Duration Models

Class Notes

Manuel Arellano

November 1989

Revised: December 3, 2008

1 Duration data

Duration data appear in a diversity of situations in economics, but I will often refer to unemployment duration to provide a focus for the presentation of the material.

Suppose we select a random sample of N persons entering unemployment and wait until they find jobs. Then we record the number of weeks each person has been unemployed: $t_1, t_2, ..., t_N$. We have observations on the duration of a spell of unemployment for each of the individuals in the sample.

In practice we are more likely to have censored duration data. That is, for some individuals we do not observe t_i, possibly because they have not found a job at the time of the interview, so that all we know is that $t_i > \overline{t}$ for some particular $\overline{t}$, or they have found a job between selection and interview and all we know is that t_i lies within a certain interval $\underline{t}< t_i < \overline{t}$.

There are enormous variations in the duration of spells of unemployment from one individual to the next (from a few weeks to five years or more) and it is important to model the causes for these differences. In particular it is important to know how the re-employment probability changes over the period of the spell and what is the impact of the level of unemployment benefits on these probabilities.

More generally, duration data measure how long individuals remain in a certain state. Other examples include: job turnover, marital instability, time to transactions in stock markets, or durations to realignments of international currencies.

The analysis of duration data has a long tradition in biometrics and medical statistics (as well as in industrial life testing and demography), from where most of the terminology and the statistical models used by econometricians originate. This is also the reason for the intimidating names that are used in this field: Failure time data, hazards, risks, survivor functions... A typical example in textbooks is of the form: you have a number of rats, which you inject with a cancer producing substance, then time to mortality is measured (e.g. Kalbfleisch and Prentice).

Remark on the notation: The standard notation for a cross-sectional explained variable in econometrics is Y_i. However, in duration analysis we often use T_i, which emphasizes the fact that the variable measures time in a state. We may also have a time series of durations (e.g. time to transaction or time to a price change), which naturally connects with the perspective of point processes.

The purpose of this note is to cover basic concepts in duration analysis. Multiple spell data, treatment effects in duration models, and point processes are discussed in separate notes.

2 The hazard function

2.1 Hazard function for a discrete random variable

Let T be a discrete duration random variable taking on values $\{1, 2, 3, ...\}$ with *pmf*:

$$p(t) = \Pr(T = t) \qquad (t = 1, 2, ...)$$

and *cdf*:

$$F(t) = \Pr(T \leq t) = p(1) + p(2) + ... + p(t).$$

The hazard function or exit rate from the state is

$$\begin{aligned} h(t) &= \Pr(T = t \mid T \geq t) = \frac{\Pr(T = t)}{\Pr(T \geq t)} = \frac{\Pr(T = t)}{1 - \Pr(T \leq t - 1)} \\ &= \frac{p(t)}{1 - p(1) - ... - p(t-1)} = \frac{F(t) - F(t-1)}{1 - F(t-1)} \quad \text{for } t > 1 \end{aligned}$$

and $h(1) = p(1) = F(1)$.

That is, the hazard gives probabilities of exit defined over the surviving population at each time.

Example: Suppose T is time to mortality (in years). Then

$$\begin{aligned} p_{100} &= \Pr(T = 100) \simeq 0 \\ h_{100} &= \Pr(T = 100 \mid T \geq 100) \simeq 1. \end{aligned}$$

In this case the hazard may look as depicted in Figure 1.

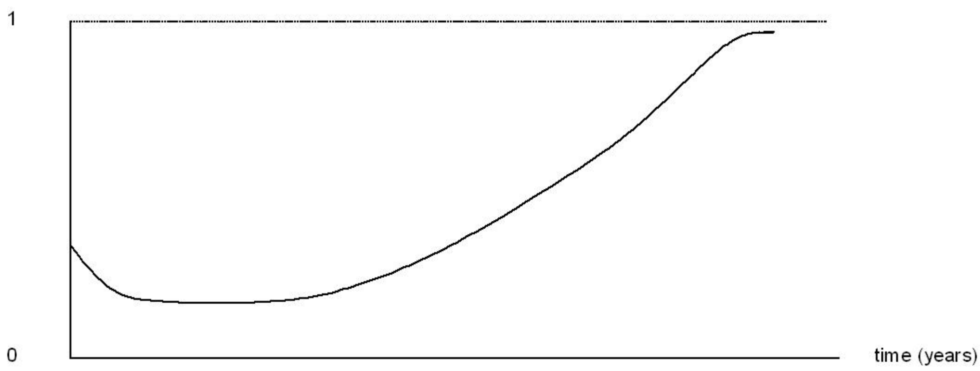

Figure 1: Mortality hazard rate

The hazard $h(t)$ provides an alternative way of characterizing the distribution of T. Let us see how we can recover $F(t)$ and $p(t)$ from $h(t)$. For shortness we use the notation $p_t = p(t)$, $F_t = F(t)$, and $h_t = h(t)$. For $t > 1$

$$\Pr(T \geq t + 1 \mid T \geq t) = 1 - h_t = \frac{1 - F_t}{1 - F_{t-1}}.$$

Using this expression recursively we get

$$\begin{aligned}
1 - F_1 &= (1 - h_1) \\
1 - F_2 &= (1 - h_1)(1 - h_2) \\
&\vdots \\
1 - F_t &= \prod_{s=1}^{t} (1 - h_s) \qquad (1)
\end{aligned}$$

Therefore,

$$F_t = 1 - \prod_{s=1}^{t} (1 - h_s) \qquad (t = 1, 2, ...), \qquad (2)$$

which shows how the *cdf* of T can be obtained from the hazards.

Similarly,

$$p_t = (1 - F_{t-1}) h_t = h_t \prod_{s=1}^{t-1} (1 - h_s) \qquad (3)$$

Note that (1) is an intuitive probability factorization. We have:

$$\begin{aligned}
1 - h_t &= \Pr(T \geq t + 1 \mid T \geq t) \\
1 - F_t &= \Pr(T \geq t + 1)
\end{aligned}$$

so that

$$\Pr(T \geq t + 1) = \Pr(T \geq t + 1 \mid T \geq t) \Pr(T \geq t).$$

Repeatedly using this factorization we obtain (1):[1]

$$\Pr(T \geq t + 1) = \Pr(T \geq t + 1 \mid T \geq t) \Pr(T \geq t \mid T \geq t - 1) ... \Pr(T \geq 2)$$

2.2 Hazard function for a continuous random variable

We know that a continuous random variable T can be characterized by the *pdf* $f(t)$ and by the *cdf* $F(t)$. An alternative way of characterizing the distribution of T is the hazard function defined as

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

which is particularly useful when T represents duration in a certain state.

Note that $h(t)$ gives the conditional density of T given $T > t$:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T > t)}{\Delta t}$$

[1] Note that $\Pr(T \geq 2) = 1 - h_1 = \Pr(T \geq 2 \mid T \geq 1)$ since we assume that $\Pr(T \geq 1) = 1$.

where

$$\Pr\left(t \leq T < t + \Delta t \mid T > t\right) = \frac{\Pr\left(t \leq T < t + \Delta t\right)}{\Pr\left(T > t\right)} = \frac{F\left(t + \Delta t\right) - F\left(t\right)}{1 - F\left(t\right)}.$$

If T represents unemployment duration, $h(t)\,dt$ is the probability of leaving unemployment during $(t, t + dt)$ given that the individual has been unemployed for t periods.

Note that $h(t)$ is the closest link to the predictions of a theory of job search. Suppose an individual has a sequence of reservation wages over time $\overline{w}(t)$, that he faces a distribution of wage offers $G(w)$, and that job offers arrive randomly at a rate $\psi(t)$ (i.e. the probability of a job offer in the interval $(t, t + dt)$ is $\psi(t)\,dt$). Then the probability of exit during $(t, t + dt)$ is the probability of getting a job offer times the probability of an offered wage greater than the reservation wage:

$$h(t)\,dt = \left[1 - G\left(\overline{w}(t)\right)\right]\psi(t)\,dt$$

It is therefore natural to start by making economic assumptions about $h(t)$ rather than other aspects of the distribution of durations such as $f(t)$, $F(t)$, or $E(T)$.

Recovering $F(t)$ and $f(t)$ from the hazard rate First let us introduce the cumulative hazard function (or integrated hazard), which is given by

$$H(t) = \int_{-\infty}^{t} h(u)\,du$$

(or $\int_0^t h(u)\,du$ if T is non-negative as it would normally be).

Moreover,

$$H(t) = \int_{-\infty}^{t} \frac{f(u)}{1 - F(u)} du) = \left[-\ln\left(1 - F(u)\right)\right]_{-\infty}^{t} = -\ln\left[1 - F(t)\right]$$

so that

$$F(t) = 1 - \exp\left[-H(t)\right]$$

and similarly,

$$f(t) = h(t)\exp\left[-H(t)\right].$$

The comparison between survival probabilities for discrete and continuous durations is as follows. In the discrete case we have

$$\ln\left[1 - F(t)\right] = \sum_{s=1}^{t} \ln\left[1 - h(s)\right]$$

whereas in the continuous case we have

$$\ln\left[1 - F(t)\right] = \int_0^t \left[-h(s)\right] ds.$$

Note that for small $h(s)$ we have $\ln\left[1 - h(s)\right] \simeq -h(s)$.

3 The hazard function of some frequently used distributions

3.1 Constant hazard: The exponential distribution

A simple model of the hazard function is to assume that it is constant:

$$h(t) = \lambda.$$

In the context of our example, a constant hazard means that the probability of leaving unemployment in a given time interval is the same regardless of how long the individual has been unemployed.

The constant hazard assumption defines a family of probability distributions indexed by one parameter (λ).

If the duration variable is continuous we obtain the class of exponential distributions. The integrated hazard is

$$H(t) = \int_0^t \lambda du = \lambda t$$

so that the log survivor function is a straight line through the origin: $\ln[1 - F(t)] = -\lambda t$. Therefore, the *cdf* and *pdf* are given by

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

The fact that an exponential random variable has a constant hazard is called the memoryless property of the exponential distribution. The hazard function is also the inverse of the expected value: $E(T) = 1/\lambda$.

If the duration variable is discrete we obtain the class of geometric distributions. The *cdf* and the *pmf* are given by

$$\begin{aligned} F(t) &= 1 - (1 - \lambda)^t \\ p(t) &= \lambda (1 - \lambda)^{t-1}. \end{aligned}$$

These expressions result from setting $h_t = \lambda$ in (2) and (3). The expected value is also $1/\lambda$.

3.2 The Weibull distribution

This is a two-parameter generalization of the exponential distribution, which allows for a hazard increasing or falling monotonically:

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

The hazard is given by

$$h(t) = \alpha\lambda^{\alpha}t^{\alpha-1}.$$

If $\alpha > 1$ $h(t)$ is monotone increasing; if $\alpha < 1$ $h(t)$ is monotone decreasing; if $\alpha = 1$ $h(t)$ is constant and reduces to the exponential case.

4 Conditional models and the proportional hazard specification

In econometric applications we are usually concerned with the relationship between duration time and explanatory variables, and therefore we look at the conditional distribution of durations given a set of exogenous variables x, $F(t \mid x)$, so that

$$h(t, x) = \frac{f(t \mid x)}{1 - F(t \mid x)}.$$

That is, we expect the hazard rate to differ between members of the population.

4.1 The proportional hazard model

The proportional hazard (PH) model (Cox, JRSS, 1972) specifies that

$$h(t, x) = \lambda(t) \exp(x'\beta).$$

That is, $h(t, x)$ factors into a function of t and a function of x, so that two different individuals have re-employment probabilities that are proportional for all t. The model is widely used because of its simplicity and straightforward interpretation.

$\lambda(t)$ is called the base-line hazard function. Common specifications for $\lambda(t)$ are the ones considered earlier: $\lambda(t)$ constant as in the exponential case (in fact, $\lambda(t) = 1$ if the constant is subsumed in the $x'\beta$ index), or

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

as in the Weibull case.

Linear representation of the proportional hazard model The *cdf* associated with the PH model satisfies

$$1 - F(t \mid x) = \exp\left[-\int_0^t h(u, x)\, du\right] = \exp\left[-e^{x'\beta}\int_0^t \lambda(u)\, du\right] = \exp\left[-e^{x'\beta}\Lambda(t)\right]$$

where $\Lambda(t)$ is the integrated baseline hazard

$$\Lambda(t) = \int_0^t \lambda(u)\, du.$$

Therefore, we also have

$$\ln\{-\ln[1-F(t \mid x)]\} = x'\beta + \ln\Lambda(t).$$

Now, if the random variable $T \mid x$ has *cdf* $F(t \mid x)$, $U \equiv F(T \mid x)$ is uniformly distributed independently of x. Moreover, the variable

$$V = \ln[-\ln(1-U)] \equiv \ln\{-\ln[1-F(T \mid x)]\}$$

is an extreme value variate, independent of x, with *cdf* $F(r) = 1 - \exp(-e^r)$.

Thus, the PH model can be written as a linear regression model for a transformation of the duration variable:

$$\ln\Lambda(T_i) = -x_i'\beta + V_i$$

where the error term is extreme value distributed independently of x (see Fourgeaud, Gourieroux, and Pradel, 1988).

In particular, if $\lambda(t) = 1$, $\Lambda(t) = t$ and the PH model reduces to the exponential regression

$$\ln T_i = -x_i'\beta + V_i.$$

On the other hand, if $\lambda(t) = \alpha t^{\alpha-1}$, $\Lambda(t) = t^\alpha$ and we obtain

$$\alpha \ln T_i = -x_i'\beta + V_i$$

or

$$\ln T_i = -x_i'\frac{\beta}{\alpha} + \frac{1}{\alpha}V_i.$$

5 Likelihood functions for censored and non-censored duration data

This section is based on Tony Lancaster's 1979 paper "Econometric Models for the Duration of Unemployment". Let $T_i \mid x_i$ be a duration variable of interest with density $f(t_i \mid x_i)$. Ideally, we would like to observe a random sample of completed durations from entrants. This is the first case.

Case I: Completed durations (inflow sample) We observe $\{t_1, t_2, ..., t_N\}$. The likelihood function of the sample is

$$\mathcal{L} = \prod_{i=1}^{N} f(t_i \mid x_i)$$

where

$$f(t_i \mid x_i) = h(t_i, x_i)\exp\left[-\int_0^{t_i} h(u, x_i)\,du\right].$$

For example, if $h(t_i, x_i)$ is PH with $\lambda(t) = 1$, then

$$f(t_i \mid x_i) = e^{x_i'\beta}\exp\left(-t_i e^{x_i'\beta}\right).$$

Case II: Censored durations (inflow sample) We have a random sample of entrants interviewed $\bar{t}$ periods later. We observe t_i if $t_i \leq \bar{t}$, otherwise we just observe that $t_i > \bar{t}$:

$$\mathcal{L} = \prod_{t_i \leq \bar{t}} f(t_i \mid x_i) \prod_{t_i > \bar{t}} \left[1 - F(\bar{t} \mid x_i)\right]$$

Case III: Censored stock sample We have a sample from the stock of unemployed individuals at a certain period and interview them h periods later. Let $\{d_1, d_2, ..., d_N\}$ be the number of weeks they have been unemployed at the time of selection. We observe t_i if $t_i \leq d_i + h$. Otherwise we just observe that $t_i > d_i + h$. The likelihood is:

$$\mathcal{L} = \prod_{d_i < t_i \leq d_i + h} \frac{f(t_i \mid x_i)}{1 - F(d_i \mid x_i)} \prod_{t_i > d_i + h} \frac{1 - F(d_i + h \mid x_i)}{1 - F(d_i \mid x_i)}.$$

In a stock sample individuals with a small d_i are less likely to be sampled. Thus, contributions to the likelihood are conditioned on $T_i > d_i$. For example, the likelihood contribution of a censored of observation is $\Pr(T_i > d_i + h \mid T_i > d_i)$.

Case IV: Interval stock sample Same as in Case III but now we never observe t_i. We only observe whether $d_i < t_i \leq d_i + h$ or $t_i > d_i + h$:

$$\mathcal{L} = \prod_{d_i < t_i \leq d_i + h} \frac{F(d_i + h \mid x_i) - F(d_i \mid x_i)}{1 - F(d_i \mid x_i)} \prod_{t_i > d_i + h} \frac{1 - F(d_i + h \mid x_i)}{1 - F(d_i \mid x_i)}.$$

In this case, the likelihood contribution of the individuals who found a job between selection and interview is $\Pr(d_i < T_i \leq d_i + h \mid T_i > d_i)$.

Clearly, the previous list of cases is not exhaustive. It is just intended to illustrate how the construction of the likelihood makes the connection between the sample design and the underlying model of interest.

6 Unobserved heterogeneity

6.1 Introduction: Unobserved heterogeneity in a simple case

Consider a discrete duration variable T and a $(0,1)$ covariate X. The conditional hazard rates are assumed constant for all t:

$$\begin{aligned} h_1(t) &= \Pr(T = t \mid T \geq t, X = 1) = h_1 \\ h_0(t) &= \Pr(T = t \mid T \geq t, X = 0) = h_0. \end{aligned}$$

Suppose that $h_1 > h_0$ and let $p = \Pr(X = 1)$. Now consider the marginal (or aggregate) hazard:

$$h(t) = h_1 \Pr(X = 1 \mid T \geq t) + h_0 \Pr(X = 0 \mid T \geq t)$$

For $t = 1$,

$$h(1) = h_1 p + h_0 (1 - p),$$

but for $t = 2$ the probability of $X = 1$ in the surviving population with $T \geq 2$ is

$$\begin{aligned}
\Pr(X = 1 \mid T \geq 2) &= \frac{\Pr(T \geq 2 \mid X = 1)\, p}{\Pr(T \geq 2)} = \frac{\Pr(T \geq 2 \mid X = 1)\, p}{\Pr(T \geq 2 \mid X = 1)\, p + \Pr(T \geq 2 \mid X = 0)(1 - p)} \\
&= \frac{(1 - h_1)\, p}{(1 - h_1)\, p + (1 - h_0)(1 - p)} = \frac{p}{1 + \left(\frac{h_1 - h_0}{1 - h_1}\right)(1 - p)} < p.
\end{aligned}$$

Therefore, we observe "spurious state dependence":

$$h(2) < h(1).$$

6.2 The PH model with unobserved heterogeneity

Lancaster (1979) argued that in the specification of the hazard there may be omitted explanatory variables, e.g. because some determinants of the hazard cannot be observed. He addressed the problem by introducing a multiplicative random effect in the PH specification:

$$h(t, x, v) = \lambda(t) \exp(x'\beta)\, v$$

where v is assumed independent of x with $E(v) = 1$ and *pdf* $g(v)$. $h(t, x, v)$ is now the hazard function conditional on x and v.

The *cdf* of unemployment duration given x and v is

$$F(t \mid x, v) = 1 - \exp\left[-\int_0^t h(u, x, v)\, du\right],$$

while the *cdf* given x only is

$$F(t \mid x) = \int_0^\infty F(t \mid x, v)\, g(v)\, dv.$$

Lancaster assumed a Gamma distribution for $g(v)$.

6.3 Identification of $\lambda(t)$ and $g(v)$ in the PH model

If $\lambda(t) \equiv 1$ there is no time dependence whereas if $v \equiv 1$ there is no unobserved heterogeneity.

In principle, one could think that the two effects are interchangeable and the only reason why we can separate them is because we impose particular functional forms for $\lambda(t)$ and $g(v)$.

The intuitive argument is as follows: Suppose $\lambda(t) \equiv 1$ and $g(v)$ is non-degenerate, so that each individual has a constant probability of leaving unemployment, but these probabilities vary between them. Individuals with high probabilities leave early, and by this selection process long durations are recorded for individuals with low probabilities of finding a job. This result is apparently the same as when probabilities are the same for everybody but declining with the time already elapsed: $\lambda'(t) < 0$ and $v \equiv 1$.

This suggests that it is not possible to distinguish between time dependence and unobserved heterogeneity. In fact when there are no exogenous covariates this is exactly the case. For example, it can be shown (c.f. Lancaster and Nickell, JRSS, 1980; Elbers and Ridder, RES, 1982) that the *cdf* given by

$$F(t) = 1 - \frac{1}{1+at}$$

can be obtained from the following two observationally equivalent specifications:

- *Specification 1*:

$$\begin{aligned} H_1(t) &= \ln(1+at) \\ G_1(t) &= \begin{cases} 1 & v \geq 1 \\ 0 & v < 1. \end{cases} \end{aligned}$$

  Note that

$$F(t) = 1 - \exp[-H_1(t)] = 1 - \exp[-\ln(1+at)] = 1 - \frac{1}{1+at}.$$

- *Specification 2*:

$$\begin{aligned} H_2(t) &= t \\ G_2(t) &= \int_0^v \frac{1}{a} e^{-u/a} du. \end{aligned}$$

  In this case v has an exponential distribution. Note that

$$F(t \mid v) = 1 - \exp[-vH_2(t)] = 1 - \exp(-vt)$$

$$F(t) = E_v[1 - \exp(-vt)] = 1 - E_v\left(e^{-vt}\right) = 1 - \frac{1}{1+at}$$

  where we are using the expression for the *mgf* of the exponential distribution.

Elbers and Ridder (1982) prove that this is not the case when there are regressors. Specifically, they prove that there are no two combinations of heterogeneity distributions and time dependence functions given the same duration distribution. This result is stronger than the more traditional results on identification in parametric models and pioneered semiparametric identification results in econometrics.

Why this is so can be seen intuitively in the previous example assuming that we can observe the duration distributions for two different known values of $\exp(x'\beta)$, say 1 and ϕ. For $\exp(x'\beta) = 1$ the two *cdf*'s are identical. The question now is whether for $\exp(x'\beta) = \phi$ the two specifications also lead to the same *cdf* for durations. It can be seen that this requires that

$$(1 + at)^{\phi} = (a\phi t + 1) \quad \text{for } t \geq 0$$

but this only holds if $\phi = 1$. So introducing systematic variation in the hazard has as a consequence that specifications 1 and 2 are distinguishable.

This identification result suggests that it may be possible to devise semiparametric estimators of PH models with heterogeneity that do not require a specific distribution for v. Heckman and Singer (1984) proposed a method along these lines.

7 Estimating discrete duration models

The usual approach has been to model discrete (or grouped) durations using density function terms in a likelihood as if the spells were observed continuously.

Suppose T_i is a discrete random variable. The conditional hazard given covariates x_i is given by

$$h(t, x_i) = \Pr(T_i = t \mid T_i \geq t, x_i) \quad (t = 1, 2, 3, ...).$$

Since this is now a probability it is natural to use discrete choice models in order to specify the exit rate:

$$\Pr(T_i = t \mid T_i \geq t, x_i) = F\left(\gamma_t + x_i'\beta_t\right)$$

where F is a *cdf* (e.g. normal or logistic). If γ_t and β_t were constant for all t, then $h(t, x_i)$ would be constant for a given value of x and there would be no state dependence.

A possible way of introducing state dependence is to specify γ_t and β_t as polynomials in t. For example,

$$\gamma_t = \gamma_0 + \gamma_1 (\ln t) + \gamma_2 (\ln t)^2 .$$

More generally, γ_t and β_t could be treated as unrestricted duration-specific parameters:

$$\gamma_t = \sum_{j=1}^{T^*} \gamma_j \mathbf{1}(T_i = j)$$

where T^* is near to the maximum spell observed in the data. Such a model amounts to a sequence of T^* binary choice models for the surviving populations at each t.

More specifically, for $t = 1$:

$$\Pr(T_i = 1 \mid T_i \geq 1, x_i) = \Pr(y_{1i} = 1 \mid w_{1i} = 1, x_i) = F\left(\gamma_1 + x_i'\beta_1\right)$$

where $y_{1i} = \mathbf{1}(T_i = 1)$ and $w_{1i} = \mathbf{1}(T_i \geq 1)$ (note that $w_{1i} = 1$ with probability one).

Now for $t = 2$:

$$\Pr(T_i = 2 \mid T_i \geq 2, x_i) = \Pr(y_{2i} = 1 \mid w_{2i} = 1, x_i) = F\left(\gamma_2 + x_i'\beta_2\right)$$

where $y_{2i} = \mathbf{1}(T_i = 2)$ and $w_{2i} = \mathbf{1}(T_i \geq 2)$, etc.

Hence, the log likelihood function for a sample of entrants can be written as:

$$L = \sum_{i=1}^{N} \sum_{t=1}^{T^*} w_{ti} \left\{ y_{ti} \ln F\left(\gamma_t + x_i'\beta_t\right) + (1 - y_{ti}) \ln\left[1 - F\left(\gamma_t + x_i'\beta_t\right)\right] \right\}.$$

There are two types of individual contributions to the likelihood:

- A completed spell $T_i = t$ contributes a term $\Pr(T_i = t \mid x_i)$:

$$\begin{aligned} \ln \Pr(T_i = t \mid x_i) &= \ln\left[h_{ti} \prod_{s=1}^{t-1} (1 - h_{si})\right] = \ln h_{ti} + \sum_{s=1}^{t-1} \ln(1 - h_{si}) \\ &= \ln F\left(\gamma_t + x_i'\beta_t\right) + \sum_{s=1}^{t-1} \ln\left[1 - F\left(\gamma_s + x_i'\beta_s\right)\right]. \end{aligned}$$

- A censored spell at t contributes a term $\Pr(T_i \geq t + 1 \mid x_i)$:

$$\begin{aligned} \ln \Pr(T_i \geq t + 1 \mid x_i) &= \ln\left[1 - \Pr(T_i \leq t \mid x_i)\right] = \ln\left[\prod_{s=1}^{t} (1 - h_{si})\right] \\ &= \sum_{s=1}^{t} \ln\left[1 - F\left(\gamma_s + x_i'\beta_s\right)\right]. \end{aligned}$$

The amount of generality in the specification of γ_t and β_t will be partly dictated by the frequency at which durations are reported but also by the shape of empirical hazards.

Thus, discrete duration models can be regarded as a sequence of binary models. This is convenient because standard logit software can be used to estimate discrete duration models (Jenkins, 1995).

Time-varying characteristics We often have explanatory variables whose values change during the course of the spell for a given individual. For example, changing levels of unemployment insurance benefits over time, or aggregate economic variables.

In such cases, the interpretation of the hazard rate changes. $h(t, x_i(t))$ can be regarded as:

$$\Pr(T_i = t \mid T_i \geq t, x_i(1), ..., x_i(\infty)) = F(\gamma_t + x_i(t)\beta_t).$$

That is, the hazard for the distribution of T_i conditioned on the entire $\{x_i(t)\}$ process. This is a case where x is taken as a "strictly exogenous" variable in duration time. Duration models with "predetermined" variables are also possible (see Bover, Arellano, and Bentolila, 2002, for a discussion).

Continuous-time duration models with time-varying x's pose difficult problems because the x's are not continuously observed. However, in discrete duration models they just involve a trivial extension of the basic models.

Unobserved heterogeneity Let u_i be an unobserved covariate distributed independently of x_i with *cdf* G. Then

$$\Pr\left(T_i = t \mid x_i\right) = \int \Pr\left(T_i = t \mid x_i, u\right) dG\left(u\right) = \int h_{ti}\left(u\right) \prod_{s=1}^{t-1} \left[1 - h_{si}\left(u\right)\right] dG\left(u\right)$$

where for example

$$h_{ti}\left(u\right) = F\left(\gamma_t + x_i'\beta_t + u\right).$$

A common specification for G is a "mass point" distribution: u is assumed to take on $\{\xi_1, ..., \xi_m\}$ different values with probabilities $\{p_1, ..., p_m\}$. Thus,

$$\Pr\left(T_i = t \mid x_i\right) = \sum_{j=1}^{m} h_{ti}\left(\xi_j\right) \prod_{s=1}^{t-1} \left[1 - h_{si}\left(\xi_j\right)\right] p_j.$$

Both the ξ_j and the p_j are treated as additional parameters to be estimated.

8 Grouped duration data: linking discrete and continuous models

In a continuous model $\{T = t\}$ was interpreted as an observation from a continuous process, hence contributing a density function term to the likelihood. In a discrete model, $\{T = t\}$ is interpreted as an observation of a discrete process. Alternatively, it can be regarded as a grouped (or time-aggregate) observation from a continuous process. That is, it may be interpreted as the observation of the event $\{t \leq T < t + 1\}$, where T is an underlying continuous duration random variable.

In such case, the contribution to the likelihood of a spell completed between t and $t + 1$ would be

$$\Pr\left(t \leq T < t + 1\right) = F\left(t + 1\right) - F\left(t\right)$$

and the contribution of a spell censored at c:

$$\Pr\left(T > c\right) = 1 - F\left(c\right)$$

where F is now the continuous *cdf* of T.

In this context, the discrete hazard rate of the previous section would be reinterpreted as a "grouped hazard":

$$\begin{aligned} \Pr\left(t \leq T < t + 1 \mid T \geq t\right) &= 1 - \Pr\left(T \geq t + 1 \mid T \geq t\right) = 1 - \frac{\Pr\left(T \geq t + 1\right)}{\Pr\left(T \geq t\right)} = 1 - \left(\frac{1 - F\left(t + 1\right)}{1 - F\left(t\right)}\right) \\ &= 1 - \frac{\exp\left[-\int_{-\infty}^{t+1} h\left(u\right) du\right]}{\exp\left[-\int_{-\infty}^{t} h\left(u\right) du\right]} = 1 - \exp\left[-\int_{t}^{t+1} h\left(u\right) du\right] \end{aligned}$$

where $h(u)$ is the continuous hazard function of T.

Proportional hazards from grouped durations Suppose that $h(u,x)$ is specified as:

$$h(u,x) = \lambda(u)\exp(x'\beta).$$

Then

$$\Pr(t \leq T < t+1 \mid T \geq t, x) = 1 - \exp\left[-e^{x'\beta}\int_t^{t+1}\lambda(u)\,du\right] = 1 - \exp\left(-e^{\gamma_t + x'\beta}\right) = F^*\left(\gamma_t + x'\beta\right)$$

where

$$\gamma_t = \ln\int_t^{t+1}\lambda(u)\,du$$

and F^* is the extreme value *cdf* $F^*(r) = 1 - \exp(-e^r)$.

Thus, the discrete-time hazard takes the form of an extreme value distribution. Note that this follows directly from the PH specification without any further distributional assumptions.

The more traditional approach was to use a Weibull specification for $\lambda(u)$, which implies a particular parametric function for γ_t (that in this case would depend on a single parameter). Bruce Meyer (Econometrica, 1990) propose to censor any ongoing observations at some large spell length T^* and then treat $\left(\gamma_0, \gamma_1, ..., \gamma_{T^*-1}\right)$ as a vector of T^* additional parameters to be estimated. From this perspective, a parametric assumption about $\lambda(u)$ can be regarded as putting restrictions on the elements of γ.

We can use LR tests to test parametric specifications of $\lambda(u)$ against the unrestricted model. Meyer does so and rejects the parametric models; however, there is no strong evidence of biases in the estimates of β due to the Weibull assumption (Meyer, 1990, p. 776).

"Proportional hazards" with time-varying covariates Consider the specification

$$h(u,x) = \lambda(u)\exp\left[x(u)'\beta\right].$$

In such case, assuming that $x(u)$ is constant for $t \leq u < t+1$, i.e. that the changes in the time-varying covariates occur at integer points, the discrete time hazard can also be written as the extreme value model:

$$\Pr(t \leq T < t+1 \mid T \geq t, x) = 1 - \exp\left[-e^{x(t)'\beta}\int_t^{t+1}\lambda(u)\,du\right] = 1 - \exp\left(-e^{\gamma_t + x_t'\beta}\right) = F^*\left(\gamma_t + x_t'\beta\right).$$

Note though, that the result requires that β does not change with t.

This is a convenient device for handling time-varying covariates in continuous time models, since the x's are only observed at given intervals.

The straightforward interpretation of the discrete hazard model with the extreme value specification as a grouped hazard for the continuous PH specification is an attractive feature of the extreme value model. However, in some applications probit and logit have been seen to outperform the extreme value probability (c.f. Narendranathan and Stewart, J. of Applied Econometrics, 1993).

Estimating a continuous time model has the advantage that the parameters are free from the level of aggregation or grouping in the data, hence making the comparison with estimates from other data sets feasible. Nevertheless, economic duration data is often more suited to discrete or grouped specifications than to continuous ones, given the nature of the data where the spells take on only a small number of different vales (as in weekly or monthly unemployment durations).

General mapping between discrete and continuous duration models Consider a generic continuous-time hazard model $h(u,x)$. Generalizing the previous argument we have

$$\Pr(t \leq T < t+1 \mid T \geq t, x) = F^*[H(t,x)] = 1 - \exp\left(-e^{H(t,x)}\right)$$

where

$$e^{H(t,x)} = \int_t^{t+1} h(u,x)\,du.$$

To explore the connection with the logistic discrete duration model, define the function

$$\varphi(t,x) = \ln\left(\frac{F^*[H(t,x)]}{1-F^*[H(t,x)]}\right) = \ln\left(\frac{1-\exp\left(-e^{H(t,x)}\right)}{\exp\left(-e^{H(t,x)}\right)}\right),$$

so that

$$\int_t^{t+1} h(u,x)\,du = \ln\left(1+e^{\varphi(t,x)}\right).$$

We have,

$$\Pr(t \leq T < t+1 \mid T \geq t, x) = \Lambda[\varphi(t,x)]$$

where $\Lambda(r)$ is the logistic *cdf*. Thus, when we estimate a logistic model of the form $\Lambda[x'\beta(t)]$, the index $x'\beta(t)$ can be regarded as an approximating model for $\varphi(t,x)$. This is of interest because, having established the connection, we can use the logistic estimates to calculate approximate derivative effects for the underlying continuous-time model.

A derivative effect for the continuous-time hazard integrated between t and $t+1$ is

$$D(t,x) = \int_t^{t+1} \frac{\partial h(u,x)}{\partial x} du.$$

Therefore,

$$\frac{\partial}{\partial x}\int_t^{t+1} h(u,x)\,du = \frac{e^{\varphi(t,x)}}{\left(1+e^{\varphi(t,x)}\right)}\frac{\partial\varphi(t,x)}{\partial x} = \Lambda[\varphi(t,x)]\frac{\partial\varphi(t,x)}{\partial x}.$$

So, if we have estimated the model $\Lambda [x'\beta (t)]$, the following holds as an approximation:

$$D(t,x) \simeq \Lambda \left[x'\beta (t)\right] \beta (t).$$

More generally, for an arbitrary *cdf* G and $\varphi (t,x) = G^{-1} [H (t,x))]$ we have

$$\int_t^{t+1} h(u,x)\, du = -\ln [1 - \Pr (t \leq T < t+1 \mid T \geq t, x)] = -\ln \{1 - G [\varphi (t,x)]\},$$

so that

$$\frac{\partial}{\partial x} \int_t^{t+1} h(u,x)\, du = \frac{g [\varphi (t,x)]}{1 - G [\varphi (t,x)]} \frac{\partial \varphi (t,x)}{\partial x} = h_g [\varphi (t,x)] \frac{\partial \varphi (t,x)}{\partial x}$$

where h_g is the hazard associated with G. The conclusion is that if we have estimated the model $G [x'\beta (t)]$, the following holds as an approximation:

$$D(t,x) \simeq h_g \left[x'\beta (t)\right] \beta (t).$$

9 Multiple-exit discrete duration models

Consider now a model in which there is more than one possible exit from unemployment. This section follows closely the presentation in Bover and Gomez (2004), and as in their paper we distinguish between exits to a permanent job and exits to a temporary job.

If we have a discrete duration variable T and two alternatives represented by the indicators D_1 and D_2, we can define the following intensities of transition to each of the states:

$$\begin{aligned} \phi_1 (t) &= \Pr (T = t, D_1 = 1 \mid T \geq t) \\ \phi_2 (t) &= \Pr (T = t, D_2 = 1 \mid T \geq t) \end{aligned}$$

such that the hazard rate from unemployment is given by:

$$\phi_H (t) = \phi_1 (t) + \phi_2 (t).$$

Likewise, in order to see the discrete duration models as discrete choice models, it is useful to introduce sequences of exit indicators at t to a given alternative:

$$Y_{1t} = \mathbf{1} (T = t, D_1 = 1), Y_{2t} = \mathbf{1} (T = t, D_2 = 1) \qquad \text{for } t = 1, 2, 3...$$

According to this notation, $\phi_1 (t) = \Pr (Y_{1t} = 1 \mid T \geq t)$ and $\phi_2 (t) = \Pr (Y_{2t} = 1 \mid T \geq t)$.

Alternatively, we can define exit rates to each of the states conditional upon not exiting to the alternative state:

$$\begin{aligned} h_1 (t) &= \Pr (Y_{1t} = 1 \mid T \geq t, Y_{2t} = 0) \\ h_2 (t) &= \Pr (Y_{2t} = 1 \mid T \geq t, Y_{1t} = 0). \end{aligned}$$

The relationship with the previous transition intensities is given by:

$$h_1(t) = \frac{\Pr(Y_{1t} = 1 \mid T \geq t)}{\Pr(Y_{2t} = 0 \mid T \geq t)} = \frac{\phi_1(t)}{1 - \phi_2(t)}$$

and similarly

$$h_2(t) = \frac{\phi_2(t)}{1 - \phi_1(t)}.$$

Thus, unlike the continuous case, in the context of discrete duration variables and multiple alternatives, we can choose between modeling the intensities $\phi_j(t)$ or the conditional hazard rates $h_j(t)$. For example, if T represents the duration of unemployment and exits 1 and 2 are permanent employment and temporary employment, respectively, $\phi_1(t)$ is the probability of exiting to permanent employment at $T = t$ among those who remain unemployed for at least $T \geq t$ periods. For its part, $h_1(t)$ is the probability of exiting to permanent employment at $T = t$ among those who remain unemployed for at least $T \geq t$ and do not exit to temporary employment at $T = t$.

A specification commonly used in multiple choice problems is the multinomial logit model. In such case, the dependence of $\phi_1(t)$ and $\phi_2(t)$ on the explanatory variables x is specified by

$$\begin{aligned} \phi_1(t) &= \frac{e^{x'\beta_1}}{1 + e^{x'\beta_1} + e^{x'\beta_2}} \\ \phi_2(t) &= \frac{e^{x'\beta_2}}{1 + e^{x'\beta_1} + e^{x'\beta_2}}. \end{aligned}$$

Note that, in accordance with the relationships given above, this specification for $\phi_1(t)$ and $\phi_2(t)$ implies that

$$\begin{aligned} h_1(t) &= \frac{e^{x'\beta_1}}{1 + e^{x'\beta_1}} \\ h_2(t) &= \frac{e^{x'\beta_2}}{1 + e^{x'\beta_2}}. \end{aligned}$$

That is, if the transition intensities are multinomial logit, the conditional exit rates are binary logit with the same parameters. As a result, the use of the logistic specification leads to the same model in both cases.

Nevertheless, having obtained estimates of the parameters β_1, β_2, we can obtain two different measurements of the effect of the explanatory variables on the probabilities of exit to a specific alternative depending on whether changes in the $\phi_j(t)$ or changes in the $h_j(t)$ are used. Specifically, for a continuous variable and for the exit to alternative 1 we can use

$$\varepsilon_{\phi_1 tk} = \frac{\partial \phi_1(t)}{\partial x_k} . \frac{x_k}{\phi_1(t)}$$

or else

$$\varepsilon_{h_1 tk} = \frac{\partial h_1(t)}{\partial x_k} . \frac{x_k}{h_1(t)}.$$

It can be easily shown that the relationship between the two elasticities is given by

$$\begin{aligned}\varepsilon_{h_1tk} &= \varepsilon_{\phi_1tk} + \frac{\phi_2(t)}{1-\phi_2(t)}\varepsilon_{\phi_2tk} \\ \varepsilon_{h_2tk} &= \varepsilon_{\phi_2tk} + \frac{\phi_1(t)}{1-\phi_1(t)}\varepsilon_{\phi_1tk}.\end{aligned}$$

In addition, in the logistic case:

$$\varepsilon_{h_1tk} = \beta_{1k}\left[1-h_1(t)\right]x_k$$

where β_{1k} denotes the k-th coefficient of the vector β_1.

The differences between the two types of elasticity may be greater when the temporal aggregation of the durations is large. An alternative way of constructing derivative effects is to regard discrete durations as grouped observations from a continuous process, and relate the grouped transitions to the underlying continuous transition intensities, along the lines of the discussion in the previous section.

Competing risk models The models for the conditional probabilities $h_1(t)$, $h_2(t)$ are usually called competing risk models. This name derives from the fact that if we consider the existence of two latent duration variables T_1^* and T_2^*, such that the observed duration is $T = \min(T_1^*, T_2^*)$ and T_1^*, T_2^* are independent, then the conditional exit rates can be interpreted as exit rates for the latent durations:

$$\begin{aligned}h_1(t) &= \Pr(T_1^* = t \mid T_1^* \geq t) \\ h_2(t) &= \Pr(T_2^* = t \mid T_2^* \geq t).\end{aligned}$$

That is, to analyze exits to alternative 1 we take the exits to alternative 2 as censored observations, and vice versa.

Note that irrespective of whether T_1^*, T_2^* correspond to well defined concepts (and in the case of exits to permanent or temporary contracts it is difficult to imagine that they do), $h_1(t)$, $h_2(t)$ generally represent useful descriptive characteristics for the durations and exits observed.

Estimation of the parameters (β_1, β_2) of the logistic model We can consider two different methods for estimating the model. The first consists in the joint estimation of β_1 and β_2 by maximum likelihood, while the second consists in separate estimation of β_1 and β_2 by conditional maximum likelihood. Both methods provide consistent and asymptotically normal estimates of the parameters, although the first estimator is in general asymptotically more efficient than the second. The advantage of the second is basically that its computation is faster. Moreover, separate estimators of the parameters corresponding to one of the alternatives are robust to specification errors in the regression index for the other alternative.

Consider a sample that includes both entrants into unemployment and the stock of the unemployed workers at the time of interview. Let c_i be an indicator that takes the value 1 if the end of the period of unemployment is observed, and 0 if not; T_i^0 denotes the observed duration, and q_i is the number of months in unemployment at the time of the first interview (for an entrant $q_i = 1$).

The joint log-likelihood function is given by:

$$\begin{aligned} L(\beta_1, \beta_2) &= \sum_{i=1}^{N} \Bigg\{ (1-c_i) \sum_{t=q_i}^{T_i^0} \ln\left[1 - \phi_{1i}(t) - \phi_{2i}(t)\right] \\ &\quad + c_i \left(\sum_{t=q_i}^{T_i^0 - 1} \ln\left[1 - \phi_{1i}(t) - \phi_{2i}(t)\right] + D_{1i} \ln \phi_{1i}\left(T_i^0\right) + D_{2i} \ln \phi_{2i}\left(T_i^0\right) \right) \Bigg\}. \end{aligned}$$

Likewise, using the sequences of indicators defined above we can express $L(\beta_1, \beta_2)$ as

$$L(\beta_1, \beta_2) = \sum_{t=1}^{\max\left(T_i^0\right)} L_t$$

where

$$\begin{aligned} L_t &= \sum_{i=1}^{N} 1\left(T_i^0 \geq t \geq q_i\right) \{ c_i Y_{1ti} \ln \phi_{1i}(t) + c_i Y_{2ti} \ln \phi_{2i}(t) \\ &\quad + (1 - c_i Y_{1ti} - c_i Y_{2ti}) \ln\left[1 - \phi_{1i}(t) - \phi_{2i}(t)\right] \}, \end{aligned}$$

which shows that $L(\beta_1, \beta_2)$ can be regarded as the log-likelihood of a multinomial logit model defined on the basis of the concatenation of the samples surviving at each duration. The joint maximum likelihood estimators $\left(\widehat{\beta}_1, \widehat{\beta}_2\right)$ are defined as the values that maximize $L(\beta_1, \beta_2)$.

In addition, the conditional log-likelihood function for exit 1 is given by:

$$\begin{aligned} L_{c1}(\beta_1) &= \sum_{i=1}^{N} \Bigg\{ c_i \left(D_{1i} \ln h_{1i}\left(T_i^0\right) + D_{1i} \sum_{t=q_i}^{T_i^0 - 1} \ln\left[1 - h_{1i}(t)\right] \right) \\ &\quad + \left[D_{2i} + (1 - c_i)\right] \sum_{t=q_i}^{T_i^0} \ln\left[1 - h_{1i}(t)\right] \Bigg\} \end{aligned}$$

with a similar expression for the likelihood corresponding to exit 2, $L_{c2}(\beta_2)$. Note that in $L_{c1}(\beta_1)$ the exits to alternative 2 are treated as censored observations, so that formally it is a function with exactly the same form as the likelihood with a single exit of the previous section. The implication is that the conditional maximum likelihood estimators, $\left(\widetilde{\beta}_1, \widetilde{\beta}_2\right)$ defined as the maximizers of $L_{c1}(\beta_1)$ and $L_{c2}(\beta_2)$, respectively, can be obtained as separate maximum-likelihood estimates of two binary logit models.

References

[1] Bover, O., M. Arellano, and S. Bentolila (2002): "Unemployment Duration, Benefit Duration, and the Business Cycle", *The Economic Journal*, 112, 223-265.

[2] Bover, O. and R. Gómez (2004): "Another Look at Unemployment Duration: Exit to a Permanent vs. a Temporary Job", *Investigaciones Económicas*, 28, 285-314.

[3] Elbers, C. and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, 49, 403-409.

[4] Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

[5] Jenkins, S. (1995): "Easy Estimation Methods for Discrete-time Duration Models", *Oxford Bulletin of Economics and Statistics*, 57, 120-138.

[6] Lancaster, T. (1979): "Econometric Models for the Duration of Unemployment", *Econometrica*, 47, 939-956.

[7] Lancaster, T. (1990): *The Econometric Analysis of Transition Data*, Cambridge.

[8] Meyer, B. (1990): "Unemployment Insurance and Unemployment Spells", *Econometrica*, 58, 757-782.

[9] Narendranathan, W. and M. Stewart (1993): "How does the benefit effect vary as unemployment spells lengthen?", *Journal of Applied Econometrics*, 8, 361–81.

[10] Van den Berg, G. (2001): "Duration Models: Specification, Identification and Multiple Durations", in Heckman and Leamer (eds.), *Handbook of Econometrics*, Vol. 5, Ch. 55.