University of Illinois
Fall 1998

Department of Economics
Roger Koenker

Economics 472

Lecture 18

Duration Models

There is considerable interest, especially among labor-economists in models of duration. These models originated in biomedical applications, insurance, and quality control but are now being applied broadly to unemployment, retirement and an array of other issues.

Survival Functions and Hazard Rates

Often duration models are described in terms of survival models of the sort that might be appropriate for biomedical clinical trials in which we are interested in evaluating the effectiveness of a medical treatment and the response variable is the length of time that the patient lives following the treatment. But there are a wide variety of other applications. I like to think of this in terms of predicting time of birth, ex ante we have some positive random variable, T, with density $f(t)$, and distribution function $F(t)$. One can then consider the conditional density of the birth date given that a birth hasn't occurred up to time t. This is rather like the computations we considered in the previous lecture. There is a considerable amount of specialized terminology which we will need to introduce. The *survival* function is simply

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

and the hazard function is

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

note that

$$\lambda(t)dt \quad = \quad P[t < T < t + dt | T > t] = P(\text{ born in hour } t + 1 | \text{ not born by hour } t]$$

clearly,

$$\int_0^t \lambda(u)du = -\log(1 - F(u))|_0^t = -\log(1 - F(t)) = -\log S(t)$$

so*

$$S(t) = \exp\{-\int_0^t \lambda(u)du\}.$$

Digression on the Mills' Ratio & Hazard Rates.

Suppose X is a positive r.v. representing life time of an individual, with density f, and df F, obviously, $P(X > x) = 1 - F(x)$ Given that survival until x what is probability of death before $x + t$

$$P(X > x + t | X > x) = \frac{P(x < X < x + t)}{P(X > x)} = \frac{F(x + t) - F(x)}{1 - F(x)}$$

*Such so-called product integrals have a rich theory which has been recently developed in Gill & Johanson *Annals*, 1990, but we will not concern ourselves with this here.

to get a death *rate* (deaths per unit time) between x and $x+t$ compute

$$\lim_{t\to 0} \frac{t^{-1}(F(x+t)-F(x))}{1-F(x)} \qquad = \frac{f(x)}{1-F(x)}$$

which is called the hazard rate. The reciprocal of the hazard rate is sometimes called the Mills ratio.

A common problem in data of this sort is that we observe T for only some observations, while for others we observe only that T is greater that some censoring time t_c, e.g., in a clinical trial, individuals may be still alive at the end of the experimental period. So we see

$$Y_i = \begin{cases} T_i & \text{if } T_i < t_c \\ t_c & \text{if } T_i \geq t_c \end{cases}$$

Maximum Likelihood Estimation of Parametric Models.

The likelihood for a fully parametric model is given by,

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(y_i,\theta)^{\delta_i} S(y_i,\theta)^{1-\delta_i}$$

where δ_i denotes the censoring indicator,

$$\delta_i = \begin{cases} 1 & T_i < t_c \\ 0 & T_i \geq t_c \end{cases}$$

so this is somewhat like the tobit model of the last lecture. Of course we now need to specify the parametric model for f and S.

Menu of Choices for the Parametric Specification

1. Exponential – this is simplest

$$\begin{aligned} \lambda(t) &\equiv \lambda > 0 \Rightarrow S(t) = e^{-\lambda t} \\ f(t) &= \lambda e^{-\lambda t} \\ E(T) &= \lambda^{-1} \\ V(T) &= \lambda^{-2} \\ \text{median} &= -\log(1/2)/\lambda \cong .69/\lambda \end{aligned}$$

2. Gamma – generalization of exponential

$$\begin{aligned} f(t) &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda_t} \qquad \alpha > 0, \lambda > 0 \\ E(T) &= \alpha/\lambda \\ V(T) &= \alpha/\lambda^2 \\ S(T) & \quad \text{is messy (involves incomplete gamma)} \end{aligned}$$

3. Weibull – another generalization of exponential model

$$\begin{aligned}
\lambda(t) &= \alpha\lambda(\lambda t)^{\alpha-1} \qquad \alpha > 0, \lambda > 0 \\
S(t) &= e^{-(\lambda t)^{\alpha}} \\
f(t) &= \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}
\end{aligned}$$

   Note that depending upon whether $\alpha \lesseqgtr 1$ you get either increasing or decreasing hazard. This model is probably the most common parametric one.

4. Rayleigh $\lambda(t) = \lambda_0 + \lambda_1 t$

5. Uniform $U[0,1]$ $\lambda(t) = \frac{1}{1-t}$

Clearly there is some *à priori* ambiguity as to which probability model should be used. This leads naturally to the next topic.

Nonparametric Methods – The Kaplan-Meier Estimator

Suppose you have a reasonably homogeneous sample like our WECO employees and we want to estimate a "survival" distribution for them – how long they stay on-the-job. We can chop the time axis into arbitrary intervals and write,

$$\begin{aligned}
S(\tau_k) &= P[T > \tau_k] \\
&= P[T > \tau_1]P[T > \tau_2 | T > \tau_1] \ldots P[T > \tau_k | T > \tau_{k-1}] \\
&= p_1 \cdot p_2 \cdots p_k
\end{aligned}$$

as an estimate of p_i we could use

$$\hat{p}_i = \left(1 - \frac{d_i}{n_i}\right) = \left(1 - \frac{\#\text{ quit in period } i}{\#\text{ left in period } i}\right)$$

Then the survival function can be estimated as,

$$\hat{S}(\tau_k) = \prod_{j=1}^{k} \hat{p}_j$$

The Kaplan Meier estimator of $S(t)$ is like the previous method except that we replace the fixed intervals with random intervals determined by the observations themselves. As above, we observe pairs: $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ where Y_i is observed duration for i^{th} subject and

$$\delta_i = \begin{cases} 1 & \text{uncensored} \\ 0 & \text{censored} \end{cases}$$

Let $(Y_{(i)}, \delta_{(i)})$ denote the ordered observation (ordered on Y's!). Then set as above

$$\begin{aligned}
n_i &= \#\text{ alive at time } Y_{(i)} - \varepsilon \\
d_i &= \#\text{ died at time } Y_{(i)} \\
p_i &= P[\ \textit{surviving} \text{ through period } I_i |\text{ alive at beginning of } I_i] \\
q_i &= 1 - p_i
\end{aligned}$$

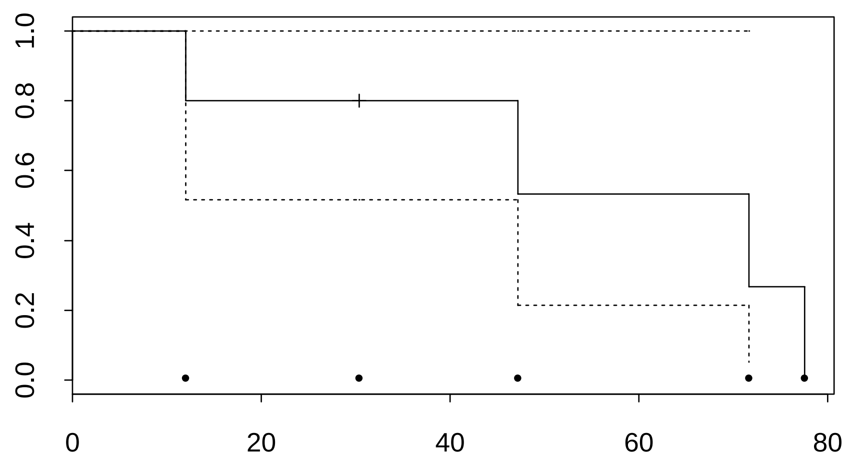

Figure 1: A Simple Kaplan Meier Plot for 5 Observations: The figure illustrates a very simple version of the Kaplan Meier estimator of the survival function for 5 observations, one of which is censored and the others of which are uncensored. The 5 observed times are represented on the horizontal axis as plotted points with vertical coordinate zero. A useful exercise is to compute the vertical ordinates of $\hat{S}(t)$ given in the figure. Note that there is no drop in the estimated function at y_2 since this observation is censored. The dotted lines denote a confidence band for $S(t)$ which, since there are so few observations is essentially uninformative.

then $\hat{q}_i = \delta_i / n_i$, so

$$\hat{p}_i \;=\; 1 - \hat{q}_i = \begin{cases} 1 - \frac{1}{n_i} & \text{if } \delta_{(i)} = 1 \\ 1 & \text{if } \delta_{(i)} = 0 \end{cases}$$

Then (drum roll!) the *product-limit* Kaplan-Meier estimate is,

$$\hat{S}(t) \;=\; \prod_{y_{(i)} \le t} \hat{p}_i = \prod_{y_{(i)} \le t} \left(1 - \frac{1}{n_i}\right)^{\delta_i} = \prod_{y_{(i)} \le t} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_i} = \prod_{y_{(i)} \le t} \left(\frac{n - i}{n - i - 1}\right)^{\delta_i}$$

This estimator satisfies several nice requirements

(i) It is consistent

(ii) It is asymptotically normal (involves weak convergence to Brownian motion argument.

(iii) It is a generalized MLE à la Kiefer-Wolfowitz.

(iv) Without censoring it is the empirical *df*, i.e. $\hat{S}(t) = 1 - \hat{F}(t)$ where $\hat{F}(t) = n^{-1} \sum I(T_i < t)$.

This is particularly good for one-and two-sample problems.

The estimates $\hat{p}_i$'s are the *conditional probabilities*, while one needs to compute the associated conditional survival probabilities to find the Survival Function Estimate, and the product accomplishes this.

In *Splus* this is done using the function `survfit`. To use it e.g., try `plot(survfit(Surv(`y, δ`)~ strata),data=some.dataframe)`. The difficulty of this approach in most econometric applications is that we can't usually rely on a simple categorization of the sample observations into a small number of groups, we have covariates which we would like to use in a way which is close to the usual linear regression model fashion. This leads to an attempt to make some compromise between the nonparametric and parametric approaches.

Semi-Parametric Models – Cox's Proportional Hazard Model

This is a common econometric approach. Let $\{T_i\}$ and $\{C_i\}$ be independent r.v's. C_i is the censoring time associated with survival times T_i. We observe $\{(Y_i, \delta_i)\}$ where

$$\begin{aligned} Y_i &= \min\{T_i, C_i\} \\ \delta_i &= I(T_i \leq C_i) \end{aligned}$$

we also observe a vector of covariates x_i for each "individual." Of course "individuals" might be firms which we are modeling bankruptcy decisions for, or some other unit of economic analysis. Recall,

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

The crucial assumption of the Cox model is,

$$\lambda(t|x) = e^{x\beta}\lambda_0(t)$$

Note that the form $h(x) = e^{x\beta}$ is far less essential than multiplicative separability of the function in x and t. We now introduce a rather high-brow definition which is useful in interpretating the essential role of the Cox assumption.

Definition: A family of df's $\mathcal{F}$ constitute a family of Lehmann alternatives if there exists $F_0 \in \mathcal{F}$ such that for any $F \in \mathcal{F}$, $1 - F(t) = (1 - F_0(t))^\gamma$ for some $\gamma > 0$ and all t. I.e. $S(t) = S_0^\gamma(t)$.

Clearly the proportion hazard model implies a family of L alternatives since,

$$\begin{aligned} S(t;x) &= \exp\{-\int_0^t \lambda(u;x)du\} \\ &= \exp\{-e^{x\beta}\int_0^t \lambda_0(u)du\} \qquad \text{Recall(!)} \quad e^{ax} = (e^x)^a \\ &= S_0(t)^\gamma \qquad \text{where} \quad \gamma = e^{x\beta}. \end{aligned}$$

Special case: if we have the two sample problem, then $x\beta =$ either 0 or 1 so $S_1(t) = S_0^\gamma(t)$ for some constant γ.

Estimation (Sketchy)

Let $\Re_{(i)}$ denote the set of individuals at risk at time $y_{(i)} - \varepsilon$, for each uncensored time, $y_{(i)}$

$$P[\underline{a} \text{ death in } [y_{(i)}, y_{(i)} + \Delta y)|\Re_{(i)}] \cong \sum_{j \in R_{(i)}} \underbrace{e^{x_j\beta}\lambda_0(y_{(i)})}_{\text{hazard}} \Delta y$$

so

$$P\{ \text{ death of } (i) \text{ at time } y_{(i)} | \text{ a death at time } y_{(i)}\} = \frac{e^{x_{(i)}\beta}}{\sum_{j\in\Re_i} e^{x_{(i)}\beta}}.$$

Note that the λ_0 effect cancels in numerator and denominator. and this gives the partial likelihood

$$\mathcal{L}(\beta) = \prod_i \left[\frac{e^{x_{(i)}\beta}}{\sum_{j\in\Re_{(i)}} e^{x_{(j)}\beta}}\right]$$

Cox's proposal was to estimate β by maximizing this partial likelihood. In what sense is this likelihood partial? This is really a question for Economics 476, but here I will just say that we have ignored the λ_0 contribution to the full likelihood by the trick used above, and we have to hope that this isn't really very informative about the parameter β. This turns out to be more or less true, of course we still might worry about the loss of efficiency entailed, and also about the plausibility of the Cox model assumption which we would like to test. This too will be left for 476. A partial explanation of why the partial likelihood doesn't sacrifice much information is the following. It *conditions* on the set of instants at which "failures" occur, since $\lambda_0(t)$ is assumed arbitrary no information *about* β is contained in those instants. Why? This mystery is revealed in recent martingale reformulations of the Cox Model.

Estimating the Baseline Hazard

It remains to discuss how to estimate λ_0 from the Cox model,

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du$$

in the Cox Model. Breslow assumes à la Cox that $\lambda_0(t)$ is constant between uncensored observations,

$$\hat{\lambda}_0(t) = \frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{j\in\Re_{u(i)}} e^{\hat{\beta}x_j}}$$

for $t \in (y_{u(i-1)}, y_{u(i)})$ and $u(i)$ index of ith censored observations. Then,

$$\hat{S}_0(t) = \prod_{\{i:y_{(i)}<t\}} \left(1 - \frac{\delta_{(i)}}{\sum_{j\in\Re_{(i)}} e^{\beta x_j}}\right)$$

Note here

$$\hat{S}_0(t) \neq e^{-\Lambda_0(t)}$$

But it has the virtue that we get Kaplan Meier when $\beta = 0$! Since $\sum e^{\beta x_j}$ in this case is just the number of observations in the risk set.

Tsiatis uses instead,

$$\hat{S}_0(t) = e^{-\Lambda_0(t)}$$

$$\Lambda_0(t) = \sum \frac{\delta_{(i)}}{\sum_{j\in\Re_{(i)}} e^{\hat{\beta}x_j}}$$

This doesn't simplify like the Breslow estimator. The relationship between Tsiatis and Breslow estimates is seen simply by noting that $-\log(1-x) \approx x$ for small x.