Duration Models

JOAN LLULL

MOVE, Universitat Autònoma de Barcelona
and Barcelona School of Economics

I.   Introduction

A.   Motivation

There are many examples in economics in which our variable of interest is a *duration*. Duration data can answer the question "how long an individual has been in a particular state when exiting from it". Examples are the number of weeks or months that an individual has been unemployed when she finds a job, how long an individual have been in a hospital before leaving it, or what is the life expectancy of an individual with certain age (i.e. how long has she been alive when she dies). In this chapter we learn the basic tools to model this kind of data.

This kind of analysis allows us to talk about why durations differ across individuals (i.e. what is the effect of individual characteristics on duration), and about how and why do exit probabilities vary over time. These techniques have a long tradition in biometrics. For this reason, it is common to find a lot of nomenclature that has been borrowed from that field: survival probabilities, hazard functions,...

B.   Duration data

Our data consist of a sample of durations. We have a sample of durations for N individuals, $t_1, t_2, ..., t_N$. Importantly, these data are typically censored. Figure 1 draws two examples of censored samples. In the left plot, we have a hypothetical sample of durations that is assumed to be obtained interviewing individuals from January 1990 to January 1992 at a monthly basis. For individuals 2 and 4, we observe complete duration spells. Individual 1 was already unemployed at the first interview date. Individual 3 is still unemployed at the last interview date. In both cases we know that the duration of their unemployment spell is larger than a certain value, but not by how much; i.e. we observe $t > \bar{t}$, but not t.

The second hypothetical sample could be collected through registries. Imagine that, in order to receive the unemployment benefit, workers have to show up at the unemployment office and prove that they are still unemployed. In the figure, individual 1 found a job in less than one year; in our data, we observe that he does not appear in the sample of January 1991, so we learn that he found a job before; as a result, we know that the duration of this unemployment spell was

FIGURE 1. TWO EXAMPLES OF CENSORED OBSERVATIONS

(a) Example 1   (b) Example 2
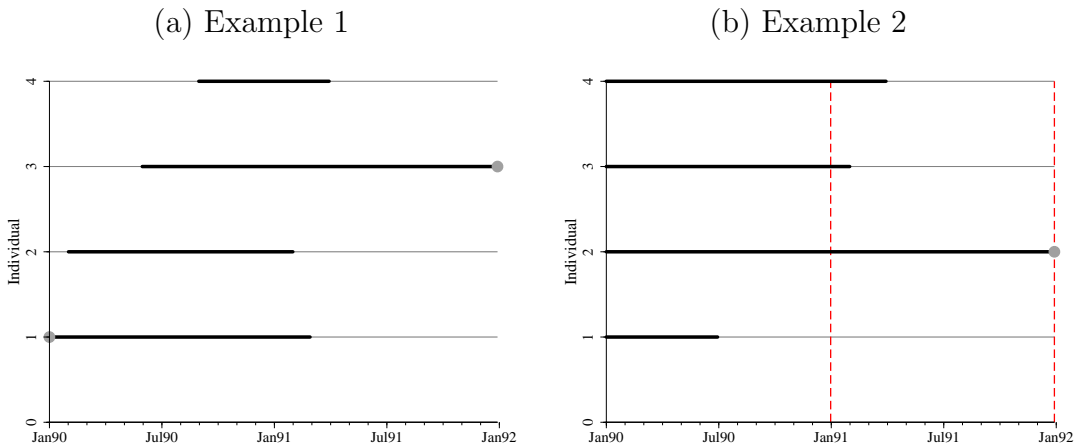
Note: Black lines represent the time when the individual was unemployed. A dot indicates that the individual is still unemployed at that date, but we do not have further information about him/her. Vertical red dashed lines in Example 2 are interview dates.

below one year, but not by how much. Similarly, individuals 3 and 4 found a job during the second year. For these three individuals data is censored in the sense that we only observe that the duration is within a given interval, but we do not know it exactly; i.e. we observe $\underline{t} < t < \bar{t}$ instead of t. Individual 2 is censored exactly in the same way as individuals 1 and 3 from the left figure.

The presence of censoring is one of the main motivations to use the techniques presented here (as opposed to, say, regression). If data are censored, sample average durations are biased. This framework incorporates censoring into the analysis without the need of additional strong assumptions.

On top of censoring, duration analysis allows for time/state dependence, i.e. the probability of terminating the current spell depends on the duration of the spell. For instance, we might be interested in analyzing whether individuals' probability of finding a job is decreasing in the time they have been unemployed in the current spell. These techniques allow us to go beyond the simple average duration and look at the shape of exit probabilities over different durations.

II. The Hazard Function

The *hazard function* is the most important object in this analysis. It is defined as the probability/density of exiting at t conditional on being alive. In the unemployment example, it is the probability that an individual finds a job at, say, the 10th month of unemployment, conditional on being unemployed in the 9th month.

This hazard function can be constant or time varying. For instance, if individuals receive offers in every period with the same probability, the hazard is constant.

FIGURE 2. MORTALITY HAZARD RATE
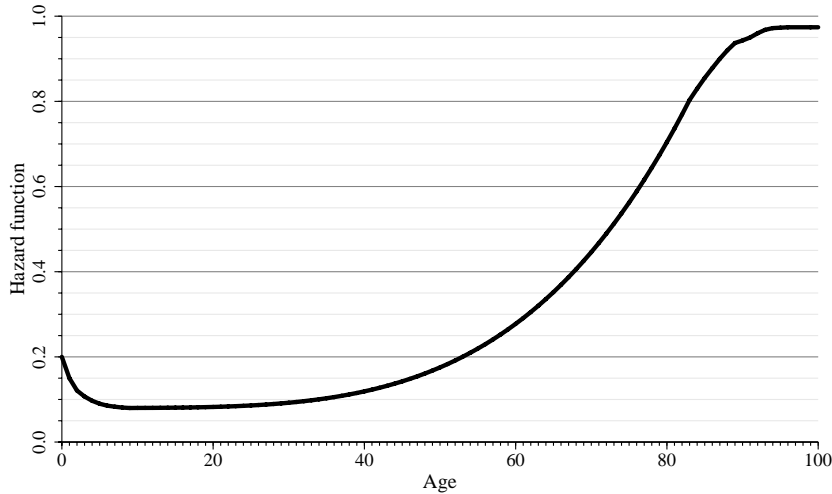
Note: The line depicts the hazard mortality rate, i.e. the probability of dying at age a conditional on survival until that age.

A time-varying example is plotted in Figure 2. The figure depicts the hazard mortality rates of a given population. Infant mortality makes it decreasing at the first five years of age. A hazard rate of around 8% at age 5 does not mean that the probability of dying at age 5 is 8% but, instead, that the probability of dying at age 5 for those individuals who survived the first 4 years is 8%. A good example to understand this distinction is the hazard rate at age 100: probability of dying at age 100 conditional on having survived until age 99 is almost 1; however, the probability of dying at age 100 is almost zero.

We consider both discrete and continuous durations. Whenever our duration of interest is discrete, the hazard function is a probability. When the duration is continuous, this hazard function is a pdf.

Our interest in the hazard function has theoretical and empirical grounds. From a theoretical perspective, it is more appealing to model the decision of exiting conditional on survival than the duration itself (e.g. modeling job offer arrival rate). Empirically, it is also convenient to model the hazard function because it implies a binomial discrete decision and delivers a very clean likelihood, and because it allows to take censoring into account without need of further strong assumptions.

In this section, we characterize the unconditional hazard function. We build bridges between hazard functions and pdfs/cdfs/probability masses, which then are used to write the likelihood function. Regressors are introduced later on in the chapter.

A. Hazard function for a discrete variable

Let t be a random variable with discrete support $\{1, 2, 3, ...\}$ with probability mass function $p(\tau) = \Pr(t = \tau)$ and cdf $F(t) = p(1) + p(2) + ... + p(t)$ for $t = 1, 2, 3, ...$ The hazard function is defined as:

$$h(\tau) \equiv \Pr(t = \tau | t \geq \tau) = \frac{\Pr(t = \tau)}{\Pr(t \geq \tau)} = \frac{p(\tau)}{1 - F(\tau - 1)} = \frac{F(\tau) - F(\tau - 1)}{1 - F(\tau - 1)}. \quad (1)$$

This hazard function is a modeling decision (e.g. with a logistic or normal cdf). We need to recover $p(t)$ and $F(t)$ in order to write the likelihood. To recover them, we proceed recursively. In the first period, we know that:

$$h(1) = \Pr(t = 1 | t \geq 1) = \Pr(t = 1) = p(1) \quad \text{and} \quad F(1) = p(1). \quad (2)$$

In the second period, we can use equation (1):

$$h(2) = \frac{p(2)}{1 - F(1)} = \frac{p(2)}{1 - h(1)} \quad \Rightarrow \quad p(2) = h(2)(1 - h(1)). \quad (3)$$

Hence:

$$F(2) = p(1) + p(2) = h(1) + h(2)(1 - h(1)) \quad \Leftrightarrow \quad 1 - F(2) = (1 - h(2))(1 - h(1)). \quad (4)$$

In the third period:

$$h(3) = \frac{p(3)}{1 - F(2)} = \frac{p(3)}{(1 - h(2))(1 - h(1))} \quad \Rightarrow \quad p(3) = h(3) \prod_{s=1}^{2} (1 - h(s))). \quad (5)$$

Hence:

$$F(3) = p(1) + p(2) + p(3) = h(1) + h(2)(1 - h(1)) + h(3) \prod_{s=1}^{2} (1 - h(s))), \quad (6)$$

which implies:

$$1 - F(3) = \prod_{s=1}^{3} (1 - h(s))). \quad (7)$$

In general, the recursion is such that:

$$p(t) = F(t) - F(t - 1) = h(t)(1 - F(t - 1)) = h(t) \prod_{s=1}^{t-1} (1 - h(s)), \quad (8)$$

and:

$$F(t) = 1 - \prod_{s=1}^{t} (1 - h(s)). \quad (9)$$

These expressions are interpretable. In particular:

$$p(\tau) = \Pr(t = \tau) = h(\tau) \prod_{s=1}^{\tau-1} (1 - h(s)) = \Pr(t = \tau | t \geq \tau) \Pr(t > \tau - 1 | t \geq \tau - 1)...,$$

(10)

or, in words, the probability of exiting at time t is equal to the probability of exiting at time t conditional on survival times the probability of survival until t. Additionally:

$$F(\tau) = \Pr(t \leq \tau) = 1 - \prod_{s=1}^{\tau} (1 - h(s)) =$$
$$= 1 - \Pr(t > \tau | t \geq \tau) \Pr(t > \tau - 1 | t \geq \tau - 1)... =$$
$$= 1 - \Pr(t > \tau).$$

(11)

B. Hazard function for a continuous variable

Consider now the case of a continuous duration t. This random variable is characterized by its pdf $f(t)$ instead of a probability mass. Therefore, its hazard function is also a density:

$$h(\tau) = \lim_{dt \to 0} \frac{\Pr(\tau \leq t < \tau + dt | t \geq \tau)}{dt} = \lim_{dt \to 0} \frac{\Pr(\tau \leq t < \tau + dt)}{\Pr(t \geq \tau)} \bigg/ dt = \frac{f(\tau)}{1 - F(\tau)}.$$

(12)

Note that this expression is analogous to equation (1).

In order to derive $f(t)$ and $F(t)$ from the hazard function (so that we can write the likelihood of our sample) we make use of the *integrated* or *cumulative hazard*:

$$H(t) = \int_0^t h(s)ds.$$

(13)

The integrated hazard can be written as a function of $f(t)$ and $F(t)$:

$$H(t) = \int_0^t \frac{f(s)}{1 - F(s)} ds = [-\ln(1 - F(s))]_0^t = -\ln[1 - F(t)].$$

(14)

Therefore, we can trivially see that:

$$F(t) = 1 - \exp(-H(t)),$$

(15)

and, similarly:

$$f(t) = \frac{\partial F(t)}{\partial t} = h(t) \exp(-H(t)).$$

(16)

The interpretation of these expressions is not as straightforward as before, but

we can make a connection with the discrete case. In the discrete case:

$$\ln[1 - F(t)] = \sum_{s=1}^{t} \ln(1 - h(s)), \tag{17}$$

and $\ln(1 - h(s)) \approx -h(s)$, which compares to the continuous case:

$$\ln[1 - F(t)] = -H(t) = \int_{0}^{t} (-h(s))ds. \tag{18}$$

C. Some frequently used hazard functions

In the discrete case, we often avoid parametric assumptions on the hazard function, and we estimate it semi-parametrically. When the duration is continuous, however, (and in some cases when it is discrete), we need to make functional form assumptions on the hazard function. These are two widely used cases.

Constant hazard This is the simplest possible hazard function. In our example of unemployment duration, this assumption is consistent with a constant job arrival rate, i.e. in every period we have the same probability of receiving a job offer, no matter how long we have been unemployed. Therefore, we assume $h(t) = \lambda$, with $\lambda > 0$.

Given this function, the integrated hazard is very easy to compute:

$$H(t) = \int_{0}^{t} \lambda du = \lambda t. \tag{19}$$

Therefore, cdf and pdf are trivially derived:

$$F(t) = 1 - e^{-\lambda t}, \tag{20}$$

and:

$$f(t) = \lambda e^{-\lambda t}, \tag{21}$$

which is the exponential distribution.

Therefore, the exponential distribution has a constant hazard (this is called the memoryless property of the exponential distribution). Additionally, the expected duration with this function is the inverse of the hazard function: $\mathbb{E}[T] = 1/\lambda$.

The discrete counterpart of this parametric family of functions is:
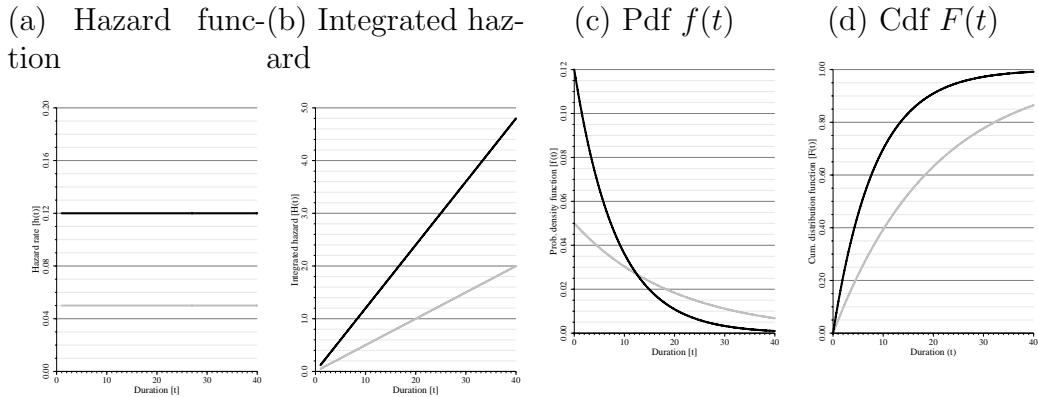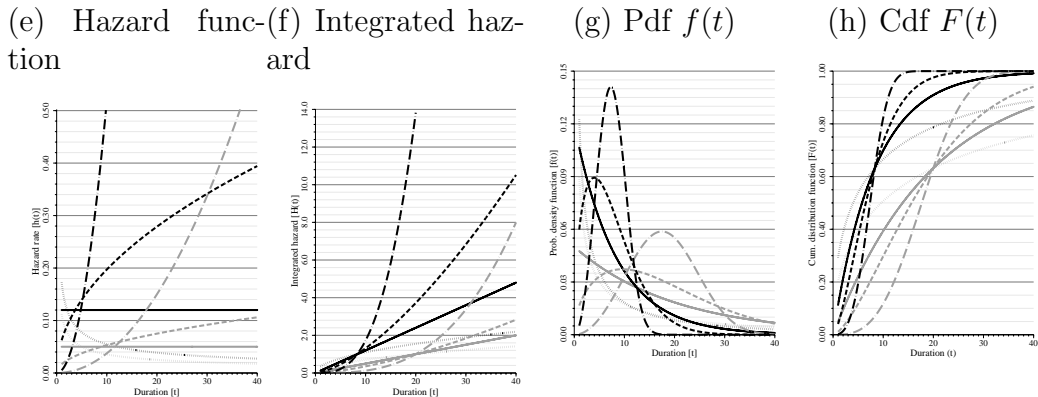
$$F(t) = 1 - (1 - \lambda)^t, \tag{22}$$

and:

$$f(t) = \lambda(1 - \lambda)^{t-1}. \tag{23}$$

FIGURE 3. EXAMPLES OF HAZARD FUNCTIONS: CONSTANT AND WEIBULL

Panel A. Constant Hazard

(a) Hazard function
(b) Integrated hazard
(c) Pdf $f(t)$
(d) Cdf $F(t)$

Panel B. Weibull Distribution

(e) Hazard function
(f) Integrated hazard
(g) Pdf $f(t)$
(h) Cdf $F(t)$

Note: Black: $\lambda = 0.12$; gray: $\lambda = 0.05$. Solid: $\alpha = 1$; dotted: $\alpha = 0.5$; dashed: $\alpha = 1.5$; dot-dashed: $\alpha = 3$. The examples in the top panel depict the hazard function (left), the integrated hazard (center-left), the pdf or probability mass function (center-right) and the cdf (right) of a constant hazard model with hazard equal to λ. The bottom panel depicts the corresponding functions for a Weibull hazard model with parameters λ and α.

Top panel in Figure 3 depicts the hazard rate, the integrated hazard, the cdf and the pdf of a constant hazard model. We can see that a constant hazard rate implies a decreasing unconditional probability of exiting and a marginally decreasing cdf. The figure also allows us to illustrate the fact that the integrated hazard does *not* have any interpretation as a conditional cdf, as it can take values above 1 as in this case.

The Weibull distribution Another hazard function that is commonly used for continuous durations is derived from a two-parameter generalization of the exponential distribution known as the Weibull distribution. This function allows for a hazard that increases or decreases monotonically.

7

The Weibull distribution is given by:

$$F(t) = 1 - e^{-(\lambda t)^\alpha} \quad \lambda > 0, \alpha > 0. \tag{24}$$

Hence, its pdf is:

$$f(t) = \frac{\partial F(t)}{\partial t} = \alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha}, \tag{25}$$

the hazard function is:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{\alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha}}{e^{-(\lambda t)^\alpha}} = \alpha \lambda^\alpha t^{\alpha-1}, \tag{26}$$

and the integrated hazard is:

$$H(t) = \int_0^t h(s)ds = (\lambda t)^\alpha. \tag{27}$$

The bottom panel of Figure 3 plots this function for different combinations of the two parameters. It can be seen that the function is very flexible in mimicking different shapes for the hazard function. Its main limitation, however, is that hazard rates are either monotonically increasing or monotonically decreasing (for instance, it cannot approximate the hazard function of mortality from Figure 2).

III.   Conditional Hazard Functions

This section discusses the introduction of covariates into the model. We need to write the hazard function conditional on covariates:

$$h(t, \boldsymbol{x}) = \frac{f(t|\boldsymbol{x})}{1 - F(t|\boldsymbol{x})}. \tag{28}$$

Below we review some of the most popular approaches for doing this.

A.   The proportional hazard model

The Proportional Hazard (PH) model (or Cox model, after Cox (1972)) is probably one of the most widely used duration models because of its simplicity. In this model, the conditional hazard function is given by:

$$h(t, \boldsymbol{x}) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}). \tag{29}$$

This is, we factorize $h(t, \boldsymbol{x})$ into a function of t and a function of $\boldsymbol{x}$, so that two different individuals have exit probabilities that are proportional for all t (hence the name). For instance, if an individual has twice the probability of another individual of exiting from unemployment in period t_1 conditional on survival, she also has twice the conditional probability of the other individual of exiting at t_2.

The function $\lambda(t)$ is called the *baseline* hazard function (as it provides the shape of the conditional hazard function, that then is scaled differently for every individual). The baseline hazard is often assumed to be given by one of the two possibilities described above. Note that if $\boldsymbol{x}'\boldsymbol{\beta}$ includes a constant term, then the scale of the corresponding hazard function needs to be normalized (e.g. $\lambda(t) = 1$ in the constant case, and $\lambda(t) = \alpha t^{\alpha-1}$ in the Weibull case).

B. Discrete durations

In discrete duration models, even though some times we might proceed with the standard PH model, it is common to select richer models. The reason is that the problem then reduces to a sequence of Probit or Logit estimations. In general, we can specify conditional hazard rates as follows:

$$h(\tau, \boldsymbol{x}) = \Pr(t = \tau | t \geq \tau, \boldsymbol{x}) = G(\gamma_\tau + \boldsymbol{x}'\boldsymbol{\beta}_\tau), \tag{30}$$

where $G(.)$ is a cdf (e.g. normal or logistic for Probit and Logit respectively). This model is richer than the previous one because it allows for different different $\boldsymbol{\beta}$s at different durations. The equivalent to the baseline hazard here, $\{\gamma_t\}_{t=0}^{T}$, where T is the maximum duration observed in the data, can be specified in many different ways. One possibility is to specify a polynomial like:

$$\gamma_t = \gamma_0 + \gamma_1 \ln t + \gamma_2 (\ln t)^2. \tag{31}$$

Another possibility is to leave these parameters free:

$$\gamma_t = \sum_{j=1}^{T^*} \gamma_j \mathbb{1}\{t = j\}. \tag{32}$$

Note that we only specify $T^* < T$ parameters instead of T. The reason for this is identification: on the one hand, we want T^* to be close enough to T, but, on the other hand, we need a critical mass of individuals being alive at T^* to be able to identify exit rates (e.g. if T is the period in which the last individual exits, then γ_T would be such that the probability of exiting at period T is equal to 1). This flexible option is very interesting, as it provides a semi-parametric estimation of $\lambda(t)$ and it allows to test other parametric assumptions.

IV. Likelihood Functions

A. Complete continuous durations

Assume that we observe $\{t_1, t_2, ..., t_N\}$. Then, the log-likelihood of this sample is:

$$\mathcal{L}_N = \sum_{i=1}^{N} \ln f(t_i | \boldsymbol{x}_i), \tag{33}$$

where:

$$f(t|\boldsymbol{x}) = h(t, \boldsymbol{x}) \exp(-H(t, \boldsymbol{x})) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}) \exp\left\{-\Lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta})\right\}. \tag{34}$$

B. Censored continuous durations

Duration data can be censored, i.e. we know that $t > \bar{t}$ or that $\underline{t} < t < \bar{t}$, but we do not observe the exact value of t. We allow the level of censoring to be observation specific (with the individual subindexes in the upper and lower limits). The contribution to the likelihood of an observation that is censored because we only observe that $t > \bar{t}$ is $\Pr(t \geq \bar{t}|\boldsymbol{x}) = 1 - F(\bar{t}|\boldsymbol{x})$. Hence, in this case, the log-likelihood boils down to:

$$\mathcal{L}_N = \sum_{i=1}^{N} \left\{ w_i \ln f(t_i|\boldsymbol{x}_i) + (1 - w_i) \ln(1 - F(\bar{t}_i|\boldsymbol{x}_i)) \right\}, \tag{35}$$

where $w_i = \mathbb{1}\{t_i < \bar{t}_i\}$, i.e. equals 1 if the observation is not censored. This is the a log-likelihood of a sample like the one generated in Example 1 of Figure 1.

Now consider the Example 2 of Figure 1. There we have observations with durations below one year $(\bar{t}^1)$, others with durations between one and two years $(\bar{t}^1 < t < \bar{t}^2)$, and others with durations above two years $(t > \bar{t}^2)$. In this case, the log-likelihood is:

$$\mathcal{L}_N = \sum_{i=1}^{N} \left\{ w_i^1 \ln F(\bar{t}_i^1|\boldsymbol{x}_i) + w_i^2 \ln \left(F(\bar{t}_i^2|\boldsymbol{x}_i) - F(\bar{t}_i^1|\boldsymbol{x}_i) \right) \right.$$
$$\left. + (1 - w_i^1 - w_i^2) \ln \left(1 - F(\bar{t}_i^2|\boldsymbol{x}_i) \right) \right\}, \tag{36}$$

where $w_i^1 = \mathbb{1}\{t_i < \bar{t}_i^1\}$, and $w_i^2 = \mathbb{1}\{\bar{t}_i^1 < t_i < \bar{t}_i^2\}$.

As a final example, consider a case like Example 2 of Figure 1 but in which at the starting point, individuals have not had 0 periods unemployed, but, instead, $d_1, d_2, ..., d_N$, with d_i known by the researcher. In this case, the log-likelihood

looks similar to (36) with the exception of conditioning on initial duration:

$$\mathcal{L}_{\mathrm{N}} = \sum_{i=1}^{N} \left\{ w_i^1 \ln \frac{F(d_i + \bar{t}_i^1 | \boldsymbol{x}_i) - F(d_i | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} + w_i^2 \ln \frac{F(d_i + \bar{t}_i^2 | \boldsymbol{x}_i) - F(d_i + \bar{t}_i^1 | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} \right.$$
$$\left. + (1 - w_i^1 - w_i^2) \ln \frac{1 - F(d_i + \bar{t}_i^2 | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} \right\}, \tag{37}$$

where $w_i^1 = \mathbb{1}\{d_i < t_i < d_i + \bar{t}_i^1\}$, and $w_i^2 = \mathbb{1}\{d_i + \bar{t}_i^1 < d_i + t_i < \bar{t}_i^2\}$.

C. Discrete durations

In the discrete duration model, we use a logistic or normal cdf (Logit or Probit) to estimate the hazard function. Also, we use the link between the probability of observing a given duration and the hazard rate seen in Section II. The log-likelihood is given by:

$$\mathcal{L}_{\mathrm{N}} = \sum_{i=1}^{N} \sum_{\tau=1}^{T^*} w_{i\tau} \{ y_{i\tau} \ln G(\gamma_\tau + \boldsymbol{x}_i' \boldsymbol{\beta}_\tau) + (1 - y_{i\tau}) \ln(1 - G(\gamma_\tau + \boldsymbol{x}_i' \boldsymbol{\beta}_\tau)) \} \tag{38}$$

where $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$, and $G(.)$ is a cdf (e.g. logistic or normal as before). This expression includes two types of contributions:

- Spells that end at time τ:

$$\ln \Pr(t = \tau | \boldsymbol{x}) = \ln h(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - h(s, \boldsymbol{x})). \tag{39}$$

- Spells that are incomplete at time T^*:

$$\ln \Pr(t > T^* | \boldsymbol{x}) = \sum_{s=1}^{T^*} \ln(1 - h(s, \boldsymbol{x})). \tag{40}$$

For every period $\tau = 1, ..., T^*$ we estimate a Probit or Logit of exiting vs not exiting conditional on survival (i.e. estimated on the sample of individuals still alive, which are those with $w_{it} = 1$). Given this, it becomes clear the importance of setting T^* small enough, so that we have enough observations alive to estimate the probit of exiting at period T^* with precision. We can set $T_i^*(= \min\{\bar{t}_i, T^*\})$ different for every individual, if there are observations censored below the maximum T^* that we are considering.

11

FIGURE 4. AN EXAMPLE WITH UNOBSERVED HETEROGENEITY
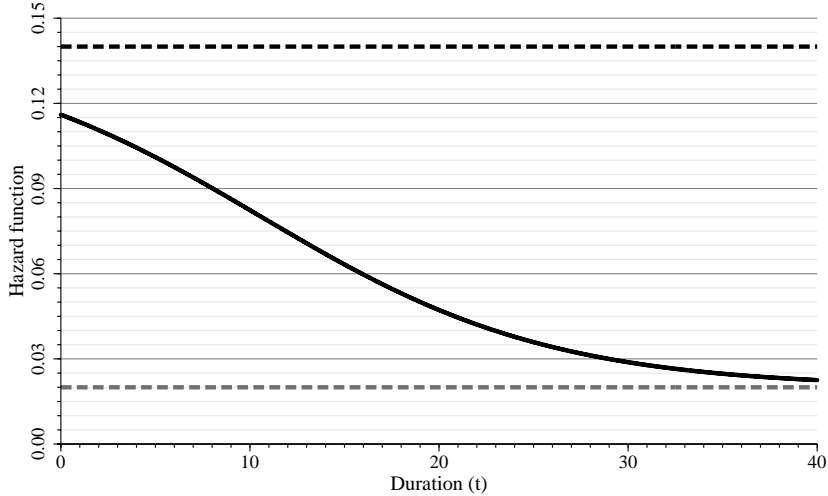
Note: Black dashed: h_1 (hazard rate when $x = 1$); gray dashed: h_0 (hazard rate when $x = 0$); gray dashed: observed (unconditional) hazard.

V. Unobserved Heterogeneity

A. *Unobserved heterogeneity vs spurious duration dependence*

Often, we cannot observe all important determinants of durations. The omission of important regressors can generate *spurious duration dependence*. To illustrate this idea, consider a regressor $x = \{0, 1\}$ (e.g. low vs high ability), and that the conditional hazard is well represented by a constant proportional hazard model in which:

$$h(t, x = 0) = h_0 \qquad h(t, x = 1) = h_1 \qquad h_1 > h_0. \tag{41}$$

Now assume that we do not observe x. The (unconditional) hazard that we identify is the following:

$$h(\tau) = h_1 \Pr(x = 1 | t \geq \tau) + h_0 \Pr(x = 0 | t \geq \tau). \tag{42}$$

The shape of this hazard is not constant anymore. Given that individuals with $x = 1$ have a higher hazard of exiting, the proportion of individuals with $x = 1$ is decreasing in the population, and the unconditional hazard converges to h_0.

Figure 4 is an example of this. In the figure, the conditional hazards are $h_1 = 0.14$ and $h_0 = 0.02$, and the initial fraction of individuals with $x = 1$ is 80%, with 20% having $x = 0$. As it emerges from the figure, after 40 periods, the unconditional hazard rate has completely converged to h_0, as all individuals with $x = 1$ exited, while there are still individuals with $x = 0$ remaining unemployed.

Hence, not being able to control for a covariate x (e.g. ability) that correlates with the hazard of exiting (e.g. high ability unemployed workers have a larger

12

hazard of exiting from unemployment) can create a spurious duration dependence. In our unemployment example, not controlling for ability leads to the conclusion that the hazard of finding a job decreases with the duration of the current unemployment spell when this hazard is indeed constant because an individual that has been unemployment for long is more likely to be of low ability.

B. Dealing with heterogeneity in continuous hazard models

Lancaster (1979) addressed this problem by introducing a multiplicative random effect in the proportional hazard specification:

$$h(t, \boldsymbol{x}, \nu) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta})\nu, \tag{43}$$

where ν is assumed independent of $\boldsymbol{x}$ with positive support, $\mathbb{E}[\nu] = 1$ and pdf $g(\nu)$. Hence, $h(t, \boldsymbol{x}, \nu)$ is a hazard function conditional on $\boldsymbol{x}$ and ν. The cdf conditional on $\boldsymbol{x}$ and ν is:

$$F(t|\boldsymbol{x}, \nu) = 1 - \exp\left(-\int_0^t h(u, \boldsymbol{x}, \nu)\right), \tag{44}$$

and the cdf of t given $\boldsymbol{x}$ only, based on which we write the integrated likelihood, is

$$F(t|\boldsymbol{x}) = \int_0^\infty F(t|\boldsymbol{x}, v)g(v)dv. \tag{45}$$

Lancaster assumed a Gamma distribution for $g(\nu)$. An important remark here is that we need to include regressors to be able to identify this model.

C. Dealing with heterogeneity in discrete hazard models

Similarly, in the discrete case, we can also write hazards conditional on ν and $\boldsymbol{x}$ and then integrate over ν:

$$\Pr(t = \tau|\boldsymbol{x}) = \int \Pr(t = \tau|\boldsymbol{x}, v)g(v)dv = \int h(\tau, \boldsymbol{x}, v)\prod_{s=1}^{\tau-1}[1 - h(s, \boldsymbol{x}, v)]g(v)dv, \tag{46}$$

where, for instance:

$$h(t, \boldsymbol{x}, \nu) = G(\gamma_t + \boldsymbol{x}'\beta_t + \nu). \tag{47}$$

A frequently used specification for $g(v)$ is a discrete-support mass point distribution: $\{\nu_1, ..., \nu_m\}$ with probabilities $\{p_1, .., p_m\}$:

$$\Pr(t = \tau|\boldsymbol{x}) = \sum_{j=1}^m \left\{ h(\tau, \boldsymbol{x}, \nu_j)\prod_{s=1}^{\tau-1}[1 - h(s, \boldsymbol{x}, \nu_j)]p_j \right\}, \tag{48}$$

where ν_j and p_j are additional parameters to be estimated.

VI. Multiple-exit discrete duration models

A. Discrete competing risk models

In this last section we discuss the case in which there are multiple exits from the current state. For instance, we consider a model in which individuals can exit from unemployment into a temporary or a permanent job.

Let duration t be discrete, and consider the two indicator functions d_j with $j = 1, 2$ that equal one if the exit is to alternative j. We can define the following intensities of transition to each state:

$$\phi_j(\tau) = \Pr(t = \tau, d_j = 1 | t \geq \tau), \quad j = 1, 2. \tag{49}$$

This expression has a direct link with the unconditional hazard rates:

$$h(\tau) = \Pr(t = \tau | t \geq \tau) = \phi_1(\tau) + \phi_2(\tau). \tag{50}$$

Conditional hazard rates are:

$$h_1(\tau) = \Pr(y_{1\tau} = 1 | t \geq \tau, y_{2\tau} = 0) \tag{51}$$

$$h_2(\tau) = \Pr(y_{2\tau} = 1 | t \geq \tau, y_{1\tau} = 0), \tag{52}$$

where $y_{j\tau} = \mathbb{1}\{t = \tau, d_j = 1\}$.

The mapping between intensities and conditional hazards is given by the definition of conditional expectation:

$$h_j(\tau) = \frac{\Pr(y_{j\tau} = 1 | t \geq \tau)}{\Pr(y_{k\tau} = 0 | t \geq \tau)} = \frac{\phi_j(t)}{1 - \phi_k(t)}, \tag{53}$$

where the numerator is the joint probability as $y_{1t} = 1$ implies $y_{2t} = 0$ and viceversa. Therefore, we can write the model in terms of either of the two. For instance, a MNL for ϕ's is equivalent to a binary Logit for h's with the same parameters.

Models presented in terms of conditional hazards $h_1(t)$ and $h(2(t)$ are also known as competing risk models. This name comes from considering two latent random variables t_1^* and t_2^* such that the observed duration is $t = \min\{t_1^*, t_2^*\}$. If t_1^* and t_2^* are independent, $h_1(t)$ and $h_2(t)$ can be interpreted as hazard rates of latent durations:

$$h_j(\tau) = \Pr(t_j^* = \tau | t_j^* \geq \tau). \tag{54}$$

This implies that the analysis of exits to 1 takes exits to 2 as censored observations.

B. Full information ML

The log-likelihood function is analogous to the discrete case:

$$\mathcal{L}_{\mathrm{N}} = \sum_{i=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau} \{ y_{i\tau} (d_{1i} \ln \phi_1(\tau, \boldsymbol{x}_i) + d_{2i} \ln \phi_2(\tau, \boldsymbol{x}_i))$$
$$+ (1 - y_{i\tau}) \ln(1 - \phi_1(\tau, \boldsymbol{x}_i) - \phi_2(\tau, \boldsymbol{x}_i)) \}, \qquad (55)$$

where $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$, as defined above. This expression includes three types of contributions:

- Spells that end at time τ exiting to option 1:

$$\ln \Pr(t = \tau, d_1 = 1 | \boldsymbol{x}) = \ln \phi_1(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})). \quad (56)$$

- Spells that end at time τ exiting to option 2:

$$\ln \Pr(t = \tau, d_2 = 1 | \boldsymbol{x}) = \ln \phi_2(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})). \quad (57)$$

- Spells that are incomplete at time T^*:

$$\ln \Pr(t > T^* | \boldsymbol{x}) = \sum_{s=1}^{T^*} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})). \qquad (58)$$

C. Limited information ML based on competing risk models

We can also estimate separately the model for each option, considering exits to the other option as censored, in a competing risk fashion. This is a LIML estimation. The likelihood for option j would be:

$$\mathcal{L}_{\mathrm{N}j} = \sum_{i=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau} \{ y_{ij\tau} \ln h_j(\tau, \boldsymbol{x}_i) + (1 - y_{ij\tau}) \ln(1 - h_j(\tau, \boldsymbol{x}_i)) \}, \qquad (59)$$

or, alternatively:

$$\mathcal{L}_{\mathrm{N}j} = \sum_{i=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau} \{ y_{i\tau} d_j \ln h_j(\tau, \boldsymbol{x}_i) + (1 - y_{i\tau} d_j) \ln(1 - h_j(\tau, \boldsymbol{x}_i)) \}, \qquad (60)$$

where $y_{ij\tau} = \mathbb{1}\{t_i = \tau, d_{ij} = 1\}$, $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$, with $t_i = \min\{t_{1i}^*, t_{2i}^*\}$, which delivers two types of contributions as in the standard discrete case (with the "censored" contributions being for both censored observations and exits to alternative $k \neq j$.

References

Cameron, A. Colin and Pravin K. Triverdi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press.

Cox, David R. (1972), "Regression Models and Life Tables (with Discussion)", *Journal of the Royal Statistical Society*, B, 34, 187-220.

Lancaster, Tony (1979), "Econometric Models for the Duration of Unemployment", *Econometrica*, 47, 939-956.

Lancaster, Tony (1990), *Econometric Analysis of Transition Data*, Cambridge.

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