Survival Analysis using the Swiss Household Panel

FORS - Swiss Centre of Expertise in the Social Sciences

Priscila Ferreira   Stephen P. Jenkins   Thomas Siedler

UMinho —— LSE —— UHamburg
e-mail: priscila@eeg.uminho.pt

4-7 July 2016, University of Lausanne

Outline

1. Introduction
2. Basic Concepts
3. Functional Forms of the Hazard rate
4. Estimation
5. Continous time multivariate Models
6. Discrete time multivariate models
7. Cox's proportional hazard model

Course Materials

http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/index.php

- reading list
- lecture notes
- do-it-yourself course on how to apply survival analysis to empirical data, illustrated using Stata

Introduction and Specification

Introduction

- Distinctive features of survival time data $\rightarrow$ distinctive methods

Part I. Specification

- Basic Concepts
  - the hazard rate, survivor function, failure function, etc.
- Functional forms for the hazard rate function

Estimation

Part II. Estimation

- Non-parametric estimators of the survivor function and hazard rate function
- Continuous time multivariate regression models
- Discrete time multivariate regression models
- Cox's proportional hazard regression model

Part III. Additional Topics

- Unobserved heterogeneity (frailty)
- Independent competing risks models
- ...

Introduction

Introduction to Survival Analysis

What Survival Analysis is about

- Modelling of {time-to-event | transition | survival time | duration, event history} data
- Consider a particular life-course *domain*, partitioned into a number of *mutually-exclusive states* at each point in time.
- With the passage of time, individuals move (or do not move) between states.
- Examples: ...

lifecourse domains	state at each t (mutually exclusive)
Marriage	married cohabiting separated widowed divorced single never-married
Receipt of security benefit(s)	receiving benefit x receiving benefit y receiving x and y receiving neither x or y
Housing Tenure	owned outright owner with mortgage renter – social housing renter – private landlord private renter other
Paid work	employed self-employed unemployed inactive retired

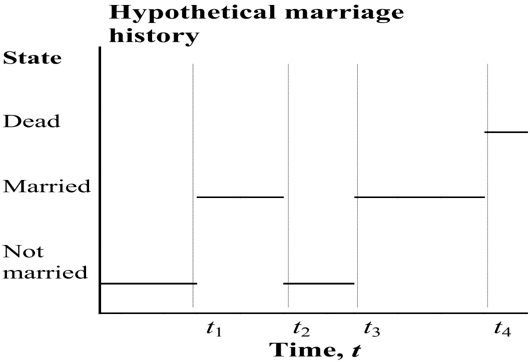
Table: Given a domain, which states to distinguish, and how many?

Transition Patterns

For each domain, the transition patterns for each individual are characterised by:

- the time spent within each state;
- the dates of any transitions made between states (if any)

Example: hypothetical marital history with 3 states for a given individual (adapted from Tuma, and Hannan, 1984, Fig. 1.)...



Length of time spent within each state $\sim$ length of horizontal line
Spells within a given state marked out by dates (start, and end)

More generally, we could have survival time data for:

- a large number of subjects (e.g. individuals, firms,) $\rightarrow$ usefulness of statistical models in order to describe data and to predict spell lengths, and also have
- other information about the subjects (their characteristics) $\rightarrow$ explanatory variables for multivariate modelling

This course is about the methods used to model this sort of data.

Note the complexity of the hypothetical marital history (and more generally):

- multi-state transitions
- repeat spells

$\rightarrow$ simplify and narrow the focus:

The Focus

- Survival time within a *single state*
- *Single spell* observed for each subject
- plus various simplifying assumptions:
  - → *No state dependence:* the chances of making a transition from current state do not depend on transition history prior to entry to current state
  - → *No initial conditions issues:* Entry to state being modelled is treated as exogeneous (otherwise we would have to model the chances of having arrived in the state in the first place)
  - → *Stationary process:* Model parameters are fixed constant, or can be characterised using explanatory variables, or parametrically

Distinctive models are used for survival time data modelling because of the datas special features:

- censoring (and truncation)
- time-varying explanatory variables ("covariates")

with two approaches to measuring time (continuous and discrete)

Survival time data collection methods

- *Stock sample:* Survey those currently in the state (and determine entry date = spell start date)
- *Inflow sample:* survey those entering a state, and follow them until some common pre-specified date or e.g. until spell ends
- *Outflow sample:* survey those leaving a state (and determine when entered)
- *Population sample:* survey population and determine dates of spells experienced

Longitudinal survey instruments for getting spell data

- Administrative records (e.g. current recipients, or ever a recipient within some observation window) + interview
- Sample survey (often one-off) of population, with retrospective questions, e.g. Women & Employment Survey, Families and Working Lives Survey, BHPS (waves 2, 3)
- Panel and cohort surveys follow a population' spell info built up from repeated obs on persons, e.g. BHPS, GSOEP, ECHP

Censoring

An individual's survival time in a state is *censored* if the date of transition into the state, or the date of transition out of the state, is not known exactly (only that before some date, or after some date).

Censoring, *ctd.*

Right censoring: spell end date not observed (only know that total time in state $\geq$ time from start of spell to end of observation period

Left censoring: spell start date not observed (cf. biostatisticians' definition: know that end date before observation date, but not exactly when)

Intervall Censoring: Event time falls in certain interval (L_i, R_i)

Truncation

Whereas censoring means that we don't know the exact length of a completed spell in total, truncation refers to whether or not we observe a spell or not in our data (sample selection on dependent variable):

Truncation

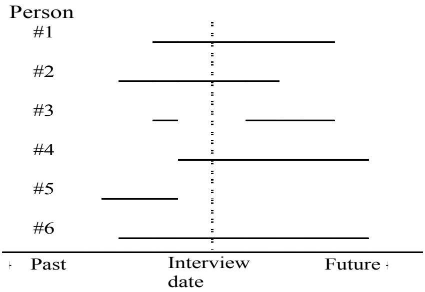
- *Left truncation ("delayed entry"):* only those surviving a sufficient amount of time are included in the sample (e.g. stock sample with follow-up).
- *Right truncation:* only those with a transition by a particular time are included in the sample (e.g. sample from the outflow from a state).

Truncation

Example for Left truncation
(Klein, Moeschberge, 2003): Survival study of residents in retirement center. Individual must survive to a sufficient age to enter retirement center.

Example for Right truncation
Lagakos et al. (1988) AIDS study: Contaminated blood transfusion, 1.4.1978, waiting time to develop AIDS by 30.6. 1986. Infected ind. who have not developed AIDS are excluded.

Censored and truncated spells: examples



Suppose population survey. At interview date, respondents asked about start date of current spell if in progress, or start and end dates of most recent spell if not in state then.

Spells for #1, #2, #4, #6 are right-censored

Left-censoring? (Suppose didn't ask about state before some fixed date in past)

Left- and right-truncation?

The nature of the survival time data: continuous versus discrete

- Is a transition out of current state something that can occur at any instant of time, or only in terms of discrete points? (Cf unemployment exits; machine cycles). Related to the *process* determining transitions.

- How are survival times *recorded*? As exact dates or only within intervals of time (hence grouped or banded data = 'interval censoring')?

Types of explanatory variable

- Characteristics of subject (person, firm) vs. characteristics of socio-economic environment
  - $\rightarrow$ not an issue analytically; may be empirically
- Fixed versus time-varying covariates (TVCs), where TVCs may vary with (i) survival time in state, &/or (ii) calendar time. (E.g. UK social assistance. Cf. local unemp rate)
  - $\rightarrow$ analytics and interpretation easier if no TVCs; estimating models with TVCs means data re-organisation ('episode-splitting')

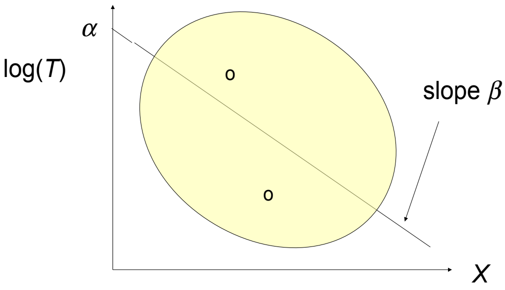
Econometric methods for survival time data: some motivation

- Why not use OLS?
  - $\rightarrow$ Regress each survival time (T or $\log T$) on covariates
- Problems with OLS:
  - $\rightarrow$ (right-)censoring of spell data
  - $\rightarrow$ time-varying covariates
  - $\rightarrow$ 'structural' modelling

OLS and right-censored spell data

Suppose that log survival times T_i are a linear function of a single characteristic $X_i, i = 1, \ldots, N$:

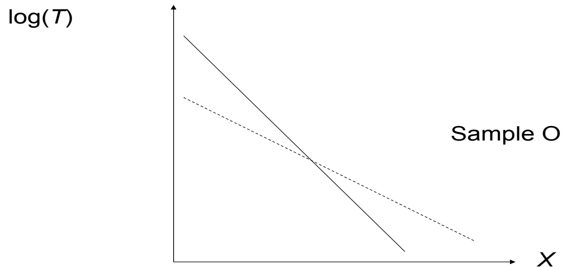
$\log(T_i) = \alpha + \beta X_i + e_i$; and $\alpha > 0, \beta < 0$



OLS: choose estimators a, b, that minimise the sum of the squared residuals (e_i)

Suppose prevalence of right-censoring greater at longer durations than shorter durations:

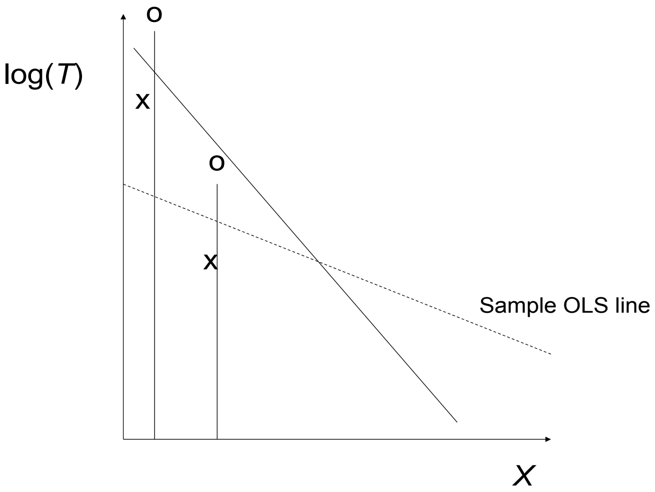
(a) *Exclude censored spells altogether from OLS estimation* → *sample* data cloud less dense everywhere, but especially so higher values of $\log(T)$ → estimated slope not as negative as true slope, i.e. over-estimate.



Like sample selection bias. Cf 'Heckman 2-stage': OLS like omitting lambda term

Suppose prevalence of right-censoring greater at longer durations than shorter durations:

(b) *Treat censored spells as if they were complete* $\rightarrow$ under-recording, especially so at higher values of $\log(T)$ = like non-random mis-measurement of depvar $\rightarrow$ estimated slope not as negative as true slope, i.e. over-estimate.


log(T)
o
x
o
x
Sample OLS line
X

OLS and TVCs

How can OLS handle them, given that each observation (one spell length) contributes one observation to the regression?

If one were to choose one value of the TVC for each person, which one would one choose?

- That just before the transition (but this varies by person, and what about censored observations?)
- Might use value of TVC at start of spell? (Consistent definition for all spell, but now fixed covariate and lose information)

OLS and ‘structural modelling’ of longitudinal processes

Most behavioural models – of e.g. job search, marital search, etc. – are framed in terms of decisions to do something ($\rightarrow$ transition event), and not a spell length per se

Why not use a binary dependent variable model?

Use logit (probit,) regression of whether or not experience a transition or not against characteristics? (This would deal with right-censored obs.)

- But would take no account of the differences in *length* of time each person was at risk of experiencing the transition, and so loses information (when left, if did so).
- How to handle TVCs?

Implications

- We need methods recognising the longitudinal (passage of time) nature of the spell data, the ability to handle censored (and truncated) spells, and also time-varying covariates
- Solution = use estimation methods other than OLS (typically ML), and also re-organise the data set (so that get likelihood right, and can handle TVCs)
- Throughout, we need tools for summarising survival time distributions (hazard rate, survivor and failure functions)

Basic Concepts

Basic Concepts

- Probability distribution functions
- Probability density functions
- Survivor functions (new?!)
- Hazard rate functions (new?!)

We need them

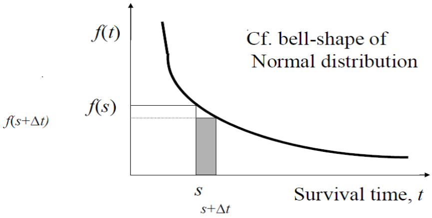
- to describe distributions of survival times (cf. description of income distributions). Special forms – data are not Normal.
- to fit models of the distributions, using specific functional forms

Continuous & discrete time Relationships between concepts

Continuous Time

The length of a spell is a realisation of a continuous random variable **T** with

- probability density function (PDF): **f(t)**

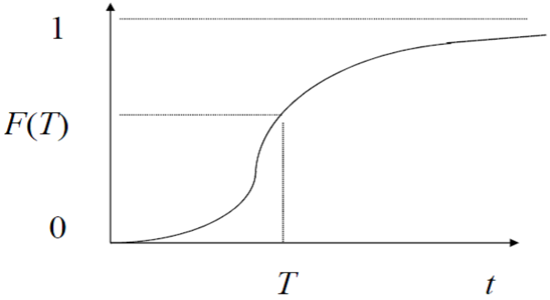


NB: areas under PDFs are probabilities

area(rectangle) = height X base $\rightarrow f(s) = height = area/base = Prob/\Delta t$

Failure function, *F(t)*

cumulative density function (CDF), probability distribution function, or “failure function”:



$F(t) = Pr(T \leq t) =$ area under *f(t)*

(PDF) up to t=T

f(t) = slope of CDF at t: $f(t) = \lim_{\Delta t \to 0} \dfrac{Pr(t \leq T \leq t + \Delta t)}{\Delta t} = \dfrac{\partial F(t)}{\partial t}$

Survivor Function, *S(t)*

$$Pr(T > t) = 1 - F(t) \equiv \overline{F}(t) \equiv S(t)$$

- The survivor function equals
  1 – failure function [NB different notations]
- *S(t)* is probability of survival (i.e. remaining in state) at least t units of time since entry at $t = 0$
- Note that:

$$f(t) = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t}$$

Survivor Function, *S(t)*

- *S(t)* and *F(t)* are probabilities, so $0 \leq S(t) \leq 1$, and $0 \leq F(t) \leq 1$
- $S(0) = 1$, $S(\infty) = 0$, $\partial S/\partial t < 0$
- f(t) is not a probability (its a density); $f(t) \geq 0$

Hazard rate function, $\theta(t)$

$\theta(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{\overline{F}(t)} = \frac{f(t)}{S(t)}$

Hazard rate at t equals the ratio of the pdf at t to the survivor function at t.

Properties of hazard rate:

$\forall \theta(t) \geq 0$, but may be > 1!

NB given an expression for *S(t)*, equivalently for *F(t)*, one could derive *f(t)*, and thence hazard rate $\theta(t)$. [More on links below.]

Interpretation of $\theta(t)$

$\theta(t)\Delta t = \frac{f(t)\Delta t}{S(t)}$

for some tiny interval of time Δt.

Numerator of RHS is like a probability (recall areas under PDFs = probabilities):

$f(t)\Delta t \approx$ Pr(leaving the state in the interval $[t,\ t + \Delta t]$

So, the expression for the hazard rate looks a bit like a conditional probability.

Recall the rules of conditional probability:

$Pr(A \mid B) = Pr(A \cap B)/\mathrm{Pr(B)}$

Interpretation of $\theta(t)$

$\forall\theta(t)$: conditional failure rate

Examples:

- Given that you smoked for t periods, what is the failure rate ('likelihood') to quit in subsequent period?
- Prisoner has been released for t months, failure rate to return to prison in subsequent months?
- Being unemployed for t months, what is the failure rate to find a job in the near future?

Interpretation of $\theta(t)$

$Pr(A \mid B) = Pr(A \cap B)/Pr(B)$
$= Pr(B \mid A)Pr(A)/Pr(B)$

Now let A: "leaving the state in the interval $[t, t + \Delta t]$"
B: "survival to time t"

So,
Pr(leaving in interval $[t, t + \Delta t]$ conditional on survival until t)
$= Pr(A \mid B) = Pr(A)/Pr(B)$, since $Pr(B \mid A) = 1$

So, the continuous-time hazard rate has similarities to a conditional probability, but isnt a 'genuine' probability!

Conditional *versus* unconditional 'probabilities'

Contrast between

1. conditional: $\theta(t)\Delta t$

   and

2. unconditional: $f(t)\Delta t$

Conditional *versus* unconditional ‘probabilities’

For example, compare

- A1 Probability for a person who has been unemployed for 120 days of leaving unemployment on the 121st day, *versus*
- A2 Probability for persons entering unemployment of having a spell length of 121 days;

- B1 Probability of dying at age 12 for someone who is aged 12, *versus*
- B2 Probability for a new-born baby of dying at age 12.

Relationship between hazard and survivor functions

Given any functional form for the hazard rate $\theta(t)$, one can derive the functional form for the survivor function *S(t)* [and *vice versa*]:

$$\begin{aligned}\theta(t) &= \frac{f(t)}{1-F(t)} \\ &= \frac{-\partial[1-F(t)]/\partial t}{1-F(t)} \\ &= \frac{\partial\{-\ln[1-F(t)]\}}{\partial t} \\ &= \frac{\partial\{-\ln[S(t)]\}}{\partial t}\end{aligned}$$

using the fact that $\partial \ln[g(x)]/\partial x = g'(x)/g(x)$, and $S(t) = 1 - F(t)$. Now integrate both sides of the expression: ...

from $\theta(t)$ to $S(t)$, ctd.

$\int_0^t \theta(u)\mathbf{d}u = -\ln[1 - F(t)] \quad |\binom{t}{0}$

But F(0)=0, and $\ln(1) = 0$,

so, ...

$$\begin{aligned} \ln[1 - F(t)] &= -\int_0^t \theta(u)du, \text{ i.e.} \\ S(t) &= \exp\left(-\int_0^t \theta(u)du\right) \\ S(t) &= \exp[-H(t)] \end{aligned}$$

where $H(t) = -\ln[S(t)]$ is the *integrated hazard function*.
Thus, in principle, given any form for $\theta(t)$, we can derive $S(t)$.

Discrete time concepts

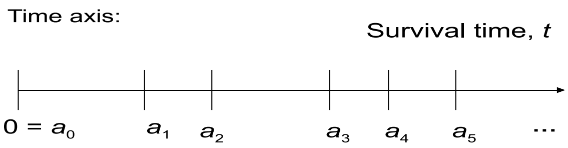
The time scale may be

a Continuous time process but survival times measured in bands (grouped data; interval censoring)

b Intrinsically discrete (e.g. time to machine breakdown measured as # machine cycles)

Consider case (a) first, and suppose underlying continuous time survival time T recorded in disjoint *intervals* (need not be of same length):

(a) Grouped data



Intervals of time, indexed by $0 = a_0, a_1, a_2, a_3, \ldots, \ldots,$ where the intervals are $[0 = a_0, a_1], (a_1, a_2], (a_2, a_3], \ldots, (a_{k-1}, a_k = \infty)$.

Survivor function at start of jth interval (just after end interval j-1):
$S(t) = 1 - F(t) = Pr(T > a_{j-1})$.

Probability of exit from state in jth interval (interval density):
$Pr(T \in (a_{j-1}, a_j]) = F(a_j) - F(a_{j-1}) = S(aj - 1) - S(a_j)$

Discrete hazard rate of exit during jth interval: $h(a_j)$

$$
\begin{aligned}
h(a_j) &= Pr(a_{j-1} < T \leq a_j \mid T > a_{j-1}) \\
&= \frac{Pr(a_{j-1}<T\leq a_j)}{Pr(T>a_{j-1})} \\
&= \frac{S(a_{j-1})-S(a_j)}{S(a_{j-1})} \\
&= 1 - \frac{S(a_j)}{S(a_{j-1})}
\end{aligned}
$$

$h(a_j)$ is a probability (cf. continuous time hazard), and so $0 \leq h(a_j) \leq 1$

Easiest to consider case in which every interval is of unit length so recorded duration intervals become $(t-1, t]$ with $t = 1, 2, 3, \ldots$ (positive integer); $T \in (t-1, t]$

Discrete time hazard rate and survivor functions

- Alternatively (and perhaps easier), instead of indexing intervals using the *date* at end of each *interval*, let us index each interval directly.
- Thus refer to a spell of length j (i.e. one lasting to end of the j^{th} interval)
- *Survivor function*: Probability of survival to the end of interval j is the product of the probabilities of not experiencing the event in each of the intervals up to and including the current one, i.e. product of discrete hazards:

Discrete time survivor and failure functions

$$
\begin{aligned}
S_j &= S(j) = (1-h_1)(1-h_2)\dots(1-h_j) \\
&= \prod_{k=1}^{j}(1-h_k)
\end{aligned}
$$

$S_j = S(j)$ now refers to a discrete *time survivor function*.
Cf. $S(a_j)$ which is a continuous time survival function (with an argument which is a date - an instant of time - rather than an interval of time)

– NB: a matter of notation, since
$S(j) = S(a_j)$

Discrete time failure function:
$F_j = F(j) = 1 - S(j)$

Discrete time hazard rate and density functions

Discrete *time hazard rate* for jth interval can be written as

$h(j) = \frac{f(j)}{S(j-1)}$

where f(j) is the *discrete time density function*, given by:

$$\begin{aligned} f(j) &= h_j S_{j-1} \\ &= h_j \prod_{k=1}^{j-1} (1 - h_k) \\ &= \frac{h_j}{1 - h_j} \prod_{k=1}^{j} (1 - h_k) \end{aligned}$$

(Later we use this expression a lot when deriving expressions for sample likelihoods.)

(b) Discrete time: time intrinsically discrete

Survival time T is now a discrete random variable with probabilities

$f_j = f(j) = Pr(T = j)$

where $j = 1, 2, 3, \ldots$ is is the set of positive integers.

NB j now indexes 'cycles' (not unit 'intervals'), but we can use same notation.

Discrete time survivor function:

$$\begin{aligned} S_j = S(j) &= (1 - h_1)(1 - h_2)\ldots(1 - h_j) \\ &= \prod_{k=1}^{j}(1 - h_k) \end{aligned}$$

(b) time intrinsically discrete, ctd.

Discrete time hazard at j, *h(j)*, is conditional probability of event at j (with conditioning on survival until completion of the previous cycle, $j1$):

$$\begin{aligned} h(j) &= Pr(T = j \mid T \geq j) \\ &= \tfrac{f(j)}{S(j-1)} \end{aligned}$$

Discrete failure function:

$$\begin{aligned} F_j = F(j) &= 1 - S_j \\ &= 1 - \prod_{k=1}^{j} (1 - h_k) \end{aligned}$$

(b) time intrinsically discrete, ctd.

Discrete density function:

$$
\begin{aligned}
f(j) &= h_j S_{j-1} \\
&= h_j \prod_{k=1}^{j-1} (1 - h_k) \\
&= \frac{h_j}{1 - h_j} \prod_{k=1}^{j} (1 - h_k)
\end{aligned}
$$

NB: Same expressions as in unit-interval grouped data case

Analogy between discrete and continuous time expressions

Discrete time:

$$\log S(j) = \sum_{k=1}^{j} \log(1 - h_k)$$

For 'small' h_k, $\log(1 - h_k) \approx -h_k$

which implies:

$$\log S(j) = -\sum_{k=1}^{j} h_k$$

Contrast with continuous time case:

$\log S(t) = -H(t) = -\int_0^t \theta(u)du$

As $h_k \to 0$, the discrete expressions $\to$ continuous time counterparts. (Cf. sums of discrete hazards and integrations of continuous hazards.)

Functional Forms of the Hazard rate

Criteria for choice of specification

- Given 1:1 relationships between hazard and density, failure, and survivor functions, we could specify our models in terms of any one of these
- But typically done in terms of the hazard rate function (more closely related to the underlying behavioural processes)

Criteria for choice of hazard rate function

- Shape that is empirically relevant, or suggested by theoretical models
  - $\rightarrow$ likely to differ between applications (cf. human mortality, unemployment spell lengths, failure times of machine tools)
- Specification with convenient mathematical properties
  - $\rightarrow$ e.g. closed form expressions for survivor function, and summary statistics such as mean, median duration
- Trade-off between parametric functional forms and more flexible ones?
  - $\rightarrow$ tractability versus fit?

Taxonomy of specifications

- Continuous time *versus* discrete time models
  - $\rightarrow$ differences in the assumptions about the survival time metric (whether underlying process, or way the data are recorded)
- Proportional Hazard (PH) *versus* Accelerated Failure Time (AFT) *versus* Proportional Odds models
  - $\rightarrow$ differences in *interpretation* of a model and its parameters
  - $\rightarrow$ some models have the PH property; others the AFT one.

Continuous *vs.* discrete

- *Continuous time parametric*
  - → Log-logistic*
  - → Log-normal
  - → Gompertz
  - → Generalized Gamma
- *Continuous time semi-parametric*
  - → Piecewise Constant Exponential (PCE)
  - → [Cox's model]*
- *Discrete time (par. & semipar.)*
  - → Logistic*
  - → Complementary log-log ('cloglog')*

* focused on in lectures (for others, see Lecture Notes)

Introducing differences in characteristics

- To allow for differences in characteristics to enter the hazard rate function, define

$$\beta' X \equiv \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k$$

and

$$\beta^{*\prime} X \equiv \beta_0^* + \beta^*{}_1 X_1 + \beta^*{}_2 X_2 + \ldots \beta^*{}_k X_k$$

Introducing differences in characteristics, *ctd.*

- The βs (and β*) are parameters (later to be estimated) and the elements of the X vector summarise observed characteristics
- For the moment, suppose that X does not vary with survival time (or calendar time).
  - $\rightarrow$ Assumption relaxed later (and we also consider effects of unobserved characteristics)

Weibull model

(continuous time)

$$\begin{aligned}\theta(t, X) &= \alpha t^{\alpha-1} \exp(\beta' X) \\ &= \alpha t^{\alpha-1} \lambda\end{aligned}$$

where $\lambda \equiv \exp(\beta' X)$, and exp(.) is the exponential function.

λ is a scaling factor: larger $\lambda \Rightarrow$ larger hazard, at each t

$\alpha > 0$ is the *shape parameter*:

$\alpha = 1$ *Exponential* model

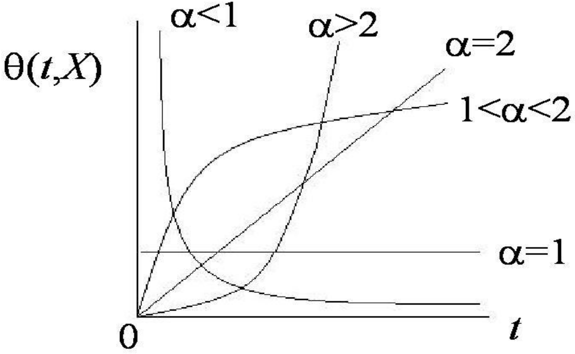
- hazard rate constant over time

$\alpha > 1$ hazard monotonically increases with survival time

$\alpha < 1$ hazard monotonically decreases with survival time

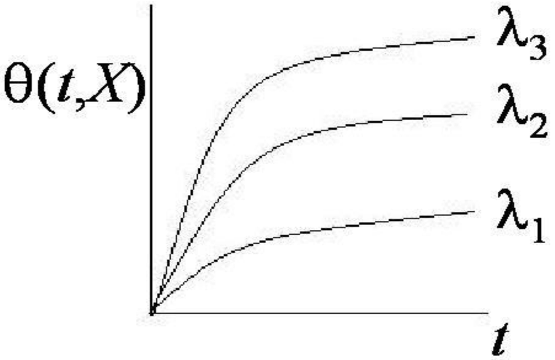
Weibull hazard function

Variations in α, for fixed λ:



Weibull hazard function, *ctd.*

Variations in λ, for fixed α:



Gompertz hazard function

(continuous time)

$$\begin{aligned} \theta(t, X) &= \lambda \exp(\gamma t) \\ \log \theta(t, X) &= \beta' X + \gamma t \end{aligned}$$

- I.e. log (hazard) is linear in survival time, where
  $\rightarrow$ $\lambda \equiv \exp(\beta' X) > 0$ (i.e. parameterization as per Weibull model)
- The larger λ is, *cet. par.*, the larger the hazard rate.

γ is the *shape parameter*. [Not the same as the loglogistic shape parameter!]

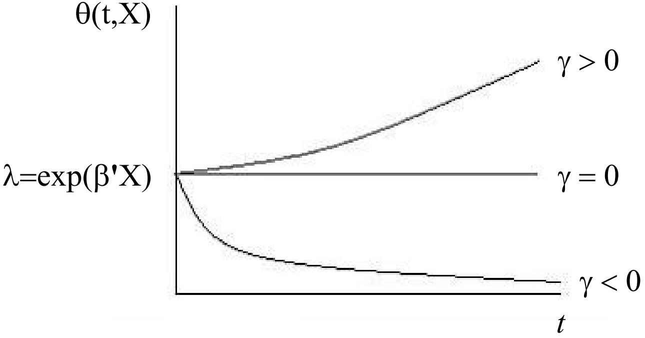
$\gamma > 0$ hazard monotonically increasing in t.

$\gamma = 0$ hazard constant for all t (Exponential)

$\gamma < 0$ hazard monotonically decreasing in t.

Gompertz hazard function, *ctd.*

(continuous time)



Loglogistic hazard function

(continuous time)

$$\theta(t, X) = \frac{\psi^{\frac{1}{\gamma}} t^{\left(\frac{1}{\gamma} - 1\right)}}{\gamma\left[1 + (\psi t)^{\frac{1}{\gamma}}\right]}$$

where

- $\psi \equiv \exp(-\beta^{*\prime} X) > 0$ is a scale factor [reason for changing parameterisation apparent later].
- The larger ψ is, *cet. par.*, the lower the hazard rate.

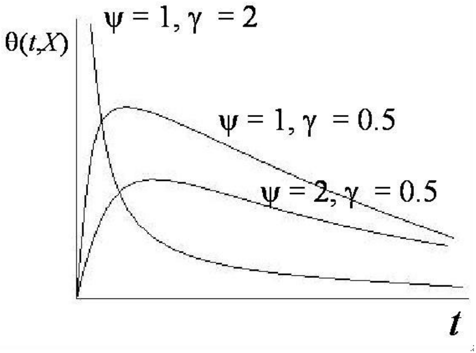
$\gamma > 0$ is the *shape parameter*:

$\gamma \geq 1$ hazard monotonically decreases in t.

$\gamma < 1$ hazard rises, then falls monotonically.

Loglogistic hazard function, *ctd.*

(continuous time)



Lognormal hazard function

(continuous time)

$$\theta(t,X) = \frac{\frac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\{\frac{\ln(t)-\mu}{\sigma}\right\}^2\right]}{1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)}$$

where $\mu \equiv \beta^{*\prime}X$

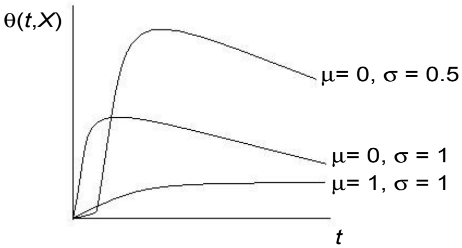
The larger μ is, *cet. par.*, the lower the hazard rate function.

$\sigma > 0$ is a scale parameter

Similar shape to loglogistic $(\gamma < 1)$ hazard.

Lognormal hazard function, *ctd.*

(continuous time)



Generalized Gamma hazard function (continuous time)

- Flexible but complicated functional form
- Encompasses others as special cases (useful for specification testing)
- Depends on two parameters:

  $\rightarrow \kappa, \sigma$

$$\kappa = 1 \text{ Weibull model}$$

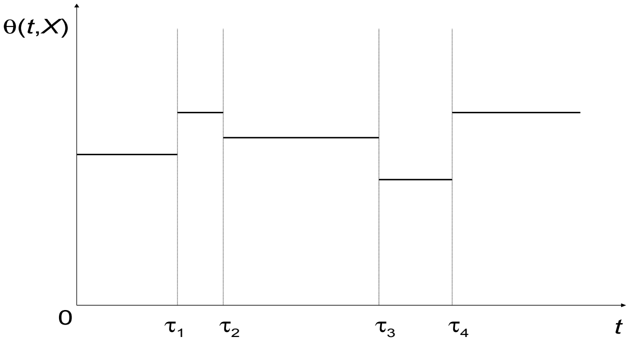
$$\kappa = 1, \sigma = 1 \text{ Exponential model}$$

$$\kappa = 0 \text{ Lognormal model}$$

Piecewise constant Exponential (PCE) hazard

(continuous time)

Allows flexibility in shape (of a sort).
Hazard constant within (user-specified) intervals, but may differ between them:



Piecewise constant Exponential (PCE) hazard, *ctd.*

(continuous time)

Hazard constant within (user-specified) intervals, but may differ between them.

$$\theta(t,X) = \begin{cases} \overline{\theta}_1 \exp(\beta' X) & t \in (0,\tau_1] \\ \overline{\theta}_2 \exp(\beta' X) & t \in (\tau_1,\tau_2] \\ \vdots & \vdots \\ \overline{\theta}_K \exp(\beta' X) & t \in (\tau_{K-1},\tau_K] \end{cases}$$

Expression for $\theta(t,X)$ may be re-written as:

$$\begin{cases} \exp[\log(\overline{\theta}_1) + \beta' X] & t \in (0,\tau_1] \\ \exp[\log(\overline{\theta}_2) + \beta' X] & t \in (\tau_1,\tau_2] \\ \vdots & \vdots \\ \exp[\log(\overline{\theta}_K) + \beta' X] & t \in (\tau_{K-1},\tau_K] \end{cases}$$

PCE model, *ctd.*

With further rewriting,

$$\theta(t, X) = \begin{cases} \exp(\tilde{\lambda}_1) & t \in (0, \tau_1] \\ \exp(\tilde{\lambda}_2) & t \in (\tau_1, \tau_2] \\ \vdots & \vdots \\ \exp(\tilde{\lambda}_K) & t \in (\tau_{K-1}, \tau_K] \end{cases}$$

- Thus, PCE model equivalent to having interval-specific intercept terms in the overall hazard rate, $\theta(t, X)$
- Simple example of a model with time-varying covariates. (Straightforward to allow X to vary with time too.)

Coxs model

(continuous time)

Hazard rate has general form

$$\begin{aligned} \theta(t, X) &= \theta_0(t)\exp(\beta' X) \\ &= \theta_0(t)\lambda \end{aligned}$$

- where $\theta_0(t)$, the 'baseline hazard' function, can take on **any** shape (hence 'semi-parametric' model)
- Parameters β estimable, but baseline hazard function not identified.
- More on this model in a separate lecture later on

Interpreting models (1): proportional hazards (PH)

- PH models = 'multiplicative hazard' models = 'log relative hazard' models (for reasons, see below)
- **All PH models satisfy a separability condition:**

$$
\begin{aligned}
\theta(t, X) &= \theta_0(t)\exp(\beta' X) \\
\Rightarrow \log[\theta(t, X)] &= \log[\theta_0(t)] + \beta' X
\end{aligned}
$$

where

$\theta_0(t)$ *baseline hazard* function depending on t, but not X. Summarises the pattern of *duration dependence*.

$\exp(\beta' X)$ non-negative function of X, but not t. (Ensures $\theta > 0$. In principle, other functions might be specified, but this is the one virtually always used.)

PH model again

- **Sometimes you see PH specification written differently**

$$\theta(t, X) = \theta_0(t)\exp(\beta' X)$$

$$\Rightarrow \text{if } X = 0:$$

$$\theta(t, X \mid X = 0) = \theta_0(t)\exp(\beta_0) \equiv \theta_0^*(t)$$

$$\Rightarrow \theta(t, X) = \theta_0^*(t)\lambda$$

where $\lambda \equiv \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$

- $\theta_0^*(t)$: *baseline hazard* function (again).
- Issue is whether or not to include the intercept term (fixed and common to each person) in baseline hazard or in the scale factor

PH model interpretations

- Absolute differences in X imply *proportionate* differences in hazard (at each t):

For some $t = \bar{t}$, and for two persons i and j with characteristics vectors X_i and X_j,

$$\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)} = \exp(\beta' X_i - \beta' X_j)$$

- Equivalently, in log relative hazard form:

$$\log\left[\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)}\right] = \beta'(X_i - X_j)$$

PH model interpretations, *(ctd.)*

- Each *regression coefficient* β_k summarises the *proportional effect on the hazard* of a unit change in the corresponding covariate X_k:

$$\beta_k = \partial \log \theta(t, X) / \partial X_k$$

NB This proportional effect does not vary with survival time. (β_k does not vary with t.)

PH model interpretations, *(ctd.)*

- *Elasticity of the hazard* with respect to X_k

$$
\begin{aligned}
&= \partial \log[\theta(t, X)]/\partial[\log(X_k)] \text{ , or} \\
X_k \partial \log \theta(t, X)/\partial X_k &= \beta_k X_k
\end{aligned}
$$

If covariate measured in logs: $X_k \equiv \log(Z_k)$, then it follows that β_k is the elasticity of the hazard with respect to Z_k.

PH model interpretations, *(ctd.)*

Proportionate change in the hazard given a change in a dummy variable covariate from zero to one (more precisely $X_{ik} = 0$ to $X_{jk} = 1$), with all other covariates held fixed, i.e.

$$\textit{Hazard ratio} \text{ for covariate } k = \exp(\beta_k)$$

PH model interpretations, *(ctd.)*

Differences in characteristics imply a scaling of the common 'baseline survivor function'.

$$\begin{aligned} S(t, X) &= \exp\left[-\int_0^t \theta(u)du\right] \\ &= \exp\left[-\lambda \int_0^t \theta_0(u)du\right] \\ &= [S_0(t)]^{\lambda} \end{aligned}$$

given a 'baseline survivor function'

$$S_0(t) \equiv \exp\left[-\int_0^t \theta_0(u)du\right]$$

PH model interpretations, *(ctd.)*

Ditto for integrated hazard function $H(t) = \ln S(t)$:

$$\begin{aligned} H(t) &= \lambda H_0(t) \\ \text{where } H_0(t) &= \textstyle\int_0^t \theta_0(u)du \text{ . Hence,} \\ \ln[H(t)] &= \beta' X + \ln[H_0(t)] \end{aligned}$$

'Parallel lines' specification check for PH

Interpreting models (2): accelerated failure time (AFT) models

- AFT models assume a linear relationship between log of completed (latent) survival time T and characteristics X :

$$\begin{aligned} \ln(T) &= \beta^{*\prime}X + z, \text{ or} \\ Y &= \mu + \sigma u, \text{ or} \\ (Y-\mu)/\sigma &= u \end{aligned}$$

[GLM or -log-linear model- formulation]

where $Y \equiv \ln(T)$, $\mu \equiv \beta^{*\prime}X$, and $u = z/\sigma$ is an 'error' term with density function *f(u)*, and $\sigma > 0$ is a scale factor

AFT models, *ctd.*

Different specifications for the distribution of the error term u lead to different models of the distribution of T:

$u \sim$	Normal	$\Rightarrow$ Lognormal model
$u \sim$	Logistic	$\Rightarrow$ Loglogistic model
$u \sim$	Extreme Value	$\Rightarrow$ Weibull (and Exponential) models
$u \sim$	3-parameter Gamma	$\Rightarrow$ Generalized Gamma model

NB Recall our opening discussion of the problems with modelling T or $\log(T)$ using OLS (implicitly assuming Normal error). Above indicates that the problems are

(a) censoring (not all T observed), and

(b) Normal not necessarily appropriate.

'AFT' interpretation

$$\text{Let} \psi \equiv \exp(-\beta^{*\prime} X) = \exp(-\mu). \text{ Then,}$$
$$\ln(T\psi) = z.$$

Term ψ acts like a time scaling factor:

$\psi > 1$ Failure is *accelerated* (survival time shortened).

$\rightarrow$ It is as if the clock ticks faster (time scale for someone with characteristics X is $T\psi$, whereas for someone with $X = 0$, time scale is T.

$\psi < 1$ Failure is *decelerated* (survival time lengthened).

$\rightarrow$ It is as if the clock ticks slower.

Can also see time-scaling property in terms of the survivor function: ...

AFT interpretation, *ctd.*

- General definition of survivor function:

$$S(t, X) = Pr[T > t \mid X]$$

- If we have an AFT model, then:

$$\begin{aligned} S(t, X) &= Pr[Y > \ln(t) \mid X] \\ &= Pr[\exp(\sigma u) > t\exp(-\mu)] \end{aligned}$$

AFT interpretation, *ctd.*

- Define 'baseline survivor function' $S_0(t)$ for case $X = 0$, in which case $\exp(-\mu) = \exp(-\beta_0^*) \equiv \psi_0$

Now,

$$\begin{aligned} S_0(t) &= Pr[T > t \mid X = 0] \\ &= Pr[\exp(\sigma u) > t\psi_0], \text{ or} \end{aligned}$$

$$S_0(s) = Pr[\exp(\sigma u) > s\psi_0], \text{ any } s.$$

Now substitute $s = t\exp(-\mu)/\psi_0$ in $S_0(t) \Rightarrow$

$$\boldsymbol{S(t, X) = S_0[t\exp(-\mu)] = S_0[t\psi]}$$

NB Longevity example: 'If one year for a dog = 7 years for a human's (dogs age 7 times faster).

Dog's survivor function = $S_0(t_\psi)$; human's = $S_0(t)$; $\psi = 7$.

AFT interpretation, *ctd.*

- Each *regression coefficient* β_k^* summarises the *proportional effect on survival time* T to a unit change in corresponding covariate X_k:

$$\beta_k^* = \frac{\partial \ln(T)}{\partial X_k}$$

- *Elasticity of latent survival time* with respect to X_k, $\partial \ln[(T)]/\partial[\ln(X_k)]$, is $\beta_k^* X_k$.

If covariate measured in logs: $X_k \equiv \log(Z_k)$, then β_k^* is the elasticity of survival time with respect to Z_k.

AFT interpretation, *ctd.*

Given log-linear specification for T in an AFT model, absolute differences in characteristics imply *proportionate differences in survival times* (cf. PH models: proportionate differences in hazards!)

$$\log(T_i/Tj) = \beta^{*'}(X_i - X_j)$$

Proportionate change in T given a one unit change in X_k, with all other covariates held fixed, i.e.

Time ratio for covariate $k = \exp(\beta^*_k)$

The Weibull model is the *only* model that has PH *and* AFT properties

Proof of this requires showing that the Weibull functional form is the only one that satisfies the PH and AFT restrictions:

$$S(t) = [S_0(t)]^\lambda = S_0[t\psi]$$

i.e.

$$[\exp(-t^\alpha)]^\lambda = \exp(-[t\exp(-\mu)]^\alpha)$$

Relationships between Weibull AFT coefficients (β_k^*) and PH (β_k) coefficients :

$$\beta_k^* = -\sigma\beta_k = -\beta_k/\alpha, \text{ for each } k$$

I.e. AFT coefficients of opposite sign to corresponding PH ones, and $\sigma = 1/\alpha$

Model summary: PH, AFT

Model	PH	AFT
Exponential	✓	✓
Weibull	✓	✓
Log-logistic	X	✓
Lognormal	X	✓
Gompertz	✓	X
Generalized Gamma	X	✓

NB ‘PH’ and ‘AFT’ refer to model interpretation, not to differences in estimation.

- Time-varying covariates can be incorporated in these models straightforwardly. [See Lecture Notes.]
- Stata can be used to estimate all the models; you can choose to report coefficients or hazard ratios for PH models; coefficients or time ratios for AFT models.

Weibull Model of U.N. Peacekeeping Missions

- Green et al. (1998), *Policy Studies Journal*
- Data on duration of U.N. peacekeeping mission, 1948-2001
- Binary covariates indicate types of conflict

1. Civil war
2. Interstate conflict
3. Internationalized civil war

Weibull Model of U.N. Peacekeeping Missions (ctd)

Variable	Exponen-tial	Weibull AFT	Weibull PH
Constant	4.35 (0.21)	4.29 (0.27)	-3.46 (0.50)
Civil war	-1.16 (0.36)	-1.10 (0.45)	0.89 (0.38)
Interstate conflict	1.64 (0.50)	1.74 (0.62)	-1.40 (0.51)
Shape parameter		σ =1.24 (0.15)	α=0.81 (0.10)
N	54	54	54

Interpretation

PH Interpretation PH Interpretation Each *regression coefficient* β_k summarises the *proportional effect on the hazard* of a unit change in the corresponding covariate X_k:

$$\beta_k = \partial \log \theta(t, X) / \partial X_k$$

AFT Interpretation Each *regression coefficient* β_k^* summarises the *proportional effect on survival time* T to a unit change in the corresponding covariate X_k:

$$\beta_k^* = \frac{\partial \ln(T)}{\partial X_k}$$

Results

Baseline Hazard Rate	The longer peacekeeping mission lasts, the risk of terminating decreases
PH	Positive coefficient implies risk of termination increases
AFT	Positive coefficient implies longer duration

Results

- Recall relationship between AFT and PH parameters in the Weibull model:

$$\beta_k^* = -\sigma\beta_k = -\beta_k/\alpha, \text{ for each } k$$

- Civil War coefficient:

$$\beta_k^* = -1.24 * 0.89 = -1.10$$

AFT interpretation

- RB absolute differences in characteristics imply *proportionate differences in survival times*

$$\log(T_i \mid T_j) = \beta^{*'}(X_i - X_j)$$

Time ratio for covariate $k = \exp(\beta_k^*)$

- Interstate Conflict: $\beta^* = 1.74$
- $\exp(1.74) = 6$
- Interstate conflicts are around 6 times longer than internationalized civil wars

PH Interpretation

with characteristics vectors X_i and X_j,

$$\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)} = \exp\left(\beta' X_i - \beta' X_j\right).$$

- Note that hazards have same shape, but different scales (λ)
- Hazard ratio between Civil war and internationalized civil war (ICW) is: $\exp(0.89) = 2.4$
- Hazard ratio between Interstate conflict and ICW: $\exp(-1.4) = 0.24$

Discrete time models

1. Complementary log-log (cloglog) model
   → Can interpret as specification of a *proportional hazards* model: underlying survival process is continuous, but survival time data are recorded in bands ('grouped').
2. Discrete time logistic model
   → Can interpret as specification of a *proportional odds* model for an intrinsically discrete survival process

NB can apply both models to discrete time data.

Cloglog model as discrete time PH model

- Underlying process described by continuous hazard $\theta(t, X)$ but data grouped into intervals.
- We estimate the parameters for a PH model of θ, taking account of the interval censoring.

Cloglog model as discrete time PH model (ctd)

Survivor function at time a_j:

$$\begin{aligned} S(a_j, X) &= \exp\left[-\int_0^{a_j} \theta_0(t)\lambda du\right] \\ &= \exp\left[-\lambda\int_0^{a_j} \theta_0(t) du\right] \\ &= \exp\left[-H_j\lambda\right] \end{aligned}$$

where $H_j \equiv H(a_j) = \int_0^{a_j} \theta_0(u, X)du$ is the integrated baseline hazard at a_j, and using the PH assumption $\theta(t, X) = \theta_0(t)\exp(\beta' X)$

Cloglog model as discrete time PH model (ctd)

Discrete time hazard, $h(a_j, X) \equiv h_j(X)$,

$$\begin{aligned} h_j(X) &= \frac{S(a_{j-1},X)-S(a_j,X)}{S(a_{j-1},X)} \\ &= 1 - \frac{S(a_j,X)}{S(a_{j-1},X)} \\ &= 1 - \exp[\lambda(H_{j-1} - H_j)] \end{aligned}$$

which implies

$$\log(1 - h_j(X)) = \lambda(H_{j-1} - H_j)$$

and, hence,

$$\log(-\log[1 - h_j(X)]) = \beta' X + \log(H_j - H_{j-1})$$

Cloglog model as discrete time PH model

Similarly, for the baseline hazard:

$$\begin{aligned}\log\left[-\log(1-h_{0j})\right] &= \log\left(H_j - H_{j-1}\right) \\ &= \log\left[\int_{a_{j-1}}^{a_j} \theta_0(u)du\right] \\ &= \gamma_j, \text{ say}\end{aligned}$$

where $\gamma_j =$ log of integrated hazard over interval $(a_{j-1}, a_j]$

Cloglog model as discrete time PH model

Substituting this back into the expression for overall hazard rate yields

$\log(-\log[1 - h_j(X)]) = \beta' X + \gamma_j$

I.e. cloglog(hazard for interval j) = linear function of characteristics, plus duration-interval-specific parameter

$h(a_j, X) = 1 - \exp[-\exp(\beta' X + \gamma_j)]$

Logistic hazard model as discrete time proportional odds model

Suppose that the relative odds of failure in interval j, conditional on survival to end of interval $j-1$, take the proportional odds form:

$$\frac{h(j,X)}{1-h(j,X)} = \left[\frac{h_0(j)}{1-h_0(j)}\right]\exp(\beta' X)$$

for discrete hazard *hj(X)* for interval j. Hence,

$$logit[h(j,X)] = \log\left[\frac{h(j,X)}{1-h(j,X)}\right] = \alpha_j + \beta' X$$

I.e. logit(hazard for interval j) = linear function of characteristics, plus duration-interval-specific parameter.

Cloglog *vs.* logit

- Interval-specific parameters in each model (γ_j and α_j) may, in principle, differ for each interval $\Rightarrow$ non-parametric duration dependence (but can't then extrapolate out of sample range)
- Alternatively, parameterize.

Cloglog *vs.* logit

- Alternatively, parameterize. E.g.
  - → $r\log(j)$: $r+1 \equiv q$ is analogous to Weibull model's shape parameter. Include $\log(j)$ as additional regressor; estimated r is coeff.
  - → $aj + bj^2 + \ldots zj^n$: n^{th} order polynomial of time (usually with $n = 2$ or 3 at most). E.g. quadratic: define new variables j and *j-squared*
  - → piecewise constant : several intervals have the same constant hazard (rather than differing in every interval). Define a set of dummy variables where each variable identifies a group of spell months. Then either include all the dummy variables in the regression and exclude the constant term, OR include all but one dummy and include constant term

Cloglog *vs.* logit

- The 2 models provide similar estimates if the hazard is 'small':

$$logit(h) = \log\left(\frac{h}{1-h}\right) = \log(h) - \log(1-h)$$

but as $h \to 0$, $\log(1-h) \to 0$ too. In which case like a PH model: $\log[\theta(t)] + \beta' X$

Cloglog *vs.* logit

Interpretation of coefficient β_k:

- *logistic model:* change in proportionate (log) odds of failure given a one unit change in X_k
- *cloglog model:* proportionate change in continuous time hazard (θ) given a one unit change in X_k, or (intrinsically discrete case), proportionate change in cloglog(hazard) given same change.

Describing the distribution of spell lengths

You have estimates of the parameters of a model but what do they imply about the spell length distributions?

- How long are spells?
  - $\rightarrow$ Median and mean spell lengths
- How do spell lengths differ for persons with different characteristics?
- What is the pattern of duration dependence?

Focus here on Weibull model examples; less on other models.

Weibull model

Recall that $\theta(t, X) = \alpha t^{\alpha-1}\lambda$, where $\lambda \equiv \exp(\beta' X)$.

Weibull is PH model, so have all the interpretations that derived for PH models

Weibull model

Weibull is PH model, so have all the interpretations that derived for PH models, e.g.

- β_k is proportionate change in hazard given unit change in X_k
- Elasticity of hazard w.r.t. X_k is $\beta_k X_k$
- Elasticity of hazard w.r.t. t is $\alpha - 1$

- Two persons with same X, at different survival times:

$\theta(t, X)/\theta(u, X) = (t/u)^{\alpha-1}$

- Two persons at same t, different X:

$\theta(t, X_1)/\theta(u, X_2) = \exp(\beta' X_1 - \beta' X_2)$

as for all PH models

Weibull survivor function

We can derive the survivor function from the hazard rate function: $S(t) = exp[-H(t)]$, where

$H(t) = \int_0^t \theta(u)du$

$$H(t) = \int_0^t \theta(u)du$$

So, substitute the Weibull hazard into general expression for S(t):

Weibull survivor function

$$
\begin{aligned}
S(t) &= \exp\left(-\int_0^t \alpha u^{\alpha-1}\lambda du\right) \\
&= \exp\left(-\lambda\alpha\left\{\left.\frac{u^\alpha}{\alpha}\right]_0^t\right\}\right) \\
&= \exp\left(-\lambda\alpha\left[\frac{t^\alpha}{\alpha}-\frac{0^\alpha}{\alpha}\right]\right)
\end{aligned}
$$

Hence,

$S(t, X) = \exp(-\lambda t^\alpha)$

Weibull density and integrated hazard

Density function:

- Since, in general, $f(t) = \theta(t)S(t)$,

$f(t, X) = \alpha t^{\alpha-1}\lambda \exp(-\lambda t^{\alpha})$

Integrated hazard function

- Since, in general, $H(t) = -\ln[S(t)]$,

$H(t, X) = \lambda t^{\alpha}$

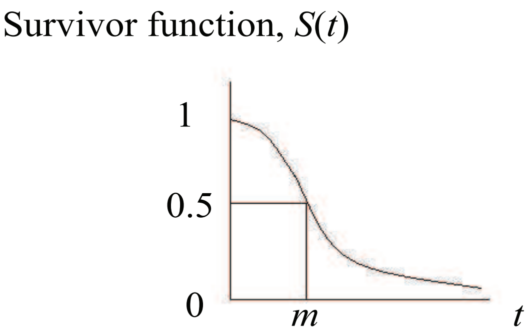
Weibull density and integrated hazard

It therefore follows that

$$\begin{aligned}\log H(t, X) &= \log(\lambda) + \alpha \log(t) \\ &= \beta' X + \alpha \log(t).\end{aligned}$$

Plot $\log(H)$ against $\log(t)$: if Weibull, should see parallel lines with common slope (α).

Weibull: median duration



Median duration: survival time m, such that $S(m) = 0.5$.

Weibull: $\ln[S(t)] = -\lambda t^{\alpha}$

Weibull: median duration

Median duration: survival time m, such that $S(m) = 0.5$.

Weibull: $\ln[S(t)] = -\lambda t^{\alpha}$

So, median satisfies $\ln(0.5) = -\lambda m^{\alpha}$

And, hence:

$$\begin{aligned} m &= \left[\frac{1}{\lambda}[-\log(0.5)]\right]^{\frac{1}{\alpha}} \\ &= \left[\frac{1}{\lambda}\log(2)\right]^{\frac{1}{\alpha}}. \end{aligned}$$

Upper and lower quartiles, etc., derived similarly

Weibull: mean duration

Mean ('expected') duration:

$$E(T) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt$$

Weibull model:

$$E(t) = \left(\frac{1}{\lambda}\right)^{\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)$$

where $\Gamma(z)$ is the Gamma function.

$\Gamma(z) = (z-1)!$ if z is an integer.

$$\begin{array}{lcccc} \text{E.g. } \Gamma(5) & = & 4*3*2*1 & = & 24 \\ \Gamma(4) & = & 3*2*1 & = & 6 \\ \Gamma(3) & = & 2*1 & = & 2 \\ \Gamma(2) & = & 1 & & \end{array}$$

Weibull: mean duration

If z non-integer, use special tables or function built into software.

So, if $\alpha = 0.5$ (negative duration dep.):

$$E(T) = 2/\lambda^2.$$

If $\alpha = 1$ (Exponential model) ,

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

Weibull: median vs. mean

Ratio of mean to the median:

$$ratio = \frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{[\log(2)]^{1/\alpha}}$$

Unless hazard increasing at particularly fast rate ($\alpha \gg 1$), mean $>$ median.

$$\begin{aligned} \text{If } \alpha &= 0.5, & \text{ratio} \approx & \ 4.2 \\ \text{If } \alpha &= 1, & \text{ratio} \approx & \ 1.4 \\ \text{If } \alpha &= 2, & \text{ratio} \approx & \ 1.1 \end{aligned}$$

Weibull: median vs. mean

Elasticity of mean w.r.t. one unit change in X_k

$=$ Elasticity of median w.r.t. one unit change in X_k!

$= \frac{-\beta_k X_k}{\alpha}$

If $X_k \equiv \log(Z_k), -\beta_k/\alpha =$ elasticity of the me(di)an with respect to Z_k.

Loglogistic model

- Loglogistic model is AFT so have all the interpretations that derived for AFT models, e.g.
  - → β_k^* is proportionate change in latent survival time given unit change in X_k
  - → Elasticity of latent survival time w.r.t. X_k is $\beta_k^* X_k$

Loglogistic model

Survivor Function:

$$S(t, X) = \frac{1}{1+(\psi t)^{1/\gamma}}$$

Integrated hazard:

$$H(t) = \log\left[1 + (\psi t)^{\frac{1}{\gamma}}\right]$$

Density function:

$$f(t) = \theta(t)S(t)$$

Loglogistic: median, mean

Median: $m = \frac{1}{\psi} = \frac{1}{\exp(-\beta^{*\prime} X)}$

Mean: $E(T) = \frac{1}{\psi}\frac{\gamma\pi}{\sin(\gamma\pi)}$, $\gamma < 1$

If $\gamma \geq 1$, no closed form.

Elasticity of median w.r.t. one unit change in X_k

$=$ Elasticity of mean w.r.t. one unit change in X_k ($\gamma < 1$ case)

$= \beta_k^* X_k / \psi$

Ratio of mean to median:

$$\frac{\gamma\pi}{\sin(\gamma\pi)}, \ \gamma < 1$$

Loglogistic model: log-odds of survival interpretation

From the definition of $S(t, X)$, the conditional odds of survival to time t are

$$\frac{S(t,X)}{1-S(t,X)} = (\psi t)^{-\frac{1}{\gamma}}$$

When $X = 0$, also true, so:

$$\frac{S(t,X|X=0)}{1-S(t,X|X=0)} = (t\psi_0)^{-\frac{1}{\gamma}}$$

Loglogistic model: log-odds of survival interpretation

Hence,

$$\frac{S(t,X)}{1-S(t,X)} = \left[\frac{S(t,X|X=0)}{1-S(t,X|X=0)}\right]\left(\frac{\psi}{\psi_0}\right)^{-\frac{1}{\gamma}}$$

$$\log\left[\frac{S(t,X)}{1-S(t,X)}\right] = \beta^{*\prime}X - \varphi\log(t)$$

Odds of survival depend on a common 'baseline' odds scaled by person-specific factor. Specification check: graph log odds of survival against $\log(t)$. Parallel lines result.

Discrete time models

Survival up to end of j^{th} interval (or completion of j^{th} cycle):

$$S(j) = S_j = \prod_{k=1}^{j} (1 - h_k)$$

where discrete time hazard, h_k , is a logistic or cloglog function of characteristics and elapsed survival time.

Discrete time models

Median duration:
closed form expressions not usually available. Usually have to derive them numerically.
E.g. calculate $S(t, X)$ and find m s.t. $S(m, X) = 0.5$.

Mean duration: *for maximum time K,*

$$E(T) = \sum_{k=1}^{K} kf(k) = \sum_{k=1}^{K} S(k)$$

Closed forms available if $h(j)$ constant, all j.

Estimation of the survivor and hazard functions

Estimation

- The empirical survivor and hazard rate functions
- Continuous time multivariate models
- Discrete time multivariate models
  - → random sample, with right censoring
  - → left-truncation ('delayed entry'; 'stock sampling with follow-up')
  - → right-truncation (outflow sample)
- Cox's PH model

Empirical survivor and hazard rate functions

- Estimators of
  - → $S(t)$ and $H(t)$ from continuous time survival time data: *Kaplan-Meier (product-limit) estimators*
  - → $S(j)$, $H(j)$, and $h(j)$, from grouped time data: *lifetable estimators*
- Shapes of these functions for population as a whole (or separately for subgroups)
- Assume random samples from population of spells (allow right censoring, but not truncation)

Kaplan-Meier estimators

Definitions:

$t_1 < t_2 < \ldots t_j < t_k < \infty$ the failure times observed in the data

d_j # persons observed to 'fail' (make transition out of state) at time t_j

m_j # persons whose spell length is censored in the interval of time $[t_j, t_{j+1})$

n_j # persons at risk of failure just immediately prior to date t_j

$$n_j = (m_j + d_j) + (m_{j+1} + d_{j+1}) + \ldots + (m_k + d_k)$$

Failure time	# failures	# censored	# at risk of failure
t_1	d_1	m_1	n_1
t_2	d_2	m_2	n_2
t_3	d_3	m_3	n_3
:	:	:	:
t_j	d_j	m_j	n_j
:	:	:	:
t_k	d_k	m_k	n_k

Kaplan-Meier estimators

Failure time	# failures	# censored	# at risk of failure
t_1	d_1	m_1	n_1
t_2	d_2	m_2	n_2
t_3	d_3	m_3	n_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_j	d_j	m_j	n_j
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_k	d_k	m_k	n_k

K-M estimators of $S(t)$ and $H(t)$

Recall that $S(t) = 1 - F(t)$ and $H(t) = -\ln[S(t)]$.

Estimate *survivor function* at each failure time observed in data set using estimator

$$\hat{S}(t_j) = \prod_{j|t_j<t} \left(1 - \frac{d_j}{n_j}\right)$$

$$1- \text{ 'exit rate' at } t_j$$
$$= 1- (\# \text{ exits})/(\# \text{ at risk})$$

$\hat{S}(t_j)$ is product of one minus the exit rate at each of the observed failure times.

K-M estimators of $S(t)$ and $H(t)$

Integrated hazard function

$$\hat{H}(t_j) = -\log \hat{S}(t_j)$$

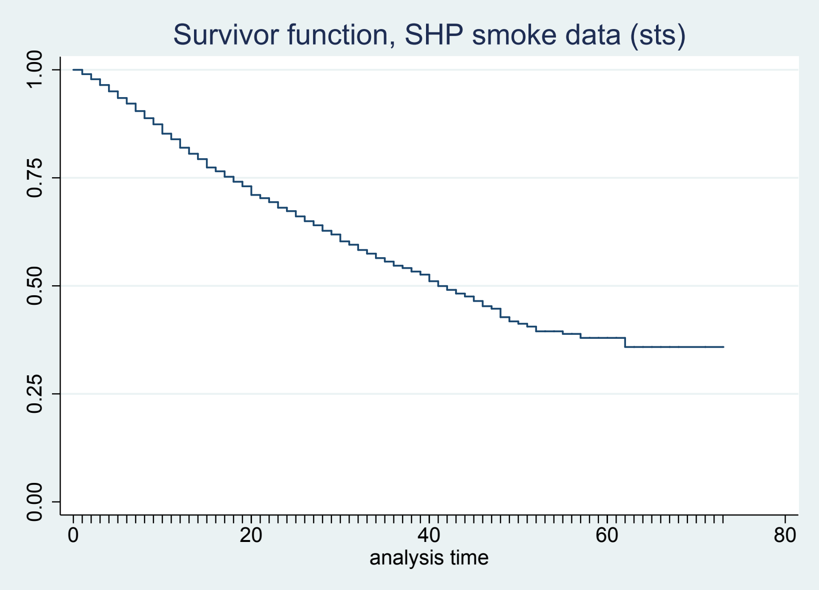
Or Nelson-Aalen estimator:

$$\hat{H}(t_j) = \sum_{j|t_j<t} \left(\frac{d_j}{n_j} \right)$$

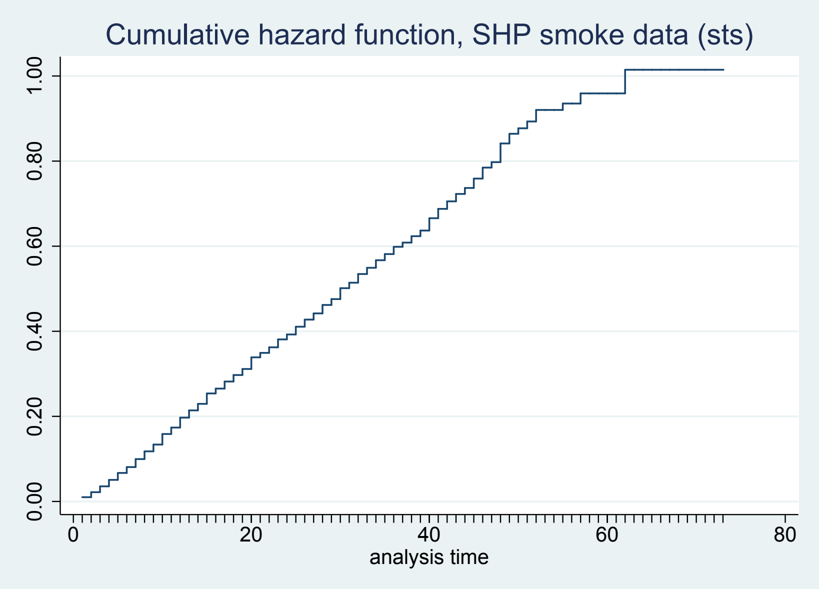
Survivor & integrated hazard functions

- Estimates possible only at the observed failure times!
- Graphs of estimated $S(t)$, $H(t)$, are step functions
- Cannot reliably estimate θ from change in estimated $H(t)$ divided by difference in times – slope of $H(t)$ not well-defined with step-function. Cf. estimators of 'smoothed' hazard rate.

Survivor & integrated hazard functions



Survivor & integrated hazard functions



Lifetable estimators of $S(j)$, $H(j)$, and $h(j)$

Uses broadly same idea, but is applied to discrete (or banded) survival time data:

- Intervals of time $I_k = [t_k, t_{k+1}), k = 1, 2, \ldots, K$, where
  - $\rightarrow$ d_k: # failures in interval I_k
  - $\rightarrow$ m_k: # censored spell endings in interval I_k
  - $\rightarrow$ N_k: # persons at risk of failure at start of I_k

Two approaches:
(1) Interval hazard and survivor functions (or if time is intrinsically discrete)
(2) Continuous time hazard and survivor functions estimated, using additional assumptions, from the grouped data ('actuarial adjustment'), and taken to refer to the time corresponding to the midpoint of each interval

Lifetable estimators of $S(j)$, $H(j)$, and $h(j)$

(1) Interval hazard and survivor functions

Exit rate for the k^{th} interval:

$$\frac{d_k}{n_k}$$

Survivor function for the k^{th} interval:

$$\hat{S}(k) = \prod_{j=1}^{k} \left(1 - \frac{d_k}{n_k}\right)$$

Lifetable estimators of $S(j)$, $H(j)$, and $h(j)$

(2) 'Actuarial adjustment' for (continuous) hazard and survivor functions

- Aim: produce estimates referring to the dates corresponding to each mid-point
- Method: some faiures occur during interval – adjust the number at risk in each interval to take account of this ⇒ 'averaged' estimate centred on interval mid-point.
- Assumption: If transitions evenly spread over interval (uniform density), 50% fail by half-way through I_k.

n_k Adjusted # persons at risk of failure used for midpoint of I_k:

$$n_k = N_k - \frac{d_k}{2}$$

Recall that $S(t) = 1 - F(t)$, $f(t) = \partial F / \partial t$ and $H(t) = -\ln[S(t)]$.

Lifetable estimators of $S(j)$, $H(j)$, and $h(j)$

(2) 'Actuarial adjustment' (ctd.)

Estimator of $S(k)$:

$$\hat{S}(k) = \prod_{j=1}^{k} \left(1 - \frac{d_k}{n_k}\right)$$

Estimator of $f(k)$:

$$\hat{f}(k) = \frac{\hat{F}(k+1) - \hat{F}(k)}{t_{k+1} - t_k} = \frac{\hat{S}(k) - \hat{S}(k+1)}{t_{k+1} - t_k}$$

Estimator of hazard:

$$\hat{\theta}(k) = \frac{\left[\hat{f}(k)\right]}{\tilde{S}(k)}$$

where $\tilde{S}(k) = \frac{\hat{S}(k) + \hat{S}(k+1)}{2}$ is taken as applying to the time corresponding to midpoint of the interval.

If data are intrinsically discrete, or *interval hazard* required, there's no need to 'adjust' (same as K-M).

Estimation using Stata

See web course Lessons

- Kaplan-Meier (and related) estimators:
  - → **stset** the survival data, and then use **sts** commands (sts, sts graph, sts generate)
- Lifetable estimators
  - → **ltable** command (by default produces actuarilly adjusted estimates)
  - → Use **noadjust** option otherwise (compare with **sts** applied to the same data)

Multivariate regression models

- Estimates for both continuous and discrete time models derived using the method of 'maximum likelihood' (ML) [except Cox's model]. Cf. problems with OLS discussed at start; ML plus basic concept allows us to derive good estimates.
- Likelihoods need to be appropriate for the data generation process (differing with sampling scheme to account for right censoring, truncation, etc.)

Maximum likelihood principle

Cf. OLS principle choose as parameter estimates, those which minimize the residual sum of squares.

ML principle choose as parameter estimates, those which maximize the likelihood of observing the distribution of data in the sample:

- Likelihood contribution for each person $i = 1, 2, \dots n$, (cf. "probability of observing i's data"): $\mathcal{L}_i(\delta)$
  - $\rightarrow$ where δ is a vector of parameters (e.g. β, and shape parameters).

Maximum likelihood principle

Sample likelihood:

$$\mathcal{L}(\delta) = \prod_{i=1}^{n} \mathcal{L}_i(\delta)$$

ML estimators $\hat{\delta}$ are the values for which $\mathcal{L}(\hat{\delta})$ or, equivalently $\log \mathcal{L}(\hat{\delta})$ are maximised.

Get standard error estimates too, from 2nd order conditions. ML estimates are consistent, and asymptotically efficient (given correct distributional assumption).

Likelihoods for continuous time data

(1) Random sample from inflow, each spell monitored until completion (no right-censoring)

Individual likelihood contributions given by the relevant density function; sample likelihood is their product:

$$\mathcal{L} = \prod_{i=1}^{n} f(T_i)$$

where T_i is length of completed spell for person i. Choice of model fixes f(t) form.

Likelihoods for continuous time data

(2) Random sample from inflow with right censoring at common time t^*

$$\mathcal{L} = \prod_{j=1}^{J} f(T_j) \prod_{k=1}^{K} S(t^*)$$

Density function Survivor function

Completed spells indexed by $j = 1, \ldots, J(T_j \leq t^*)$ censored spells indexed by $k = 1, \ldots, K(T_k \succ t^*)$

Likelihoods for continuous time data

(3) Random sample from population, right censoring but varies across persons (most common case in lit.)

$$\mathcal{L} = \prod_{j=1}^{J} f(T_j) \prod_{k=1}^{K} S(T_k)$$

L is often written differently in this case (to facilitate the estimation). Taking logs,

Likelihoods for continuous time data

$$
\begin{aligned}
\ln \mathcal{L} &= \sum_{j=1}^{J} \ln f(T_j) + \sum_{k=1}^{K} \ln S(T_k) \\
&= \sum_{j=1}^{J} \ln \left[\left(\frac{f(T_j)}{S(T_j)} \right) S(T_j) \right] + \sum_{k=1}^{K} \ln S(T_k) \\
&= \sum_{j=1}^{J} \ln \left[\theta(T_j) S(T_j) \right] + \sum_{k=1}^{K} \ln S(T_k) \\
&= \sum_{j=1}^{J} \ln \theta(T_j) + \sum_{i=1}^{N} \ln S(T_i)
\end{aligned}
$$

Likelihoods for continuous time data

$$= \sum_{j=1}^{J} \ln \theta(T_j) + \sum_{i=1}^{N} \ln S(T_i)$$

$$= \sum_{i=1}^{N} [c_i \ln \theta(T_i) + \ln S(T_i)]$$

$\ln S(T_i) = -H(T_i)$

Censoring indicator: $c_i = 1$ if spell completed
$= 0$ if spell censored.

Likelihoods for continuous time data

(4) Left truncated spell data ('delayed entry'), with right censoring

- Most common social science example is where one has a *sample from the stock* of persons in the state of interest at one date, plus follow-up of sample to some later date.
- Spell start dates are assumed known (they're before the sampling date.)
- 'Delayed entry' because observation of subjects starts some time after first at risk of event.
- Non-random sample. We have to condition on the fact that a person survived sufficiently long in the state in order to be at risk of being sampled from the stock. Else have a 'selection bias' related to spell length.
- 'Left-truncation': short spells under-represented; long spells over-represented

Left-truncated data

- We need to derive the contribution to the likelihood for each person, *conditioning* on survival up to the date of truncation
- Recall rule of conditional probability: $\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$
  - A: "extra time observed in state after truncation date"
  - B: "survival to truncation date"
- Letting
  - T_i spell length for i at truncation date
  - Z_i length of time between truncation date and interview (= extra time observed if a right-censored case)
  - Δt extra time observed if a completed spell case

Sample likelihood is

$$\mathcal{L} = \prod_{j=1}^{J} \frac{f(T_j + \Delta t_j)}{S(T_j)} \prod_{k=1}^{K} \frac{S(T_k + Z_k)}{S(T_k)}$$

completed spells   censored spells

Left-truncated data

(5) Sample from stock with no follow-up

- E.g. data for people in stock in cross-section survey (for whom ask spell start date). Used to be common (lack of longitudinal surveys).
- Have no information on which to condition survival in state. So, to derive the likelihood, have to write down probability of observing a given spell taking account of the different chances of entering state at different dates.
- Very complicated (see e.g. Nickell 1979, and Lecture Notes).

Likelihoods for continuous time data

(6) Right-truncated spell data

- Most common social science example is where one has a *sample of the outflow* of persons from the state of interest
- Spell start dates are known, and all spells are complete (outflow sample!)
- Non-random sample. We have to condition on the fact, of all those beginning a spell at some date in past, outflow likely to have an over-representation of relatively short spells (long-stayers are still in the state). A 'selection bias' related to spell length.

Likelihoods for continuous time data

- ‘Right-truncation’: long spells under-represented; short spells over-represented.
- Condition on failure at outflow date:

$$\mathcal{L} = \prod_{i=1}^{n} \frac{f(T_i)}{F(T_i)}$$

Estimation using Stata

- Models for all the sampling schemes discussed above (with the exception of the right-truncation, and stock-sample-no-follow-up) can be estimated using **streg** command:
- **stset** the survival time data
  - → includes setting entry date for left-truncated data
- use **streg**, with relevant options to choose e.g.
  - → the model (Weibull through Gamma, etc.)
  - → metric for reporting parameters (coeffs. *vs* hazard ratios, etc)

"Episode splitting" to incorporate time-varying covariates

Return to case (2) = Random sample from inflow, with right censoring.
We assumed

- all explanatory variables constant
- data set organised so that have one row for each individual at risk of failure.

Incorporating time-varying covariates requires *episode splitting*:

- Split the survival time for each person into sub-periods within which each TVC is constant.
- Create multiple records for each person, with one record for each subperiod

What is the logic behind this?

Episode splitting, ctd.

Consider a person i with 2 different values for covariate X:

$$X = X_1 \text{ if } t < u$$
$$X = X_2 \text{ if } t \geq u$$

Log-likelihood contribution for person in sample of type (2) is

$$\ln \mathcal{L}_i = c_i \ln[\theta(T_i)] + \ln[S(T_i)]$$

with censoring indicator $c_i = 1$ if complete spell, 0 otherwise.

Episode splitting, ctd.

$\ln \mathcal{L}_i = c_i \ln[\theta(T_i)] + \ln[S(T_i)]$

But,log survival to T_i:

$$\begin{aligned} \ln[S(T_i)] &= \ln\left[S(u)\frac{S(T_i)}{S(u)}\right] \\ &= \ln[S(u)] + \ln\left[\frac{S(T_i)}{S(u)}\right] \end{aligned}$$

Log probability of survival from entry until time u: **create new record with $c_i = 0$, $t = u$**

Log probability of survival to T_i *conditional on entry at time u*. **One record with 'delayed entry' at** u, c_i set at 0 or 1 (depending on whether censored)

Episode splitting, ctd.

Episode splitting reorganises the data so that records yield the correct likelihood contributions

Record #	Censoring indicator, c_i	Survival time	Entry time	TVC value
Single record for i				
1	0 or 1	T_i	0	–
Multiple records for i, after episode splitting				
1	0	u	0	X_1
2	0 or 1	T_i	u	X_2

Episode splitting, ctd.

Record #	Censoring indicator, c_i	Survival time	Entry time	TVC value
Single record for i				
1	0 or 1	T_i	0	–
Multiple records for i, after episode splitting				
1	0	u	0	X_1
2	0 or 1	T_i	u	X_2

Likelihood contributions:

$$\begin{aligned} \text{Record 1: } \ln \mathcal{L}_i &= S(u, X_1) \\ \text{Record 2: } \ln \mathcal{L}_i &= c_i \ln[\theta(T_1, X_2)] + \ln[S(T_1, X_2)/S(u, X_1)] \end{aligned}$$

Stata and TVCs

- episode split using **stsplit** (thereby updating **stset**)
- create the relevant TVCs
- estimate (as above)

Likelihoods for discrete time data

(1) Random sample, with right censoring.

- Likelihood contribution for a censored spell is:

$$\begin{aligned} \mathcal{L}_i &= \Pr(T_i > j) = S_i(j) \\ &= \prod_{k=1}^{j} (1 - h_{ik}) \end{aligned}$$

- Likelihood contribution for a completed spell is:

$$\begin{aligned} \mathcal{L}_i &= \Pr(T_i = j) = f_i(j) \\ &= h_{ij} S_i(j-1) \\ &= \tfrac{h_{ij}}{1-h_{ij}} \prod_{k=1}^{j} (1 - h_{ik}) \end{aligned}$$

Likelihoods for discrete time data

censored spell

$$\mathcal{L}_i = \prod_{k=1}^{j} (1 - h_{ik})$$

completed spell

$$\mathcal{L}_i = \frac{h_{ij}}{1-h_{ij}} \prod_{k=1}^{j} (1 - h_{ik})$$

Likelihoods for discrete time data

Hence likelihood contribution for whole sample is . . . :

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [\Pr(T_i = j)]^{c_i} [\Pr(T > j)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{h_{ij}}{1 - h_{ij}} \right) \prod_{k=1}^{j} (1 - h_{ik}) \right]^{c_i} \left[\prod_{k=1}^{j} (1 - h_{ik}) \right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{h_{ij}}{1 - h_{ij}} \right)^{c_i} \prod_{k=1}^{j} (1 - h_{ik}) \right] \\
\Rightarrow \log \mathcal{L} &= \sum_{i=1}^{n} c_i \log \left(\frac{h_{ij}}{1 - h_{ij}} \right) + \sum_{i=1}^{n} \sum_{k=1}^{j} \log(1 - h_{ik})
\end{aligned}$$

Likelihoods for discrete time data

Now define a new binary indicator variable $y_{ik} = 1$ if person i experiences event in interval k, and $y_{ik} = 0$ otherwise.

I.e.

$$c_i = 1 \implies y_{ik} = 1 \text{ for } k = T_i,\ y_{ik} = 0 \text{ otherwise}$$
$$c_i = 0 \implies y_{ik} = 0 \text{ for all } k$$

Hence, ...

Likelihoods for discrete time data

Sample likelihood

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{n}\sum_{k=1}^{j} y_{ik}\log\left(\frac{h_{ik}}{1-h_{ik}}\right) + \sum_{i=1}^{n}\sum_{k=1}^{j}\log(1-h_{ik}) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{j}[y_{ik}\log h_{ik} + (1-y_{ik})\log(1-h_{ik})]\end{aligned}$$

Expression has same form as standard likelihood function for a binary regression model in which y_{it} is the dependent variable, and data structure reorganized from having one record per spell to one one record for each interval that each person is at risk of transition from state (person-period data structure)

Likelihoods for discrete time data

Example of person and person-period data:

Person data structure			*Person-period data structure*				
Person *i*, id #	c_i	T_i	Person *i*, id #	c_i	T_i	y_{it}	Person -month *k*, id #
1	0	2	1	0	2	0	1
			1	0	2	0	2
2	1	3	2	1	3	0	1
			2	1	3	0	2
			2	1	3	1	3
:	:	:	:	:	:		

Likelihoods for discrete time data

"Easy estimation" method:

1. Reorganize data into person-period format (= episode splitting at each interval).
   In Stata, use **expand** or **stsplit**
2. Create any time-varying covariates; at minimum this includes a variable to describe duration dependence
3. Choose functional form for hazard (logistic or cloglog)
4. Estimate using relevant binary depvar program
   (In Stata, **logit**, **cloglog**)

Likelihoods for discrete time data

(2) Left-truncated spell data ('delayed entry')

- (analogous to continuous time case) need to condition on survival up to truncation point. Call this u_i for person i:
- I.e. divide likelihood contribution derived in previous case by Pr(survival to time u_i)

$$\mathcal{L}_i = \frac{\left(\frac{h_{ij}}{1-h_{ij}}\right)^{c_i} \prod_{k=1}^{j} (1-h_{ik})}{S(u_i)}$$

But

$$S(u_i) = \prod_{k=1}^{u_i} (1-h_{ik})$$

Left truncation case, *ctd.*

Hence, get 'convenient cancelling' result:

$$\begin{aligned}\mathcal{L}_i &= \left(\tfrac{h_{ij}}{1-h_{ij}}\right)^{c_i} \left[\frac{\prod\limits_{k=1}^{j}(1-h_{ik})}{\prod\limits_{k=1}^{u_i}(1-h_{ik})}\right] \\ &= \left(\tfrac{h_{ij}}{1-h_{ij}}\right)^{c_i} \prod_{k=u_i+1}^{j}(1-h_{ik})\end{aligned}$$

Left truncation case, *ctd.*

Taking logs, we have:

$$\log \mathcal{L}_i = \sum_{k=u_i+1}^{j} [y_{ik} \log h_{ik} + (1 - y_{ik}) \log(1 - h_{ik})]$$

which is very similar to the expression in the no-truncation case, except that the summation now runs over the intervals from month of truncation to month when last observed. So implement . . .

Left truncation case, *ctd.*

"Easy estimation" method with left-truncated data:

1. Reorganize data into person-period format (= episode splitting at each interval).
   In Stata, use **expand** or **stsplit**
2. *Drop the records for all records corresponding to intervals before truncation period u*
3. Create any time-varying covariates; at minimum, this includes a variable to describe duration dependence
4. Choose functional form for hazard (logistic or cloglog)
5. Estimate using relevant binary depvar program. (In Stata, **logit**, **cloglog**.)

The only difference from before is Step 2 (throwing away some data).

Likelihoods for discrete time data, *ctd.*

(3) Right-truncated spell data (outflow sample)

- Parallels continuous time data case again.
- No censored spells (all completed), by construction, but
- need to account for selection bias arising from sampling: condition each individual's likelihood contribution on failure at observed failure time.

$$\mathcal{L}_i = \frac{\left(\frac{h_{ij}}{1-h_{ij}}\right)\prod_{k=1}^{j}(1-h_{ik})}{1-\left[\prod_{k=1}^{j}(1-h_{ik})\right]} \quad \frac{\text{Density}}{\text{failure}}$$

Unfortunately no 'convenient cancelling' result in this case: special programs required.

Cox's PH model

- General PH specification

$$\begin{aligned}\theta(t, X_i) &= \theta_0(t)\exp(\beta' X_i)\\ &= \theta_0(t)\lambda_i\end{aligned}$$

- Cox model: Unspecified (non-parametric) baseline hazard function $\theta_0(t)$
- Estimated using method of 'partial likelihood' (PL), not ML
- Intuitive illustration here of how model works
- Assume: random sample with right-censoring (but no truncation), and
- no time-varying covariates
- only one event (max.) at each survival time

Cox's PH model

PL works in terms of ordering of events, and their occurrence (cf. focus in ML on persons)

Example data

Person i	Time t_i	Event # k
1	2	1
2	4	2
3	5	3
4	5*	
5	6	4
6	9*	
7	11	5
8	12*	

Sample Partial Likelihood

$$PL = \prod_{k=1}^{K} L_k$$

K ← # events (total)

$k = 1$ ← indexes events

$\Rightarrow$ What is each L_k?

Cox's PH model

Example data

Person i	Time t_i	Event # k
1	2	1
2	4	2
3	5	3
4	5*	
5	6	4
6	9*	
7	11	5
8	12*	

*censored spell

Sample Partial Likelihood

$$PL = \prod_{k=1}^{K} L_k$$

K ← # events (total)

$k = 1$ ← indexes events

$\Rightarrow$ What is each L_k?

$$\begin{aligned} L_k &= \text{Pr(person } i \text{ has event at } t = t_i\text{, conditional on being in the risk set at } t = t_i\text{), i.e.} \\ &= \text{Pr(this particular person has event at this time, given that there is one obs} \end{aligned}$$

Cox's PH model

Evaluate using rules of conditional probability and fact that $f(t) = \theta(t)S(t)$.

Pr(event in tiny interval) $(t, t + \Delta t] = f(t)dt = \theta(t)S(t)dt$

Example data

Person i	Time t_i	Event # k
1	2	1
2	4	2
3	5	3
4	5*	
5	6	4
6	9*	
7	11	5
8	12*	

*censored spell

Sample Partial Likelihood

$$PL = \prod_{k=1}^{K} L_k$$

K ← # events (total)

$k = 1$ ← indexes events

$\Rightarrow$ What is each L_k?

Consider event $k = 5$ with risk set $i \in \{7, 8\}$.

Cox's PH model

Consider event $k = 5$ with risk set $i \in \{7, 8\}$.

$$\begin{aligned} \text{Let A} &= \text{Pr(event experienced by } i = 7 \text{ and not } i = 8) \\ &= [\theta_7(11) S_7(11) dt]\ [S_8(11) dt] \end{aligned}$$

$$\begin{aligned} \text{Let B} &= \text{Pr(event experienced by } i = 8 \text{ and not } i = 7) \\ &= [\theta_8(11) S_8(11) dt]\ [S_7(11) dt] \end{aligned}$$

Expression for probability A conditional on the probability of either A or B (sum of probabilities A and B)

Cox's PH model

PL(event # 5) $= \frac{\theta_7(11)}{\theta_7(11)+\theta_8(11)}$

NB Survivor function terms cancel!

- Can apply same idea to derive PL contributions for all the other events. E.g.

PL(event # 1) $= \frac{\theta_1(2)}{\theta_1(2)+\theta_2(2)+...+\theta_8(2)}$

(Everyone is in the risk set for the first event.)

Now let us apply the PH assumption about shape of the hazard, i.e. $\theta(t, X) = \theta_0(t) \exp(\beta' X)$

Cox's PH model

Substituting in,

$$\text{Pl(event \# 5)} = \frac{\theta_0(11)\lambda_7}{\theta_0(11)\lambda_7 + \theta_0(11)\lambda_8} = \frac{\lambda_7}{\lambda_7 + \lambda_8}$$

The baseline hazard contributions cancel! (So too does the intercept term in λ, i.e. $\exp(\beta_0)$)

Similarly,

$$\text{PL(event \# 1)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_8}$$

and so on for all the other events.

- Given the PL for whole sample, maximise it to derive estimates of the slope coefficients in β. (Estimator has nice properties.)...

Cox's PH model

- NB baseline hazard $\beta_0(t)$ not identified. Can take any form. (flexibility = advantage ... or disadvantage if baseline shape is of intrinsic interest!)
- Estimate in Stata using **stcox** (after **stset**)
- With tied survival times, various approximations to PL available
- Can incorporate TVCs: only need to episode split (**stsplit**) at failure times, since estimates based only on risk pool at those times (and not anything that happens in between)
- NB expression for each PL contribution depends only on order of occurrence (not precise survival times) Exercise: double all survival times, and repeat derivations above.

Additional Topics

(a) Unobserved heterogeneity ('frailty')

(b) Independent competing risks models

(c) not at present

... (e.g. model specification tests, modelling repeated spells, more complicated data set-up examples, case studies; etc.)

Unobserved heterogeneity (‘frailty’)

- In regression models so far, all differences between individuals assumed to be captured using the measured X
- Generalize to allow for unobserved individual effects (‘frailty’), because e.g.:
  - $\rightarrow$ omitted variables (unobserved or unobservable)
  - $\rightarrow$ measurement errors in observed survival times or regressors

Unobserved heterogeneity ('frailty')

What if important, but ignored?

The literature suggests:

1. 'no frailty' model over-estimates the degree of negative duration dependence (under-estimates the degree of positive duration dependence)
2. Proportionate response of the hazard to a unit change in X_k is no longer constant ($= \beta_k$ in earlier PH models): declines with time
3. one gets an under-estimate of the true proportionate response of the hazard to a change in a regressor k from the no-frailty-model β_k

Unobserved heterogeneity ('frailty')

Now examine frailty specifications, and these findings, in more detail ...

For convenience we suppress the subscript indexing individuals, and assume there are no time-varying covariates.

We consider the model $\theta(t, X|v) = v\theta(t, X)$

- v is unobservable individual effect; it scales the no-frailty hazard rate, where
- $v > 0$ $\mathrm{E}(v) = 1$, finite variance $\sigma^2 > 0$, and is distributed independently of t and X

Unobserved heterogeneity ('frailty')

- Relationship between frailty and non-frailty survivor function

$$S(t, X|v) = [S(t, X)]^v$$

- Interpretation: Individuals with above-average v exit relatively fast (higher hazard, shorter survival time). Vice-versa for those with below-average v
- if $\theta()$ has PH form, then

$$\log[\theta(t, X|v)] = [\log \theta_0(t)] + \beta' X + u$$

where $u \equiv \log(v)$. I.e. frailty model for log-hazard adds an additive 'error' term (random intercept model)

Estimation with frailty models

- How does one estimate frailty models, given that the individual effect is unobserved? (We cannot estimate each v since, by construction, they are unobserved.)
- Suppose the distribution of v has a shape whose functional form is summarised in terms of only a few key parameters, then we can estimate those parameters with the data available.

Estimation with frailty models

The steps are:

1. Specify a distribution for the random variable v, where this 'mixing' distribution has a particular parametric functional form, e.g. summarising var(v)
2. Derive the 'frailty' survivor function corresponding to this 'mixture' distribution (i.e. 'integrate out' the random individual effect).
3. Using this, and the frailty hazard, derive the likelihood function: it refers to the mixing distributional parameter(s), and the original parameters

Frailty survival function

- One works with some function

$$S_\upsilon(t, X) = S_\upsilon(t, X|\beta, \sigma^2)$$

rather than $S_\upsilon(t, X|\beta, \upsilon)$

- Given some probability density $g(\upsilon)$ for υ, then

$$S_\upsilon(t, X) = \int_0^\infty [S(t, X)]^\upsilon g(\upsilon) d\upsilon$$

Frailty survival function

- Most commonly-used $g(\upsilon)$ is Gamma distribution, implying *frailty survivor function*:

$$\begin{aligned} S(t, X|\beta, \sigma^2) &= [1 - \sigma^2 \ln S(t, X)]^{-\frac{1}{\sigma^2}} \\ &= [1 + \sigma^2 H(t, X)]^{-\frac{1}{\sigma^2}} \end{aligned}$$

- And *frailty hazard function*

$$\begin{aligned} \theta(t, X|\upsilon) &= \theta(t, X)[1 - \sigma^2 \ln S(t, X)]^{-1} \\ &= \theta(t, X)[1 + \sigma^2 H(t, X)]^{-1} \end{aligned}$$

Frailty in a discrete time model

- Discrete time PH (cloglog) hazard model (Meyer 1990):

$$\text{cloglog}[h(j, X|\upsilon)] = D(j) + \beta' X + u$$

  where $u \equiv \log(\upsilon)$, as before.
- Assuming Gamma mixture for frailty, then there is a closed-form frailty survivor function (as earlier), and one uses these to construct the likelihood contributions of persons with a censored spell (frailty survivor function), and with a completed spell (discrete time density = difference between frailty survival probabilities over the interval).
- Alternatively, suppose $u \sim$ Normal distribution. Integrating out is then done numerically
- Non-parametric approach: see over

A non-parametric approach

- Heckman & Singer (1984): fit an arbitrary discrete distribution characterized by a set of 'mass points' along the support, and corresponding probabilities of being located at each of these points. Cf. sociologists' latent class model.

- Example using discrete time PH model hazard. Suppose individuals belong to one of two possible unobserved 'types' (fast and slow leavers). Allow intercept in hazard to differ between classes:

A non-parametric approach

Allow intercept in hazard to differ between classes:

$$h_1(j, X) = 1 - \exp[-exp(\mu_1 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \gamma_j)]$$
$$h_2(j, X) = 1 - \exp[-exp(\mu_2 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \gamma_j)]$$

Likelihood contribution for individual:

$$\mathcal{L} = \pi \mathcal{L}(\mu_1) + (1 - \pi)\mathcal{L}(\mu_2)$$

with L_1 constructed from h_1, L_2 from h_2, and π is Pr(person is Type 1).

Can generalize to more types.

What if frailty is important, but ignored?

- Suppose 'true' model, with no omitted regressors, takes PH form, and continuous mixing distribution. Then...
- Proportionate response of hazard to an observed variable X_k is β_k:

$$\frac{\partial \log[\theta(t,X|\upsilon)]}{\partial X_k} = \beta_k$$

Does not depend on t or X.

- Also proportionate change in hazard with time depends only on the baseline hazard

$$\frac{\partial \log[\theta(t,X)\upsilon]}{\partial t} = \frac{\partial \log \theta_0(t,X)}{\partial t}$$

What if frailty is important, but ignored?

- Suppose observed model, i.e. with omitted regressors, takes PH form

$$\theta(t, X|\upsilon) = \upsilon\theta_0(t)\lambda_1$$

where $\lambda_1 = e^{\beta_1' X_1}$, X_1 is a subset of X.

- Assuming $\upsilon \sim \text{Gamma}(1, \sigma^2)$, then

$$S_1(t|\sigma^2) = [1 - \sigma^2 S(t)]^{-\frac{1}{\sigma^2}} = [1 + \sigma^2 H(t)]^{-\frac{1}{\sigma^2}}$$

$$\theta_1(t|\sigma^2) = [S_1(t|\sigma^2)]^{\sigma^2}\theta_0(t)\lambda_1$$

Implications of ignoring frailty: duration dependence effect

- Ratio of hazards from the observed and true models:

$$\frac{\theta_1}{\theta} = \frac{\theta_0(t)\lambda_1[S_1(t|\sigma^2)]^{\sigma^2}}{\theta_0(t)\upsilon}$$

$$\infty[S_1(t|\sigma^2)]^{\sigma^2}$$

  which is monotonically decreasing with t.
- I.e. the hazard rate from a model with omitted regressors increases less fast, or falls faster, that does the ‘true’ hazard (from the model with no omitted regressors)

Implications of ignoring frailty: duration dependence effect

Intuition: a selection or ‘weeding out’ effect.

- Controlling for observable differences, people with unobserved characteristics associated with higher exit rates leave the state more quickly than others.
- Hence ‘survivors’ at longer t increasingly comprise those with low v which, in turn, implies a lower hazard, and the estimate of hazard is an underestimate of ‘true’ one.
- Bergström & Edin (1992) illustration using AFT representation of Weibull model:

Implications of ignoring frailty: duration dependence effect

- Bergström & Edin (1992) illustration using AFT representation of Weibull model:

$$\log(T) = \beta^{*\prime}X + \sigma u$$

with variance of ‘residuals’ $= \sigma^2 \text{var}(u)$.

If one added heterogeneity to the systematic part of model, would expect smaller error variance, i.e. smaller σ.

But $\sigma \equiv 1/\alpha$, where α is shape parameter.
So $\sigma \downarrow$ like $\alpha \uparrow$, and ‘true’ model has more positive duration dependence than model without this heterogeneity

Implications of ignoring frailty: proportionate response of hazard

- Consider the proportionate response of the hazard to a variation in X_k where X_k is an included regressor (part of X_1).
- Proportionate response in the *true model*:

$$\frac{\partial \log \theta}{\partial X_k} = \beta_k$$

- Proportionate response in the *observed model* (see Lecture Notes for derivation):

$$\frac{\partial \log \theta_1}{\partial X_k} = \beta_k \left[S_1(t|\sigma^2)^{\sigma^2} \right]$$

NB $0 \leq S_1() \leq 1$, and tends to 0 as $t \to \infty$. So, ...

Implications of ignoring frailty: proportionate response of hazard

1. Omitted-regressor model provides under-estimate (in modulus) of the 'true' proportionate response;
2. With omitted regressors, proportionate effect tends to zero as $t \to \infty$

'Weeding out' effect again (see Notes; Lancaster 1990)

Frailty in practice

- The results suggest that taking accounting of unobserved heterogeneity is a potentially important
- The ‘early’ empirical social science literature found that conclusions about whether or not frailty was ‘important’ (effects on estimate of duration dependence and estimates of β) appeared to be sensitive to choice of shape of the mixing distribution.
- Some argued that the choice of distributional shape was essentially ‘arbitrary’, and this stimulated the development of non-parametric methods (Heckman-Singer, etc.).

Frailty in practice

- Subsequent empirical work suggests, however, that the effects of unobserved heterogeneity are mitigated, and thence estimates more robust, if the analyst uses a flexible baseline hazard specification.

  → Earlier literature had typically used specifications, often the Weibull one, that were not flexible enough.

- All in all, the topic underscores the importance of getting good data, including a wide range of explanatory variables that summarize well the differences between individuals.

Independent competing risks (ICR) models

Independent competing risks (ICR) models

- Until now, we have modelled a *single risk* (exit from current state to any other).
- Now we consider possibility of exit to one of several destination states (*competing risks*)
- Suppose, for illustration, just two exit states, but the arguments generalize to any number
- Continuous and discrete time models considered
- We will see that the assumption of *independence in competing risks* aids estimation of models

ICR model (Continuous time)

Two destination states: A, B (C *censored*)

Define

$\theta_A =$ (latent) hazard rate of exit to state A, with survival times characterized by density function $f_A(t)$, and latent failure time T_A;

$\theta_B =$ (latent) hazard rate of exit to state B, with survival times characterized by density function $f_B(t)$, and latent failure time T_B;

ICR model (Continuous time)

- Observed failure time $T = min\{T_A, T_B, T_C\}$
- Assume θ_A, θ_B are independent
  $\Rightarrow \theta(t) = \theta_A(t) + \theta_B(t)$

If probabilities A,B are independent, then $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$.
Ditto $\theta(t)dt = \theta_A(t)dt + \theta_B(t)dt$

ICR model (Continuous time)

Independence $\Rightarrow$ survivor function for exit to any destination state can be factored into a product of destination-specific survivor functions:

$$
\begin{aligned}
S(t) &= \exp\left[-\int_0^t \theta(u)du\right] \\
&= \exp\left[-\int_0^t [\theta_A(u)+\theta_B(u)]du\right] \\
&= \exp\left[-\int_0^t \theta_A(u)du\right]\exp\left[-\int_0^t \theta_B(u)du\right] \\
&= S_A(t)S_B(t)
\end{aligned}
$$

ICR model (Continuous time)

An individual's contribution to the sample likelihood for ICR model with two destination states is one of 3 types:

- **exit to A:** $L^A = f_A(T)S_B(T)$
- **exit to B:** $L^B = f_B(T)S_A(T)$
- **censored spell:** $L^C = S^A(T)S^B(T)$

L^A summarises chances of transition to A combined with no transition to B; similarly for L^B.

ICR model (Continuous time)

Now define new censoring indicators:

δ^A $= 1$ if i exists to A, 0 otherwise

(exit to B or censored)

δ^B $= 1$ if i exits to B, 0 otherwise

(exit to A or censored)

ICR model (Continuous time)

The overall likelihood contribution from individual i is:

$$\begin{aligned}
\mathcal{L} &= (\mathcal{L}^A)^{\delta^A}(\mathcal{L}^B)^{\delta^B}(\mathcal{L}^C)^{1-\delta^A-\delta^B} \\
&= [f_A(T)S_B(T)]^{\delta^A}[f_B(T)S_A(T)]^{\delta^B}[S_A(T)S_B(T)]^{1-\delta^A-\delta^B} \\
&= \left[\frac{f_A(T)}{S_A(T)}\right]^{\delta^A} S_A(T)\left[\frac{f_B(T)}{S_B(T)}\right]^{\delta^B} S_B(T) \\
&= \left\{[\theta_A(T)]^{\delta^A} S_A(T)\right\}\left\{[\theta_B(T)]^{\delta^B} S_B(T)\right\} \\
\Rightarrow \ln\mathcal{L} &= \{\delta^A \ln\theta_A(T) + \ln S_A(T)\} + \{\delta^B \ln\theta_B(T) + \ln S_B(T)\}
\end{aligned}$$

ICR model (Continuous time)

$$\Rightarrow \ln \mathcal{L} = \{\delta^A \ln \theta_A(T) + \ln S_A(T)\} + \{\delta^B \ln \theta_B(T) + \ln S_B(T)\}$$

- Thus, the (log)likelihood for continuous time ICR with 2 destination states factors into two parts, each of which depends only on parameters specific to that destination.

ICR model
(Continuous time)

$\rightarrow$ Easy estimation method for ICR:

1. define new censoring variables (δ^A, δ^B), and
2. estimate separate models for each destination state.
3. Overall model likelihood = sum of likelihoods for each of the destination-specific models

- Problem: if want to test restrictions across destination-specific hazards, need to return to estimating jointly?
- But see Narendranathan & Stewart tests

ICR model (Intrinsically discrete time)

- In continuous time, exits to only one destination state are feasible at any instant. Hence ICR assumption $\Rightarrow \theta(t) = \theta^A(t) + \theta^B(t)$; separability result
- Intrinsically discrete time process : exits to each destination only feasible at each cycle, so $h(j) = h_A(j) + h_B(j)$
- but separability result no longer holds!

ICR model (Intrinsically discrete time)

Likelihood contributions ...

$$
\begin{aligned}
\mathcal{L}^A &= h_A(j)S(j-1) \\
&= \left[\frac{h_A(j)}{1-h(j)}\right] S(j) \\
&= \left[\frac{h_A(j)}{1-h_A(j)-h_B(j)}\right] S(j)
\end{aligned}
$$

Similarly for $\mathcal{L}^B$, and

$$
\mathcal{L}^C = S(j) = \prod_{k=1}^{j} [1 - h_A(k) - h_B(k)]
$$

ICR model (Intrinsically discrete time)

The overall likelihood contribution of an individual with an spell length of j cycles is:

$$\begin{aligned}\mathcal{L} &= (\mathcal{L}^A)^{\delta^A}(\mathcal{L}^B)^{\delta^B}(\mathcal{L}^C)^{1-\delta^A-\delta^B} \\ &= \left[\frac{h_A(j)}{1-h_A(j)-h_B(j)}\right]^{\delta^A}\left[\frac{h_B(j)}{1-h_A(j)-h_B(j)}\right]^{\delta^B}\prod_{k=1}^{j}[1-h_A(k)-h_B(k)]\end{aligned}$$

ICR model (Intrinsically discrete time)

Another way of writing the likelihood, which we refer back to later on, is

$$\mathcal{L} = S(j) \left[\frac{h(j)}{1-h(j)}\right]^{\delta^A+\delta^B} \left[\frac{h_A(j)}{h(j)}\right]^{\delta^A} \left[\frac{h_B(j)}{h(j)}\right]^{\delta^B}$$

- No neat separability result ⇒ have to estimate jointly
- But there is one 'easy' estimation method if you assume the destination-specific hazard rates have a specific form: . . .

ICR model (Intrinsically discrete time)

$$
\begin{aligned}
h_A(k) &= \frac{\exp(\beta_A' X)}{1+\exp(\beta_A' X)+\exp(\beta_B' X)} \\
h_B(k) &= \frac{\exp(\beta_B' X)}{1+\exp(\beta_A' X)+\exp(\beta_B' X)}
\end{aligned}
$$

and hence

$$
1 - h_A(k) - h_B(k) = \frac{1}{1+\exp(\beta_A' X)+\exp(\beta_B' X)}
$$

ICR model (Intrinsically discrete time)

- Substitution of these into the expression for L (above) ⇒likelihood contribution for individual has same shape as likelihood for a *multinomial logit model* applied to reorganised data (Allison 1982)
- Estimation in 4 steps:
  1. Expand data to person-period form
  2. Create new categorical depvar to identify the destinations
  3. Create any other vars (eg duration dependence and other TVCs)
  4. Estimate using an MNL program (can test joint hypotheses directly)

ICR model (Interval-censored data)

1. Could estimate using MNL model just discussed, or
2. Estimate a model in which we relate the discrete hazard to the underlying continuous time hazard (cf earlier). If do this, then:
   - Likelihood not separable
   - Shape of the continuous time hazard *within* each interval cannot be identified from grouped data
     ...so to construct likelihood need assumptions about this shape.

NB more than one latent event is possible in each *interval* (though only one can be observed) $\Rightarrow$when considering Pr(exit to a given destination during given interval) related to probability of exit to that destination *and* that exit occurred before exit(s) to other possible destinations.

ICR model (Interval-censored data)

Relationship between discrete overall and destination-specific hazards, and underlying continuous time hazards for jth interval $(a_j - 1, a_j]$?

$$
\begin{aligned}
h(j) &= 1 - \frac{S(a_j)}{S(a_j - 1)} \\
&= 1 - \frac{\exp\left[-\int_0^{a_j}[\theta_A(t) + \theta_B(t)]dt\right]}{\exp\left[-\int_0^{a_j - 1}[\theta_A(t) + \theta_B(t)]dt\right]} \\
&= 1 - \exp\left[-\int_{a_j - 1}^{a_j}[\theta_A(t) + \theta_B(t)]dt\right]
\end{aligned}
$$

ICR model (Interval-censored data)

using the result that $\theta(t) = \theta_A(t) + \theta_B(t)$

$$
\begin{aligned}
h_A(j) &= 1 - \exp\left[-\int_{a_j-1}^{a_j} \theta_A(t)dt\right] \\
h_B(j) &= 1 - \exp\left[-\int_{a_j-1}^{a_j} \theta_B(t)dt\right]
\end{aligned}
$$

ICR model (Interval-censored data)

- it follows that

$$h(j) = 1 - \{[1 - h_A(j)][1 - h_B(j)]\}$$

- or

$$1 - h(j) = [1 - h_A(j)][1 - h_B(j)]$$

- Thus the overall discrete hazard is *not* the sum of the destination-specific hazards (cf earlier)! Separability result as in cts time no longer holds in general!
- Overall hazard for each interval $= 1 - \{$probability not exited in the interval by either a or $b\}$

ICR model (Interval-censored data)

- Multiplying out the terms in expression, we have

$$\begin{aligned} h(j) &= h_A(j) + h_B(j) + h_A(j)h_B(j) \\ &\approx h_A(j) + h_B(j) \text{ if } h_A(j)h_B(j) \approx 0 \end{aligned}$$

- More like cts time case, the smaller the h

ICR model
(Interval-censored data)

- Now consider relationship between the survivor function for exit to *any* destination, and survivor functions for exits to *each* destination.

$$
\begin{aligned}
S(j) &= (1-h_1)(1-h_2)(\ldots)(1-h_j) \\
&= (1-h_{A1})(1-h_{B1})(1-h_{A2})(1-h_{B2}) \\
&\quad \times \ldots \times (1-h_{A2})(1-h_{Bj}) \\
&= (1-h_{A1})(1-h_{B2})(\ldots)(1-h_{Aj}) \\
&\quad \times (1-h_{B1})(1-h_{B2})(\ldots)(1-h_{Bj})
\end{aligned}
$$

ICR model (Interval-censored data)

- in other words,

$$S(j) = S_A(j)S_B(j)$$

- So there is a factoring of the overall survivor function, exactly as in the continuous time case!
- What, then, is expression for each likelihood contribution?

ICR model (Interval-censored data)

- 3 types of contribution (as before)
- *Censored spell:*

$$\mathcal{L}^C = S(j) = S_A(j)S_B(j)$$
$$= \prod_{k=1}^{j}[1 - h_A(k)][1 - h_B(k)]$$

ICR model (Interval-censored data)

Exit to destination A:

- We need an expression for the *joint* probability that had completed spell of type exit-to-A in interval j, and that latent exit time to destination B was after that for A.
- For an exit to A in the jth interval, the expression is . . .

ICR model (Interval-censored data)

$$
\begin{aligned}
\mathcal{L}^A &= \Pr(a_j - 1 < T_A \leq a_j, T_B > T_A) \\
&= \int_{a_j-1}^{a_j} \int_u^{\infty} f(u,v) dv du \\
&= \int_{a_j-1}^{a_j} \int_u^{\infty} f_A(u) f_B(v) dv du \\
&= \int_{a_j-1}^{a_j} \left[\int_u^{a_j} f_A(u) f_B(v) dv + \int_{a_j}^{\infty} f_A(u) f_B(v) dv \right] du
\end{aligned}
$$

ICR model (Interval-censored data)

- where $f(u, v)$ is the joint probability density for latent spell lengths T_A, T_B, and
- u, lower integration point in second integral is (unobserved) time within the interval when exit to A occurred, and we also assumed independence of competing risks, so $f(u, v) = f_A(u)f_B(v)$
- We cannot proceed further without making some assumptions about the shape of the within-interval density functions (or hazards)!

ICR model (Interval-censored data)

- **Five main assumptions** about shape of within-interval density/hazard:
  1. transitions can only occur at the boundaries of the intervals.
  2. destination-specific density functions are constant within each interval (though may vary between intervals),
  3. destination-specific hazard rates are constant within each interval (though may vary between intervals).
  4. the hazard rate takes a particular proportional hazards form (‘proportional intensities’)
  5. the log of the integrated hazard changes linearly over the interval.
- Focus here on (1), (2), and (3)

ICR model (Interval-censored data)

1. Transitions only occur at interval boundaries (Narendranathan/Stewart, 1993)

- If transitions can only occur at interval boundaries then, if a transition to A occurred in interval $j = (a_j - 1, a_j]$, it occurred at date a_j , and it must be the case that $T_B > a_j$ (i.e. after interval j).

→ This, in turn, means that $fB(v) = 0$ between dates u and a_j. ⇒ big simplification of $\mathcal{L}^A$ (ie second term = 0):

ICR model (Interval-censored data)

$$
\begin{aligned}
\mathcal{L}^A &= \int_{a_j-1}^{a_j}\int_{a_j}^{\infty} f_A(u)f_B(v)dvdu \\
&= \int_{a_j-1}^{a_j} f_A(u)du \int_{a_j}^{\infty} f_B(v)dv \\
&= [F_A(a_j) - F_A(a_j-1)][1 - F_B(a_j)] \\
&= h_A(j)S_A(j-1)S_B(j) \\
&= \left[\frac{h_A(j)}{1-h_A(j)}\right] S_A(j)S_B(j)
\end{aligned}
$$

- by similar arguments, we may write

$$\mathcal{L}^B = \left[\frac{h_B(j)}{1-h_B(j)}\right] S_A(j)S_B(j)$$

ICR model (Interval-censored data)

- Overall likelihood contribution for person with spell length of j intervals:

$$\begin{aligned} \mathcal{L} &= (\mathcal{L}^A)^{\delta^A}(\mathcal{L}^B)^{\delta^B}(\mathcal{L}^C)^{1-\delta^A-\delta^B} \\ &= \left[\frac{h_A(j)}{1-h_A(j)}\right]^{\delta^A} S_A(j) \left[\frac{h_B(j)}{1-h_B(j)}\right]^{\delta^B} S_B(j) \end{aligned}$$

ICR model
(Interval-censored data)

- Separability result analogous to that for continuous time ICR model! So can estimate using destination-specific models.
- But how realistic is the assumption of transitions only occurring at the interval boundaries?

2. Destination-specific densities constant within intervals (Dolton & van der Klaauw, REStat, 1999)

$$f_A(u)f_B(v) = \overline{f}_{Aj}\overline{f}_{Bj}$$

given u, v in interval j

ICR model (Interval-censored data)

- Lengthy manipulations show that $\mathcal{L}^A =$:

$$\begin{aligned}
\mathcal{L}^A &= \tfrac{1}{2}[\Pr(a_j - 1 < T_A \leq a_j)\Pr(T_B > a_j - 1)] \\
&\quad + \tfrac{1}{2}[\Pr(a_j - 1 < T_A \leq a_j)\Pr(T_B > a_j)] \\
&= \left[\frac{h_A(j)}{1-h_A(j)}\right] S_A(j) \times \tfrac{1}{2}[S_B(j-1) + S_B(j)] \\
&= \left[\frac{h_A(j)}{1-h_A(j)}\right] S_A(j) \times \frac{S_B(j)}{2}\left[\frac{1}{1-h_B(j)} + 1\right] \\
&= \left[\frac{h_A(j)}{1-h_A(j)}\right] S_A(j) S_B(j) \left[\frac{1-\frac{h_B(j)}{2}}{1-h_B(j)}\right]
\end{aligned}$$

ICR model (Interval-censored data)

- I.e. assumption leads to an expression with nice interpretations:

  (1) the two component probabilities in $\mathcal{L}^A$ provide bounds on the joint probability of interest and we simply take the average of them;

  (2) This in turn involves a simple averaging of survival functions that refer to the beginning and end of the relevant interval

ICR model (Interval-censored data)

Overall likelihood contribution for individual:

$$\begin{aligned}
\mathcal{L} &= (\mathcal{L}^A)^{\delta^A}(\mathcal{L}^B)^{\delta^B}(\mathcal{L}^C)^{1-\delta^A-\delta^B} \\
&= \left[\frac{h_A(j)}{1-h_A(j)}\right]^{\delta^A}\left[\frac{1-\frac{h_B(j)}{2}}{1-h_B(j)}\right]^{\delta^A} S_A(j) \\
&\quad \times \left[\frac{h_B(j)}{1-h_B(j)}\right]^{\delta^B}\left[\frac{1-\frac{h_A(j)}{2}}{1-h_A(j)}\right]^{\delta^B} S_B(j)
\end{aligned}$$

- Not separable; need special program
- NB similarity/difference with earlier case: an extra term here.

ICR model (Interval-censored data)

3. Destination-specific hazards constant within intervals (i.e. survival times Exponential within each interval)

$$\begin{aligned}
\theta_A(t) &= \overline{\theta}_{Aj} \text{ if } a_j - 1 < t \leq a_j \text{, and} \\
\theta_B(t) &= \overline{\theta}_{Bj} \text{ if } a_j - 1 < t \leq a_j \text{, implying}
\end{aligned}$$

$$\theta(j) = \overline{\theta}_{Aj} + \overline{\theta}_{Bj} \text{, if } a_j - 1 < t \leq a_j$$

ICR model (Interval-censored data)

One obvious parameterisation would be

$$\overline{\theta}_{Aj} = \exp(\beta_{0Aj} + \beta_{1A}X_1 + \beta_{2A}X_2 + \ldots + \beta_{KA}X_K), \text{ and}$$

$$\overline{\theta}_{Bj} = \exp(\beta_{0Bj} + \beta_{1B}X_1 + \beta_{2B}X_2 + \ldots + \beta_{KB}X_K)$$

I.e. each destination-specific hazard rate has a piece-wise constant exponential (PCE) form.

ICR model
(Interval-censored data)

In this case, the likelihood has a very similar shape to that for the ‘multinomial logit’ model for intrinsically discrete time:

$$
\begin{aligned}
\mathcal{L}_3 \quad &= \quad S(j)\left(\frac{h(j)}{1-h(j)}\right)^{\delta^A+\delta^B} \\
&\quad \times \left(\frac{\bar{\theta}_{Aj}}{\bar{\theta}_{Aj}+\bar{\theta}_{Bj}}\right)^{\delta^A}\left(\frac{\bar{\theta}_{Bj}}{\bar{\theta}_{Aj}+\bar{\theta}_{Bj}}\right)^{\delta^B}
\end{aligned}
$$

ICR model (Interval-censored data)

- If intervals are short, or interval-hazards are relatively small then, the more likely it is that

$$h_A(j) \approx \overline{\theta}_{Aj},\ h_B(j) \approx \overline{\theta}_{Bj}$$

- In this case, the 'multinomial model' will provide estimates that are similar to the interval-censoring model assuming a constant hazard within intervals (and also assuming that each uses the same specification for duration dependence in the discrete hazard).

ICR model (Interval-censored data)

Extensions:

1. Left-truncated spells
   - ICR models for interval-censored and left-truncated survival data can be easily-estimated using the same programs as for random samples of spells, applied to data sets in which data rows corresponding to the intervals prior to the truncation point have been excluded

2. Correlated risks
   - The models discussed so far can be extended to allow for unobserved heterogeneity: individual-specific error terms in each destination-specific risk that are correlated across risks (see Lecture Notes)