University of Illinois
Fall 1998

Department of Economics
Roger Koenker

Economics 472
Lecture 19
Duration Models and Binary Response

This lecture 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

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

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

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

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.

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

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

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.