariates affect the mean or median duration, but interest often centers on the hazard function.

- We must deal with the problem of censored data, how to include covariates, including those that change over time.
- Another important issue concerns the introduction of heterogeneity into duration models. Even when heterogeneity is assumed independent of covariates – as it almost always is – the presence of heterogeneity can lead to wrong conclusions about the nature of the duration distribution.

2. Hazard Functions

Hazard Functions without Covariates

- Let $T \geq 0$ now denote a random variable, which is the amount of time elapsed before an event occurs. It is helpful to distinguish between the random variable T and a generic possible value, t.

- T has some distribution in the relevant population. Let its cdf be

$$F(t) = P(T \leq t), t \geq 0.$$

- The **survivor function** is

$$S(t) = 1 - F(t) = P(T > t).$$

• Typical to assume that the underlying duration is a continuous random variable. In fact, assume that $F(t)$ is continuously differentiable, with density

$$f(t) = \frac{dF}{dt}(t).$$

• For $h > 0$ and $t \geq 0$, we can define the conditional probability

$$P(t \leq T < t + h | T \geq t),$$

which is the probability of leaving the intial state during the interval $[t, t + h)$ given "survival" up through time t.

- The **hazard function** of T is defined as

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h | T \geq t)}{h},$$

which is the instantaneous rate of leaving the state per unit of time. For "small" h,

$$P(t \leq T < t + h | T \geq t) \approx \lambda(t)h,$$

so the hazard function can be used to approximate a conditional probability much as the height of the density of T can be used to approximate an unconditional probability. [$P(t \leq T < t + h) \approx f(t)h$.]

• As an example, if T is the number of weeks unemployed, then $\lambda(20)$ is approximately the probability of becoming employed between weeks 20 and 21. "Becoming employed" entails having been unemployed through week 20 (which is why it is a conditional probability.)

• Can write the hazard in terms of the density and survivor function:

$$P(t \leq T < t + h | T \geq t) = \frac{F(t+h) - F(t)}{1 - F(t)}$$

and, dividing by h and taking the limit as $h \downarrow 0$, gives

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

- Other useful representations:

$$f(t) = -\frac{d \log S(t)}{dt}$$

and, using $F(0) = 0$,

$$F(t) = 1 - \exp\left[-\int_0^t \lambda(r)dr\right].$$

- Differentiating this last expression gives

$$f(t) = \lambda(t)\exp\left[-\int_0^t \lambda(r)dr\right].$$

- All probabilities can be obtained from the hazard function, such as

$$P(a_1 \leq T < a_2 | T \geq a_1) = 1 - \exp\left[-\int_{a_1}^{a_2} \lambda(r) dr \right]$$

- The shape of $\lambda(t)$, $t \geq 0$, is important in applications. Simplest case is a constant hazard:

$$\lambda(t) = \lambda, \text{ all } t \geq 0.$$

Such a process is *memoryless* because the probability of exiting the state in the next interval does not depend on how much time has been spent in the initial state.

- The cdf of such a process is $F(t) = 1 - \exp(-\lambda t)$, which is the cdf of an exponential distribution.

• When $\lambda(t)$ is not constant, the process exhibits **duration dependence**. If $d\lambda(t)/dt > 0$ then there is positive duration dependence at t; if $d\lambda(t)/dt < 0$ there is negative duration dependence.

EXAMPLE: Weibull Distribution. If T has cdf

$$F(t) = 1 - \exp(-\gamma t^{\alpha})$$

where $\gamma, \alpha \geq 0$, then $f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^{\alpha})$ and so

$$\lambda(t) = \frac{\gamma \alpha t^{\alpha-1} \exp(-\gamma t^{\alpha})}{\exp(-\gamma t^{\alpha})} = \gamma \alpha t^{\alpha-1}.$$

• $\alpha = 1$ is the exponential distribution. $\alpha > 1$ is positive duration dependence (for all t); $\alpha < 1$ is negative duration dependence.

EXAMPLE: Log-Logistic Hazard. Here we state the hazard function:

$$\lambda(t) = \frac{\gamma\alpha t^{\alpha-1}}{1+\gamma t^{\alpha}}$$

- Allows a variety of shapes. For example, if $\alpha > 1$, $\lambda(t)$ increases until $t = [(\alpha-1)/\gamma]^{1-\alpha}$ after which it declines to zero.
- Can show the density is

$$f(t) = \gamma\alpha t^{\alpha-1}(1+\gamma t^{\alpha})^{-1}.$$

Can use this to show that $\log(T)$ has a logistic distribution with mean $-\alpha^{-1}\log(\gamma)$ and variance $\pi^2/(3\alpha^2)$.

Hazard Functions Conditional on Time-Invariant Covariates

- Very rarely interested in an unconditional hazard function. Generally, we want to know how the hazard function changes with observed explanatory variables.
- With covariates that do not change over time – so, we observe them prior to entering the initial state – the extensions of the previous definitions are straightforward.

- For example, the survivor function conditional on **x** is

$$S(t|\mathbf{x}) = 1 - F(t|\mathbf{x}) = P(T > t|\mathbf{x})$$

and the hazard function is

$$\lambda(t;\mathbf{x}) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h|T \geq t, \mathbf{x})}{h},$$

- It is easily shown that

$$\lambda(t;\mathbf{x}) = \frac{f(t|\mathbf{x})}{1 - F(t|\mathbf{x})} = \frac{f(t|\mathbf{x})}{S(t|\mathbf{x})}.$$

• An important class of hazards with covariates: **proportional hazard models**. These have the form

$$\lambda(t; \mathbf{x}) = \kappa(\mathbf{x})\lambda_0(t)$$

for $k(\bullet) \; > 0$. The function $\lambda_0(t)$ is called the **baseline hazard**; it is the part of the hazard common to all units in the population. Each individuals hazard function is proportional to $\lambda_0(t)$ based on a function of the observed covariates.

• Most commonly, parameterize $\kappa(\mathbf{x})$ as $\exp(\mathbf{x}\boldsymbol{\beta})$ where $\mathbf{x}$ contains unity. Then

$$\log \lambda(t; \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \log \lambda_0(t),$$

so β_j measures the proportionate increase in the hazard when x_j increases by one unit. If $x_j = \log(z_j)$, β_j is the elasticity of the hazard with respect to z_j.

• Turns out that $\boldsymbol{\beta}$ can be estimated without specifying $\lambda_0(t)$, but, in practice, the shape of the baseline hazard is of interest, along with how the covariates shift the hazard.

Hazard Functions Conditional on Time-Varying Covariates

• The case with time-varying covariates is more difficult to handle, both conceptually and technically.

• Let $\mathbf{x}(t)$ denote the set of covariates at time t; we think of these as being defined at any time, even though we do not collecte data that often. Let $\mathbf{X}(t)$ be the covariate path up through time t, so $\mathbf{X}(t) = \{\mathbf{x}(r) : 0 \leq r \leq t\}$.

• Now define the hazard conditional on covariates as

$$\lambda(t; \mathbf{X}(t)) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h | T \geq t, \mathbf{X}(t + h))}{h}$$

• There are technical issues about when this limit is well defined. Easiest to think of $\mathbf{x}(t+h)$ being constant for h in some neighborhood of zero. As a practical matter, we have to assume this in estimation (as we will see).

• Notice that $\{\mathbf{x}(t)\}$ is, by construction, sequentially exogenous because we are defining the hazard conditional on $\mathbf{X}(t)$. But sometimes we need to actually assume a kind of strict exogeneity. Lancaster (1990) proposes

$$D(\mathbf{X}(t+h)|T \geq t+h, \mathbf{X}(t)) = D(\mathbf{X}(t+h)|\mathbf{X}(t))$$

for all $t \geq 0$, $h > 0$.

- This means that, conditional on the past history of the covariates, the future distribution does not depend on events involving T.
- Later, for estimation with grouped duration data (and especially with heterogeneity), we will use a somewhat different notion of strict exogeneity.

3. Estimation with Time-Invariant Covariates

- Focus here on **single-spell data**: each unit starts in the initial state, and we observe that unit until it leaves the state or the duration is censored (more later).
- Need to carefully define the population. Individuals enter the initial state during the interval $[0, b]$ where $b > 0$ is known. For example, if unemployment duration is measured in weeks, and we consider people becoming unemployed during a particular calendar year, then $b = 52$.
- For now, we ignore the discreteness of the measured durations and treat them as continuous.

Flow Sampling

- Here we sample units that enter the state at some point during $[0, b]$, and then we record the amount of time they are in the initial state. Covariates are collected at the time they enter the state.
- For example, suppose we are interested in the population of workers who become unemployed at some point during 2005. Initially, we observe the population of people who are working at the beginning of 2005, and then we observe some of them "flowing" into the unemployment state.

- One way to collect the data: obtain a large sample from the large population of people employed at the beginning of 2005. Some will become unemployed during that year, and those who become unemployed during the calendar year comprise our sample.
- Another way is to use retrospective sampling: in 2006 or later, we have access to unemployment records in 2005. We can see who flowed into unemployment during 2005.

• Flow data are usually subject to right censoring: after a certain amount of time, we stop following the units in the sample. We might decide to follow each individual who becomes unemployed in 2005 for two years. This means the censoring value is the same for all individuals, 104.

• Alternatively, we might set a fixed calendar date at which point we need to analyze the data – say, the end of 2006. In this case, the censoring value varies with the date the individual enters the initial state (unemployment): in this case, between one and two years.

MLE with Right Censored Flow Data

- For each i, let $a_i \in [0, b]$ be the date at which unit i enters the initial state. So, if $b = 52$, $a_i = 14$ means person i became unemployed during the 14^{th} week if the year.

- Let t_i^* denote the length of time in the initial state. (As a shorthand, we can call this the "true" duration or survival time.) Without right censoring, we would observe random draws $(a_i, t_i^*, \mathbf{x}_i)$.

- Assume t_i^* has a continuous conditional density, $f(t|\mathbf{x};\boldsymbol{\theta})$, $t \geq 0$, where $\boldsymbol{\theta}$ is the vector of unknown parameters.

- Let c_i be the censoring time for individual i.

- Then we observe

$$t_i = \min(t_i^*, c_i).$$

- If $b = 52$ and we censor all units two years after the first date that people can become unemployed, c_i ranges from 52 to 104.
- Key assumption: conditional on $\mathbf{x}_i$, the true duration is independent of the starting date and the censoring time:

$$D(t_i^*|\mathbf{x}_i, a_i, c_i) = D(t_i^*|\mathbf{x}_i).$$

- Might need to include certain variables in $\mathbf{x}_i$ to make this true. In an unemployment duration example, $\mathbf{x}_i$ might need to include seasonal dummies to indicate the time of the year the person became unemployed.
- Cannot have c_i change with t_i^*. So cannot extend the censoring time when you see a duration is lasting a long time.
- To apply MLE, we need to have $D(t_i|\mathbf{x}_i, a_i, c_i)$. First, if the duration is not censored, so that $t_i = t_i^*$, then the density of t_i given $(\mathbf{x}_i, a_i, c_i)$ is just $f(t|\mathbf{x}_i; \boldsymbol{\theta})$. We covered this in censored regression.

- The conditional probability that $t_i = c_i$ is just

$$P(t_i = c_i|\mathbf{x}_i, a_i, c_i) = P(t_i^* > c_i|\mathbf{x}_i, a_i, c_i) = 1 - F(c_i|\mathbf{x}_i;\boldsymbol{\theta}).$$

- Now let $d_i = 1[t_i < c_i]$, so $d_i = 1$ for uncensored observations. Then the log likelihood for observation i is

$$\begin{aligned}\ell_i(\boldsymbol{\theta}) &= d_i \log[f(t_i|\mathbf{x}_i;\boldsymbol{\theta})] + (1 - d_i)\log[1 - F(t_i|\mathbf{x}_i;\boldsymbol{\theta})] \\ &= d_i \log[f(t_i|\mathbf{x}_i;\boldsymbol{\theta})] + (1 - d_i)\log[1 - F(c_i|\mathbf{x}_i;\boldsymbol{\theta})]\end{aligned}$$

- As usual, sum this across $i = 1, \ldots, N$ and maximize the resulting function. The density function is usually very smooth, so MLE has its usual properties of consistency and $\sqrt{N}$-asymptotic normality. Inference is straightforward.
- This is just like a censored regression model, except we may not be primarily interested in estimating the conditional mean.

- The log likelihood does not depend on a_i or b. (And, these values are often not reported in data sets.) But we need the conditional independence assumption on the censoring time to justify this.
- The censoring values, c_i, do appear, unless there are no uncensored observations.
- Unlike with top coding, it is fairly common for c_i to vary with i.
- Note that we only need to have data on t_i. In particular, if $t_i = t_i^*$, we do not need to know the censoring value, c_i.

• A common distribution in applications is the Weibull with covariates. Its hazard is of the proportional hazard form:

$$\lambda(t; \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})\alpha t^{\alpha-1}$$

• The value of α tells us whether there is postive ($\alpha > 1$), negative ($\alpha < 1$), or no ($\alpha = 1$) duration dependence. Can easily test the null $H_0 : \alpha = 1$.

• The lognormal distribution is also used for t_i^* conditional on $\mathbf{x}_i$, which is the same as saying that $\log(t_i^*)$ given $\mathbf{x}_i$ has the $Normal(\mathbf{x}_i\boldsymbol{\delta}, \sigma^2)$ distribution. The hazard is not of the proportional hazard form:

$$\lambda(t;\mathbf{x}) = h[(\log(t) - \mathbf{x}\boldsymbol{\delta})/\sigma]/\sigma t$$
$$h(z) \equiv \frac{\phi(z)}{1 - \Phi(z)}$$

• For fixed $\mathbf{x}\boldsymbol{\delta}$ and σ, this function is not monotonic in t. Therefore, the lognormal may not work as well as other models when the hazard is monotonically increasing or decreasing.

• The parameters $\boldsymbol{\delta}$ are easy to interpret because we have the underlying classical regression model

$$\log(t_i^*) = \mathbf{x}_i\boldsymbol{\delta} + e_i$$
$$e_i|\mathbf{x}_i \sim Normal(0,\sigma^2)$$

• The δ_j are semielasticities or elasticities on the mean duration: $E(t_i^*|\mathbf{x}_i) = \exp(\mathbf{x}_i\boldsymbol{\delta} + \sigma^2/2)$.

• Estimation is the same as the censored normal regression model applied to the log of the duration, with variable right censoring.

Reporting the Results

- Is the shape of the hazard function of interest, or are the effect of covariates on the mean duration?

- In the lognormal case, get the semi-elasticities (or elasticities) on the mean duration directly.

- For the Weibull, can show that

$$\alpha \log(t_i^*) = -\mathbf{x}_i\boldsymbol{\beta} + u_i$$

where u_i is independent of $\mathbf{x}_i$ with density $g(u) = \exp(u)\exp[\exp(-u)]$. The mean of u_i is not zero, but u_i is independent of $\mathbf{x}_i$.

• In Stata, the default is to report the estimates of $\boldsymbol{\beta}$. To get the effects on the mean, namely $-\hat{\beta}_j/\hat{\alpha}$, easiest to use the "nohr" ("no hazard ratio") option.

• If $\beta_j > 0$, an increase in x_j increases the hazard (the "probability" of exiting the initial state, conditional on still being in the state). But an increase in x_j decreases the expected time in the initial state.

- In plotting the hazard for the proportional hazard case, plot

$$\left[N^{-1} \sum_{i=1}^{N} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}) \right] \hat{\lambda}_0(t)$$

or

$$\exp(\bar{\mathbf{x}} \hat{\boldsymbol{\beta}}) \hat{\lambda}_0(t)$$

as functions of t. In the Weibull, case, $\hat{\lambda}_0(t) = \hat{\alpha} t^{\hat{\alpha}-1}$.

• In something like the lognormal case, can insert the mean values of the covariates:

$$\frac{\phi[(\log(t) - \bar{\mathbf{x}}\hat{\boldsymbol{\delta}})/\hat{\sigma}]}{1 - \Phi[(\log(t) - \bar{\mathbf{x}}\hat{\boldsymbol{\delta}})/\hat{\sigma}]} \cdot \frac{1}{\hat{\sigma}t}$$

(which is what Stata does).

• One can also make a case for the average of the estimated hazards:

$$N^{-1} \sum_{i=1}^{N} \left\{ \frac{\phi[(\log(t) - \mathbf{x}_i\hat{\boldsymbol{\delta}})/\hat{\sigma}]}{1 - \Phi[(\log(t) - \mathbf{x}_i\hat{\boldsymbol{\delta}})/\hat{\sigma}]} \cdot \frac{1}{\hat{\sigma}t} \right\}$$

EXAMPLE: Effects of a North Carolina prison work program on Criminal Recidivism.

```
. use recid

. des

Contains data from recid.dta
  obs:         1,445
 vars:            19                          23 Sep 2002 13:37

variable name   type   format      label      variable label
-------------------------------------------------------------------------------
black           byte   %9.0g                  =1 if black
alcohol         byte   %9.0g                  =1 if alcohol problems
drugs           byte   %9.0g                  =1 if drug history
super           byte   %9.0g                  =1 if release supervised
married         byte   %9.0g                  =1 if married when incarc.
felon           byte   %9.0g                  =1 if felony sentence
workprg         byte   %9.0g                  =1 if in N.C. pris. work prg.
property        byte   %9.0g                  =1 if property crime
person          byte   %9.0g                  =1 if crime against person
priors          byte   %9.0g                  # prior convictions
educ            byte   %9.0g                  years of schooling
rules           byte   %9.0g                  # rules violations in prison
age             int    %9.0g                  in months
tserved         int    %9.0g                  time served, rounded to months
follow          byte   %9.0g                  length follow period, months
durat           byte   %9.0g                  min(time until return,follow)
cens            byte   %9.0g                  =1 if duration right censored
ldurat          float  %9.0g                  log(durat)
```

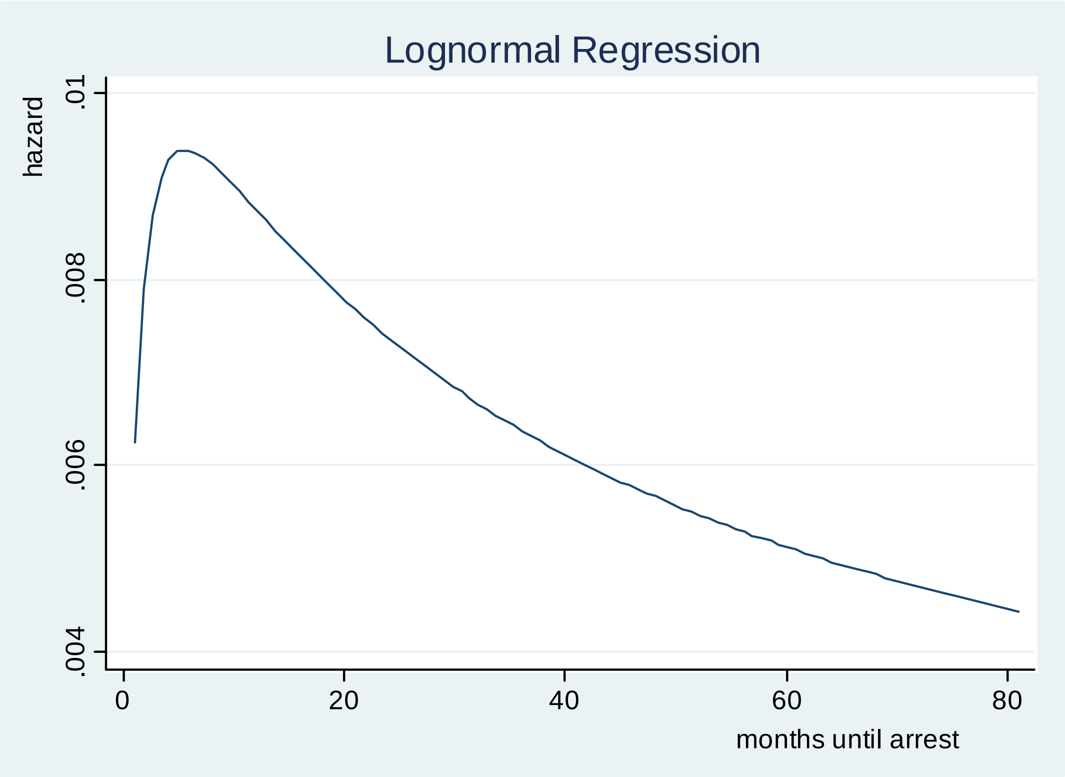
```
. tab follow

     length |
     follow |
    period, |
     months |      Freq.     Percent        Cum.
------------+-----------------------------------
         70 |        169       11.70       11.70
         71 |        148       10.24       21.94
         72 |        139        9.62       31.56
         73 |        161       11.14       42.70
         74 |        116        8.03       50.73
         75 |         71        4.91       55.64
         76 |        157       10.87       66.51
         77 |         91        6.30       72.80
         78 |         88        6.09       78.89
         79 |         91        6.30       85.19
         80 |        124        8.58       93.77
         81 |         90        6.23      100.00
------------+-----------------------------------
      Total |      1,445      100.00

. tab cens

     =1 if |
   duration |
      right |
   censored |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        552       38.20       38.20
          1 |        893       61.80      100.00
------------+-----------------------------------
      Total |      1,445      100.00
```

```
. gen nocens = 1 - cens

. stset durat, failure(nocens)

     failure event:  nocens != 0 & nocens < .
obs. time interval:  (0, durat]
 exit on or before:  failure

------------------------------------------------------------------------------
     1445  total obs.
        0  exclusions
------------------------------------------------------------------------------
     1445  obs. remaining, representing
      552  failures in single record/single failure data
    80013  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        81
```

```
. * Estimate the semi-elasticities on the mean duration, first using
. * the lognormal distribution. The default for the lognormal is
. * to report the semielasticities on the mean duration.

. streg workprg priors tserved felon alcohol drugs black married educ age,
         d(logn)

         failure _d:  nocens
   analysis time _t:  durat

Log-normal regression -- accelerated failure-time form

No. of subjects =         1445                     Number of obs   =      1445
No. of failures =          552
Time at risk    =        80013
                                                   LR chi2(10)     =    166.74
Log likelihood  =    -1597.059                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     workprg |  -.0625714   .1200369    -0.52   0.602    -.2978394    .1726965
      priors |  -.1372528   .0214587    -6.40   0.000     -.179311   -.0951946
     tserved |  -.0193305   .0029779    -6.49   0.000    -.0251671   -.0134939
       felon |   .4439944   .1450865     3.06   0.002     .1596302    .7283586
     alcohol |  -.6349088   .1442165    -4.40   0.000    -.9175681   -.3522496
       drugs |  -.2981599   .1327355    -2.25   0.025    -.5583168   -.0380031
       black |  -.5427175   .1174427    -4.62   0.000     -.772901    -.312534
     married |   .3406835    .139843     2.44   0.015     .0665962    .6147707
        educ |   .0229195   .0253974     0.90   0.367    -.0268584    .0726975
         age |   .0039103   .0006062     6.45   0.000     .0027221    .0050984
       _cons |   4.099386   .3475349    11.80   0.000      3.41823    4.780542
```

```
-------------+----------------------------------------------------------------
     /ln_sig |   .5935861   .0344122    17.25   0.000     .5261395    .6610327
-------------+----------------------------------------------------------------
       sigma |   1.810469   .0623022                      1.692386    1.936791
------------------------------------------------------------------------------

. * The above is the same as applying censored normal regression to log(durat).

. * The default in Stata for the lognormal (and other distributions where the)
. * hazard does not have the proportional hazard form) is to evaluate the
. * covariates at mean values.
```

```
. stcurve, haz
```



Lognormal Regression
hazard
.004
.006
.008
.01
0
20
40
60
80
months until arrest

```
. * Now use the Weibull distribution. The coefficients must be transformed
. * to get the semielasticities on the mean.
. * In Stata, p is our alpha.

. streg  workprg priors tserved felon alcohol drugs black married educ age,
         d(weibull) nohr

         failure _d:  nocens
   analysis time _t:  durat

Weibull regression -- log relative-hazard form

No. of subjects =         1445                     Number of obs   =      1445
No. of failures =          552
Time at risk    =        80013
                                                   LR chi2(10)     =    165.48
Log likelihood  =   -1633.0325                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     workprg |   .0908893   .0906478     1.00   0.316    -.0867772    .2685558
      priors |   .0887867   .0134355     6.61   0.000     .0624535    .1151198
     tserved |   .0135625   .0016808     8.07   0.000     .0102682    .0168567
       felon |  -.2994775    .105974    -2.83   0.005    -.5071826   -.0917723
     alcohol |   .4473611   .1057353     4.23   0.000     .2401236    .6545985
       drugs |   .2814605   .0978644     2.88   0.004     .0896499    .4732711
       black |   .4537147   .0883037     5.14   0.000     .2806426    .6267867
     married |  -.1515864   .1092454    -1.39   0.165    -.3657035    .0625307
        educ |  -.0232984   .0194196    -1.20   0.230    -.0613601    .0147633
         age |  -.0037246    .000525    -7.09   0.000    -.0047536   -.0026956
       _cons |  -3.402094   .3010177   -11.30   0.000    -3.992077    -2.81211
```

```
-------------+----------------------------------------------------------------
       /ln_p |  -.2158398   .0389149    -5.55   0.000    -.2921115   -.1395681
-------------+----------------------------------------------------------------
           p |   .8058644   .0313601                      .7466852    .8697338
         1/p |   1.240904   .0482896                      1.149777    1.339252
------------------------------------------------------------------------------
```

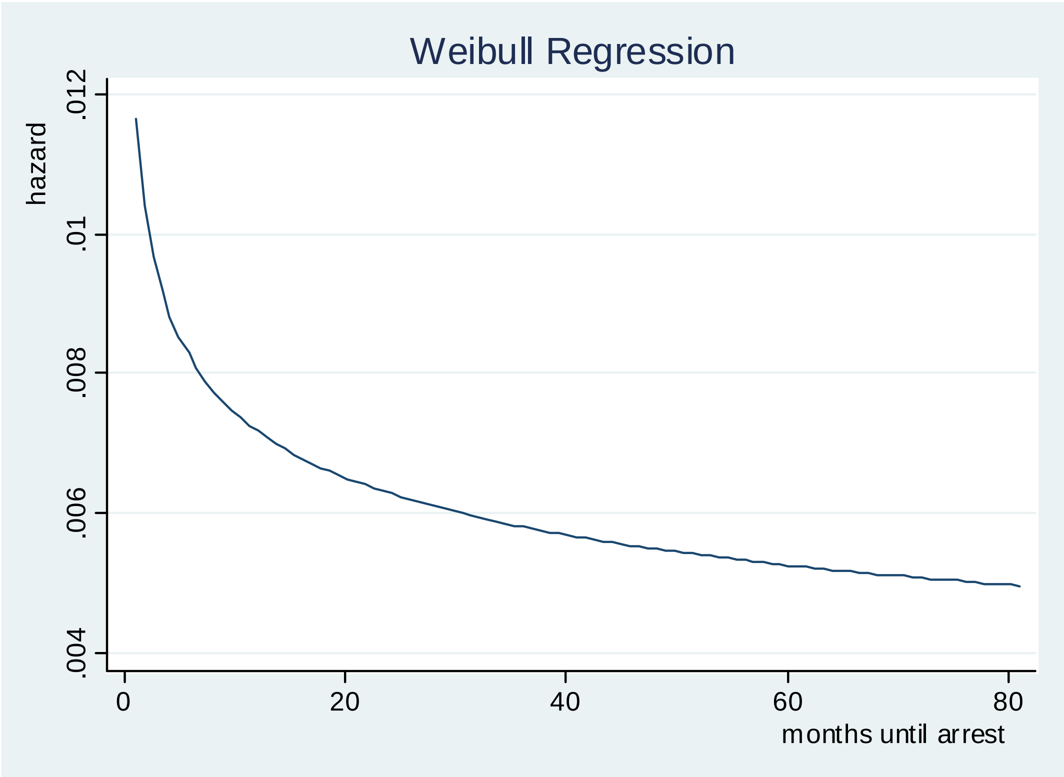
• The estimate of α is about $.806$, and so there is negative duration dependence.

```
. * Compute the semielasticity on the mean for workprg:

. di -.091/.806
-.11290323

. * This estimate of the effect of workprg is somewhat larger in magnitude than
. * the lognormal estimate, -.063.

. * Interestingly, the lognormal distribution fits substantially
. * better than the Weibull: -1,597.059 for lognormal versus -1,633.033
. * for the Weibull.

. * With some work, we could compute a Vuong model specification statistic.
. * We need to obtain the log likelihood values for each i.
```

```
. * Now report the exp(betaj) for the Weibull.

. streg workprg priors tserved felon alcohol drugs black married educ age,
         d(weibull)

         failure _d:  nocens
   analysis time _t:  durat

Weibull regression -- log relative-hazard form

No. of subjects =         1445                     Number of obs   =      1445
No. of failures =          552
Time at risk    =        80013
                                                   LR chi2(10)     =    165.48
Log likelihood  =   -1633.0325                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     workprg |   1.095148   .0992728     1.00   0.316     .9168814    1.308074
      priors |   1.092848    .014683     6.61   0.000     1.064445    1.122008
     tserved |   1.013655   .0017037     8.07   0.000     1.010321       1.017
       felon |   .7412054   .0785485    -2.83   0.005     .6021898    .9123128
     alcohol |   1.564179    .165389     4.23   0.000     1.271406     1.92437
       drugs |   1.325064   .1296765     2.88   0.004     1.093791    1.605237
       black |   1.574149   .1390031     5.14   0.000      1.32398    1.871587
     married |   .8593436   .0938794    -1.39   0.165     .6937084    1.064527
        educ |   .9769709   .0189724    -1.20   0.230     .9404845    1.014873
         age |   .9962823    .000523    -7.09   0.000     .9952577     .997308
```

```
-------------+----------------------------------------------------------------
       /ln_p |  -.2158398   .0389149    -5.55   0.000    -.2921115   -.1395681
-------------+----------------------------------------------------------------
           p |   .8058644   .0313601                      .7466852    .8697338
         1/p |   1.240904   .0482896                      1.149777    1.339252
------------------------------------------------------------------------------

. * For example, for workprg, the 1.095 means that the hazard for workprg = 1
. * is 1.095 times the hazard when workprg = 0. That is, it is the ratio
. * of hazards with the other covariates fixed and xj
. * increased by one unit. The estimate 9.51% is a more accurate estimate in
. * the percentage increase in the hazard starting at workprg = 0 and
. * then setting workprg = 0. (Compare 9.01%.)

. * The shape of the Weibull hazard is notably different from the lognormal
. * (and we know the Weibull is either monotonically increasing,
. * monotonically decreasing, or flat).
```

```
. stcurve, haz
```



Weibull Regression
hazard
.004
.006
.008
.01
.012
0
20
40
60
80
months until arrest

Unobserved Heterogeneity with Single-Spell Flow Data

• We can easily add multiplicative unobserved heterogeneity to hazard specifications. (Heterogeneity in this context is also called *frailty*.) Let $\kappa(t;\mathbf{x})$ be a hazard function that need not be of the proportional hazard form. For a random draw i, let $v_i > 0$ be the unobserved heterogeneity. Then we can specify

$$\lambda(t;\mathbf{x}_i,v_i) = v_i\kappa(t;\mathbf{x}_i)$$

• Normalization is almost always $E(v_i) = 1$, so $\kappa(\bullet;\mathbf{x})$ is the average hazard for an individual with observed characteristics $\mathbf{x}$.

• The hazard conditional on v is usually called the **conditional hazard**. (The conditioning on $\mathbf{x}$ is always implicit.) Plots of the conditional hazard set $v = 1$ and then evaluate $\mathbf{x}$ at its sample mean value, $\bar{\mathbf{x}}$, and then plots $\kappa(t;\bar{\mathbf{x}})$ as a function of t.

• We can find the cdf of t_i^* conditional on $(\mathbf{x}_i, v_i)$ as

$$F(t|\mathbf{x}_i, v_i) = 1 - \exp\left[-v_i \int_0^t \kappa(r;\mathbf{x}_i)dr\right] \equiv 1 - \exp[-v_i \xi(t;\mathbf{x}_i)]$$

where $\xi(t;\mathbf{x}_i) \equiv \int_0^t \kappa(r;\mathbf{x}_i)dr$.

• It is almost always assumed that v_i is independent of $\mathbf{x}_i$, in which case the cdf of t_i^* given $\mathbf{x}_i$ can be found by "integrating out" the heterogeneity using the the density of v_i:

$$G(t|\mathbf{x}_i) = \int_0^\infty F(t|\mathbf{x}_i, v)h(v)dv,$$

where $h(v)$ is the density of v_i. It is parametric versions of $G(t|\mathbf{x}_i)$ that we use in standard MLE.

• Notation here assumes v_i is continuous. Not critical, but common.

• The hazard function associated with $G(t|\mathbf{x}_i)$ us often called the **unconditional hazard** because we are not conditioning on v.

- A leading case is when v_i has a $Gamma(\delta, \delta)$ distribution – so $E(v_i) = \delta/\delta = 1$ and $Var(v_i) = \delta/\delta^2 = 1/\delta \equiv \eta$. Then then $G(t|\mathbf{x}_i)$ has a fairly simple form. Namely,

$$G(t|\mathbf{x}_i) = 1 - [1 + \eta\xi(t; \mathbf{x}_i)]^{-(1/\eta)}$$

- Once we have a parametric form for $\xi(t; \mathbf{x})$, MLE is relatively straightforward (though we must still account for right censoring).

• Suppose we use the Weibull hazard with gamma heterogeneity. Then the conditional hazard has the form

$$\lambda(t; \mathbf{x}, v) = v \cdot \exp(\mathbf{x}\boldsymbol{\beta})\alpha t^{\alpha-1}$$

• Because of the proportional hazard form, the shape of $\lambda(\cdot; \mathbf{x}, v)$ for given $(\mathbf{x}, v)$ is the shape of the baseline hazard, $\lambda_0(t) = \alpha t^{\alpha-1}$.

- Then $\xi(t;\mathbf{x}_i) = \exp(\mathbf{x}_i\boldsymbol{\beta})\left(\int_0^t \alpha r^{\alpha-1}dr\right) = \exp(\mathbf{x}_i\boldsymbol{\beta})t^\alpha$ and so

$$G(t|\mathbf{x}_i;\boldsymbol{\theta},\delta) = 1 - [1 + \exp(\mathbf{x}_i\boldsymbol{\beta})\eta t^\alpha]^{-(1/\eta)},$$

which is known as the **Burr distribution**. The Burr cdf without covariates is $1 - (1 + \eta\gamma t^\alpha)^{-(1/\eta)}$. (As $\eta = Var(v_i) \rightarrow 0$ the Burr cdf converges to $1 - \exp(-\gamma t^\alpha)$, which is the Weibull.)

- This leads to a very tractable analysis, and is pretty common.
- A key focus is the nature of duration dependence conditional on $(\mathbf{x}_i, v_i)$. Is α greater than, less than, or equal to unity?

- But the idea of trying to learn about state dependence conditional on heterogeneity is questionable. One can simply start with, say, the Burr distribution as the correct duration distribution without ever introducing heterogeneity.

- Consider a regression model with a random slope, written as $E(y_i|x_i, v_i) = \beta_0 + \beta_1 x_i + v_i x_i$ and suppose v_i and x_i are independent with $E(v_i) = 0$. Then

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i + E(v_i|x_i)x_i = \beta_0 + \beta_1 x_i$$

- How can we tell if there is heterogeneity in the partial effect of x_i? If we also assume $Var(y_i|x_i, v_i) = \sigma_u^2$, then

$$Var(y_i|x_i) = \sigma_u^2 + \sigma_v^2 x_i^2$$

and so heteroskedasticity in $Var(y_i|x_i)$ is a consequence of the random slope. But heteroskedasticity can arise for many reasons, and a finding that $Var(y_i|x_i)$ depends on x_i^2 essentially tells us nothing about random coefficients.

• Importantly, the shape of the hazard for the Burr distribution is often quite different from the Weibull. In fact, the survivor function for the Burr distribution is

$$S(t) = (1 + \eta\gamma t^{\alpha})^{-(1/\eta)}$$

and the hazard is

$$\lambda(t) = \frac{\gamma\alpha t^{\alpha-1}}{(1 + \eta\gamma t^{\alpha})}$$

• When $\eta = 0$, Weibull. When $\eta = 1$, log-logistic.

• With covariates, the hazard for the Burr distribution – that is, a Weibull hazard with gamma multiplicative heterogeneity – is

$$\frac{\exp(\mathbf{x}\boldsymbol{\beta})\alpha t^{\alpha-1}}{[1+\eta\exp(\mathbf{x}\boldsymbol{\beta})t^{\alpha}]}$$

This is the unconditional hazard, which we contrast with the conditional hazard with v set to one:

$$\exp(\mathbf{x}\boldsymbol{\beta})\alpha t^{\alpha-1}.$$

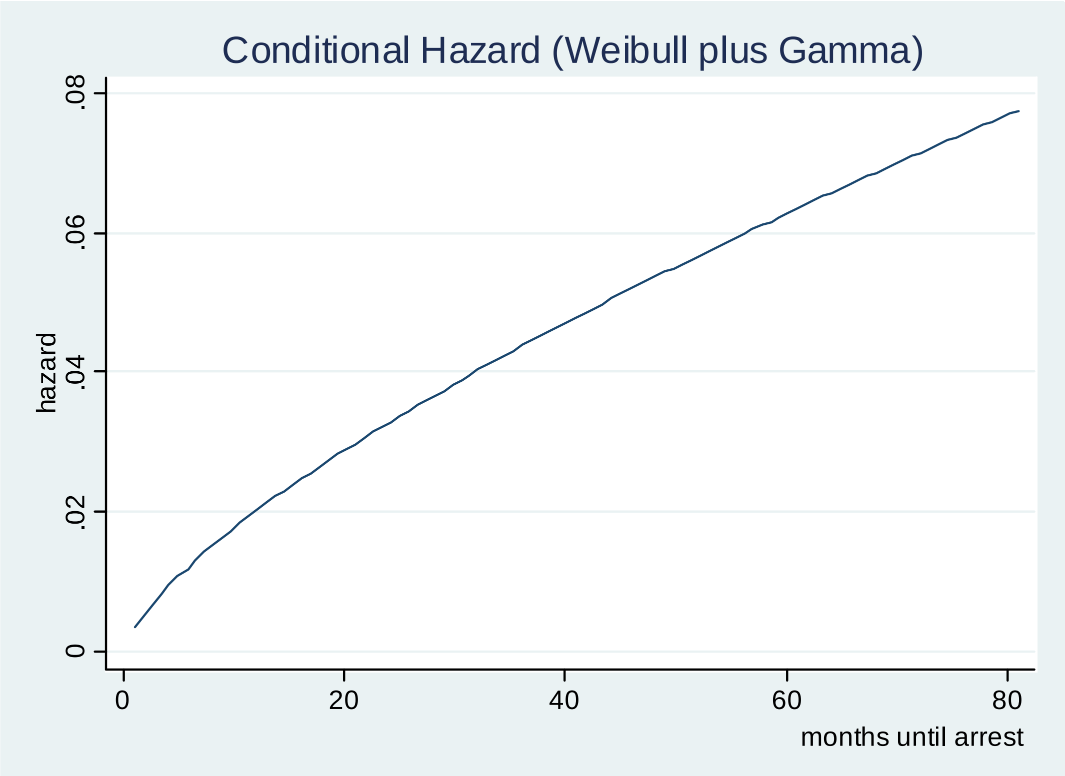
- For fixed $\mathbf{x}$, the conditional hazard is either flat ($\alpha = 1$), monotically increasing ($\alpha > 1$), or monotonically decreasing ($\alpha < 1$). By contrast, the sign of derivative of the unconditional hazard generally changes with t.
- General problem with interpreting hazard models with unobserved heterogeneity: cannot distinguish between the existence of heterogeneity and a functional form for the hazard that is not flexible enough. A Weibull hazard at the individual level with gamma heterogeneity is observationally equivalent to a Burr hazard without heterogeneity.

```
. streg  workprg priors tserved felon alcohol drugs black married educ age,
         d(weibull) fr(gamma) nohr

         failure _d:  nocens
   analysis time _t:  durat


Weibull regression -- log relative-hazard form
                      Gamma frailty

No. of subjects =         1445                    Number of obs   =      1445
No. of failures =          552
Time at risk    =        80013
                                                  LR chi2(10)     =    143.82
Log likelihood  =   -1584.9172                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     workprg |   .0073827   .2038775     0.04   0.971    -.3922099    .4069753
      priors |   .2431142   .0421543     5.77   0.000     .1604933    .3257352
     tserved |   .0349363   .0070177     4.98   0.000     .0211818    .0486908
       felon |  -.7909533   .2666084    -2.97   0.003    -1.313496   -.2684104
     alcohol |   1.173558   .2805222     4.18   0.000     .6237451    1.723372
       drugs |   .2847665   .2233072     1.28   0.202    -.1529074    .7224405
       black |   .7715762   .2038289     3.79   0.000      .372079    1.171073
     married |  -.8057042   .2578214    -3.13   0.002    -1.311025   -.3003834
        educ |  -.0271193    .044901    -0.60   0.546    -.1151237     .060885
         age |  -.0052162   .0009974    -5.23   0.000    -.0071711   -.0032613
       _cons |  -5.393658    .720245    -7.49   0.000    -6.805312   -3.982004
```

```
-------------+----------------------------------------------------------------
       /ln_p |   .5352553   .0951206     5.63   0.000      .3488225    .7216882
     /ln_the |   1.790243   .1788498    10.01   0.000      1.439703    2.140782
-------------+----------------------------------------------------------------
           p |   1.707884   .1624549                       1.417398    2.057904
         1/p |   .5855198    .055695                       .4859312    .7055184
       theta |   5.990906   1.071472                       4.219445    8.506084
------------------------------------------------------------------------------
Likelihood-ratio test of theta=0: chibar2(01) =    96.23 Prob>=chibar2 = 0.000
```

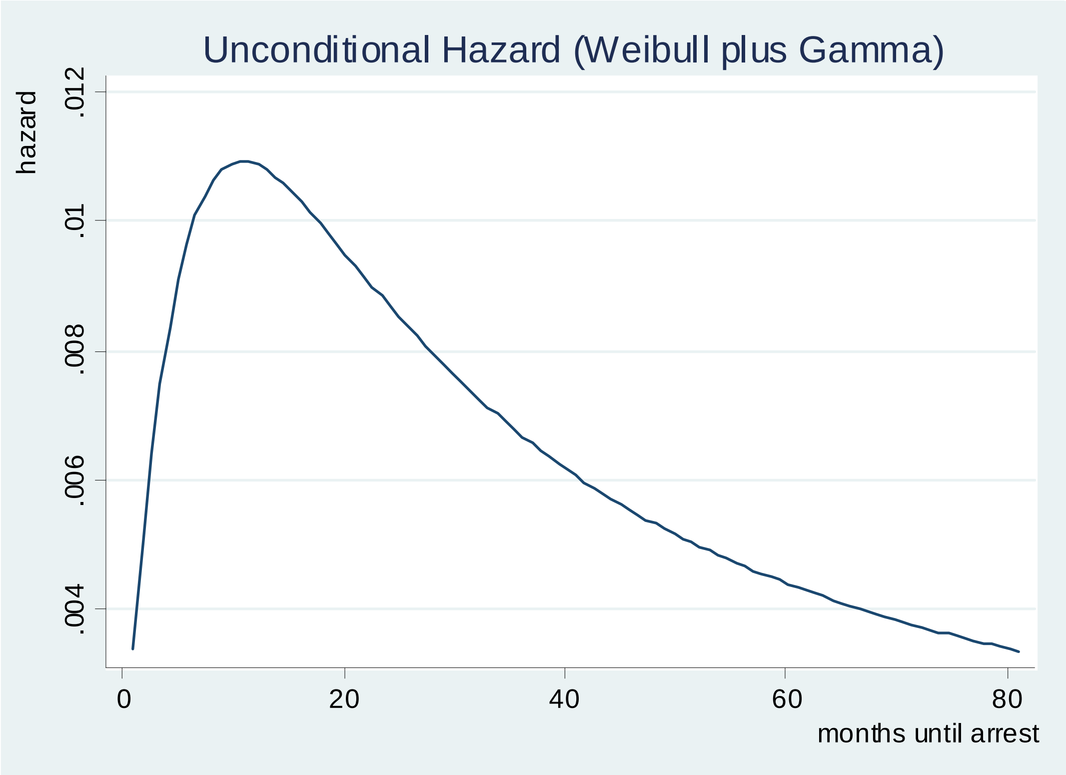
• The (conditional) Weibull hazard, with heterogeneity (ν_i in our notation, α_i in Stata's notation) set to unity:

```
. stcurve, haz alpha1
```



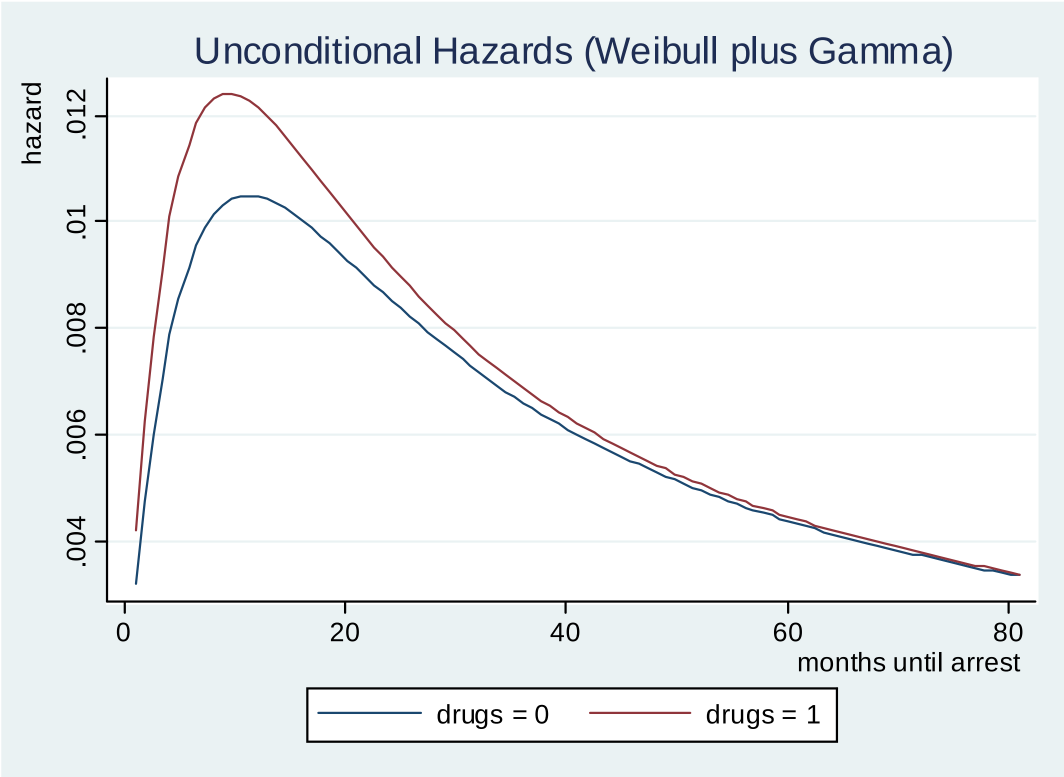
• The (unconditional) Burr hazard, evaluated at the mean value of the covariates:

```
. stcurve, haz
```



- The unconditional hazard has the same shape as the lognormal hazard obtained earlier, but its peak is at around 12 months, rather than six months.
- The Burr distribution fits better than the lognormal: $-1,584.92$ versus $-1,597.06$. I suspect a Vuong test would reject the lognormal in favor of the Burr, but the calculation needs to be done.

```
. stcurve, haz at1(drugs = 0) at2(drugs = 1)
```



- One advantage of the Weibull compared with the lognormal is that the latter forces a turning point in the hazard; it cannot be monotomically decreasing. By contrast, if negative duration dependence is the best description, the Weibull allows that.

Stock Sampling

• The population is still the same: individuals entering the initial state during the interval $[0, b]$.

• Unlike flow sampling, where we essentially sample individuals when they enter the intial state, stock sampling obtains a random sample of of individuals that are in the initial state at time b. In other words, we have a random sample from the "stock" of people unemployed at time b.

- For example, suppose we measure unemployment durations dailay, and we are interested in people who become unemployed at some point in the year 2005. However, we only obtain a random sample of people who are actually unemployed on the last day of 2005.
- Stock sampling introduces a clear sample selection bias: for a given start date $a_i \in [0, b]$, we are more likely to see someone unemployed on the last day of 2005 the longer they have been unemployed.

- The sample selection problem induced by stock sampling is is sometimes called **length-biased sampling**. The label **left truncation** also fits because, given a_i, we do not see someone in our sample if their duration, t_i^*, is less than $b - a_i$. (Even if we observe the a_i for the people we sample at time b, we still have a truncation problem.)

• First assume that a_i is observed for eveyone we sample at time b. Let $(a_i, c_i, \mathbf{x}_i, t_i)$ denote a random draw, where $t_i = \min(t_i^*, c_i)$ is the right censored duration, as before.

• To obtain the MLE, we need to account for the truncated sampling.

• The condition for being observed in our sample is

$$t_i^* \geq b - a_i$$

and, under $D(t_i^*|a_i, c_i, \mathbf{x}_i) = D(t_i^*|\mathbf{x}_i)$,

$$P(t_i^* \geq b - a_i|a_i, c_i, \mathbf{x}_i) = 1 - F(b - a_i|\mathbf{x}_i; \boldsymbol{\theta}).$$

• We can apply estimation for truncated samples. Recall that the density that accounts for right censoring is

$$f(t|\mathbf{x}_i;\boldsymbol{\theta})^d[1 - F(c_i|\mathbf{x}_i;\boldsymbol{\theta})]^{(1-d)}.$$

To account for the left truncation, we divide it by $1 - F(b - a_i|\mathbf{x}_i;\boldsymbol{\theta})$.

- Plugging in the data and taking logs gives, for a raw i satisfying $t_i^* \geq b - a_i$,

$$d_i \log[f(t_i|\mathbf{x}_i;\boldsymbol{\theta})] + (1 - d_i)\log[1 - F(t_i|\mathbf{x}_i;\boldsymbol{\theta})] - \log[1 - F(b - a_i|\mathbf{x}_i;\boldsymbol{\theta})]$$

- As before, this log likelihood depends on c_i because $t_i = c_i$ when $d_i = 0$. But it also depends explicitly on b and the starting date, a_i (actually, only $b - a_i$).
- Estimation is again straightforward. So, provided that when we sample the units at time b we find out the starting dates a_i, it is fairly easy to use stock data.

• Notice that if t_i^* is right censored at calendar data b – which means the censoring time is $c_i = b - a_i$ – the log likelihood is zero: $d_i = 0$ and $t_i = c_i = b - a_i$, and so

$$\ell_i(\boldsymbol{\theta}) = \log[1 - F(b - a_i|\mathbf{x}_i;\boldsymbol{\theta})] - \log[1 - F(b - a_i|\mathbf{x}_i;\boldsymbol{\theta})] = 0.$$

• Because the log likelihood does not depend on $\boldsymbol{\theta}$, observations that are right censored at the calendar date when the data are collected provide no information for estimating $\boldsymbol{\theta}$.

- If we sample from the units in the initial state at time b and do not follow any of the units after this time then $t_i = c_i = b - a_i$ for all i. Consequently, $\boldsymbol{\theta}$ is not identified by the previous log likelihood.
- Intuitively, the log likelihood is conditional on $(a_i, \mathbf{x}_i)$, but t_i and c_i are deterministic functions of a_i: there is no randomness conditional on $(a_i, \mathbf{x}_i)$.

• If we are willing to make assumptions about the distribution of starting times, $D(a_i|\mathbf{x}_i)$, we can generally identify $\boldsymbol{\theta}$. Denote its density $k(a|\mathbf{x}_i;\boldsymbol{\eta})$ for parameters $\boldsymbol{\eta}$.

• Now let s_i be the selection indicator: we observe an observation from those entering the intial state in $[0,b]$ if $s_i = 1$ where $s_i = 1[t_i^* \geq b - a_i]$. Estimation of $\boldsymbol{\theta}$ (and $\boldsymbol{\eta}$) can proceed by obtaining the density of a_i conditional on $\mathbf{x}_i$ and $s_i = 1$. (This is the only density we can hope to estimate, as we observe $(a_i, \mathbf{x}_i)$ only when $s_i = 1$.)

• The density is informative for estimating $\boldsymbol{\theta}$ even though $\boldsymbol{\eta}$ is not functionally related to $\boldsymbol{\theta}$.

- It can be shown (see Problem 22.6) that

$$p(a|\mathbf{x}_i, s_i = 1) = k(a|\mathbf{x}_i;\boldsymbol{\eta})[1 - F(b - a|\mathbf{x}_i;\boldsymbol{\theta})]/P(s_i = 1|\mathbf{x}_i;\boldsymbol{\theta},\boldsymbol{\eta})$$

$$P(s_i = 1|\mathbf{x}_i;\boldsymbol{\theta},\boldsymbol{\eta}) = \int_0^b [1 - F(b - u|\mathbf{x}_i;\boldsymbol{\theta})]k(u|\mathbf{x}_i;\boldsymbol{\eta})du$$

- So the log likelihood for an observation actually in the sample is

$$\log[k(a_i|\mathbf{x}_i;\boldsymbol{\eta})] + \log[1 - F(b - a_i|\mathbf{x}_i;\boldsymbol{\theta})] - \log[P(s_i = 1|\mathbf{x}_i;\boldsymbol{\theta},\boldsymbol{\eta})]$$

and this generally depends on $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

- Another situation sometimes arises with stock sampling: the a_i are not observed. That is, when we sample at time b, we know people who are in the initial state (such as being unemployed) but we do not know where in $[0, b]$ the spell started. This is known as the problem of **left censoring**.

• To get anywhere, we have to follow some units after the initial sampling at time b. If there were no right censoring, this means we can follow the length of time in the state after date b. If we call this r_i, we have

$$r_i = t_i^* + a_i - b.$$

We observe r_i but, even without right censoring, we do not observe t_i^* and a_i separately.

• If we also allow for right censoring, and account for the left truncation problem, we can obtain a likelihood that identifies $\boldsymbol{\theta}$. See Problem 22.8.

4. Grouped Duration Data

- In the previous sections we have acted as if we observe the duration continuously, even if this is only a rough approximation.
- Alternatively, we can recognize that, even though time in principal can be measured continuously, our observations are always in discrete intervals. When we approach the problem from this perspective, we say we have **grouped duration data**.
- For example, the recidivism duration in the previous example is reported monthly. In effect, we only know that recidivism occurred at some time during the month.

- Assume that the nonnegative real line, $[0,\infty)$, is divided into intervals:

$$[0,a_1),\ [a_1,a_2),\ \ldots,[a_{M-1},a_M),\ [a_M,\infty),$$

where the a_m are known constants. For example, it is common to have $a_1 = 1, a_2 = 2, \ldots, a_M = M$. (For example, a duration is measured in months and we only know which monthly interval the initial state was exited.)

- For a generic unit in the population, let y_m be a binary variable $y_m = 1[a_{m-1} \leq T < a_m)$ for $m \leq M$, where we again use T to denote the random duration. Therefore, y_m is one if and only if the unit left the initial state in the interval $[a_{m-1}, a_m)$.
- If the state is not left prior to time a_M, $y_m = 0$, $m = 1, \ldots, M$ and $y_{M+1} = 1$.
- In this sense, all durations are censored at time a_M.

• As before, units may enter the initial state at different calendar dates. But with flow data, provide that date is exogenous – and we may have to put date dummies in $\mathbf{x}_i$ to ensure that – the starting dates play no role. So we do not explicitly allow for them.

• Because we assume single spell data, $y_{m+1} = 1$ if $y_m = 1$. Once a unit exits, the unit is not longer followed.

- Without right censoring the log-likelihood is straightforward to construct. What we need is the joint distribution (conditional on $\mathbf{x}$), $D(y_1, \ldots, y_M, y_{M+1}|\mathbf{x})$.
- But the only possible outcomes are $M + 1$ sequences of the form $(0, 0, \ldots, 0, 1, 1, \ldots, 1)$.
- Remember, $y_{M+1} = 1$ always, with $(0, 0, \ldots, 0, 0, 1)$ indicating that exit occurs at time a_M or later (and that is all we know).

• If unity first appears in the m^{th} position for $m \leq M$ then the spell ends before time a_M. The probability of observing an exit in interval $m \leq M$ can be written using a standard relationship between conditional and unconditional probabilities:

$$P(a_{m-1} \leq T < a_m|\mathbf{x}) = P(a_{m-1} \leq T < a_m|T \geq a_{m-1}, \mathbf{x}) \cdot P(T \geq a_{m-1}, \mathbf{x}).$$

By recursive subsitution, and using $P(T \geq a_0|\mathbf{x}) = P(T \geq 0|\mathbf{x}) = 1$,

$$P(T \geq a_{m-1}, \mathbf{x}) = \prod_{h=1}^{m-1} P(T \geq a_h|T \geq a_{h-1}, \mathbf{x}).$$

Now use

$$P(T \geq a_h | T \geq a_{h-1}, \mathbf{x}) = 1 - P(a_{h-1} \leq T < a_h | T \geq a_{h-1}, \mathbf{x}),$$

and so

$$\prod_{h=1}^{m-1} P(T \geq a_h | T \geq a_{h-1}, \mathbf{x}) = \prod_{h=1}^{m-1} [1 - P(a_{h-1} \leq T < a_h | T \geq a_{h-1}, \mathbf{x})].$$

• We have shown that $P(a_{m-1} \leq T < a_m|\mathbf{x})$ is

$$P(a_{m-1} \leq T < a_m|T \geq a_{m-1}, \mathbf{x}) \prod_{h=1}^{m-1}[1 - P(a_{h-1} \leq T < a_h|T \geq a_{h-1}, \mathbf{x})]$$

• For $m = 1, \ldots, M,$ we use the relationship between conditional probabilities and the hazard function from earlier:

$$\begin{aligned} &P(a_{m-1} \leq T < a_m|T \geq a_{m-1}, \mathbf{x}) \\ &\quad = 1 - \exp\left[-\int_{a_{m-1}}^{a_m} \lambda(r; \mathbf{x}, \boldsymbol{\theta})dr\right] \equiv 1 - \alpha_m(\mathbf{x}, \boldsymbol{\theta}) \end{aligned}$$

and so

$$1 - P(a_{m-1} \leq T < a_m|T \geq a_m, \mathbf{x}) = \alpha_m(\mathbf{x}, \boldsymbol{\theta}).$$

- It follows that, for a spell that ends in an interval $m \in \{1, 2, \ldots, M\}$,

$$P(a_{m-1} \leq T < a_m|\mathbf{x}) = [1 - \alpha_m(\mathbf{x}, \boldsymbol{\theta})]\left[\prod_{h=1}^{m-1} \alpha_h(\mathbf{x}, \boldsymbol{\theta})\right].$$

- For spells not finished by time a_M – that is, $y_m = 0$, $m \leq M$, we need

$$P(T \geq a_M|\mathbf{x}) = \prod_{h=1}^{M} \alpha_h(\mathbf{x}, \boldsymbol{\theta})$$

• Therefore, the log likelihood for a random draw i where the transition occurs in interval $m_i \leq M$ is

$$\sum_{h=1}^{m_i-1} \log[\alpha_h(\mathbf{x}_i, \boldsymbol{\theta})] + \log[1 - \alpha_{m_i}(\mathbf{x}, \boldsymbol{\theta})],$$

and for incompleted spells it is

$$\sum_{h=1}^{M} \log[\alpha_h(\mathbf{x}_i, \boldsymbol{\theta})].$$

• Now bring back censoring that can happen before time a_M. Define a sequence of censoring dummies, censoring dummies, $\{c_1, \ldots, c_M\}$, where $c_m = 1$ if the duration was censored in interval $[a_{m-1}, a_m)$. Spells not completed by time a_M are censored, so $c_{M+1} = y_{M+1} = 1$.

• A sufficient condition for exogeneity of the censoring mechanism is exogenous

$$D(T|c_1, \ldots, c_M, \mathbf{x}) = D(T|\mathbf{x})$$

• If censoring occurs in interval m_i, we only know the duration lasted at least to time a_{m_i-1}, so the log likelihood for a censored duration is simply

$$\sum_{h=1}^{m_i-1} \log[\alpha_h(\mathbf{x}_i, \boldsymbol{\theta})]$$

• Let d_i be a dummy variable equal to unity for an uncensored observation. Then the log likelihood for observation i is

$$\sum_{h=1}^{m_i-1} \log[\alpha_h(\mathbf{x}_i, \boldsymbol{\theta})] + d_i \log[1 - \alpha_{m_i}(\mathbf{x}, \boldsymbol{\theta})]$$

• As a practical matter, any $m_i = 1$ observations do not contribute to the log likelihood. They are uninformative because all they imply is $T \geq 0$, which we already know.

- $m_i = M + 1$ corresponds to observations censored in the last interval.

But observations can be censored before then, too.

- The log likelihood for the entire sample is

$$\sum_{i=1}^{N}\left\{\sum_{h=1}^{m_i-1}\log[\alpha_h(\mathbf{x}_i,\boldsymbol{\theta})] + d_i\log[1-\alpha_{m_i}(\mathbf{x}_i,\boldsymbol{\theta})]\right\}$$

• To use this for estimation, we need to specify a hazard function. A convenient formulation (and flexible if there are many intervals) is a **piecewise-constant proportional hazard**: for $m = 1, \ldots, M$,

$$\lambda(t; \mathbf{x}, \boldsymbol{\theta}) = \kappa(\mathbf{x}, \boldsymbol{\beta})\lambda_m,\ a_{m-1} \leq t < a_m$$

where, typically, $\kappa(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})$.

• When we allow the hazard to be unrestricted across intervals, we cannot estimate the hazard for $t \geq a_M$ (because there is no way to extrapolate).

• With $\kappa(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})$,

$$\begin{aligned}\alpha_m(\mathbf{x}, \boldsymbol{\theta}) &= \exp\left[-\exp(\mathbf{x}\boldsymbol{\beta})\left(\int_{a_{m-1}}^{a_m} \lambda_m dr\right)\right] \\ &= \exp[-\exp(\mathbf{x}\boldsymbol{\beta})\lambda_m(a_m - a_{m-1})]\end{aligned}$$

• Usually, $a_m = m$ and so $\alpha_m(\mathbf{x}, \boldsymbol{\theta}) = \exp[-\exp(\mathbf{x}\boldsymbol{\beta})\lambda_m]$. With unrestricted λ_m, $\mathbf{x}$ does not include a constant.

• With this specification, the hazard is discontinuous at the endpoints of the intervals. The theory easily extends to handle this case.

• A different way to look at the piecewise-constant proportional hazard case is that we estimate $\boldsymbol{\beta}$ along with parameters

$$\eta_m = \int_{a_{m-1}}^{a_m} \lambda_0(r)dr$$

• Either way, the estimation is the same, and it is easier to think of estimating the baseline hazard directly.

• Given the $\hat{\lambda}_m$, can plot these as a function of m; usually plot $\hat{\lambda}_m$ at the midpoints of the intervals $[a_{m-1}, a_m)$, and then smooth the graph.

• Of course, the $\hat{\beta}_j$ are often of interest, too.

• What if there are no covariates? The MLEs $\hat{\lambda}_m$ lead to a well-known estimate of the survivor function. For $m = 1, \ldots, M$,

$$S(a_m) = P(T > a_m) = \prod_{r=1}^{m} P(T > a_r | T > a_{r-1}).$$

• For $r = 1, \ldots, M$, let N_r denote the number of units in the **risk set**: they have neither left the initial state nor been censored at time a_{r-1}. ($N_1 = N$ is the size of the initial random sample, N_2 is the units who did not leave the initial state during the first interval, less those who were censored in $[a_1, a_2)$, and so on.)

• Let E_r denote the number of units observed to leave in $[a_{r-1}, a_r)$. A consistent estimator of $P(T > a_r | T > a_{r-1})$ is

$$(N_r - E_r)/N_r,\ r = 1, \ldots, M.$$

So

$$\hat{S}(a_m) = \prod_{r=1}^{m} (N_r - E_r)/N_r,\ m = 1, \ldots, M.$$

• Called the **Kaplan-Meier estimator**. Lancaster (1990, Section 8.2) shows the MLE of the λ_m produces the KM estimator.

- If the intervals $[a_{m-1}, a_m)$ are coarser than the data – for example, unemployment duration is measured weekly but the intervals are four weeks wide – then we can specify nonconstant hazards within each interval. For example, could specify a Weibull within each interval.

Time-Varying Covariates

• For a generic unit from the population, let $\mathbf{x}_m$ denote the vector of covariates for the m^{th} interval, that is, $[a_{m-1}, a_m)$. Because we only observed grouped durations, assume the covariates are constant within each interval. So $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ be the sequence of covariates.

• In effect, we have a panel data set: $\{(y_m, \mathbf{x}_m, c_m) : m = 1, \ldots, M\}$, where $c_m = 1$ if the observation is censored in interval m (we only know $T \geq a_{m-1}$). Recall $y_{M+1} = c_{M+1} = 1$.

• Exogeneity of censoring is now stated as

$$D(T|T \geq a_{m-1}, \mathbf{x}_m, c_m) = D(T|T \geq a_{m-1}, \mathbf{x}_m),\ m = 1, \ldots, M.$$

• The probabilities for the likelihood function are similar to before, but now we condition on the covariates in the stated time interval:

$$\begin{aligned} P(y_m &= 1|y_{m-1} = 0, \mathbf{x}_m, c_m = 0) \\ &= 1 - \exp\left[-\int_{a_{m-1}}^{a_m} \lambda(r; \mathbf{x}_m, \boldsymbol{\theta})dr\right] = 1 - \alpha_m(\mathbf{x}_m, \boldsymbol{\theta}) \end{aligned}$$

• The log likelihood takes the same value as before, except we indicate that the covariates might change across intervals:

$$\sum_{i=1}^{N}\left\{\sum_{h=1}^{m_i-1}\log[\alpha_h(\mathbf{x}_h,\boldsymbol{\theta})] + d_i\log[1-\alpha_{m_i}(\mathbf{x}_{m_i},\boldsymbol{\theta})]\right\}$$

• Note that this is only a *partial* log likelihood. It is not necessarily based on

$$D(y_1,\ldots,y_M|\mathbf{x}_1,\ldots,\mathbf{x}_M,c_1,\ldots,c_M)$$

because, in particular, we are not assuming here that the $\mathbf{x}_m$ are strictly exogenous.

- This is very similar to pooled binary response analysis that we discussed before: we need only have the sequence of distributions conditional on any set of covariates be correctly specified for partial (pooled) MLE to work. The main difference here is the need to distinguish between censored and uncensored observations.
- If the covariates are strictly exogenous in the sense that

$$D(T|T \geq a_{m-1}, \mathbf{x}_1, \ldots, \mathbf{x}_M) = D(T|T \geq a_{m-1}, \mathbf{x}_m)$$

and the censoring is strictly exogenous, then the partial log likelihood is a full log likelihood (conditional on $\{\mathbf{x}_1, \ldots, \mathbf{x}_M, c_1, \ldots, c_M\}$).

- The proportional hazard specification is still attractive:

$$\lambda(t; \mathbf{x}_m, \boldsymbol{\theta}) = \exp(\mathbf{x}_m \boldsymbol{\beta})\lambda_m, \; a_{m-1} \leq t < a_m.$$

- Meyer (1990, *Econometrica*) popularized this approach in economics, applying it to estimate the effects of unemployment insurance on unemployment spells.

Unobserved Heterogeneity with Time-Varying Covariates

• The previous analysis is attractive because it does not require strictly exogenous covariates. But it does not allow unobserved heterogeneity.

• Now the strict exogeneity assumption is conditional on the unobserved heterogeneity, v:

$$D(T|T \geq a_{m-1}, \mathbf{x}_1, \ldots, \mathbf{x}_M, v) = D(T|T \geq a_{m-1}, \mathbf{x}_m, v)$$

and we assume the heterogeneity is independent of the covariates and censoring:

$$D(v|\mathbf{x}_1, \ldots, \mathbf{x}_M, c_1, \ldots, c_M) = D(v).$$

- The hazard function is

$$\lambda(t;\mathbf{x}_m,\boldsymbol{\theta}) = v\exp(\mathbf{x}_m\boldsymbol{\beta})\lambda_m,\ a_{m-1} \leq t < a_m.$$

in which case the density of $(y_{i1},\ldots,y_{iM})$ given $(v_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iM},c_{i1},\ldots,c_{iM})$ is

$$\left[\prod_{h=1}^{m_i-1}\alpha_h(v_i,\mathbf{x}_{m_i},\boldsymbol{\theta})\right][1-\alpha_{m_i}(v_i,\mathbf{x}_{m_i},\boldsymbol{\theta})]^{d_i}$$

which is the same as before except for the presence of v_i.

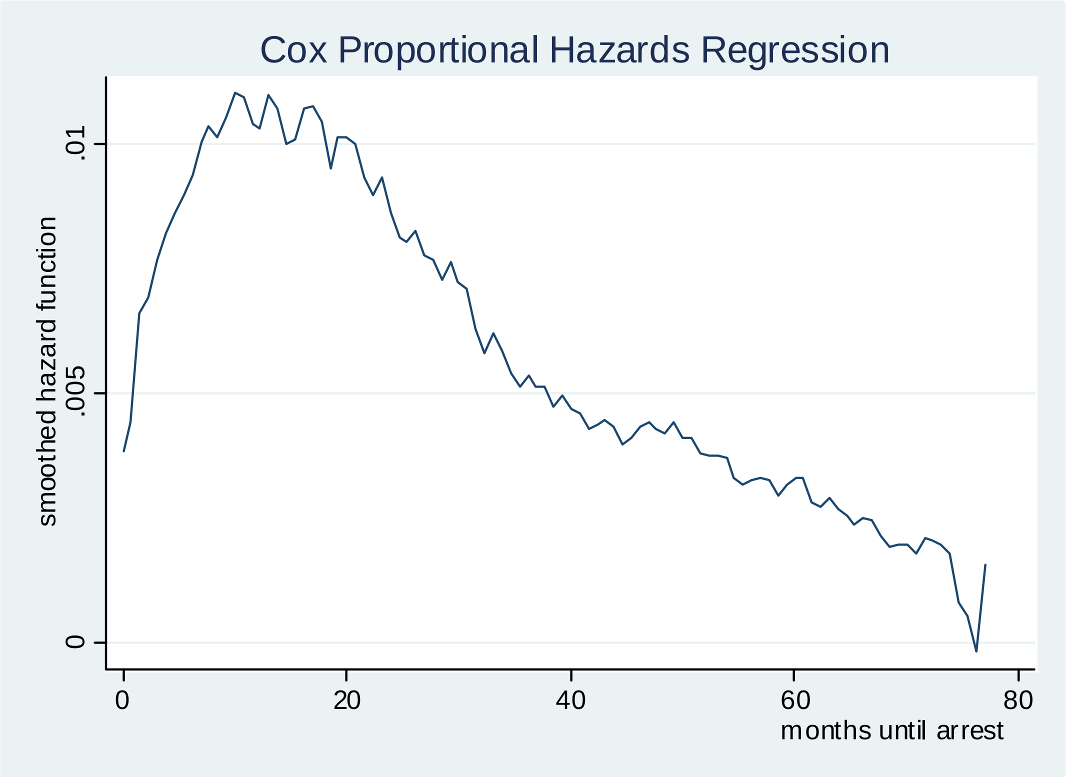
- If v_i had the *Gamma*(δ,δ) distribution, the log likelihood has a closed form; see Meyer (1990, *Econometrica*). McCall (1994, *Journal of Applied Econometrics*) provides an extension.

- Cox’s proportional hazard model is very similar to allowing an unrestricted hazard function with grouped duration data. Here we assume no heterogeneity.
- Cox’s original motivation was to estimate the parameters $\boldsymbol{\beta}$ without specifying a baseline hazard. But it is effectively the same as using a piecewise constant baseline hazard in a proportional hazard analysis.

```
. stcox workprg priors tserved felon alcohol drugs black married educ age,
         nohr basehc(h)

         failure _d:  nocens
   analysis time _t:  durat

Cox regression -- Breslow method for ties

No. of subjects =         1445                     Number of obs   =      1445
No. of failures =          552
Time at risk    =        80013
                                                   LR chi2(10)     =    155.60
Log likelihood  =   -3816.3799                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     workprg |   .0840734    .090814     0.93   0.355    -.0939189    .2620656
      priors |   .0875916   .0134774     6.50   0.000     .0611764    .1140068
     tserved |   .0129505   .0016855     7.68   0.000      .009647     .016254
       felon |  -.2827356   .1061569    -2.66   0.008    -.4907992   -.0746719
     alcohol |   .4306549   .1057194     4.07   0.000     .2234487    .6378611
       drugs |   .2756129   .0978637     2.82   0.005     .0838036    .4674222
       black |    .432593   .0883817     4.89   0.000     .2593679     .605818
     married |  -.1548816   .1092132    -1.42   0.156    -.3689356    .0591724
        educ |  -.0213195    .019447    -1.10   0.273    -.0594348    .0167959
         age |  -.0035816   .0005223    -6.86   0.000    -.0046053    -.002558
------------------------------------------------------------------------------
```

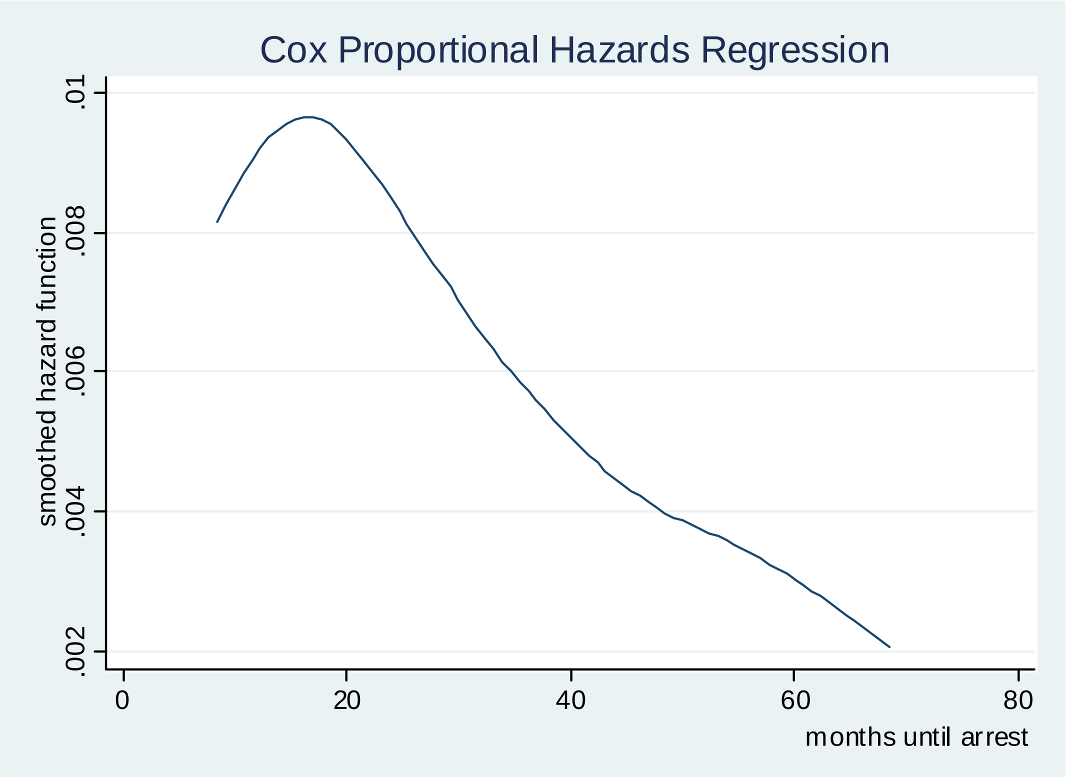
- A rectangular kernel (not much smoothing):

```
. stcurve, hazard kernel(rec)
```



- A smoother version (the default in Stata):

```
. stcurve, hazard
```



```
. * Note that the smoothed version is only graphed for a range of about 10 months
. * to 70 months. That’s because the smoother does not work well near the
. * endpoints where there is little or no data on the other side to average.
```