Duration Models

by

Jean-Pierre Florens[*], **Denis Fougere**[**] **and Michel Mouchart**[***]

February 1995

Abstract

This work has been prepared as chapter 19 for the volume **Econometrics of Panel Data** (2nd ed., Mátyás and Sevestre eds., Kluwer Academic Publischers). The main emphasis is on modelling whereas inference problems are only sketched. After a short review on the basic issues on marginal models (i.e. models involving a duration variable only) conditional models are reviewed, namely the proportional hazard model and the accelerated time model. Problems of agregation and of heterogeneity are next examined. Competing risks models are presented with a particular emphasis on the identification of dependence among latent durations. Some examples of dependent durations are given. Finally, right-censoring is treated in the framework of competing risks models.

[*] IDEI, GREMAQ, Université des Sciences Sociales de Toulouse, Toulouse, France.
[**] CNRS, CREST-ENSAE, Paris, France.
[***] Institut de Statistique and CORE, Université Catholique de Louvain, Voie du Roman Pays 34, 1348 Louvain-la-Neuve, Belgium.

1 Introduction

During the last fifteen years, statistical duration models have been increasingly used by econometricians to analyze various economic problems. The first intensive application of these models was probably the analysis of individual unemployment (and then employment) duration data. Papers by *Lancaster* [1979], *Nickell* [1979], *Lancaster and Nickell* [1980], *Heckman and Borjas* [1980] and *Heckman and Singer* [1982], which were all devoted to this topic, played obviously a prominent part in the diffusion of such models among econometricians. Econometric duration models were also used in many other contexts, for example as:

- the analysis of time intervals between successive child–births *Newman and Mac Cullough* [1984]; *Heckman and Walker* [1990]);
- the duration of strikes (*Kennan* [1985]; *Cramton and Tracy* [1992]);
- the retirement process (*Butler, Anderson and Burkhauser* [1986], [1989]; *Stock and Wise* [1990]);
- the renewal of patents (*Pakes and Simpson* [1989]);
- the timing of equipment replacement (*Rust* [1987]);
- the analysis of business cycles (*Hamilton* [1989]; *Diebold and Rudebusch* (1990]);
- the timing of prepayment for mortgage–backed securities (*Schwarz and Torous* [1989]).

But the main application field of duration models still remains the analysis of unemployment duration data, as it is emphasized in surveys by *Heckman and Singer* [1984a], *Kiefer* [1988] or *Lancaster* [1990]. This is probably because, in this case, duration models can be naturally extended to more general processes, namely individual transition processes (see Chapter 29, this Volume), which register sequences of unemployment and employment durations. Such transition processes are formalized using the general theory of point processes (see Chapter 20, this Volume).

This chapter gives a general presentation of statistical models for single–spell duration data. It must be considered as an introduction to a hudge and always increasing statistical and econometric literature which can't be exhaustively reviewed in a relatively short paper. Consequently, this chapter puts more emphasis on modelization rather than on estimation issues.

Basically a duration measures the length of time spent by an individual or a firm in a given state, or equivalently, the length of time between two consecutive events or changes of state. Obviously, a duration is a non–negative random variable whose probability law is the main question of interest. In order to define unambiguously a duration, the time origin ($t = 0$) and the time scale (the time unit) should be defined carefully.

In fact, as we shall see in the beginning of the second section, the more elementary duration model is generated by a stochastic process X_{it} taking its value, at any time t and for any individual i, in a two elements discrete state space $\{E_0, E_1\}$ and experiencing at most one transition from state E_0 to state E_1. So, a duration model is generated by a doubly indexed variable, analogous, from this point of

view, to a panel variable.

The second section of this survey concentrates on marginal models of durations, *i.e.* models without explanatory variables. It presents the main functions characterizing the distribution of a duration variable, among which the more important ones are the survivor and hazard functions. A particularly useful concept is the one of temporal dependence of the hazard function: it can be handled with the help of some common distributions which are carefully presented here.

Section three is devoted to the presentation of conditional models of duration, and more particularly, proportional hazards and accelerated life models which include the effects of explanatory variables in two different ways. A special emphasis is put on the problem of unobserved individual heterogeneity in such models.

The more elementary duration model treats a single spell (of unemployment, for example) ending with a given kind of transition (from unemployment to employment, for example). But, in general, as a death could be due to various causes, an individual could exit from unemployment to enter one among different states: a stable job, an unstable one, the out–of–labor–force state or a training program, for example. When a single–spell duration has many (at least two) outcomes, the duration model may be modelled by means of a so–called competing risks model: those models are members of the class of models where a stochastic process X_t takes its value in the discrete state space $\{E_0, E_1, \dots E_K\},\ K > 1$, and experiences at most one transition from E_0 to one of the states $\{E_1, \dots E_K\}$. Such models are presented in the last section. The right–censoring issue is modelized here as a particular outcome (for instance, state E_K) of a competing risks duration model.

2 Marginal models

2.1 Introduction

A duration is a positive random variable, denoted T, representing the length of a time period spent by an individual or a firm in a given state. For simplicity, we shall assume that the distribution of T is not defective, *i.e.* $P(T = \infty) = 0$. This variable is also called a failure time when it measures the length of time before a state change in a system for which this change can occur at most once.

The more elementary duration models is a "death process" $\{X_t, t \in \mathbb{R}_+\}$, for which X_t takes its values in the discrete state space $\{E_0, E_1\}$; at the time origin, called the birth date, the process is in state E_0, *i.e.* $X_0 = E_0$. Trajectories of the process X_t have at most a unique transition from state E_0 to state E_1, which occurs at the time τ called the death date. Consequently, the duration T generated by a trajectory of the death process X_t is defined as follows (see Figure 1):

$$T = g(X) = \tau - 0 = \inf\{t \mid X_t = E_1\}.$$

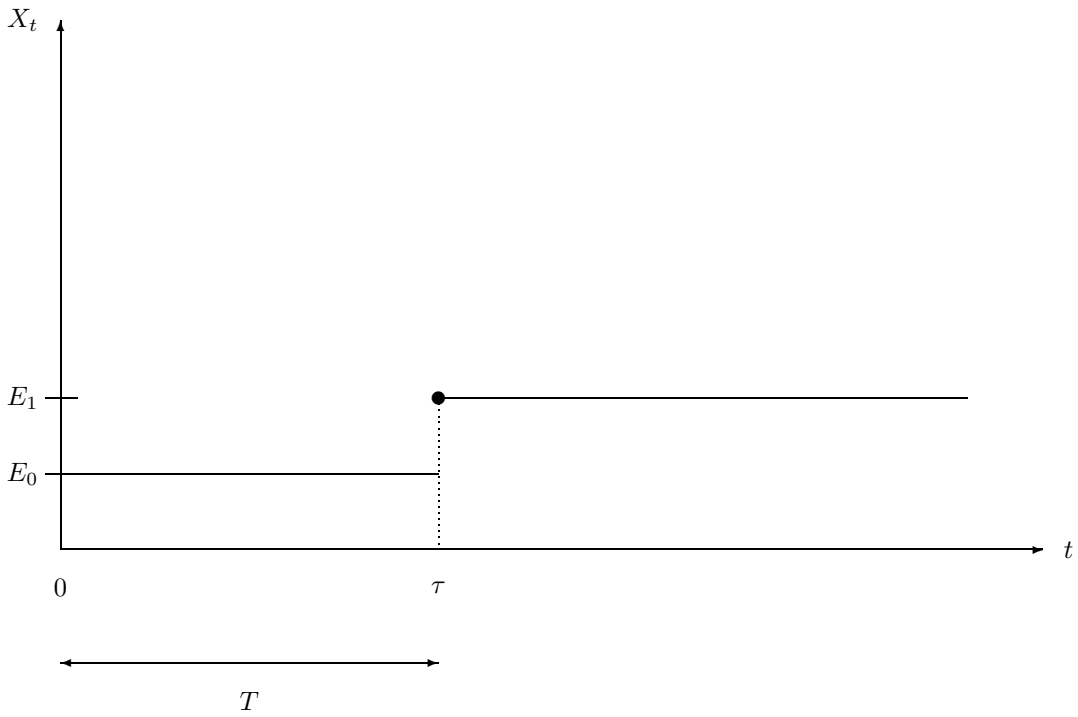

Figure 1: A typical trajectory of an elementary duration process

Thus, by definition:

$$\begin{aligned} X_t &= E_0, \quad t < \tau \\ &= E_1, \quad t \geq \tau, \end{aligned}$$

making therefore the trajectories of X_t cadlag, *i.e.* right–continuous with left limits (almost surely); in particular:

$$\begin{aligned} X_{\tau-} &= E_0 \\ X_{\tau+} &= X_\tau = E_1 \end{aligned}$$

Clearly, $P(T \geq 0) = 1$ and, in most models, we also assume that $P(T = 0) = 0$ (but see remark (i) in 2.2.d).

Often T is a continuous random variable, *but* the empirical distribution function is a discrete time process and non–parametric methods are often based on (functional) transformations of the empirical distribution function, considered as the best estimation of the "true" distribution function; this is, in particular, motivated by the Glivenko–Cantelli theorem. Therefore in this chapter we shall explicitly consider both continuous and discrete durations.

2.2 Distribution and Survivor Functions

a) General Definitions and Properties

a.1. Introduction

We first recall the general definition of the distribution function and of its complement, the survivor function. Next, we give more details for the continuous and the discrete cases, particularly from the point of view of the continuity of these functions.

a.2. Distribution function

Definition

The distribution function of the duration variable T is denoted F and is defined as:

$$F(t) = P(T \leq t), \quad t \geq 0.$$

■

This function gives the probability that a duration T is less than or equal to t. In the framework of a death process, it is the probability that an individual dies before age t. The characteristic properties of the distribution function F are:

(i) $F(t) \in [0,1], \quad \forall t \in I\!R_+$
(ii) F is right–continuous, *i.e.* $F(t_+) = F(t), \quad t \in I\!R_+$
(iii) F is monotone non–decreasing, *i.e.* : $t_1 > t_2 \Rightarrow F(t_1) \geq F(t_2)$,
(iv) $F(0_-) = 0$ and $\lim_{t\uparrow\infty} F(t) = 1$.

Here "characteristic" means that, to any function verifying those four properties, one may associate a probability measure on $I\!R_+$.

a.3. Survivor function

Definition

The survivor function of the duration variable T is denoted S and is defined as:

$$S(t) = P(T \geq t) = 1 - F(t) + P(T = t)$$

■

This function gives the probability that a duration T is greater than t. For a death process, this is the probability that an individual survives at age t. Its characteristic properties are:

(i) $S(t) \in [0,1], \quad t \in I\!R_+$,
(ii) S is left–continuous, *i.e.* $S(t_-) = S(t), \quad t \in I\!R_+$,
(iii) S is monotone decreasing, *i.e.* $t_1 > t_2 \Rightarrow S(t_1) \leq S(t_2)$,
(iv) $S(0) = 1$ and $\lim_{t\uparrow\infty} S(t) = 0$.

b) (Absolutely) Continuous Case

Definition

$$\begin{aligned}
\exists f : \mathbb{R}_+ \to \quad & \mathbb{R}_+ \text{ such that:} \quad F(t) = \int_0^t f(u)du, \\
& \text{or equivalently:} \quad S(t) = \int_t^\infty f(u)du = 1 - \int_0^t f(u)du, \\
& \text{or equivalently:} \quad f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.
\end{aligned}$$

The function $f(t)$ is called the *"density"* function of the duration variable T. Conversely,

$$f(t) = \lim_{\Delta \uparrow 0} \frac{P[t < T \leq t + \Delta]}{\Delta}$$

■

Thus, the density function may be interpreted as the "instantaneous probability" of failure, death or exit (from unemployment, for instance).

Remember that in the continuous case, there exists a value of t such that $F(t) = S(t) = 0.5$; that value is the median of the distribution.

c) Discrete Case

Definition

$$\begin{aligned}
& \exists (f_j, a_j), \quad j \in J \subseteq \mathbb{N}, \\
& f_j > 0, \quad \sum_{j \in J} f_j = 1 \\
& a_j \geq 0, \quad a_{j+1} > a_j \ldots
\end{aligned}$$

such that:

$$\begin{aligned}
F(t) &= \sum_{j \in J} f_j \mathbb{1}_{\{t \geq a_j\}} = \sum_{\{j | a_j \leq t\}} f_j, \\
S(t) &= \sum_{j \in J} f_j \mathbb{1}_{\{t \leq a_j\}} = \sum_{\{j | a_j \geq t\}}
\end{aligned}$$

or equivalently:

$$\begin{aligned}
f_j &= F(a_j) - F(a_j-) = F(a_j) - F(a_{j-1}), \\
&= S(a_j) - S(a_{j_-}) = S(a_j) - S(a_{j+1}).
\end{aligned}$$

Note in particular that:

$$F(a_j) = \sum_{\{i \leq j\}} f_i$$

$$S(a_j) = \sum_{\{i \geq j\}} f_i$$

■

In the framework of a death process, the event $\{T = a_j\}$ means "alive up to age a_{j-} and dead at age a_j" and that event has probability f_j.

d) Remarks

(i) Most models are such that: $P(T = 0) = 0$; this means that: $F(0) = F(0_-) = 0$ or equivalently, $S(0) = S(0_+) = 1$. The plausibility of such an assumption crucially depends on the specification of the time origin or on the definition of the state space. For instance, in a survey conducted to assess how students enter the labour market, the time origin (*i.e.* the calendar time for which $t = 0$) for a given individual can be specified to be the time of her (his) graduation. If T is the duration of unemployment immediately after leaving the school, the event $\{T = 0\}$ does realize for a student who has found a job before graduating and eventually starts working as early as at time $t = 0$. Modelling such a situation would require $P(T = 0) > 0$. If, however, T is defined as the duration of the unemployment of those who have failed to find a job before graduating, the time origin may be defined as the time of entrance in the state of unemployment making the property $P(T = 0) = 0$ acceptable. In other words, the specification whether or not $P(T = 0) = 0$ is essentially a matter of modelling a given situation; the usual distributions of duration display such a property and may therefore be interpreted as distributions conditional to $\{T > 0\}$, leaving the possibility of modelling apart the probability of the complementary event $\{T = 0\}$.

(ii) Defective distributions, *i.e.* distributions for which $P(T = \infty) > 0$, may be useful when modelling, for instance, individuals who never enter unemployment for statutory reasons or when analyzing some asymptotic properties. This survey however never treats those distributions.

2.3 Truncated Distributions and Hazard Functions

a) Motivations

(i) *Problem of time dependence.* Consider the question: what is the "instantaneous" probability of dying at time t given you are still living at time t_-? In other words, this is the "instantaneous" probability of leaving state E_0 at time t, given one has been in state E_0 from time 0 to t. More generally, this is the problem of the probability law of duration T, *conditionnally* to $T \geq t$ (Remember that the

event $\{T \geq t\}$ means "still alive at time t"). Such conditional distributions are "truncated" distributions.

(ii) The preceding question is often so "natural" that modelling those truncated distributions may be *economically more meaningful* than modelling the untruncated distributions. For instance, in job search models, the reservation wage, at a given instant, may be made a function of the duration of unemployment up to that instant.

(iii) *Censoring* (see section 4.2.) makes truncated distributions particularly useful.

b) Hazard Functions

b1) Integrated Hazard Function

Definition

The integrated hazard function of the duration variable T is denoted H and is defined as:

$$\begin{aligned} H: \quad & \mathbb{R}_+ \to \mathbb{R}_+ \\ & t \to H(t) = E\left[\frac{\mathbb{1}_{[o,t[}(T)}{S(T)}\right] = \int_{[0,t[} \frac{1}{S(u)} dF(u) \end{aligned}$$

■

Its main properties are the following:

(i) H is left–continuous, *i.e.* $H(t_-) = H(t), t \in \mathbb{R}_+$,
(ii) H is monotone non–decreasing, *i.e.*
$$t_1 < t_2 \Rightarrow H(t_1) \leq H(t_2)$$
(iii) $H(0) = 0, \quad H(\infty) = \infty$

As we shall see later, the integrated hazard function is a useful tool for characterizing some duration distributions.

b2) Hazard Function (or age–specific failure rate)

(i) Continuous Case

Remember that, in the continuous case, $S(u_+) = S(u)$ and there exists a density function $f(t)$ such that:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

Therefore:

$$H(t) = \int_0^t \frac{f(u)}{S(u)} du = -\int_0^t \frac{1}{S(u)} dS(u) = -ln\, S(t)$$

Definition

The hazard function of the duration variable T is denoted h and is defined as:

$$h(t) = dH(t)/dt = f(t)/S(t) = -d\, ln\, S(t)/dt$$

■

The function $h(t)$ may be viewed as the "instantaneous probability" of leaving the present state, indeed:

$$h(t) = \lim_{\Delta \downarrow 0} \frac{P[t < T \leq t + \Delta \mid T \geq t]}{\Delta}$$

Thus, $h(t)$ is also called the "age–specific failure rate" or the "age–specific death rate". The main properties of the hazard function h are:

(i) $h : I\!R_+ \rightarrow I\!R_+$

(ii) $\int_0^t h(u)du < \infty \quad \forall t \in I\!R$ but $\int_0^\infty h(u)du = \infty$

Note that h is *not* necessarily monotone. Relationships between the distribution, survivor and hazard functions can be deduced from the following equation:

$$H(t) = -ln\, S(t),$$

or equivalently

$$S(t) = exp\ -H(t).$$

Indeed, it is straightforward to show that:

$$\begin{aligned} S(t) &= exp\ -\int_0^t h(u)du, \\ H(t) &= \int_0^t h(u)du, \\ f(t) &= h(t)\ exp\ [-\int_0^t h(u)du], \\ F(t) &= 1 -\ exp\ \{-\int_0^t h(u)du\}. \end{aligned}$$

(ii) Discrete Case

Remember that, in the discrete case, for any (integrable) function $g(u)$ we have:

$$\int_{[0,t[} g(u)dF(u) = \sum_{\{j|a_j<t\}} g(a_j)f_j = \sum_j g(a_j)f_j\, I\!I_{\{a_j<t\}}$$

Therefore:

$$H(t) = \sum_{\{j|a_j<t\}} \frac{f_j}{S(a_j)} = \sum_{\{j|a_j<t\}} \frac{f_j}{f_j + f_{j+1} + \dots}$$

So, we obtain the discrete version of the (instantaneous) hazard function:

$$h_j = H(a_j+) - H(a_j) = \frac{f_j}{f_j + f_{j+1} + f_{j+2} + \dots} = \frac{f_j}{S(a_j)}$$

In particular: $h_1 = f_1$. The last formula may also be interpreted as follows :

$$h_j = P(T = a_j | T \geq a_j)$$

To deduce relationships between survivor and hazard functions in the discrete case, firstly we may write the survivor function as follows.

$$S(t) = \prod_{\{j|a_j<t\}} (1 - h_j)$$

Indeed, let us recall the familiar identity:

$$\begin{aligned}
a_0 + a_1 &= a_0(1 + \tfrac{a_1}{a_0}) \\
a_0 + a_1 + a_2 &= a_0(1 + \frac{a_1}{a_0})(1 + \frac{a_2}{a_0 + a_1}) \\
&\dots\dots \\
\sum_{0\leq j<k} a_j &= a_0 \prod_{1\leq j<k} (1 + \frac{a_j}{\sum_{0\leq m\leq j-1} a_m})
\end{aligned}$$

and apply to:

$$S(t) = 1 - \sum_{\{j|a_j<t\}} f_j$$

We also obtain the following relationships:

$$\begin{aligned}
ln\, S(t) &= \sum_{\{j|a_j<t\}} ln\,(1 - h_j) = \sum_j \mathbb{1}_{\{a_j<t\}} ln\,(1 - h_j) \\
&\approx - \sum_{\{j|a_j<t\}} h_j = -H(t) \\
&\qquad \text{(if } \quad h_j \text{ is "small", } \textit{i.e.} \quad - ln\,(1 - h_j) \approx h_j)
\end{aligned}$$

Thus, in the discrete case, $H(t)$ is *approximately* equal to $-ln\, S(t)$ if all h_j are small, while in the continuous case, $H(t)$ is exactly equal to $-ln\, S(t)$. Therefore, in the discrete case:

$$f_j = h_j \prod_{1\leq i\leq j-1} (1 - h_i)$$

c) Truncated Survivor Function

Definition

The truncated survivor function at time t_0 is given by:

$$S(t \mid t_0) = P(T \geq t + t_0 \mid T \geq t_0) = \frac{P(T \geq t + t_0)}{P(T \geq t_0)} = \frac{S(t + t_0)}{S(t_0)}$$

In the continuous case:

$$\begin{aligned} S(t|t_0) &= exp[H(t_0) - H(t + t_0)] \\ &= exp\left[-\int_{t_0}^{t+t_0} h(u)du\right] \end{aligned}$$

■

The main properties of the truncated survivor function are the following:

(i) $S(\cdot \mid \cdot) : I\!R_+ \times I\!R_+ \to [0, 1]$,
(ii) this function is monotone non–increasing in t for any t_0,
(iii) it is left–continuous in t for any t_0,
(iv) $S(0 \mid t_0) = 1, \quad \forall t_0$ and $\lim_{t \uparrow \infty} S(t \mid t_0) = 0, \quad \forall t_0$.

Note that $S(t \mid 0) = S(t), \quad \forall t \in I\!R_+$. Truncated survivor functions may be useful when dealing with grouped duration data, *i.e.* data for which we only know that the duration lies in a given interval. Consider indeed a given grid on the time axis:

$$0 < t_1 < t_2 < \ldots < t_I < \infty,$$

and define:

$$c_j = t_j - t_{j-1} \quad \text{with} \quad t_0 = 0 \quad \text{and} \quad j = 1, \ldots I.$$

Note that:

$$c_1 = t_1 \quad \text{and} \quad t_k = \Sigma_{1 \leq j \leq k} c_j.$$

Then we may write:

$$P[t_k \leq T < t_{k+1}] = [1 - S(c_{k+1} \mid t_k)][\Pi_{1 \leq j \leq k} S(c_j \mid t_{j-1})].$$

d) Truncated expected duration (expected residual life).

Definition

The truncated expected duration at time t is defined as:

$$\begin{aligned} r(t) &= E[T - t \mid T \geq t] = E[T \mid T \geq t] - t \\ &= \frac{1}{S(t)} \int_t^{\infty} S(u)du. \end{aligned}$$

■

Main properties of the function r are the following:

(i) $r : I\!R_+ \rightarrow I\!R_+$
(ii) $r(0) = E(T) > 0$.

The relationship between the survivor function and the truncated expected duration is given, in the continuous cases by:

$$S(t) = \frac{r(0)}{r(t)} \ exp \ \left[- \int_0^t \frac{1}{r(u)} du \right].$$

2.4 Some Useful Distributions for Duration Data

a) Introduction

(i) In principle, any distribution of a non–negative random variable may be used; in practice, the most useful ones are given below.

(ii) As a general remark, let us consider the following simple fact. If the random variable U has a distribution function denoted F_U with support $\{u \mid f_U(u) > 0\} = I\!R$, then the distribution function F_T of the random variable $T =$ exp (U) has for support $I\!R_+$. Furthermore, we have:

$$\begin{aligned} S_T(t) &= S_U(ln\, t), \\ f_T(t) &= (1/t) f_U(ln\, t), \\ h_T(t) &= (1/t) h_U(ln\, t), \\ H_T(t) &= -ln\, S_U(ln\, t) = H_U(ln\, t) \end{aligned}$$

(iii) In this section, we only consider continuous distributions. Remember that in such a case, the so–called *"integral transform theorem"* applies: the random variable defined by the distribution function of any continuous random variable is uniformally distributed on the interval $[0, 1]$ and the same is true for the random variable defined by its survivor function.

(iv) In most econometric applications, modelling is based on the shape of the hazard function. In principle any non–negative locally integrable function may be used as a hazard function, *i.e.* any function $h : I\!R_+ \rightarrow I\!R_+$ such that

$$\int_0^t h(u)du < \infty \qquad \forall t \in I\!R$$

but, for non defective distributions :

$$\lim_{t\uparrow\infty} \int_0^t h(u)du = \infty$$

Properties of basic interests are : is the hazard function constant, monotone increasing or decreasing, unimodal or multimodal ? In particular one has the following definition.

Definition

The hazard function of the duration T has the property of *temporal independence* if and only if it is constant over time, *i.e.* $h(t) = h$, $\forall\ t \in I\!R_+ (h > 0)$ ■

Thus we shall start by the simplest case a (short) catalogue of the mostly used distributions for duration date, namely the case of constant hazard function.

b) Exponential distribution

Definition

The duration T is exponentially distributed with parameter h $(h > 0)$ if its distribution is alternatively characterized by one of the following functions:

hazard function:	$h(t) = h, \quad t \in I\!R_+, \quad h > 0$,
integrated hazard function:	$H(t) = ht$,
survivor function:	$S(t) = \ exp\ (-ht)$,
density function:	$f(t) = h\ exp\ (-ht)$,
truncated expected duration:	$r(t) = h^{-1}$.

■

Let us note that

$$E(T) = r(0) = h^{-1}$$

and

$$V(T) = h^{-2} \quad \text{and} \quad \sigma(T) = \sqrt{V(T)} = h^{-1}.$$

This implies a unitary coefficient of variation :

$$\sigma(T)/E(T) = 1.$$

A useful property of the exponential distribution is that
cov $(T, \log T) = E(T)$. Note also that the integrated hazard function is linear in t. This eventually suggests a natural way of performing an empirical test of exponentiality by examining whether the plot of $(-ln\,\widehat{S}(t), t)$ is close to an homogenous

line, where $\widehat{S}(t)$ is the empirical (or a non–parametric estimate of the) survivor function

Remarks

1) When modelling durations, the exponential distribution has often a role similar to that of the normal distribution for real–valued random variables: it is a reference distribution. Thus, when specifying a model, if an assumption of exponentiality does not appear to be suitable, one may look either for a structural explanation or for a transformation of the time that would,for instance, render the (instantaneous) hazard function flat. As a matter of fact, it is easy to show, that the integrated hazard function $H(T)$ of any continuous random variable has an exponential distribution with unit parameter. Conversely, several of the distributions to be examined now may be obtained from simple transformations of an exponential variable. The points to be stressed are the following: firstly, those transformations are generally parametric, eventually introducing richer families than the exponential one; secondly, several of those families may also be viewed as embedding the exponential family into a larger one, a member of which being the exponential: this will provide easy tests for exponentiality with specific alternatives.

2) Later on we shall use the fact that the log of a unit exponential variable is distributed as a double–exponential distribution with density and survivor function :

$$f(w) = e^{w-e^w}$$
$$S(w) = e^{-e^w}$$

and moments

$$E(W) = -0,5722$$

$$V(W) = \frac{\pi^2}{6}$$

where $W = logY$ with $Y \sim Exp(1)$. Note that 0,5722 is known as Euler Constant.

c) Gamma distribution

Definition

The duration T has a gamma distribution with parameters α and $\beta(\alpha > 0,\ \beta > 0)$, and we denote $T \sim \gamma(\alpha, \beta)$, if and only if its distribution is characterized by one of the following functions:

hazard function: $$h(t) = \frac{t^{\beta-1}e^{-\alpha t}}{\int_t^\infty u^{\beta-1}e^{-\alpha u}du}$$

survivor function: $S(t) = 1 - I_\beta(\alpha t)$

density function: $f(t) = \dfrac{\alpha(\alpha t)^{\beta-1}e^{-\alpha t}}{\Gamma(\beta)} = \dfrac{\alpha^\beta t^{\beta-1}e^{-\alpha t}}{\Gamma(\beta)}$

where $\Gamma(\beta)$ and $\Gamma(. \mid \beta)$ (alternatively, $I_\beta(.)$) are the complete and the incomplete gamma functions, respectively defined as:

$$\Gamma(\beta) = \int_0^\infty u^{\beta-1}e^{-u}du$$

$$\Gamma(t \mid \beta) = \int_0^t u^{\beta-1}e^{-u}du = \Gamma(\beta)I_\beta(t)$$

■

With the transformation $v = u - t$, we obtain

$$\frac{1}{h(t)} = \int_0^\infty (1 + \frac{v}{t})^{\beta-1}e^{-\alpha v}dv,$$

from which we conclude that the gamma distribution is asymptotically exponential in the following sense:

$$\lim_{t\to\infty} h(t) = \alpha.$$

Furthermore, if $\beta > 1$ (respectively, if $\beta < 1$), $h(t)$ is monotone increasing with $h(0) = 0$ (respectively, monotone decreasing with $h(0) = \infty$). If $\beta = 1$, then $h(t) = \alpha,\ \forall t \in I\!R_+$. Mean and variance of T are:

$$E(T) = \beta/\alpha \quad \text{and} \quad V(T) = \beta/\alpha^2.$$

Note also that : $T \sim\ exp\ (h) \Leftrightarrow T \sim \gamma(h, 1), \quad h > 0$.

Practically, the distribution $\gamma(\alpha, \beta)$:

(i) is more flexible than the exponential distribution,
(ii) is asymptotically exponential as far as the hazard function is concerned,
(iii) has a monotone hazard function,

d) The Weibull distribution

Definition

The duration T has a Weibull distribution with parameters $\alpha > 0$ and $\beta > 0$, and we write $T \sim W(\alpha, \beta)$, if and only if $T^\beta \sim\ Exp(\alpha)$.

■

Consequently, the Weibull distribution is alternatively characterized by one of the following functions:

survivor function: $S(t) = exp\ -(\alpha t^{\beta})$,
hazard function: $h(t) = \alpha\beta t^{\beta-1}$,
integrated hazard function: $H(t) = \alpha t^{\beta}$,
density function: $f(t) = \alpha\beta t^{\beta-1}\ exp\ (-\alpha t^{\beta})$.

Noting that the first derivative of the hazard function is:

$$dh(t)/dt = \alpha\beta(\beta-1)t^{\beta-2};$$

it follows that the hazard function of the Weibull distribution is monotone incerasing (respectively, decreasing) if and only if $\beta > 1$ (respectively, $\beta < 1$).

Remarks

(i) Obsiously, $T \sim\ exp\ (h) \Leftrightarrow T \sim W(h,1),\ h > 0$.
(ii) Notice also that $ln\ [-ln\ S(t)] = ln\ \alpha + \beta ln\ t$. Thus, an empirical test for the Weibull distribution would be to check whether the plot $\{ln\ [-ln\ \widehat{S}(t)], ln\ t\}$ gives approximatively a straight line.

Practically, the Weibull distribution:

(i) displays simple analytical forms of $f(t), S(t)$ and $h(t)$, *but* is not a member of the exponential family for unknown β,
(ii) has some flexibility in the form of $h(t)$ which is nevertheless always monotone,
(iii) is not asymptotically exponential.

e) The log–normal distribution

The interest of this specification is to provide a unimodal (and non monotonic) hazard function.

Definition

The duration T has a log–normal distribution with parameters μ and σ^2, and we write $T \sim \Lambda(\mu, \sigma^2)$, if and only if $ln\ T \sim N(\mu, \sigma^2)$.

■

The log–normal distribution $\Lambda(\mu, \sigma^2)$ is alternatively characterized by any one of the following functions:

density function:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}}\ exp\ -\frac{1}{2\sigma^2}(ln\ t - \mu)^2$$

$$= \tfrac{1}{\sigma t}\varphi(\tfrac{ln\ t-\mu}{\sigma}),$$

survivor function:

$$S(t) = 1 - \Phi(\frac{ln\ t - \mu}{\sigma}),$$

hazard function:

$$h(t) = \frac{1}{\sigma t} \frac{\varphi(\frac{ln\, t-\mu}{\sigma})}{1-\Phi(\frac{ln\, t-\mu}{\sigma})},$$

where ϕ and Φ denote the density and distribution functions of a standardized normally distributed random variable, respectively.

Note that:

$$\lim_{t\to 0} h(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} h(t) = 0$$

This implies that the log–normal distribution has always a non–monotonic but unimodal hazard function.

f) The log–logistic distribution

Let us first *recall* the density (ψ) and the distribution (Ψ) functions of the standard logistic distribution $L(0,1)$. We write that:

$$X \sim L(0,1)$$

if and only if

$$\begin{aligned} \psi(x) &= \frac{e^{-x}}{(1+e^{-x})^2} \\ &= \Psi(t)[1-\Psi(t)], \\ \Psi(x) &= \frac{1}{1+e^{-x}}, \end{aligned}$$

Taking advantage of the stability of the logistic distribution for the linear transformations, the log–logistic distribution is defined similarly to the log–normal distribution.

Definition

The duration T has a log–logistic distribution with parameters μ and σ^2, and we write $T \sim LL(\mu, \sigma^2)$, if and only if $(ln\, T - \mu)/\sigma^2 \sim L(0,1)$. ■

The log–logistic distribution $LL(\mu, \sigma^2)$ is alternatively characterized by one of the following functions.

density function:

$$f(t) = \frac{1}{\sigma t}\psi(\frac{ln\, t-\mu}{\sigma}) = \frac{1}{\sigma t} \frac{exp\ -(\frac{ln\, t-\mu}{\sigma})}{\{1+\ exp\ -(\frac{ln\, t-\mu}{\sigma})\}^2},$$

survivor function:

$$S(t) = 1-\Psi(\frac{ln\, t-\mu}{\sigma}) = \frac{1}{1+\ exp\ (\frac{ln\, t-\mu}{\sigma})} = \frac{exp\ -(\frac{ln\, t-\mu}{\sigma})}{1+\ exp\ -(\frac{ln\, t-\mu}{\sigma})},$$

hazard function:

$$h(t) = \frac{1}{\sigma t}\frac{\psi(\frac{ln\, t-\mu}{\sigma})}{1-\Psi(\frac{ln\, t-\mu}{\sigma})} = \frac{1}{\sigma t}\Psi\left[\frac{ln\, t-\mu}{\sigma}\right]$$

$$= \frac{1}{\sigma t}\frac{1}{1+\ exp\ -(\frac{ln\, t-\mu}{\sigma})}.$$

It may be useful to reparametrize the log–logistic distribution as follows:

$$e^{-\frac{\mu}{\sigma}} = \alpha \ \text{ and } \ \frac{1}{\sigma} = \beta, \ \ i.e. \ \ \sigma = \frac{1}{\beta} \ \text{ and } \ \mu = -\frac{ln\, \alpha}{\beta}.$$

We then obtain:

$$f(t) = \beta\frac{\alpha t^{\beta-1}}{(\alpha t^{\beta}+1)^2},$$

$$S(t) = \frac{1}{\alpha t^{\beta}+1},$$

$$h(t) = \beta\frac{\alpha t^{\beta-1}}{\alpha t^{\beta}+1}.$$

We notice that for any value of α or β, we have $h(\infty) = 0$. Furthermore:

$$\begin{array}{lll} h(0) & = \alpha\beta & \text{if} \quad \beta = 1, \\ & = 0 & \text{if} \quad \beta > 1, \\ & = \infty & \text{if} \quad \beta < 1. \end{array}$$

The hazard function is monotone decreasing if $\beta \leq 1$; if $\beta > 1$, the hazard function is unimodal : firstly increasing and next decreasing.

g) Other distributions

Statistical analysis of duration sometimes rely on the use of other distributions, such as the Gompertz-Makeham, inverse gaussian or generalized F distributions. These distributions are extensively presented in textbooks by *Kalbfleisch and Prentice* [1980], *Lawless* [1982] or *Cox and Oakes* [1984].

2.5 Derived Distributions

a) Basic ideas

(i) From any of the distributions presented above, one may "derive" other distributions either by transforming the random variable (here, the duration) or

by transforming one of the functions characterizing those distributions. As the only requirement for a function to be used as a hazard function is to be non–negative locally integrable with non–converging integral and because the hazard function is often structurally meaningful, it may be conjectured that one natural way of transforming a distribution into another one is to transform its hazard function and derive therefrom the other characteristics of the transformed distribution.

(ii) In general, a so–called "baseline" duration T_0 is taken as a starting point along with the characteristics of its distribution, *i.e.* the functions $S_0(t \mid \alpha), f_0(t \mid \alpha), h_0(t \mid \alpha)$ and $H_0(t \mid \alpha)$. The derived distribution becomes the distribution of the duration T, characterized by the functions $S_T(t \mid \alpha), f_T(t \mid \alpha), h_T(t \mid \alpha)$ and $H_T(t \mid \alpha)$. In many applications, the baseline duration T_0 may be interpreted as the duration for individuals who have not been subjected to a "treatment" while the duration T is interpreted as the duration for those who have been subjected to a "treatment". In section 3, the "treatment" is typically the effect of exogenous variables.

(iii) In some cases, the baseline distribution is completely known, making α a known constant whereas, in other cases, α represents an unknown parameter.

b) Homethetic transformation of the Hazard Function

The simplest transformation of a non–negative function is the homothetic transformation. This produces a new model, the hazard function of which is constructed as follows:

$$h_T(t \mid \theta) = \beta h_0(t \mid \alpha)$$

where $h_0(t \mid \alpha)$ is a "baseline" hazard function with parameter α, and $\theta = (\alpha, \beta)$ is the new parameter with $\beta > 0$.

Equivalently, the proportional hazard model may also be characterized, in terms of the baseline distribution, as follows:

$$\begin{aligned}
H_T(t \mid \theta) &= \int_0^t h_T(u \mid \theta) du = \beta H_0(t \mid \alpha), \\
S_T(t \mid \theta) &= e^{-\beta H_0(t|\alpha)} = [S_0(t \mid \alpha)]^\beta, \\
f_T(t \mid \theta) &= h_T(t \mid \theta) S_T(t \mid \theta) \\
&= \beta h_0(t \mid \alpha)[S_0(t \mid \alpha)]^\beta \\
&= \beta f_0(t \mid \alpha)[S_0(t \mid \alpha)]^{\beta-1}.
\end{aligned}$$

c) Transformation of the Time

c1) A basic lemma

Let us consider a function $k : t \to k(t, \beta) = k_\beta(t)$ strictly increasing with t, *i.e.* $\frac{\partial}{\partial t} k(t, \beta) > 0$, and a "baseline" time T_0 such as:

$$T = k_\beta(T_0) \quad \text{or} \quad T_0 = k_\beta^{-1}(T).$$

Then

$$\begin{aligned}
S_T(t \mid \theta) &= S_0(k_\beta^{-1}(t) \mid \alpha), \\
f_T(t \mid \theta) &= \left[\frac{d}{dt} k_\beta^{-1}(t)\right] f_0\left(k_\beta^{-1}(t) \mid \alpha\right), \\
h_T(t \mid \theta) &= \left[\frac{d}{dt} k_\beta^{-1}(t)\right] f_0\left(k_\beta^{-1}(t) \mid \alpha\right) \{S_0(k_\beta^{-1}(t) \mid \alpha)\}^{-1} \\
&= \left[\frac{d}{dt} k_\beta^{-1}(t)\right] h_0\left(k_\beta^{-1}(t) \mid \alpha\right), \\
H_T(t \mid \theta) &= H_0(k_\beta^{-1}(t) \mid \alpha).
\end{aligned}$$

■

In particular, if $k_\beta(t_0) = H_0(T), T \sim Exp(1)$ for whatever baseline distribution.

Remark:

In this lemma, "strictly increasing" may be replaced by "strictly monotone", provided the derivative with respect to t is taken in absolute value.

c2) Exponential baseline time

In the particular case where the baseline time is exponentially distributed, *i.e.* :

$$h_0(t \mid \alpha) = \alpha$$

we have:

$$h_T(t \mid \theta) = \alpha \left[\frac{d}{dt} k_\beta^{-1}(t)\right]$$

Hence, the behaviour of $h_T(t \mid \theta)$ is completely determined by the behaviour of $k(t, \beta)$; in particular:

$$\left[\frac{d}{dt} h_T(t \mid \theta)\right] \cdot \left[\frac{d^2}{dt^2} k_\beta^{-1}(t)\right] > 0$$

Therefore, $h_T(t \mid \theta)$ is increasing when $k_\beta^{-1}(t)$ is convex and $h_T(t \mid \theta)$ is decreasing when $k_\beta^{-1}(t)$ is concave. This feature allows one to construct, in principle, an arbitrary hazard function.

c3) Scale transformation: accelerated time.

Let us consider the homothetic transformation:

$$T = \beta^{-1} \cdot T_0, \quad \text{where} \quad \beta > 0.$$

The transformed distribution can be characterized as follows:

$$\begin{array}{ll} S_T(t \mid \theta) & = S_0(t\beta \mid \alpha), \\ f_T(t \mid \theta) & = \beta f_0(t\beta \mid \alpha), \\ h_T(t \mid \theta) & = \beta h_0(t\beta \mid \alpha), \\ H_T(t \mid \theta) & = H_0(t\beta \mid \alpha). \end{array}$$

This transformation is also called an "accelerated time" model: the homothetic transformation may indeed be viewed as a change of scale of the time variable. If $\beta > 1$, the transformation of the model consists of "accelerating" the time whereas, if $\beta < 1$, it consists of "decelerating" the time.

c4) Power transformation

Let us consider the power transformation:

$$T = (T_0)^{-\beta}, \quad \text{where} \quad \beta > 0.$$

The transformed distribution can be characterized as follows:

$$\begin{array}{ll} S_T(t \mid \theta) & = S_0(t^\beta \mid \alpha), \\ F_T(t \mid \theta) & = \beta t^{\beta-1} f_0(t^\beta \mid \alpha), \\ h_T(t \mid \theta) & = \beta t^{\beta-1} h_0(t^\beta \mid \alpha), \\ H_T(t \mid \theta) & = H_0(t^\beta \mid \alpha). \end{array}$$

As seen earlier, if the distribution of T_0 is exponential, the distribution of T is Weibull. This is an example of the use of the power transformation.

3 Conditional Models

3.1 Introduction

a) Two levels of analysis can be considered

(i) For a *descriptive (or exploratory) data analysis*, covariates may be used to take into account observable factors of heterogeneity by realizing separate analyses. For example, let us consider a sample of one hundred unemployed workers and measure the following variables: the duration of their unemployment spell (T),

their gender (Z_1) which is a two level variable, their education level (Z_2) supposed to be a 3–level variable, and their qualification (Z_3) supposed to be a 5–level variable. For analyzing the distribution of T is this sample, we could make separate analyses (for instance, plotting histograms or conducting non–parametric tests) by considering the thirty (2 x 3 x 5) strata one by one. But it is clearly unsuitable with only 100 observations. The way variables (Z_1, Z_2, Z_3) do influence the distribution of T should therefore be parametrized.

(ii) When the analysis is concerned by *structural modelling*, the parameter of interest may be such that the (marginal) process generating some covariates may be uninformative for the parameter of interest which, at the same time, is a function of a parameter sufficient to parametrize the process conditional on those covariates. Those covariates are then called "exogenous variables" and will generally be denoted by "Z"whereas the other variables, denoted by "Y" (or "T", in case of a duration variable), will be called "endogenous", because the model describes the way they are generated conditionnally on the exogenous variable.

In such a case, it is admissible to specify only the process conditional on those exogenous variables, leaving the marginal process generating those exogenous variables virtually unspecified. In other words, for the parameter of interest, $p(t \mid z, \theta)$ is as informative as $p(t, z \mid \theta)$. According to a general principle of parcimony, the conditional model is therefore prefered.

b) How to specify conditional models?

(i) In general, a natural way of specifying conditional models is to make the parameters of a distribution dependent on the conditioning variable. Thus, in $F_T(t \mid \theta)$, one would transform θ into $g(z, \theta)$ where g would be a known function. For example, $Y \sim N(\mu, \sigma^2)$ could be transformed into $(Y \mid Z) \sim N(\alpha + \beta Z, \sigma^2)$. Similarly, $T \sim$ Exp (θ) could be transformed into $(T \mid Z) \sim$ Exp $[g(Z, \theta)]$ where, *e.g.,* $g(Z, \theta) = \exp(-Z'\theta)$.

(ii) When modelling individual data (and, in particular, duration data), a frequently used strategy consists of starting with a so–called "base line" distribution for the reference individual, *i.e.* either a "non–treatment" individual (*e.g.,*an individual for which $Z = 0$) or a "representative" individual (*e.g.,*an individual for which $Z = E(Z)$) and thereafter modelling, in the spirit of section 2.5., what makes the other individuals different from that individual of reference. Typical examples are the following:

- in the proportional hazard model, the effect of each regressor Z is to multiply the hazard function by a scale factor,
- in the accelerated life model, the effect of each regressor Z is to rescale the duration variable.

c) Time–varying and time–constant covariates must be distinguished

The covariates may represent:

• individual characteristics such as gender, level of education, and so on, which are fixed over time,

• other individual characteristics, such as marital status, number of children, eligibility to social benefits or programs, which are typically varying through time,

• but also characteristics of the macro–economic environment, unemployment rate, job vacancy rate, employment structure, and so on, which are also time–varying.

Some variables may also represent interactions between several covariates. The dynamic properties of the model and the estimation procedures crucially depends on whether the covariates are time dependent or not.

d) Interpretation of the parameters.

Most models are typically non linear in the sense that partial derivatives (of interest) are not constant, but are functions of the value of the covariates and/or of the duration. This feature clearly makes interpreting the coefficients more difficult. Often, those partial derivatives are, furthermore, not those of conditional expectations (as in regression analysis) but of hazard functions (*i.e.* of "instantaneous probabilities").

e) Parameters of interest and semi–parametric modelling

Let us reconsider the approach of "derived" distributions as seen in section 2.5, where a baseline hazard function $h_0(t \mid \alpha)$ is modified into $h_T(t \mid \theta)$ and where θ is an enlarged parameter: $\theta = (\alpha, \beta)$. In conditional models, $h_0(t \mid \alpha)$ is often interpreted as the hazard function of the duration for a "representative" individual, and it is therefore natural to modify β into $g(z, \beta)$ which captures the effect of the exogenous variables Z (or of the process generaling Z when Z is time–dependent) on the baseline distribution of the duration. In such a case, β is an economically meaningful parameter of finite dimension and of interest, whereas α is actually a nuisance parameter about which economic theory has not much to say. This may cause specification errors on $h_0(t \mid \alpha)$ as well as a scarcity of prior information on α. These are two motivations for developping an interest for semi–parametric modelling where the parameter space for α becomes a functional space such as, for instance, the set of non–negative continuous functions; in order to stress this feature, the baseline hazard function is then written as $\alpha(t)$ instead of $h_0(t \mid \alpha)$.

3.2 Proportional hazard or Cox Model

a) Definition

In the proportional hazard model, the effect of the exogenous variable is specified as multiplying a baseline hazard function by a function that depends on the exogenous variable only; it is accordingly defined as follows:

$$h_T(t \mid z, \theta) = h_0(t \mid \alpha) g(z, \beta), \quad \theta = (\alpha, \beta),$$

where $h_0(t \mid \alpha)$ is a baseline hazard function and g is a known function. Let us define the other characteristics of the baseline distribution along with their usual relationships:

$h_0(t \mid \alpha)$ is the baseline hazard function,
$S_0(t \mid \alpha)$ is the baseline survivor function,
$f_0(t \mid \alpha)$ is the baseline density function,
$H_0(t \mid \alpha)$ is the baseline integrated hazard function,

where, as usual:

$$h_0(t \mid \alpha) = \frac{f_0(t, \alpha)}{S_0(t, \alpha)},$$

$$H_0(t \mid \alpha = -ln\, S_0(t \mid \alpha) = \int_0^t h_0 \mid \alpha) du.$$

When Z is not time–dependent, the proportional hazard model is equivalently characterized as follows:

$$\begin{aligned}
H_T(t \mid z, \theta) &= g(z, \beta) \int_0^t h_0(u \mid \alpha) du = g(z, \beta) H_0(t \mid \alpha); \\
S_T(t \mid z, \theta) &= exp\, \{-g(z, \beta) \int_0^t h_0(u \mid \alpha) du\} \\
&= exp\, \{-g(z, \beta) H_0(t \mid \alpha)\} \\
&= [S_0(t \mid \alpha)]^{g(z, \beta)}, \\
f_T(t \mid z, \theta) &= h_T(t \mid z, \theta) \cdot S_T(t \mid z, \theta) \\
&= g(z, \beta) \cdot h_0(t \mid \alpha) [S_0(t \mid \alpha)]^{g(z, \beta)} \\
&= g(z, \beta) \cdot f_0(t \mid \alpha) [S_0(t \mid \alpha)]^{g(z, \beta)}.
\end{aligned}$$

b) Identification

The problem of identifying separately the functions g and h_0 raises from the simple remark that for any $k > 0 : g \cdot h_0 = gk \cdot k^{-1}h_0$. A rather natural solution consists of defining an individual of reference, *i.e.* a particular value z_0 of Z for which $g(z_0, \beta) = 1$ and, consequently, $h_T(t \mid z_0, \theta) = h_0(t \mid \alpha)$. Two typical normalizations are:

(i) $g(0, \beta) = 1$ when $Z = 0$ is meaningful,
(ii) $g(m_\gamma, \beta) = 1$ where $m_\gamma = E(Z \mid \gamma)$ is the mathematical expectation in the exogenous process generating $(Z \mid \gamma)$: in this case, $h_0(t \mid \alpha)$ is the hazard function of the "average individual".

c) A remark:

It is important to notice that, in the proportional hazard model, we have:

$$\frac{\partial}{\partial z} ln\, h_T(t \mid z, \theta) = \frac{\partial}{\partial z} ln\, g(z, \beta),$$

which depends on z and β only and is therefore independent of t if z is not time–dependent.

d) Semi–parametric modelling:

When interest is focused on the role of the exogenous variables, α is often a nuisance parameter and only β is of interest. In such a case, modelling often relies on one of the following two extreme possibilities:

(i) $h_0(t \mid \alpha)$ is specified in a most simplest way such as $h_0(t \mid \alpha) = h_0(t)$, *i.e.* is completely known, or $h_0(t \mid \alpha) = \alpha$, *i.e.* the baseline distribution is exponential and therefore depends on only one unknown parameter;

(ii) $h_0(t \mid \alpha)$ is specified in the most general way: $h_0(t \mid \alpha) = \alpha(t)$, *i.e.* a functional parameter (α is a non–negative function such that its integral on the positive real line diverges). This is a "semi–parametric model" with parameter $\theta = (\alpha, \beta)$ where α takes its value in a functional space, whereas β takes its value in a (finite dimensional) Euclidean space. This approach is particularly attractive in situations where economic theory would not give much information on the structure of $h_0(t \mid \alpha)$. Estimation of the semi–parametric Cox model has been initially treated by Cox [1972], [1975] and by Tsiatis [1981].

e) A particular case:

The function $g(z, \beta)$ should clearly be non–negative. An easy way to obtain that property without restriction on β is the log–linear specification, viz.:

$$g(z,\beta) = \ exp\ (z'\beta), \quad \beta \in I\!R^k.$$

That specification has a number of interesting properties:

$$\frac{\partial}{\partial z} ln\ h_T(t \mid z,\theta) = \frac{\partial}{\partial z} ln\ g(z,\beta) = \beta,$$

i.e. z has a constant proportional effect on the instantaneous conditional probability of leaving state E_0.

When z is not time–dependent, one may also write:

$$\begin{array}{ll} S_T(t \mid z,\theta) & = \ exp - \{H_0(t \mid \alpha)\ exp\ (z'\beta)\} = [S_0(t \mid \alpha)]^{exp(z'\beta)} \\ f_t(t \mid z,\theta) & = h_0(t \mid \alpha)\ exp\ - \{H_0(t \mid \alpha)\ exp\ (z'\beta)\}. \end{array}$$

Let us define:

$$\varepsilon_t = -ln\ H_0(t \mid \alpha) - z'\beta$$

Using the remarks at the end of 2.4.b), one may check that ε_t has a completely specified distribution, independent of α, z or β, namely a unit double–exponential distribution, and we may write:

$$-ln\ H_0(t \mid \alpha) = z'\beta + \varepsilon_t$$

This is a (non–normal) *non–linear* regression unless α is known. This feature of the proportional hazard model has been used by Han and Hausman [1990] for conducting a semi–parametric estimation of grouped duration data.

3.3 Accelerated life

a) Basic idea:

In the accelerated time model, the effect of the exogenous variable is specified as modifying the time scale, exactly as in section 2.5. For the ease of exposition, we assume that the exogenous variables are not time-dependent, the accelerated time model is accordingly defined as follows:

$$T = [g(z,\beta)]^{-1} T_0 \quad \text{or} \quad T_0 = g(z,\beta)T,$$

or, equivalently, :

$$\begin{array}{ll} h_T(t \mid z,\theta) & = g(z,\beta) \cdot h_0(t \cdot g(z,\beta) \mid \alpha) \\ H_T(t \mid z,\theta) & = H_0(t \cdot g(z,\beta) \mid \alpha) \\ S_T(t \mid z,\theta) & = S_0(t \cdot g(z,\beta) \mid \alpha) \\ f_T(t \mid z,\theta) & = g(z,\beta) \cdot f_0(t \cdot g(z,\beta) \mid \alpha) \end{array}$$

with, as usual, $\theta = (\alpha, \beta)$. This specification may be particularly attractive when the base–line distribution admits a scale parameter.

b) Empirical test for the accelerated time model

Let us consider the quantile functions, *i.e.* the inverse of the survivor (rather than, as more usually, the distribution) function:

$$\begin{array}{ll} q_T(p \mid z, \theta) & = S_T^{-1}(p \mid z, \theta), \quad 0 \leq p \leq 1, \\ q_0(p \mid \alpha) & = S_0^{-1}(p \mid \alpha), \quad 0 \leq p \leq 1. \end{array}$$

Because of the strict monotonicity (in the continuous case) of the survivor functions, we have:

$$q_0(p \mid \alpha) = g(z, \beta) \cdot q_T(p \mid z, \theta)$$

In the $(q_0(p \mid \alpha), q_T(p \mid z, \theta))$–space, this gives, for a fixed value of z, an homogenous straight line, the gradient of which is given by $g(z, \beta)$. This feature suggests that an easy empirical test for the accelerated time model may be obtained through an examination of the so–called "Q–Q–plot" (*i.e.* plot of the two quantiles) for a fixed value of Z and a fixed (typically, estimated) value of $\theta = (\alpha, \beta)$.

c) Regression representation of the accelerated time model

The accelerated time model may also be written, in logarithmic terms, as follows:

$$ln\, T = ln\, T_0 - ln\, g(z, \beta).$$

If we define:

$$\begin{array}{lll} \mu_0 & & = E[ln\, T_0] \\ \text{and} & \varepsilon & = ln\, T_0 - E[ln\, T_0], \end{array}$$

we may also write:

$$ln\, T = \mu_0 - ln\, g(z, \beta) + \varepsilon.$$

In particular:

(i) if $ln\, T_0 \sim N(\mu, \sigma^2)$, *i.e.* $T_0 \sim LN(\mu, \sigma^2)$, then $\varepsilon \sim N(\mu, \sigma^2)$. Thus we obtain a normal regression model (*if* there is no censoring; see subsection 4.2);

(ii) if $g(z, \beta) = exp\ (z'\beta)$, we obtain a linear regression model: $ln\, T = \mu_0 - z'\beta + \varepsilon$.

3.4 Aggregation and Heterogeneity

Aggregating over heterogenous individuals may create complicated structures of the hazard function. The analytical aspect is shown, for the general case, in next lemma. An example illustrates a simple application of this lemma. Next it is shown that, in particular, aggregation destroys the exponentiality of a duration.

A Basic lemma:

$$\begin{array}{llll} \text{Let} & (T \mid Z) \sim F_T^Z, & \textit{i.e.} & P(T \leq t \mid Z) = F_T^Z(t) \\ \text{and} & Z \quad\quad \sim F_Z, & \textit{i.e.} & P(Z \leq z) \quad = F_Z(z) \end{array}$$

Then

$$\begin{aligned} f_T(t) &= \int f_T(t \mid z) dF_z(z), \\ S_T(t) &= \int S_T(t \mid z) dF_z(z), \\ h_T(t) &= \frac{f_T(t)}{S_T(t)} = \frac{\int f_T(t \mid z) dF_z(z)}{\int S_T(t \mid z) dF_z(z)} \\ &= \int h_T(t \mid z) \cdot \frac{S_T(t \mid z)}{\int S_T(t \mid z) dF_z(z)} \cdot dF_z(z) \\ &= \int h_T(t|z) dF_Z(z|T \geq t) \end{aligned}$$

■

This lemma may be interpreted as follows: aggregating over heterogenous individuals, characterized by z, produce a duration distribution the hazard function $h_T(t)$ of which is a *weighted* average of the individual hazard functions $h_T(t \mid z)$. This possibly complicated weighting scheme may eventually account for complex hazard functions to be expected when analyzing aggregate data. A simple example illustrates this point.

Example.
Let $Z = 0$ for individuals with low educational level,
and $Z = 1$ for individuals with high educational level. The distribution of this variable over the whole population is defined by:

$$P(Z = z) = \theta^z (1 - \theta)^{1-z}.$$

Moreover, we suppose that:

$$(T \mid Z = j) \sim F_T^j \quad j = 0, 1$$

Then we can deduce:

$$
\begin{aligned}
f_T(t) &= \theta f_T(t \mid z=1) + (1-\theta) f_T(t \mid z=0), \\
S_T(t) &= \theta S_T(t \mid z=1) + (1-\theta) S_T(t \mid z=0), \\
h_T(t) &= \frac{f_T(t)}{S_T(t)} = \frac{\theta F_T(t \mid z=1) + (1-\theta) f_T(t \mid z=0)}{\theta S_T(t \mid z=1) + (1-\theta) S_T(t \mid z=0)} \\
&= \theta \frac{f_T^1(t)}{\theta S_T^1(t)(1-\theta) S_T^0(t)} + (1-\theta) \frac{f_T^0(t)}{\theta S_T^1(t) + (1-\theta) S_T^0(t)} \\
&= \theta \cdot \frac{f_T^1(t)}{S_T^1(t)} \cdot \frac{S_T^1(t)}{\theta S_T^1(t) + (1-\theta) S_T^0(t)} \\
&\quad + (1-\theta) \cdot \frac{f_T^0(t)}{S_T^0(t)} \cdot \frac{S_T^0(t)}{\theta S_T^1 + (1-\theta) S_T^0(t)} \\
&= h_T^1(t) \frac{\theta \cdot S_T^1(t)}{\theta S_T^1(t) + (1-\theta) S_T^0(t)} + h_T^0(t) \cdot \frac{(1-\theta) S_t^0(t)}{\theta S_T^1(t) + (1-\theta) S_t^0(t)}
\end{aligned}
$$

Lemma ("mover–stayer" lemma):

If $(T \mid Z) \sim$ Exp $(h_0(Z))$ and $Z \sim F_z$ arbitrary, then $h_T(t)$ is monotone decreasing. ■

Indeed, we successively obtain:

$$
\begin{aligned}
S_T(t) &= \int_0^\infty S_T(t \mid Z) dF_Z(Z) \\
&= \int_0^\infty exp\ \{-t h_0(Z)\} dF_Z(Z); \\
f_T(t) &= -\frac{d}{dt} S_T(t) = \int_0^\infty h_0(Z)\ exp\ \{-t h_0(Z)\} dF_Z(Z), \\
h_T(t) &= \frac{f_T(t)}{a S_T(t)} = \frac{\int_0^\infty h_0(Z)\ exp\ \{-t h_0(Z)\} dF_Z(Z)}{\int_0^\infty exp\ \{-t h_0(Z)\} dF_Z(Z)}
\end{aligned}
$$

It is then easy to check that:

$$\frac{d}{dt} h_T(t) < 0\ \forall t, \forall F_Z(Z),\ \forall h_0(Z)$$

(see for example, Fourgeaud, Gouriéroux and Pradel [1990]).

This lemma may be interpreted as follows. Individuals are characterized by their value of z. Large values of $h_0(Z)$ represent so–called "mover" individuals: they will leave first, while individuals represented by small value of $h_0(Z)$ are "stayer" individuals: they will leave (in probability) later. This explains why $h_T(t)$ will be decreasing because being determined at each t by the remaining individuals with decreasing values of $h_0(Z)$ This lemma also shows that although every individual has exponential duration, the aggregate population does not have exponential duration, except in the case where individuals would happen to be identical, a not–to–be–expected event.

Note that agregating over heterogenous individuals may be viewed as a modelling by means of mixtures of individual duration distributions weighted by the distribution of the individual characteristics. A crucial issue in mixture models concerns the identifiability of the components of the mixture from the observations of the mixed distribution. There is an important literature on this issue. Path-breaking papers are Elbers and Ridder (1982) and Heckmann and Singer (1982).

4 Competing risks and multivariate duration models

4.1 Multivariate duration

Multivariate distributions of durations (*i.e.* of positive random variables) are used in very different situations. A first use is in multivariate elementary point process ; this is the case when we observe simultaneously the length of life of several individuals or the first spell of unemployement of several individuals. This is also the case when, for a given individual, we define a multivariate point process corresponding, for instance, to its trajectories in the professional life and in the family life. Another use is in point processes with more than one transition as in the analysis of biographical data on unemployment. Still another use is in situations where the vector of durations is latent and some sampling scheme allows one to observe only partly the vector; this will be the case in competing risks models to be presented later on.

In this section we focus the attention on general issues :basic definition and properties, and methods of construction. For expository purposes we limit the presentation to bivariate distributions ; extensions to more than two dimensions are fairly obvious although, for some problems, notation may become cumbersome.

We start by the *multivariate survivor function* defined as follows:

$$S_{T_1,T_2}(t_1,t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$$

In what follows we assume S_{T_1,T_2} twice differentiable but in the last section we show how to treat continuous but not everywhere differentiable survivor function. The *multivariate density* is defined and denoted as:

$$f_{T_1,T_2}(t_1,T_2) = \frac{\partial^2}{\partial t_1 \partial t_2} S_{T_1,T_2}(t_1,t_2)$$

The *marginal survivor and density functions* are defined and denoted as:

$$\begin{aligned} S_{T_1}(t_1) &= S_{T_1,T_2}(t_1,0) \\ f_{T_1}(t_1) &= -\tfrac{d}{dt_1} S_{T_1}(t_1) \end{aligned}$$

and similarly for T_2. Often we shall write, for simplicity, $S_{1,2}, f_{1,2}$ or $S_j (j = 1,2)$ instead of S_{T_1,T_2} etc.

Conditional distributions occur in different contexts : different concepts are eventually useful and should be carefully distinguished according to the relevant conditioning. Thus we shall need both $S_{1|2}(t_1|T_2 = t_2), f_{1|2}(t_1|T_2 = t_2)$ and $S_{1|2}(t_1|T_2 \leq t_2), f_{1|2}(t_1|T_2 \leq t_2)$.

They are defined as follows:

$$\begin{aligned} S_{1|2}(t_1|T_2 \geq t_2) &= P(T_1 \geq t_1|T_2 \geq t_2) \\ &= \frac{S_{1|2}(t_1,t_2)}{S_2(t_2)} \\ f_{1|2}(t_1|T_2 \geq t_2) &= -\frac{\partial}{\partial t_1} S_{1|2}(t_1|T_2 \geq t_2) \\ &= \frac{-\frac{\partial}{\partial} \partial t_1 S_{1,2}(t_1,t_2)}{S_2(t_2)} \end{aligned}$$

Furthermore:

$$\begin{aligned} S_{1|2}(t_1|T_2 = t_2) &= P(T_1 \geq t_1|T_2 = t_2) \\ &= -\frac{\frac{\partial}{\partial t_2} S_{1,2}(t_1,t_2)}{f_2(t_2)} \\ f_{1|2}(t_1|T_2 = t_2) &= -\frac{\partial}{\partial t_1} S_{1|2}(t_1|T_2 = t_2) \\ &= \frac{f_{1,2}(t_1,t_2)}{f_2(t_2)} \end{aligned}$$

To everyone of thes univariate conditional distributions, there corresponds a respective hazard function. Marginal hazard functions are defined as:

$$
\begin{aligned}
h_j(t_j) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P\left[t_j \leq T_j \leq t_j + \Delta\right] | T_j \geq t_j \big] \\
&= -\frac{d ln\, S_j(t)}{dt_j} = \frac{f_j(t_j)}{S_j(t_j)}
\end{aligned}
$$

Conditional hazard functions are respectively defined as:

$$
\begin{aligned}
h_{1|2}(t_1|T_2 = t_2) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P\left[t_1 \leq T_1 \leq t_1 + \Delta\right] | T_1 \geq t, T_2 = t_2 \big] \\
&= \frac{f_{1|2}(t_1|T_2 = t_2)}{S_{1|2}(t_1|T_2 = t_2} \\
&= -\frac{f_{1,2}(t_1, t_2)}{\dfrac{\partial}{\partial t_2} S_{1,2}(t_1, t_2)} \\
&= -\frac{\partial}{\partial t_1}\left[ln\, - \frac{\partial}{\partial t_2} S_{1,2}(t_1, t_2)\right] \\
h_{1|2}(t_1|T_2 \geq t_2) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P\left[t_1 \leq T_1 \leq t_1 + \Delta\right] | T_1 \geq t, T_2 \geq t_2 \big] \\
&= \frac{f_{1|2}(t_1|T_2 \geq t_2)}{S_{1|2}(t_1|T_2 \geq t_2)} \\
&= -\frac{\partial}{\partial t_1} ln\, S_{1,2}(t_1, t_2)
\end{aligned}
$$

Several techniques for constructing multivariate distributions are worth mentioning. The most trivial one is the case of independent components in which case the joint survivor or density functions are the products of (arbitrary) corresponding marginal functions and the conditional survivor, density and hazard functions coïncide with the corresponding marginal functions.

For the dependent case, two general procedures are (i) take two univariate distributions, choose one to be marginal and take the other one to be conditional to the firstly chosen by making its parameters to be function of the conditioning variable (ii) take a joint distribution with survivor $S(t_1, t_2, y)$ where y is an auxiliarity variable such that $S(t_1, t_2|y)$ is economically meaningful and marginalize into $S_{1,2}(t_1, t_2)$.

When modelling duration data those general procedures are often based on a proportional hazard type of argument. Indeed, if $S_2(t_2)$ and $h_{1|2}\ (t_1|T_2 = t_2)$ are specified, one may deduce the conditional density as follows:

$$f_{1|2}(t_1|T_2 = t_2) = h_{1|2}(t_1|T_2 = t_2)exp - \int_0^{t_1} h_{1|2}(u|T_2 = t_2)du$$

whereas if $S_2(t_2)$ and $h_{1|2}(t_1|T_2 \geq 2)$ are specified, the joint survivor function may be constructed from:

$$- \int_0^{t_1} h_{1|2}(u|T_2 \geq t_2)du - ln\ S_{1,2}(t_1, t_2) = ln\ S_2(t_2)$$

Thus a first procedure would consist of taking two (arbitrary) unvariate distributions, choose, say, $S_2(t_2)$ to be marginal and specify the conditional hazard function as:

$$h_{1|2}(t_1|T_2 = t_2) = g(t_2)h_1(t_1)$$

where g is some non-negative value and $h_1(t_1)$ is the hazard function corresponding to $S_1(t_1)$. Another procedure consists to specifying that conditionally on some non-negative Y, T_1 and T_2 are independent and that given Y their hazard functions would take the form:

$$h_{j|Y}(t_j) = Yh_j(t_j)$$

In this case, Y is typically unobservable and may possibly specify an heterogeneity effect. The last subsection of this section will illustrate some of these procedures.

4.2 Competing risks models: definitions

We have seen in Section 2 that the most elementary duration model is the "death" process in which a duration of sojourn in state E_0 initially occupied at time $t_0 = 0$ is defined as the time interval between t_0 and the first unique transition from state E_0 to state E_1.

Competing risks duration models apply to situations where the state space has more than two elements, namely to situations where it has the form $\{E_0, E_1, \ldots E_J\}$, $J > 1$. Such models involve specifying not only "when" the process leaves the state E_0 where it stands initially, but also "which" state among the states $\{E_1, \ldots E_J\}$ will be entered.

Consider, for instance, a medical trial where a patient is submitted to a "treatment" for a supposedly known disease and where the survival time is observed. Typically, the cause of death is multiple; in particular, it may be different from the disease for which the treatment was originally designed and is possibly associated to the treatment itself: one says that several risks "compete" to cause the death of the patient. Similarly, in the labour market, when the initial state, E_0, is unemployment, it may be relevant to distinguish more than two states, for example: fully employed, part–time employed, early retired, unemployed for one among several reasons (without any job, temporarily laid–off, unemployed for health reasons,

etc.). The relevance of these distinctions is based on the fact that economic, social and legal factors may be most important to explain both durations and transitions of the individual trajectories; in other words, they are particularly important when analyzing biographical data.

Competing risk models are typically based on the following idea. For simplicity, we consider a process X_t with a unique transition, but with a state space involving more than two states. Let T be the duration that the process remains in the initial state E_0 untill it leaves for one of the "competing" states E_j with $1 \leq j \leq J, T_j$ be the duration (of sojourn in the initial state) if E_j were the only state alternative to the initial one and K be the index of the arrival state. In the competing risk models, these random variables are connected by the following relationships:

$$\begin{array}{rcl} T & = & \min_j\{T_j\}, \quad j = 1, \cdots J, \\ K & = & \arg\min_j\{T_j\}. \end{array}$$

Typically, the T_j's are latent duration variables, because only the minimum is observed. This structure permits to write easily the marginal laws of T and K:

$$\begin{array}{ll} S_T(t) & = P(T \geq t) = P(\cap_j\{T_j \geq t\}), \\ P(K & = k) = P(T_k \leq T_j, \quad 1 \leq j \leq J). \end{array}$$

Specific independence assumptions clearly make their joint distribution easily tractable. In particular, if the T_j's are assumed to be independent, we obtain:

$$S_T(t) = \Pi_j S_j(t),$$

where $S_j(t)$ stands for the survivor function of T_j. Similarly, $P(K = k)$ is easily obtained under assumptions similar to those familiar in the literature on discrete choice models. In the sequel, we will consider situations in which the T_j's may be dependent. For that purpose, we need to recall some preliminary definitions.

Similarly to the presentation of section 4.1, we start by the joint survivor function using * as an upper index to stress that those durations of latent:

$$S^*(t_1, \ldots t_J) = \ P(T_1 \geq t_1, \ldots, T_J \geq t_J),$$

for any $(t_1, \ldots t_J) \ \in I\!R_+^J$. The survivor function of the observed duration $T = \min_j(T_j)$ satisfies:

$$S_T(t) = S^*(t, \ldots t), \quad t \in I\!R^+$$

The marginal survivor function of the latent duration $(1 \leq j \leq J)$ is denoted S_j and defined as:

$$S_j(t_j) = S^*(0, \ldots 0, t_j, 0, \ldots 0)$$

In the case where the T_j's are independent, we have:

$$S^*(t_1, \ldots t_J) = \Pi_{j=1}^J S_j(t_j).$$

Now let us suppose that the functions S^*, and consequently S_T and S_j, are continuously differentiable. The marginal and relevant conditional hazard functions of the latent duration $T_j (1 \leq j \leq J)$ are denoted h_j and h_j^c respectively, and defined as:

$$\begin{aligned} h_j(t) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(t \leq T_j \leq t + \Delta \mid T_j \geq t) \\ &= -dln\, S_j(t)/dt, \quad t \in I\!R_+, \\ h_j^c(t) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(t \leq T_j \leq t + \Delta \mid T \geq t) \\ &= -\frac{\partial}{\partial t_j} ln\, S^*(t_1, \ldots t_J) \mid_{t_1 = t_2 = \cdots = t_J = t} \quad t \in I\!R_+. \end{aligned}$$

When the T_j's are mutually independent, it is obvious that:

$$h_j^c(t) = h_j(t), \qquad \text{for any } \; t \in I\!R_+.$$

The hazard function of the observed duration $T = \min_j T_j$ is denoted h_T and defined as:

$$\begin{aligned} h_T(t) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(t \leq T \leq t + \Delta \mid T \geq t) \\ &= -dln\, S_T(t)/dt \; t \in I\!R_+. \end{aligned}$$

Note first that

$$h_T(t) = \sum_{j=1}^{J} h_j^c(t)$$

because, in the definition of $h_T(t)$, the derivative of $S_T(t)$ is a directional derivative (in the direction of the main diagonal $(1, 1, \cdots 1)$) of $S^*(t_1 \cdots t_J)$. By definition, the generic contribution to the likelihood function may be evaluated as:

$$\begin{aligned} l_{T,K}(t,k) &= -\frac{d}{dt} P(T \geq t, K = k) \\ &= -\frac{d}{dt} P[T_j \geq T_k \geq t \quad \forall j] \end{aligned}$$

Remember that a basic result of differential calculus is:

$$S^*(t_1 \cdots t_J) = \int_{t_j}^{\infty} -\frac{\partial}{\partial t_j} S^*(t_1 \cdots t_J)\, dt_j$$

and, similarly, a basic result of conditional probability is:

$$S^*(t_1 \cdots t_J) = \int_{t_j}^{\infty} S_{\bar{j}|j}(t_1 \cdots, t_{j-1},\ t_{j+1} \cdots t_J | T_j = t_j)\ f_j(t_j)\ dt_j$$

where

$$\begin{array}{l} S_{\bar{j}|j}(t_1 \cdots, t_{j-1},\ t_{j+1}, \cdots, t_J | t_j) \\ = P\left(T_1 \geq t_1 \cdots T_{j-1} \geq t_{j-1},\ T_{j+1} \geq t_{j+1} \cdots T_J \geq t_J | T_j = t_j\right) \\ \bar{j} = \{1, 2, \cdots J\} \setminus \{j\} \end{array}$$

Therefore

$$S_{\bar{j}|j}(t_1 \cdots, t_{j-1},\ t_{j+1}, \cdots t_J | t_j) = -\frac{\frac{\partial}{\partial t_j} S^*(t_1 \cdots t_J)}{f_j(t_j)}$$

In the sequel we shall use the following simplified notation:

$$S_{\bar{j}|j}(t) = S_{\bar{j}|j}(t, t, \cdots t | t)$$

Therefore

$$\begin{aligned} l_{T,K}(t,k) &= -\frac{d}{dt}\int_t^{\infty} S_{\bar{k}|k}(u) f_k(u) du \\ &= S_{\bar{k}|k}(t) f_k(t) \\ &= -\frac{\partial}{\partial t_j} S^*(t, \cdots t) \\ &= -S^*(t, \cdots t)\frac{\partial}{\partial t_j} ln\ S^*(t, \cdots t) \\ &= h_k^c(t) S_T(t) \end{aligned}$$

Using a disjunctive coding of the exit, namely

$$A = (A_1 \cdots A_J) \quad A_j = \mathbb{1}_{\{K=j\}}$$

we may also write:

$$l_{T,A}(t,a) = \prod_{j=1}^{J} \left[f_j(t) S_{\bar{j}|j}(t)\right]^{a_j} = S_T(t) \prod_{j=1}^{J} \left[h_j^c(t)\right]^{a_j}$$

In case of independent latent durations, we have:

$$\begin{aligned} l_{T,K}(t,k) &= f_k(t) \prod_{j \neq k} S_j(t) \\ &= h_k(t) S_T(t) \end{aligned}$$

4.3 Identifiability of competing risks models

The basic idea of competing risks models is to interpret the data (T, K) on the duration in the initial state and the label of the exist state as the observation of a minimum of a random vector along with the coordinate where the minimum is obtained. Intuition suggests that the observations will give no information on the question whether the coordinate of the random vectors, *i.e.* the latent durations, are independent or not. This intuition is confirmed by next theorem

Theorem

Let

$$\mathcal{S} = \{S^*(t_1 \cdots t_J) : J - \text{dimensional continuously differentiable survivor function}\}$$

$$\mathcal{S}_I = \{S^* \in \mathcal{S} | S^*(t_1 \cdots t_J) = \textstyle\prod_j S_j(t_j)\}$$

$l_*(t,k)$: likelihood for a model in $\mathcal{S}$

$l_I(t,k)$: likelihood for a model in $\mathcal{S}_I$

Then : $\forall S^* \in \mathcal{S} \quad \exists!\ S_I \in \mathcal{S}_I$ such that $l_*(t,k) = l_I(t,k)$

In particular:

$$h^c_{j,*}(t) = h_{j,I}(t)$$

The proof of this theorem simply rests on the remark that in the general case $l_*(t,k) = h^c_k(t) S_T(t)$ and that $h_T(t) = \sum_j h^c_j(t)$, *i.e.* the distribution of the observed duration depends only on the sum of the conditional hazard functions; therefore the equality $h^c_{k,*}(t) = h_{k,I}(t)$ ensures the equality of likelihood functions.

This theorem means that to any competing risks model with dependent latent duration one way associate an observationally equivalent model with independent latent durations; the rule of association is simply to build the joint latent distribution with marginal hazard functions, of the independent model, equal to the conditional hazard functions of the dependent case. To illustrate this point, we can consider the following bivariate example. Suppose that the joint survivor function of the two latent durations (T_1, T_2) is given by:

$$S^*(t_1, t_2) = \ exp\ \{1 - \alpha_1 t_1 - \alpha_2 t_2 - \ exp\ [\alpha_{12}(\alpha_1 t_1 + \alpha_2 t_2)]\}$$

where $\alpha_1, \alpha_2 > 0$ and $\alpha_{12} > -1$. Here the parameter α_{12} measures the dependence between the two latent durations T_1 and T_2. The conditional and marginal hazard functions of this model are respectively:

$$h^c_j(t) = \alpha_j\{1 + \alpha_{12}\ exp\ [\alpha_{12}(\alpha_1 + \alpha_2)t]\},\ j = 1, 2,$$

and

$$h_j(t) = \alpha_j[1 + \alpha_{12}\ exp\ (\alpha_j \alpha_{12} t)], \quad j = 1, 2.$$

Marginal survivor functions are then:

$$S_j(t_j) = \ exp\ [1 - \alpha_j t_j - \ exp\ (\alpha_{12}\alpha_j t_j)], \quad j = 1, 2,$$

from which it is obvious that:

$$S^*(t_1, t_2) \neq S_1(t_1)S_2(t_2), \quad (t_1, t_2) \in I\!R^2_+,$$

except if $\alpha_{12} = 0$.

The likelihood element of an observation (t, k) may be written as:

$$\begin{aligned} l_{S^*}(t,k) = \alpha_k\{1 + \alpha_{12}\ exp\ [\alpha_{12}(\alpha_1 + \alpha_2)t]\} \\ exp\ \{1 - (\alpha_1 + \alpha_2)t - \ exp\ [\alpha_{12}(\alpha_1 + \alpha_2)t]\}, \end{aligned}$$

The observationally equivalent model (*i.e.* with same likelihood function) with independent latent durations has marginal hazard functions given by $h^c_j(t)$ above and eventually marginal and joint survivor functions of latent durations given by:

$$S_{j,I}(t_j) = exp\left\{ \frac{\alpha_j}{\alpha_1 + \alpha_j} - \alpha_j t_j - \frac{\alpha_j}{\alpha_1 + \alpha_2} exp\ \alpha_{12}(\alpha_1 + \alpha_2)t_j \right\} \ j = 1, 2$$

$$S_I(t_1, t_2) = exp(1 - \alpha_1 t_1 - \alpha_2 t_2 - \frac{1}{\alpha_1 + \alpha_2}\big[\alpha_1 exp\ \alpha_{12}(\alpha_1 + \alpha_2)t_1$$

$$+\alpha_2 exp\ \alpha_{12}(\alpha_1 + \alpha_2)t_2\big]$$

Note that the latent models are clearly different unless $\alpha_{12} = 0$, *i.e.* $S^*(t_1, t_2) \neq S_I(t_1, t_2)$, but the statistical models are observationally equivalent, *i.e.* $l_*(t, k) = l_I(t, k)$. Note also that both latent models have been identifiably parametrized but the parameters have very different meaning in the two latent models; in particular α_{12} measures the association among the latent variables in the case of dependence whereas α_{12} is a common parameter of the two marginal distributions in the case of independent latent variables.

4.4 Right–censoring

One usual feature of duration data is that the sampling scheme often produces right–censored observations, *i.e.* observations which have not yet exited from the initial state E_0 at the end of the sampling period. For example, in the case of single spell unemployment duration data, the sampling scheme is often the following: individual observations are sampled from the flow of individuals entering unemployment at time t_0 and followed up until date $C(> t_0)$, which is possibly determined by the analyst. Some observations correspond to individuals exiting from unemployment before C, in which case they generate complete unemployment durations. Other sampled individuals have not exited from unemployment at date C and so generate right–censored unemployment durations. Rather than sampling

from the flow into unemployment at a given date t_0, the analyst may sample from inflows considered at many staggered dates $t_0^1, t_0^2 \ldots$ and follow up observations once again up to a censoring time C.

Right–censoring can be modelized using the framework of the competing risks models with state space $\{E_0, E_1, \ldots E_K\}, K > 1$. For example, let us suppose that last state E_K denotes the right–censored situation: in this case, E_K is associated to a latent variable T_K which represents the right–censoring time. To illustrate this kind of formalization, let us consider a bivariate competing risks model (T_1, T_2) with state space $\{E_0, E_1, E_2\}$, E_0 labeling unemployment, E_1 employment and E_2 right–censoring. Thus $T_2 = C$.

Suppose first that all individual observations are sampled from a same time t_0. Without loss of generality, one may write (after due time translation) $t_0 = 0$. From the previous section, the model is specified from a latent survivor function $S_{i,2}(t_1, t_2|\theta)$ with parameter θ and a generic element of the likelihood function may be written as

$$l_{T,A}(t, a) = \left[f_1(t|\theta)S_{2|1}(t|\theta)\right]^a \left[f_2(t|\theta)S_{1|2}(t|\theta)\right]^{1-a}$$

where $A = \mathbb{1}_{\{T_1 \leq T_2\}}$. In view of the problem of identification and because it so happens that many censoring mechanisms act independently of the unemployement process, it is often assumed that T_1 and T_2 are independent.
Then

$$l_{T,A}(t, a) = \left[f_1(t|\theta)S_2(t|\theta)\right]^a \left[f_2(t|\theta)S_1(t|\theta)\right]^{1-a}$$

If, moreover, θ may be factorized into $\theta = (\theta_1, \theta_2)$ such that θ_1 characterizes the distribution of T_1 and θ_2 the distribution of T_2 the likelihood reduces to:

$$l_{T,A}(t, a) = L_1(\theta_1)L_2(\theta_2)$$

where

$$L_1(\theta_1) = f_1(t|\theta_1)^a S_1(t|\theta_1)^{1-a}$$

The parameters of interest are in general those of the distribution of duration T_1, and their estimation could be deduced from $L_1(\theta_1)$ only. Then the generic element of the relevant factor of the likelihood function is $f_1(t|\theta_1)$ (resp. $S_1(t|\theta_1)$ for the censored (resp. uncensored) observation.

Another model generating censored data may be the following. Let T_0 be the age of an individual entering unemployment. This age is randomly generated. The duration of the unemployment spell is T_1 and the age at the end of unemployment is then $T_0 + T_1$. The econometric model specifies the joint distribution of (T_0, T_1) and these two random variables are not, in general, assumed to be independent. A natural specification could be a sequential one: the (marginal) distribution of T_0 is first specified and a conditional distribution of T_1 given T_0 completes the model.

Let us now assume that all the individuals are observed at a same date T_*. In general this date is also random but, for simplicity, we consider T_* as fixed (the

model is conditional to T_*). Let us also assume that the sample is constructed in such a way that $T_0 \leq T_*$ (all the individuals have entered unemployment). Then T_0 is always observed but T_1 is not censored if $T_0 + T_1 \leq T_*$. Otherwise, the duration of unemployment is censored.

Let us define $T_2 = T_* - T_0$. From the distribution of (T_0, T_1) we obtain a distribution of (T_1, T_2) and we may consider the observations as generated by a censored duration model: T_1 is observed only if $T_1 \leq T_2$. But a specification of a likelihood based on the generic element:

$$l_{T,A}(t, a) = f_1(t)^a S_1(t)^{1-a}$$

where $T = \min(T_1, T_2)$, $A = \mathbb{1}(T_1 \leq T_2), f_1$ and S_1 are the density and the survivor functions of T_1, is incorrect for two reasons:

i) First if T_0 and T_1 are dependent, T_1 and T_2 are also dependent and the likelihood fonction must considrer the joint distribution.

ii) The censoring mechanism is different from the usual competing risks model because T_0 or T_2 is always observed and the likelihood of the actual data must be the density o (T_2, T, A). The generic element of this likelihood is the following:

$$l_{T_2,T,A}(t_2, t, a) = f_2(t_2) f_{1|2}(t|T_2 = t_2)^a \; S_{1|2}(t|T_2 = t_2)^{1-a}$$

using our previous notation.

Finally note that the identification result of section 4.2 does not applied in this case because the censoring mechanism is different from the competing risks model.

4.5 Dependent bivariate duration distributions

This section aims at presenting some commonly used dependent duration distributions. For expository purposes we stick to bivariate distributions. We underline their main properties which make them appropriate to particular econometric problems, with two different kinds of use in mind : either as distributions for multivariate observable durations, a theme presented in section 4.1, or as distributions for latent durations in dependent competing risks models.

a) Marshall-Olkin class of distributions

Marshall and Olkin (1967) introduced the following bivariate distribution whose particularity is to incorporate the possibility of equality with positive probability. Such a model could be appropriate, for instance, for situations where two strikes may have a same duration. The joint survivor function is given by:

$$S_{T_1,T_2}(t_1, t_2) = exp - \{\lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 \max(t_1, t_2)\}$$

Note that this function is everywhere continuous, is not continously differentiable on the main diagonal $\{T_1 = T_2\}$ but is directionally differentiable on the main diagonal.

One way of obtaining such a distribution would be to define $T_j = \min(X_j, X_3)$ $j = 1, 2$ with X_i independently exponentially distributed with parameters $\lambda_i (i = 1, 2, 3)$. The marginal survivor functions are:

$$S_{T_j}(t_j) = exp - (\lambda_j + \lambda_3)t_j$$

Thus T_1 and T_2 are independent only if $\lambda_3 = 0$. Furthermore:

$$\begin{aligned} P(T_1 < T_2) &= \frac{\lambda_1}{\lambda_+} \\ P(T_2 < T_1) &= \frac{\lambda_2}{\lambda_+} \end{aligned}$$

where $\lambda_+ = \lambda_1 + \lambda_2 + \lambda_3$; therefore

$$P(T_1 = T_2) = \frac{\lambda_3}{\lambda_+}$$

Thus $P(T_1 = T_2) > 0$ as soon as $\lambda_3 \neq 0$. Suppose now that (T_1, T_2) is not completely observable as is the case, for instance, when (T_1, T_2) represents latent durations. Let us define:

$$\begin{aligned} T &= \min(T_1, T_2) \\ A_1 &= \mathbb{1}_{\{T_2 < T_1\}} \\ A_2 &= \mathbb{1}_{\{T_1 = T_2\}} \end{aligned}$$

Suppose first that we observe (T, A_1, A_2). The generic element of the likelihood function is

$$\begin{aligned} l_{T,A_1,A_2}(t, a_1, a_2) = & \left[\frac{\lambda_1}{\lambda_+}(\lambda_1 + \lambda_3)exp - (\lambda_1 + \lambda_3)t\right]^{a_1(1-a_2)} \\ & \left[\frac{\lambda_2}{\lambda_+}(\lambda_2 + \lambda_3)exp - (\lambda_2 + \lambda_3)t\right]^{a_2 a_1} \\ & [\lambda_3 exp - \lambda_+ t]^{(1-a_1)a_2} \end{aligned}$$

Suppose now that we only observe (T, A_1) as in a competing risks model where ties are not observed but are taken into account, because of their strictly positive probability. Thus we now assume that the exit state is E_1 in case $T_1 = T_2$. In such a case, the generic element of the likelihood function becomes:

$$l_{T,A_1}(t_1, a_1)$$

$$= \left\{ \frac{1}{\lambda_+} [\lambda_1(\lambda_1 + \lambda_3)exp - (\lambda_1 + \lambda_3)t + \lambda_3\lambda_+ exp - \lambda_+ t] \right\}^{1-a_1}$$

$$\cdot \left\{ \frac{\lambda_2(\lambda_2 + \lambda_3)}{\lambda_+} exp - (\lambda_2 + \lambda_3)t \right\}^{a_1}$$

This provides an identified model of dependent competing risks. Note that in some cases one could observe T only if $T_1 \neq T_2$ *i.e.* the data would be $(A_1, A_2, (1 - A_2)T)$. Another possibility is that in case $T_1 \neq T_2$, the exit state is randomized. Thus, when ties may occur with strictly positive probability, the likelihood function crucially depends on a precise specification of the sampling scheme.

Marshall and Olkin [1967] mention the possibility of obtaining other similar bivariate distributions through transformations of variables T_1 and T_2. If $T_1' = g_1(T_1)$ and $T_2' = g_2(T_2)$ are monotonic functions, say increasing, with respective inverse functions g_1^{-1} and g_2^{-1}, then:

$$\begin{aligned} S^*_{T_1', T_2')}(t_1', t_2') &= P\ (T_1' > t_1', T_2' > t_2') \\ &= exp\ -[\lambda_1 g_1^{-1}(t_1') + \lambda_2 g_2^{-1}(t_2') \\ &\qquad +\lambda_3 \max\ (g_1^{-1}(t_1'), g_1^{-1}(t_2'))]. \end{aligned}$$

Note that $g_1(0) \geq 0$ and $g_2(0) \geq 0$ are also required if T_1' and T_2' are to be non–negative random variables. The particular case $T_1' = (T_1)^{1/c_1}$ and $T_2' = (T_2)^{1/c_2}$ leads to a bivariate Weibull distribution.

b) Gamma mixture models

These models, introduced by Clayton [1978], are basically proportional hazards models with an unobserved heterogeneity component, denoted Y, whose distribution is gamma with parameters δ and $\theta(\delta > 0, \theta > 0)$ and with density:

$$f_Y(y) = \theta^\delta \ \ y^{\delta - 1} \quad exp(-\theta y)/\Gamma(\delta).$$

Once again, we choose the bivariate case (T_1, T_2) for simplicity. Conditionally on Y, T_1 and T_2 are supposed to be independent with hazard functions:

$$h_j(t \mid Y) = Y\ \lambda_j(t), \quad \lambda_j(t) > 0, \ \ t \in I\!R^+ \quad j = 1, 2.$$

where the $\lambda_j(t)$'s have the properties of hazard functions. The joint survivor function conditional on Y is accordingly:

$$S^*(t_1, t_2 \mid Y) = \ exp[-Y\{\Lambda_1(t_1) + \Lambda_2(t_2)\}],$$

where

$$\Lambda_j(t) = \int_0^t \lambda_j(u)du, \quad j = 1, 2.$$

Because Y is typically unobserved, it must be integrated out to give the unconditional joint survivor function:

$$S^*(t_1, t_2) = [1 + \theta^{-1}\{\Lambda_1(t_1) + \Lambda_2(t_2)\}]^{-\delta},$$

implying that the marginal survivor functions have the form:

$$S_j(t) = \{1 + \theta^{-1}\Lambda_j(t)\}^{-\delta} \quad j = 1, 2.$$

The relation between the unconditional joint survivor function and the unconditional marginal survivor functions is then:

$$S^*(t_1, t_2) = \{S_1(t_1)^{-1/\delta} + S_2(t_2)^{-1/\delta} - 1\}^{-\delta}.$$

Independence is obtained in the limiting case, $\delta \to \infty$; when δ is finite, there is positive dependence between T_1 and T_2. Conditional hazard functions of T_1 given $T_2 > t_2$ or $T_2 = t_2$ can be obtained from S^* respectively as:

$$\begin{aligned} h_{1|2}(t_1 \mid T_2 \geq t_2) &= -\frac{d\ ln\ \ S^*(t_1, t_2)}{dt_1} \\ &= \delta\{\theta + \Lambda_1(t_1) + \Lambda_2(t_2)\}^{-1}\lambda_1(t_1), \\ h_{1|2}(t_1 \mid T_2 = t_2) &= -\frac{d\ ln\ \ \left(-\dfrac{dS^*(t_1, t_2)}{dt_2}\right)}{dt_1} \\ &= (\delta + 1)\{\theta + \Lambda_1(t_1) + \Lambda_2(t_2)\}^{-1}\lambda_1(t_1). \end{aligned}$$

Note that these two conditional hazard functions are proportional:

$$\frac{h_{1|2}(t_1 \mid T_2 = t_2)}{h_{1|2}(t_1 \mid T_2 \geq t_2)} = 1 + \delta^{-1}.$$

This relation is still valid when T_1 and T_2 are interchanged.

Covariates can be incorporated into the model using a proportional hazard specification (see Clayton and Cuzick [1985]. In this case, marginal hazard functions given Y and a vector Z_j of time–independent covariates can be written:

$$h_j(t \mid Z_j, Y) = Y\ exp\ (\beta' Z_j)\ \lambda_0(t),$$

where β is a vector of parameters to be estimated.

c) Positive stable mixture models

These models, introduced by Hougaard [1986], differ from the previous case only through the distribution of Y which is assumed to be a positive stable distribution of index $\alpha \in (0,1]$, rather than a gamma distribution. A positive stable distribution with parameter α, say $P(\alpha)$, is characterized by its Laplace transform:

$$E\ exp\ (-sY) = \ exp\ -s^{\alpha}, \quad \alpha \in (0,1].$$

For $\alpha = 1$, this is degenerate at 1, and for $\alpha = 1/2$, this a special case of the inverse gaussian distribution.

When Y is integrated out, one obtains the joint survivor function:

$$S^*(t_1,t_2) = \ exp[-\{\Lambda_1(t_1) + \Lambda_2(t_2)\}^{\alpha}].$$

Marginal survivor functions can be deduced as:

$$S_j(t) = \ exp\ [-\{\Lambda_j(t)\}^{\alpha}], \quad j = 1,2,$$

which implies that:

$$S^*(t_1,t_2) = \ exp\ [-\{-ln\ S_1(t_1)]^{1/\alpha} + [-ln\ S_2(t_2)]^{1/\alpha}\}^{\alpha}].$$

The latent durations T_1 and T_2 are independent if $\alpha = 1$. The conditional hazard functions of T_1 given $T_2 \geq t_2$ or $T_2 = t_2$ are given by:

$$h_{1|2}(t_1 \mid T_2 \geq t_2) = \alpha\{\Lambda_1(t_1) + \Lambda_2(t_2)\}^{\alpha-1}\ \lambda_1(t_1),$$

and

$$\begin{aligned} h_{1|2}(t_1 \mid T_2 = t_2) \quad &= \lambda_1(t_1)[(1-\alpha)\{\Lambda_1(t_1) + \Lambda_2(t_2)\}^{-1} \\ &\quad +\alpha\{\Lambda_1(t_1) + \Lambda_2(t_2)\}^{\alpha-1}]\ \lambda_1(t_1). \end{aligned}$$

As an example, let us consider the case where the $\lambda_j(t)$ are derived from Weibull distributions with a common scale parameter:

$$\lambda_j(t) = \beta_j \gamma t^{\gamma-1} \quad , \beta_j > 0, \gamma > 0, j = 1,2,$$

and, as before,:

$$h_j(t \mid Y) = Y\ \ \lambda_j(t).$$

In that case, the joint survivor function becomes:

$$S^*(t_1,t_2) = \ exp\ \{-[\beta_1 t_1^{\gamma} + \beta_2 t_2^{\gamma}]^{\alpha}\}.$$

The marginal distribution of T_j, $j = 1, 2$, is Weibull $(\alpha, \gamma, \beta_j^\alpha)$, *i.e.*:

$$S_j(t) = \ exp \ -[\beta_j^\alpha t^{\alpha\gamma}] \quad j = 1, 2.$$

The observed duration, $T = \ \min\ (T_1, T_2)$, is also Weibull distributed with parameters $\{\alpha\gamma, (\beta_1 + \beta_2)^\alpha\}$. Then it is easy to show that:

$$P \quad \{\arg\min_{1,2}(T_1, T_2) = j\} = \int_0^\infty h_j(t) \quad S_T(t)dt$$

$$= \beta_j/(\beta_1 + \beta_2).$$

Only three parameter functions can be identified: $\alpha\gamma, (\beta_1 + \beta_2)^\alpha$ and $\beta_1/(\beta_1 + \beta_2)$. The latent durations T_1 and T_2 are independent if $\alpha = 1$, as in the preceding general case, but here a nice result about the dependence is that the correlation between $\log T_1$ and $\log T_2$ is $(1 - \alpha^2)$: this dependence is necessarily positive, as in the gamma mixture model.

Conditional moments of the bivariate Weibull model are given by:

$$E(T_j^\gamma \mid T_{3-j}) = (\alpha\beta_j)^{-1} \ \beta_{3-j}^{1-\alpha} \ \ T_{3-j}^{\gamma(1-\alpha)}, \ j = 1, 2.$$

General results on conditional moments given that $T_{3-j} > t_{3-j}$ are described by Hougaard [1986]. A bivariate exponential model is obtained by restricting $\gamma = 1/\alpha$. In this case, both the marginal distributions and the distribution of the minimum are exponential, as in the Marshall and Olkin model, but here the joint survivor function is everywhere continuously differentiable.

Let us return to the general model. Covariates can be included in this model as they are is the gamma mixture model, by setting:

$$\lambda_j(t \mid Z_j) = \lambda_0(t) \quad exp(\beta' Z_j),$$

which gives directly the marginal hazard function of T_j as:

$$h_j(t \mid Z_j) = \ exp\ (\alpha\beta' Z_j)\ \omega_0(t),$$

where $\omega_0(t) = \alpha\lambda_0(t)\Lambda_0(t)^{\alpha-1}$. The marginal model is still a proportional hazard model and it is then possible to identify $(\alpha\beta)$ and $\omega_0(t)$ from the marginal distribution alone.

4.6 Conclusions

This chapter has focused on statistical models for duration data. Special emphasis has been put on characterization of duration distributions through hazard functions, definition of the two broad classes of duration models (proportional and accelerated hazards models) and presentation of the more commonly used parametric distributions for single–spell (univariate) and dependent competing risks duration models. Non–parametric, semi–parametric or fully parametric estimation of these models has been extensively treated in textbooks by Kalbfleisch and Prentice [1980], Lawless [1982], Cox and Oakes [1984] or Lancaster [1990].

To conclude, let us recall that identification of duration models with unobserved individual heterogeneity has been object of interest by econometriciams through the last decade. A survey on main researchs concerning this topic (Elbers and Ridder [1982], Heckman and Singer [1984 b], Heckman and Honoré [1989], Ridder [1990], Honoré [1992]) is contained in a recent paper by Heckman [1991], extending substantially what has been sketched above.

References

Buttler, J.S., Anderson, K.H., and Buhkauser, R.V. [1986]: "Testing the relationship between work and health: a bivariate hazard model" *Econommic Letters*, 20.

Clayton, D. [1978]: "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence", *Biometrika,* 65, 141–151.

Clayton, D. & Cuzick, J. [1985]: "Multivariate generalizations of the proportional hazards model (with discussion)", *Journal of the Royal Statistical Society, Series A.,* 148, 82–117.

Cox, D.R. [1972]: "Regression models and life tables (with discussion)", *Journal of the Royal Statistical Society, series B,* 34, 187–220.

Cox, D.R. [1975]: "Partial likelihood", *Biometrika,* 62, 269–276.

Cox, D.R. and Oakes, D. [1984]: *Analysis of Survival Data.* London: Chapman and Hall.

Cramton, P.C. and Tracy, J.S. [1992]: "Strikes and holdouts in wage bargaining: theory and data", *American Economic Review,* 82, 100–121.

Crowder, M. [1985]: "A distributional model for repeated failure time measurements", *Journal of the Royal Statistical Society, series, B,* 47, 447–452.

Diebold, F.X. and Rudebusch, G.D. [1990]: "A nonparametric investigation of duration dependence in the american business cycle", *Journal of Political Economy,* 98, 596–616.

Elbers, C. & Ridder, G. [1982]: "True and spurious dependence: the identifiability of the proportional hazard model", *Review of Economic Studies,* XLIX, 403–409.

Fourgeaud, C., Gouriéroux, C. and Pradel, J. [1990]: "Heterogeneity and hazard dominance in duration data models" (in french), *Annales d'Economie et de Statistique,* 18, 1–24.

Hamilton, J. [1989]: "A new approach to the economic analysis of non–stationary time series and the business cycle", *Econometrica,* 57, 357–384.

Han, A. and Hausman, J.A. [1990]: "Specification and semi–parametric estimation of duration models", *Journal of Applied Econometrics,* 5, 1–28.

Heckman, J.J. [1991]: "Identifying the hand of past: distinguishing state dependence from heterogeneity", *American Economic Review,* 81, 3, 79–79.

Heckman, J.J. and Borjas, G. [1980]: "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence", *Economica,* 47, 247–283.

Heckman, J.J. and Honoré, B. [1989]: "The identifiability of the competing risks model", *Biometrika,* 76, 325–330.

Heckman, J.J. and Singer, B. [1982]: "The identification problem in econometric models for duration data", in: *Advances in Econometrics*, ed. by W. Hildenbrand, Cambridge University Press, 39–77.

Heckman, J.J. and Singer, B. [1984 a]: "Econometric duration analysis", *Journal of Econometrics,* 24, 63–132.

Heckman, J.J. and Singer, B. [1984 b]: "The identifiability of the proportional hazards model", *Review of Economic Studies,* 51, 231–243.

Heckman J.J. and Walker J.R. [1990]: "The relationship between wages and income and the timing and spacing of births: evidence from longitudinal data", *Econometrica,* 58, 1411–1441.

Honoré B. [1992]: "Identification results for duration models with multiple spells", mimeo, Northwestern University.

Hougaard P. [1986]: "A class of multivariate failure time distributions", *Biometrika,* 73, 671–678.

Kalbfleisch and Prentice [1980]: *The statistical analysis of failure time data.* New York: Wiley.

Kennan J. [1985]: "The duration of contract strikes un U.S. manufacturing", *Journal of Econometrics,* 28, Annals 1–1985, 5–28.

Kiefer N. [1988]: "Economic duration data and hazard functions", *Journal of Economic Literature,* XXVI, 646–679.

Lancaster T. [1979]: "Econometric methods for the duration of unemployment", *Econometrica*, 47, 939–956.

Lancaster T. [1990]: *The Econometric Analysis of Transition Data*, Econometric Society Monographs. Cambridge Uneversity Press: New York.

Lancaster T. and Nickell S. [1980]: "The analysis of re–employment probabilities for the unemployed", *Journal of the Royal Statistical Society*, Series A, 143 (part 2), 141–165.

Lawless J.F. [1982]: *Statistical models and methods for lifetime data.* New York: Wiley.

Lumsdaine R.L., Stock J.H. and Wise D.A. [1990]: "Windows and retirement" (in french), *Annales d'Economie et de Statistique,* 20–21, 219–242.

Marshall A.W. and Olkin I. [1967]: "A multivariate exponential distribution", *Journal of the American Statistical Association,* 62, 30–44.

Newman J. and Mc Cullough C. [1984]: "A hazard rate approach to the timing of births", *Econometrica,* 52, 939–962.

Nickell S. [1979]: "Estimating the probability of leaving unemployment", *Econometrica,* 47, 1249–1266.

Pakes A. and Simpson M. [1989]: "Patent renewal data", *Brooking Papers on Economic Activity,* Microeconomics 1989, 331–401.

Ridder G. [1990]: "The nonparametric identification of generalized accelerated failure–time models", *Review of Economic Studies,* 57, 167–181.

Rust J. [1987]: "Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher", *Econometrica,* 55, 999–1033.

Schwartz E.S. and Torous W.N. [1989]: "Prepayment and the valuation of mortgage backed securities", *Journal of Finance,* 44, 375–392.

Stock, J.H. and Wise D. [1990]: "Pensions, the option value of Work, and retirement", *Econometrica*, J8.5.

Tsiatis A.A. [1981]: "A large sample study of Cox's regression model", *Annals of Statistics,* 9, 93–108.