University of Illinois
Fall 2012

Department of Economics
Roger Koenker

Economics 508
Lecture 22
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, finance 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 $\lambda(t)dt = P[t < T < t + dt | T > t] = P($ born in hour $t + 1|$ 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*

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

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

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 $x + t$

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

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

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

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

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

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

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

This estimator satisfies several nice requirements

(i) It is consistent

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

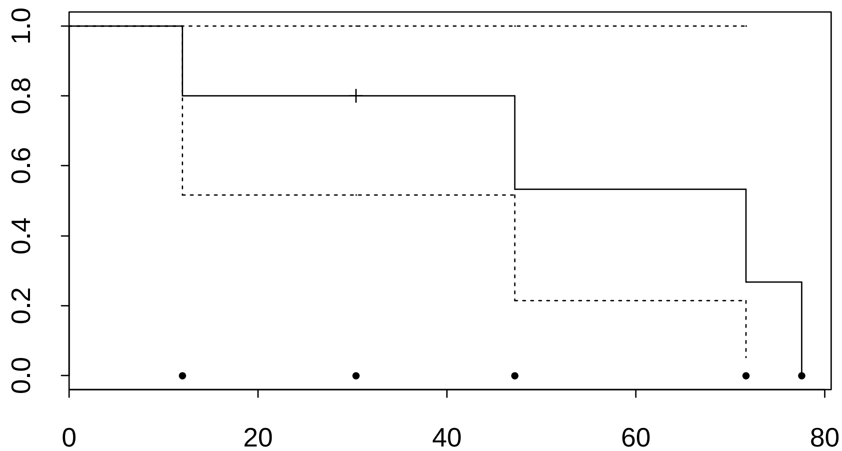

Figure 1: A Simple Kaplan Meier Plot for 5 Observations: The figure illustrates a very simple version of the Kaplan Meier estimator of the survival function for 5 observations, one of which is censored and the others of which are uncensored. The 5 observed times are represented on the horizontal axis 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.

(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.

The Kaplan Meier estimator is particularly good in situations in which we have a small number of groups and we would like to ask: do they have similar survival distributions. An example of this sort of question is addressed in the next figure. Using data from Meyer (1990) we consider the survival distributions estimated by the Kaplan-Meier technique for individuals who have more than \$100 per week in unemployment benefits versus those with less than \$100 per week in benefits.

As the figure indicates, those with higher benefits appear to stay unemployed longer. The median unemployment spell for the high benefits group is roughly 2 weeks longer than for the low benefits group. Note, however, that the difference is unclear in the right tail of the distribution; the higher benefit group appears to have a somewhat lower probability of a spell greater than 35 weeks. This plot was produced ty the splus command,

```
plot (surv.fit(dur,cens,strata=exp(ben) > 100, type='kaplan-meier'))
```

where `dur` is the observed durations, `cens` is the censoring indicator, and `ben` is the log of weekly benefits.

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.

Accuracy of the Kaplan Meier Estimator

Large Sample theory for the KM estimator is complicated but the following heuristic approach is possibly comprehensible. Let $\hat{\lambda}_i = d_i/n_i$ denote the proportion dying in bin i, for some fixed set of bins, then,

$$\log(\hat{S}(t)) = \sum_{i:t_i<t} \log(1 - \hat{\lambda}_i)$$

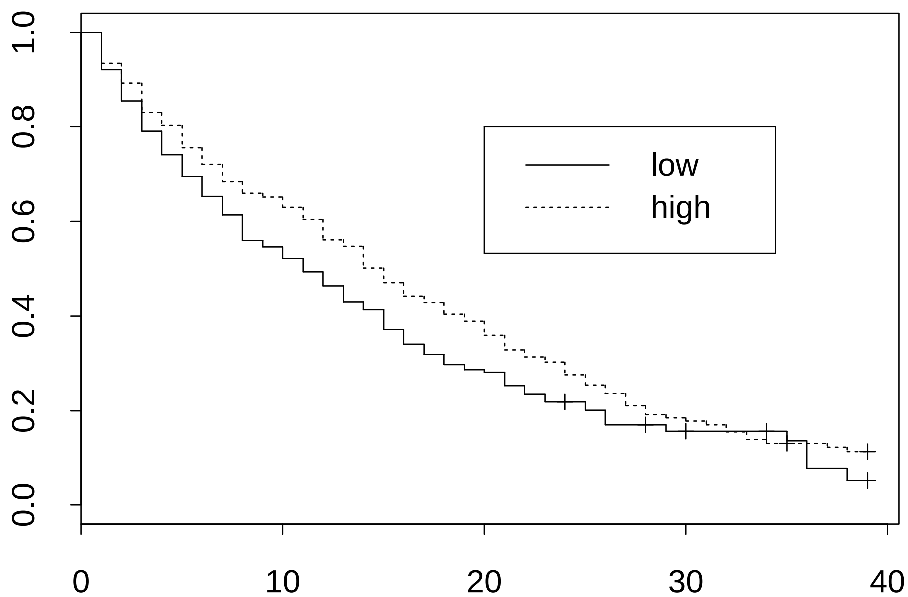

Figure 2: Two Kaplan Meier Survival Curves for the length of unemployment spells: The figure plots Kaplan-Meier estimates of the duration of unemployment function for two groups of individuals. One is a high UI-benefits group (those with weekly benefits more than 100 dollars, the other with weekly UI benefits less than 100. The data is taken from Meyer(1990).

so by δ-method

$$\begin{aligned} V(\log \hat{S}(t)) &= \sum (1-\hat{\lambda})_i^{-2} V(1-\hat{\lambda}_i) \\ &= \sum \left(1 - \frac{d_i}{n_i}\right)^{-2} V(\hat{\lambda}_i) \end{aligned}$$

but

$$\begin{aligned} V(\hat{\lambda}_i) &= \lambda(1-\lambda)/n_i \\ \hat{V}(\hat{\lambda}_i) &= \frac{d_i}{n_i}\left(1 - \frac{d_i}{n_i}\right)/n_i \end{aligned}$$

so

$$\hat{V}(\log \hat{S}(t)) = \sum \frac{d}{n_i(n_i - d_i)}$$

and therefore, again by δ-method,

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i:t_i<t} \frac{d_i}{n_i(n_i - d_i)}$$

Why? $[\hat{S}(t) = \exp(\log \hat{S}(t))]$ This variance formula is called Greenwood's formula and is what is used to construct the confidence bands for the KM estimate in `R` and other software.

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. This is obviously quite restrictive. In particular it prohibits covariate effects that are favorable for a while and then unfavorable, or vice-versa.

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,

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

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,

$$\begin{aligned} \hat{S}_0(t) &= e^{-\Lambda_0(t)} \\ \Lambda_0(t) &= \sum \frac{\delta_{(i)}}{\sum_{j \in \Re_{(i)}} e^{\hat{\beta} x_j}} \end{aligned}$$

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.

Duration Models and Binary Response

This section is based mainly on Doksum and Gasko (1990, Intl Stat Review). We can think of the usual binary response model as a survival model in which we fix the time of survival and ask, what is the probability of surviving up to time t. For example, in the problem set we can ask what is the probability of not quitting up to time 6 months. By then varying t we get a nice 1-1 correspondence between the two classes of models. We can specify the general failure-time distribution,

$$F(t|x) = P(T < t|x)$$

and fixed t so we are simply modeling a survival probability, say $S(t|x) = 1-F(t|x)$ which depends on covariates. We will consider two leading examples to illustrate this, the logit model, and the Cox proportional hazard model.

Logit

In the logit model we have,

$$\text{logit}(S(t|x)) = \log(S(t|x)/(1 - S(t|x)) = x'\beta$$

where $F(z) = (1 + e^{-z})^{-1}$ the df of the logistic distribution. In survival analysis this would correspond to the model

$$\text{logit}(S(t|x)) = x'\beta + \log\Gamma(t)$$

where $\Gamma(t)$ is a baseline odds function which satisfies the restriction that $\Gamma(0) = 0$, and $\Gamma(\infty) = \infty$. For fixed t we can simply absorb $\Gamma(t)$ into the intercept of $x'\beta$. This is the proportional-odds model. Let

$$\Gamma(t|x) = S(t|x)/(1 - S(t|x)) = \Gamma(t)\exp\{x'\beta\}$$

and by analogy with other logit type models we can characterize the model as possessing the property that the ratio of the odds-on-survival at any time t don't depend upon t, i.e.

$$\Gamma(t|x_1)/\Gamma(t|x_2) = \exp(x_1'\beta)/\exp(x_2'\beta).$$

Now choosing some explicit functional form for $\Gamma(t)$ for example $\log\Gamma(t) = \gamma\log(t)$, ie. $\Gamma(t) = t^\gamma$, gives the survival model introduced by Bennett (1983).

It may help to interpret the Bennett model and those that follow to recall that if we transform a random variable T by its distribution function, so $U = F_T(T)$ we know that $U \sim U[0,1]$. so

$$\text{logit}(S(T|X)) = x^\top\beta + \log\Gamma(T)$$

so we see that some transformation of the random survival times, T, can be written as an iid error model with linear predictor $x^\top\beta$ and error term $\log\Gamma(T)$.

Proportional Hazard Model

One can, of course, model not S, as above, but some other aspect of S which contains equivalent information, like the hazard function,

$$\lambda(t|x) = f(t|x)/(1 - F(t|x))$$

or the cumulative hazard,

$$\Lambda(t|x) = -\log(1 - F(t|x)).$$

In the Cox model we take

$$\lambda(t|x) = \lambda(t)e^{x'\beta},$$

so

$$\Lambda(t|x) = \Lambda(t)e^{x'\beta},$$

which is equivalent to

$$\log(-\log(1 - F(t|x)) = x'\beta + \log\Lambda(t).$$

This looks rather similar to the the logit form,

$$\text{logit}(F(t|x)) = x'\beta + \log\Gamma(t).$$

but it is obviously different. This form of the proportional hazard model could also be written as,

$$F(t|x) = \Psi(x'\beta + \log\Lambda(t)).$$

where $\Psi(z) = 1 - e^{-e^z}$ is the Type I extreme value distribution. For fixed t we can again absorb the $\log \Lambda(t)$ term into the intercept of the $x'\beta$ contribution and we have the formulation,

$$\log(-\log(1 - \theta(x))) = x'\beta$$

this is sometimes called the complementary $\log - \log$ model in the binary response literature. So this would provide a binary response model which would be consistent with the Cox proportional hazard specification of the survival version of the model. In general, this strategy provides a useful way to go back and forth between binary response and full-blown survival models, but I will leave a full discussion of this to 478.

Accelerated Failure Time Model

A third alternative, which also plays an important role in the analysis of failure time data is the accelerated failure time (AFT) model, where we have

$$\log(T) = x'\beta + u$$

with the distribution of u unspecified, but typically assumed to be iid. A special case of this model is the Cox model with Weibull baseline hazard, but in general we have

$$P(T > t) = P(e^u > te^{-x'\beta}) = 1 - F(te^{-x'\beta})$$

where F denotes the df of e^u and therefore in this model,

$$\lambda(t|x) = \lambda_0(te^{-x'\beta})e^{x'\beta}$$

where λ_0 denotes the hazard function corresponding to F. In effect the covariates are seen to simply rescale time in this model. An interesting extension of this model is to write,

$$Q_{h(T)}(\tau|x) = x'\beta(\tau)$$

and consider a family of quantile regression models. This allows the covariates to act rather flexibly with respect to the shape of the survival distribution. Portnoy (2004) and Peng and Huang (2008) have more recently extended these methods to censored survival data, greatly expanding the utility of these methods.

References

Kalbfleisch, J.D. and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, Wiley.

Koenker, R. and O. Geling, (2001) Reappraising Medfly Longevity, *JASA* 96, 458-468.

Koenker, R. and Y. Bilias (2001) Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments, *Empirical Economics*, 26, 199-220.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*, Cambridge U. Press.

Meyer, B. (1990). Unemployment insurance and unemployment spells, *Econometrica*, 57, 757-782.

Miller, R.G. (1981). *Survival Analysis*, Wiley.

Peng L, Huang Y (2008). Survival Analysis with Quantile Regression Models. *Journal of American Statistical Association*, 103, 637–649.

Portnoy S (2003). Censored Quantile Regression. *Journal of American Statistical Association*, **98**, 1001–1012.

Cox, D.R. (1972). Regression models and life tables, *JRSS(B)*, 34, 187-200.