Empirical models for duration data

Partial likelihood, unobserved heterogeneity, multiple spells and competing risks

(m.lindeboom@vu.nl)

PARTIAL LIKELIHOOD

(Limited information inference)

Cox (1972,75), see also Lancaster (1990)

<u>Result:</u>

If the hazard is of the Proportional Hazard (PH) type:

$$\theta(t|x) = \varphi(t)\theta_1(x'\beta)$$

Then the parameters of the regression function $\theta_1(X;\beta)$ can be obtained without making any assumption about the form of the baseline hazard $\theta_0(t)$

Maximum likelihood based on observed durations (and censoring)

The duration data can be split in 2 pieces of information:

- *Rank* information.
  (Who has the shortest spell, the 2^{nd} shortest spell etc)
- The *durations* associated with a ranks

Person	Spell length
1	20 days
2	12 days
3	16 days

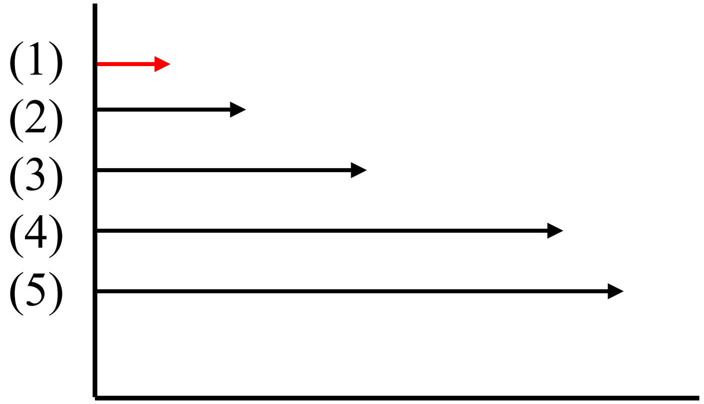
<u>Rank info</u>: 2^{nd} is shortest, 3th is 2^{nd} shortest, longest is 1^{st}

<u>Duration info</u>: shortest = 12, 2^{nd} shortest=16, longest =20

- The idea (Cox, 1972, 1975) is to base the likelihood on rank order of data alone
- More specifically, if we rank the durations in ascending order:

  $t_1 < t_2 < t_3 < \ldots < t_N$
- Then base likelihood on:
  - P=Pr(item i fails | R at risk of failing & one fails)

- This is an odds ratio (concerns relative probabilities)



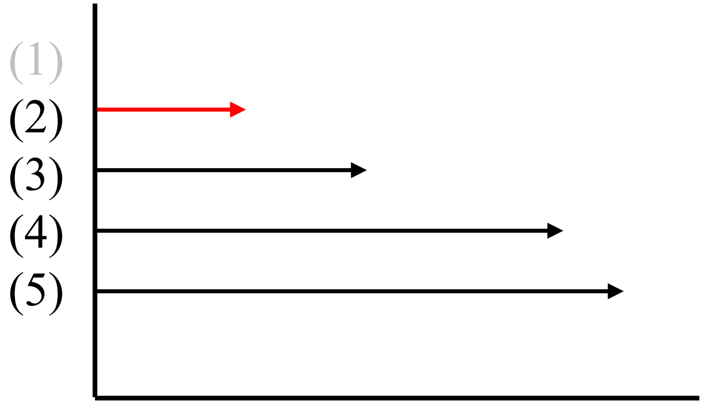
Pr((1) fails | (1)-(5) at risk & one fails)=

$$= \frac{\theta^{(1)}}{\theta^{(1)} + \theta^{(2)} + \theta^{(3)} + \theta^{(4)} + \theta^{(5)}}$$

$$= \frac{\varphi(t)e^{x_{(1)}\beta}}{\varphi(t)e^{x_{(1)}\beta} + \varphi(t)e^{x_{(2)}\beta} + \varphi(t)e^{x_{(3)}\beta} + \varphi(t)e^{x_{(4)}\beta} + \varphi(t)e^{x_{(5)}\beta}}$$

$$= \frac{e^{x_{(1)}\beta}}{e^{x_{(1)}\beta} + e^{x_{(2)}\beta} + e^{x_{(3)}\beta} + e^{x_{(4)}\beta} + e^{x_{(5)}\beta}}$$

Which does not depend on the baseline hazard φ(t) !



Pr((2) fails | (2)-(5) at risk & one fails)=

$$= \frac{\theta^{(2)}}{\theta^{(2)} + \theta^{(3)} + \theta^{(4)} + \theta^{(5)}}$$

$$= \frac{\varphi(t)e^{x_{(2)}\beta}}{\varphi(t)e^{x_{(2)}\beta} + \varphi(t)e^{x_{(3)}\beta} + \varphi(t)e^{x_{(4)}\beta} + \varphi(t)e^{x_{(5)}\beta}}$$

$$= \frac{e^{x_{(2)}\beta}}{e^{x_{(2)}\beta} + e^{x_{(3)}\beta} + e^{x_{(4)}\beta} + e^{x_{(5)}\beta}}$$

Etcetera

More generally the odds ratio

- P^i =Pr(item *i* fails | *R* at risk of failing & one fails) reads as:

$$P^i = \frac{\theta^i}{\sum_{j \epsilon R(i)} \theta^j} = \frac{\varphi(t) e^{x_{(i)}\beta}}{\sum_{j \epsilon R(i)} \varphi(t) e^{x_{(j)}\beta}} = \frac{e^{x_{(i)}\beta}}{\sum_{j \epsilon R(i)} e^{x_{(j)}\beta}}$$

- So that the log-likelihood equals:

$$Log\ L = \sum_{i=1}^{N} P^i$$

Remarks

- Indeed, from the previous expression it follows that it is only evaluated at points where exits take place
- With multiple exits, the most pragmatic solution is to pick randomly an order of the durations (rather than evaluating all possible combinations)
- No assumptions are made wrt the form of the duration dependence function (non-parametric)!
- Less efficient than maximum likelihood
(only rank order info is used)

Once again, methods only works when the hazard is PH!

- Baseline estimate not provided directly. However, Breslow (1974) gives a procedure to retrieve the baseline hazards after estimation of the partial likelihood ℓ_p
- The estimator for the integrated baseline hazard is:

$$\hat{\Lambda}(t) = \sum_{i=1}^{N} \frac{(1-c_i) I(t_i \leq t)}{\sum_{j=1}^{N} I(t_j > t_i) \exp\{x_j^{'} \hat{\beta})}$$

Evaluation takes place at exit points

Cf Kaplan -Meier estimate!

- And the baseline hazard in some interval $t_{k\text{-}1}$ and t_k can be approximated by:

$$\hat{\varphi}^k = \frac{\widehat{\Lambda}(t_k) - \widehat{\Lambda}(t_{k-1})}{t_k - t_{k-1}}$$

Stata codes for the partial likelihood method

(the Cox model)

```
stcox age male femage tenured married lowgroup
bigclass, cl(schoolid) basehc(haz1) nohr

stcurve, hazard
graph export "c:\cox_graph1.wmf", as(wmf)
```

Note: It is also possible to include time-varying regressors $x(t)$

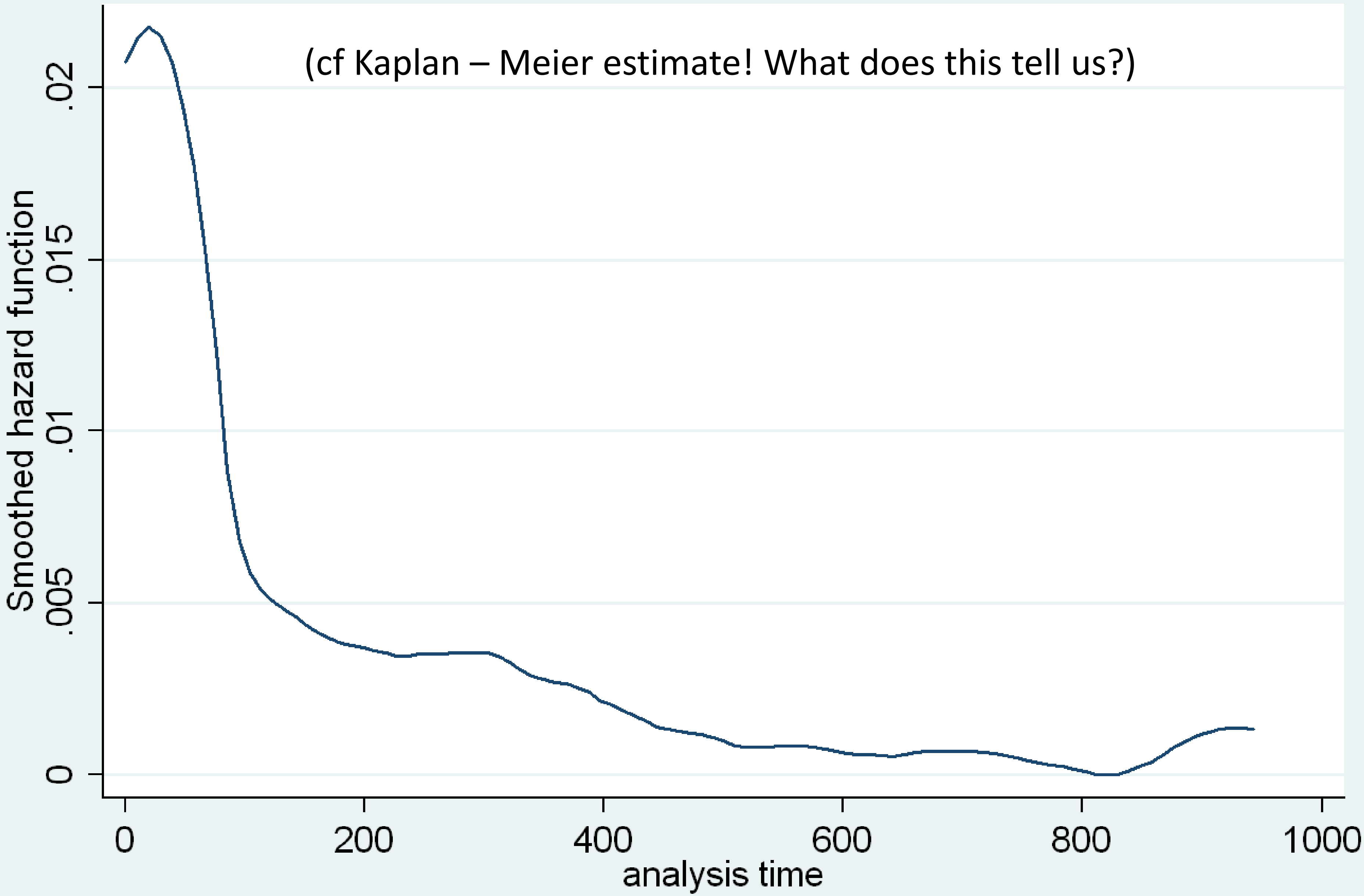
```
Cox regression -- Breslow method for ties

No. of subjects      =          6520     Number of obs   =      6520
No. of failures      =          6291
Time at risk         =        114124
                                          Wald chi2(7)    =     80.88
Log pseudolikelihood =    -50010.698      Prob > chi2      =    0.0000

                        (Std. Err. adjusted for 410 clusters in schoolid)
--------------------------------------------------------------------------
          |               Robust
       _t |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
----------+---------------------------------------------------------------
      age |  -.0152177   .0025789    -5.90   0.000    -.0202722   -.0101631
     male |   .1072319   .0472964     2.27   0.023     .0145327    .1999312
   femage |  -.0628115   .0548287    -1.15   0.252    -.1702737    .0446508
  tenured |  -.1435033   .0621016    -2.31   0.021    -.2652201   -.0217865
  married |   .0281031   .0294724     0.95   0.340    -.0296617     .085868
 lowgroup |  -.0426757   .0352766    -1.21   0.226    -.1118166    .0264653
 bigclass |  -.0244898   .0354517    -0.69   0.490    -.0939739    .0449943
--------------------------------------------------------------------------
```

No constant!

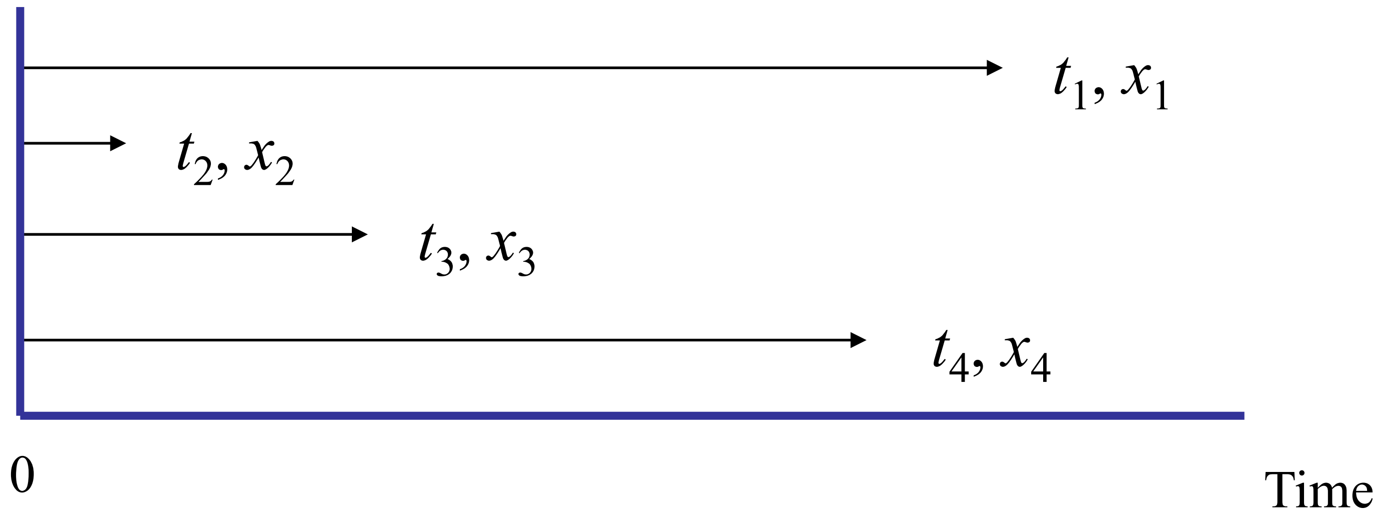
Cox proportional hazards regression



Stock & flow sampling

<u>Flow sample:</u>

Those starting into the state of interest at calendar time 0



This is a random sample from the population of distributions starting at time 0, $F(t/x)$

- Recall that the log likelihood for a simple flow sample with completed and censored spells equals:

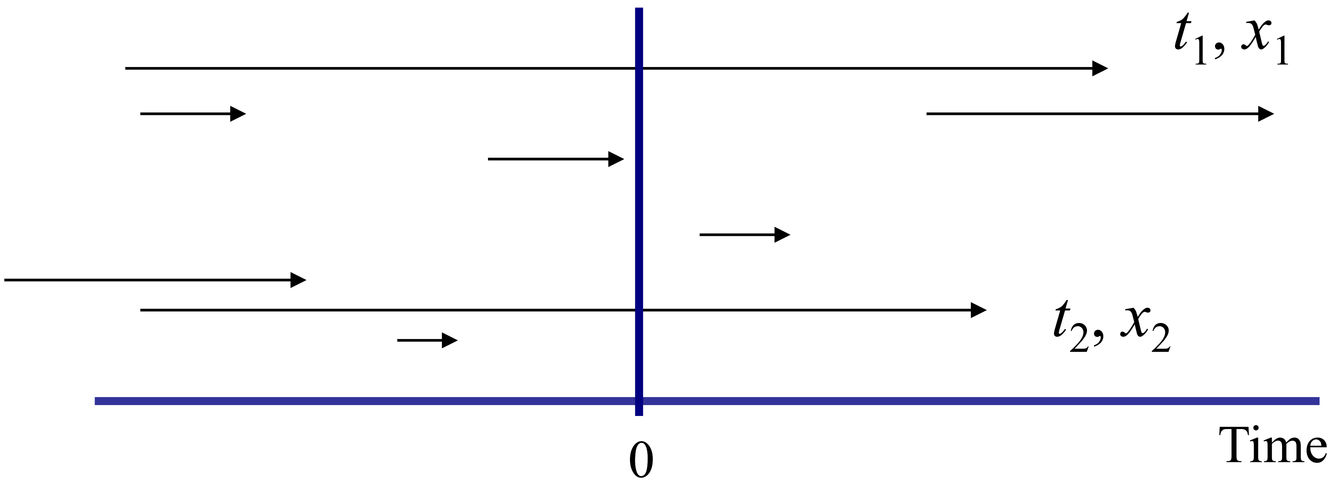
$$\log L = \sum_{i=1}^{N} (1 - c_i)\log f(t_i) + c_i log\, S(t_i)$$

- Flow sample with outcome observed in intervals
  - Sometimes spell is known to end in interval <*k*-1,*k*>
  - If interval is ‘small’, mostly treated as exact date
  - If interval is ‘long’, likelihood needs to be adjusted

$$\log L = \sum_{i=1}^{N} (1 - c_i)\log(\, S\left(t_i^{(k(i)-1}\right) - S\left(t_i^{(k(i)}\right)\,) + c_i log\, S(t_i)$$

Stock sampled durations:

- Those in state of interest at calendar time 0



- **Length biased sampling**

The distribution of stock sampled spells?

- Population distribution $\neq$ stock sampled spells

  $\downarrow$

  In general, the hazard rate estimated from stock sampled spells is different form the population hazard

- For an outstanding analysis of sampling in the context of duration models:
  - See <u>Salant (1977), QJE</u>
  - <u>Ridder (1984),</u> ‘The analysis of single spell duration data’, *Studies in Labor market dynamics,* Neuman & Westergard-Nielsen eds, Springer, Berlin

A little bit more on censoring

- There are situations where the censoring mechanism is informative on the duration outcome T . E.g.:

  i) Clinical trial where terminally ill are removed

  ii) Job search outcomes where job finders leave the sample

- In such cases the relationship between the censoring mechanism C and duration T should be specified explicitly

- Stated differently, the likelihood should be based on the joint distribution of *T* and *C*

$$f(t\,,\,c) = f(t \mid c).f(\,c\,)$$

- When the process *c* does not contain info on the distribution of *T* (it does not share pars with *T*), then the censoring can be treated as exogenous and likelihood can be based on f($t \mid c$)
- The likelihood separates in two parts that do not share common parameters
- This is what we normally do

UNOBSERVED HETEROGENEITY

(The Mixed Proportional Hazard Model)

Often, do not observe all relevant regressors *x*:

1 Let's label the unobserved component *v*, with $v \perp x$

2 $G(v)$ is the distribution of *v* in the inflow ($g(v)$ density)

The mixed proportional hazard model:

$$\theta(t|x,v) = \varphi(t)e^{x'\beta}v = \varphi(t)e^{x'\beta+v}$$

- We do not observe *v*, so the 'observed' hazard in sample is mixed wrt to *v:* $\theta(t|x)$

- The observed hazard is a weighted average of the hazards of the different types *v*   (it is a mixture!)
- As soon as time proceeds, individuals with more ‘favorable’ characteristics (here high *v*) will leave the sample
- Cf the classical mover – stayer problem
- Or, more generally suppose 10000 unemployed of which 5000 are high educated and 5000 low educated:

time	high-educated			low-educated			all workers		
	unempl.	outflow	exit rate	unempl.	outflow	exit rate	unempl.	outflow	exit rate
0	5000	2000	0.40	5000	500	0.10	10,000	2500	0.25
1	3000	1200	0.40	4500	450	0.10	7500	1650	0.22
2	1800	720	0.40	4050	105	0.10	5850	1125	0.19
3	1080	432	0.40	3645	365	0.10	4725	797	0.17
4	648	259	0.40	3280	328	0.10	3928	587	0.15
5	389	156	0.40	2952	295	0.10	3341	451	0.13

- As time proceeds, individuals with more ‘favorable’ characteristics (here high *v*) will leave the sample
- At any point in time the ‘observed’ hazard is a weighted average of the hazards of the different types of *v* that are still in the sample (it is a mixture!)

- The mean of *v* in the sample will fall as time proceeds:
  - $E(v|T>0) < E(v|T=0)$
  - $G(v \mid x, T \geq t) \neq G(v)$

- There is an issue of *dynamic selection*

- We only observe *x* and *t* and it can be shown that the Observed hazard is mixed wrt to *v* and satisfies:

$$\theta(t|x) = \varphi(t)e^{x'\beta}E(v|T > t, x)$$

See Lancaster (1990)

- Consequence of neglected unobserved heterogeneity is that hazard rate falls faster than true hazard rate
- More formally (Lancaster, 1979, see also Lancaster (1990)) :

$$\frac{\partial \log(\theta(t|x)}{\partial t} = \frac{\varphi'(t)}{\varphi(t)} + \frac{\partial E(v|T > t, x)}{\partial t} < \frac{\varphi'(t)}{\varphi(t)} = \frac{\partial \log(\theta(t|x, v)}{\partial t}$$

- Lancaster (1990) furthermore shows that the estimates of β are biased towards zero
- This discussion raises the issue of identification:

  Can we disentangle the two sources of duration dependence? (true vs spurious duration dependence)

- Elbers & Ridder 1982: the MPH model is identified if v has finite mean, if $v \perp x$ and if x takes on at least two values
- Identification means that a given infinite sample of (t_i, x_i), can only be generated by a unique set of model parameters $\varphi(t)$, β and $G(v)$.

Consequences for the likelihood function

- Standard practice with unobserved heterogeneity is to 'integrate out'/weight wrt unobservables *v*.
- In the context of the 'mover-stayer'problem with two types of individuals, Movers (v^M)and Stayers (v^S), with prevalence *p* and (1-*p*), the 'mixture'(unconditional) survivor function:

$$S(t \mid x) = pS(t \mid x, v^M) + (1-p)S(t \mid x, v^S)$$

- And for 3 types:

$$S(t \mid x) = p_1 S(t \mid x, v^1) + p_2 S(t \mid x, v^2) + (1 - p_1 - p_2) S(t \mid x, v^3)$$

- More generally, for *v* with density function *g*(*v*) the mixture survivor functions (unconditional survivor function) becomes:

$$S(t \mid x) = \int_0^{\infty} S(t \mid x, v) g(v) dv = \int_0^{\infty} \left(\exp\{-\int_0^{t} \overline{\theta}(s \mid x) v ds \right) g(v) dv$$

- The hazard rate of this distribution:

$$\theta(t \mid x) = -\frac{d \log S(t \mid x)}{dt}$$

In general this hazard rate is not proportional anymore!

Example:

- Suppose the hazard is exponential:

$$\theta(t \mid x, v) = \exp\{x'\beta\}v$$

- If the distribution of v is gamma(1,η), mean 1 and variance 1/η:

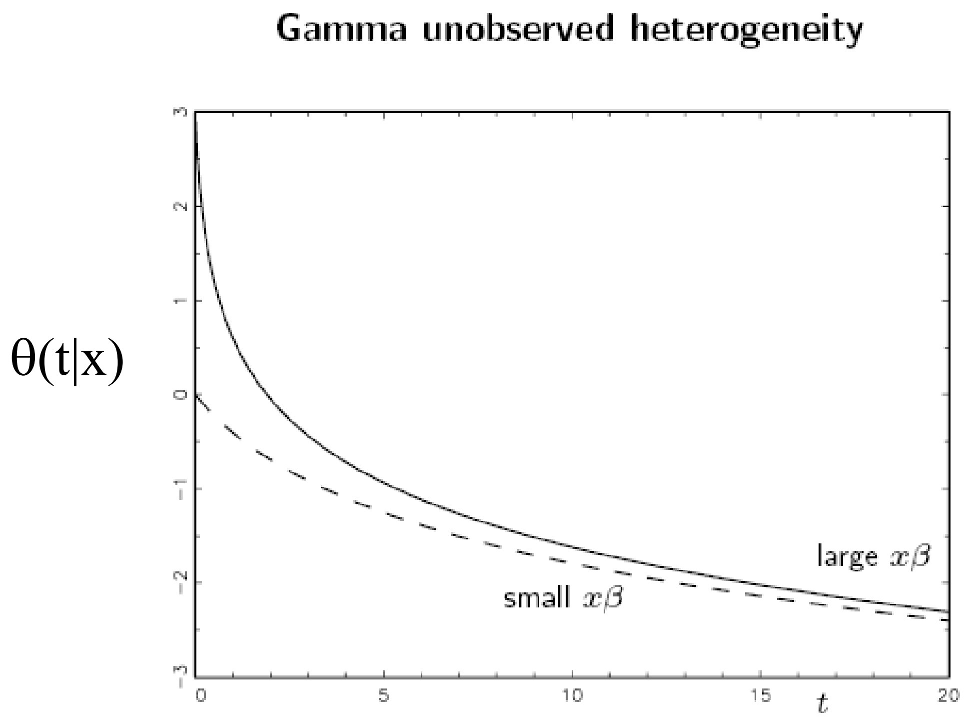
$$g(v) = \frac{\eta^{\eta} v^{\eta-1} \exp(-\eta v)}{\Gamma(\eta)}$$

- Then:

$$\theta(t \mid x) = \frac{\exp\{x'\beta\}}{1 + \exp\{x'\beta\} t / \eta}$$

Try yourself, or see e.g. Lancaster (1990)

Time path of observed hazard, when the individual hazard is exponential and unobservables are gamma distributed



Determined by the time path of E(*v*|*T*>*t*,*x*) !

- The streg command of STATA has an option “frailty” for unobserved heterogeneity:

```
streg age male femage tenured married lowgroup
bigclass, distribution(weibull) frailty(gamma)
cl(schoolid) nohr
```

- Also other frailty/mixing distributions can be chosen such as the inverse Gaussian.
- Convergence of the model is not guaranteed!

Weibull regression Gamma frailty

No. of subjects = 6520 Number of obs = 6520
Wald chi2(7) = 22.94
Log pseudolikelihood = -10582.588 Prob > chi2 = 0.0017

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0226649	.0103883	-2.18	0.029	-.0430255	-.0023043
male	.1983005	.1875746	1.06	0.290	-.1693388	.5659399
femage	-.1961088	.2029667	-0.97	0.334	-.5939162	.2016986
tenured	-.3162641	.2149364	-1.47	0.141	-.7375316	.1050035
married	.1760326	.1132995	1.55	0.120	-.0460303	.3980956
lowgroup	-.0895324	.1453775	-0.62	0.538	-.374467	.1954022
bigclass	-.0818315	.1578517	-0.52	0.604	-.3912151	.2275521
_cons	-1.287636	.5189313	-2.48	0.013	-2.304723	-.2705492
/ln_p	.9661935	.0820545	11.78	0.000	.8053698	1.127017
/ln_the	1.272894	.1307774	9.73	0.000	1.016575	1.529213
p	2.627922	.2156328			2.237524	3.086437
1/p	.3805287	.0312241			.3239982	.4469226
theta	3.571174	.4670287			2.763714	4.614545

Remarks:

1) We have to make assumptions to identify the model ($\theta(t|x,v)$ proportional; $v \perp x$; $E(v)<\infty$; at least one x)

2) Results are sensitive to parametric form assumptions concerning baseline hazard $\lambda(t)$ and distribution of v ($G(v)$)

- See unpublished WP on website Geert Ridder ("The sensitivity of duration models to misspecified heterogeneity and duration dependence"):

  Concerning G(v):

  - Regression pars sensitive to some extent under heavy censoring

  Concerning $\varphi(t)$:

  - Misspecification usually causes a significant bias

Suggested in literature:

- Use flexible functional forms (see next slide for mixing distribution and duration dependence)
- Validate if possible the results with external information
- Perform robustness checks and specification tests
- Use, if possible, multiple spell data

  (identification: no functional form assumption, nor $v \perp x$ required, see later slides)

Flexible distributions

<u>Baseline hazard:</u>

- Take flexible PWC or Cox model
  - Note that Cox method requires proportional hazards (with unobserved heterogeneity this property is destroyed)

<u>Mixing distribution:</u> Take a discrete mixing distribution

- As before, assume different types in the population (v^k), with fractions p_k = Pr(v=v_k) ,

$$\sum_{k=1}^{K} p_k = 1$$

$$S(t \mid x) = \sum_{k=1}^{K} p_k S(t \mid x, v^k)$$

- K should be fixed in advance and the parameters v^i and p_i , i=1,2,.., K are estimated along with the other parameters of the model
- In practice: start with K=2 and gradually increase K
- Use starting points around the constant of the model without unobserved heterogeneity
- During optimization it can be that:
  - Two mass-points converge to each other : $v^k \rightarrow v^{k'}$
  - The probability associated to a mass point p_k goes to zero or one
- This is an indication that M is chosen too large

Multiple spells: Fixed effects duration models

- The data sometimes provides multiple durations for a given subject (multiple spells per individual)

- The subject has a given value of v and the duration of the spells are independent drawings from the univariate duration distribution $F(t|x,v)$

- The subject can also be a group (*stratum*) of individuals, e.g. spells within school, or lifetimes of individuals within a hospital, or family etc

Message:

- With multiple spells data the MPH model is identified under much weaker assumptions
- Convenient transformation exists so that a fixed effect duration model can be estimated using standard statistical software like STATA
- The transformations bear resemblance with those used in linear models (and the Logit model and the Poisson model)

Handling fixed effects in panel data model

Consider the linear panel data model

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \varepsilon_{it} \qquad (1)$$

Interest is in the estimation of β in the presence of individual parameters α_1 ,…, α_N (fixed effects)

- One possible solution is to jointly estimate β , α_1 ,.., α_N
- But in short panels asymptotic theory relies on $N \to \infty$, but then so does the number of fixed effects

Incidental parameters problem

- It is still possible to consistently estimate β by providing a <u>suitable transformation</u> that eliminates the 'nuisance parameters' α_1 ,.., α_N from (1)

For instance, the first difference transformation:

$$y_{it} - y_{it-1} = (x_{it} - x_{it-e})'\beta + (\varepsilon_{it} - \varepsilon_{it-1}) \qquad (1')$$

Or the within transformation:

$$y_{it} - \overline{y_i} = (x_{it} - \overline{x}_i)'\beta + (\varepsilon_{it} - \overline{\varepsilon}_i) \qquad (1'')$$

Note that both transformations eliminate *all* time constant variables

Handling fixed effects in duration models

First, comment on the identification of the MPH model

The integrated hazard is exponentially (1) distributed:

$$\int_0^t \theta(s)ds = \varepsilon \qquad\qquad f(\varepsilon) = \exp\{-\varepsilon\}$$

So that if $\theta(t) = \varphi(s)\exp(x'\beta' + \nu)$, the following hold:

$$log \int_0^t \varphi(s)ds = -x'\beta - v + \mu \qquad , \mu\text{=}log\ \varepsilon\ , \qquad \mu \sim \text{EV(1)}$$

Note that the regression model and the PH model are not nested and have their flexibility from different sources:

- The right hand side is restrictive because μ is completely specified (mean and variance of EV(1) are fixed)
- The random variable *v* reflect purely random variation in the duration outcome, not unobserved individual characteristics
- The regression duration model has a more general specification of the left hand side (unknown transformation)

If we have more observations on *T*, say we observe sickness spells t_1 and t_2 of the same individual (and assuming exponential distributions $\varphi(t)$=*constant*):

$$\log t_1 = -x_1'\beta - \log v - \mu_1$$
$$\log t_2 = -x_2'\beta - \log v - \mu_2$$

This resembles the standard linear panel data set up with fixed effects and one can find a transformation to eliminate the fixed effects:

$$\log t_2 - \log t_1 = -(x_2 - x_1)'\beta - (\mu_2 - \mu_2)$$

- So no assumptions on distribution of v required, nor need $v \perp x$
- The model is even identified if x and t interact (See van den Berg, 2001)
- As a consequence empirical duration analysis with multiple spells is preferred
- With multiple spells one may also estimate random effects duration models
  - See previous slides on unobserved heterogeneity
  - With multiple spells and multiple states it is not trivial to estimate random effects models (see later slides)

Multiple spells and fixed effects in duration data: Stratified partial likelihood

Differencing does not make sense in presence of censoring $\Rightarrow$ have to look for alternative transformations

Define the hazard for an individual / group/stratum *i* as:

$$\theta^i(t|x, v_i) = \varphi^i(t, v_i)e^{x'\beta}$$

x can vary within the stratum

Note that the specification of the baseline function is more flexible as no factorization of *t* and *v* is imposed

One could see the total sample as a sum of groups that consist of spells of that group

The probability that item/indiv *j* of group *i* exits from the state, given that one item/individual exits equals:

$$P_j^i = \frac{\theta^i}{\sum_{k \epsilon R^{\mathrm{i}}(j)} \theta^k} = \frac{\varphi(t, v_i) e^{x_j \beta}}{\sum_{k \epsilon R^{\mathrm{i}}(j)} \varphi(t, v_i) e^{x_k \beta}} = \frac{e^{x_j \beta}}{\sum_{k \epsilon R^{\mathrm{i}}(j)} e^{x_k \beta}} \qquad (4)$$

With R^j (*i*) the Risk set of group *i* at the point that *j* fails

- Again, the unit/group specific baseline hazard (‘nuisance parameters’) cancels from the expression and indeed no specific relationship between *t* and *v* imposed

- The likelihood function is obtained by multiplying expressions like (4) for all exits in group *i* and subsequently multiply over all groups *I*

- Estimate with STATA via the ‘strata( )’ option within **stcox**

- Again, proportionality of the hazard is essential (flow samples or (not discussed here) look at the conditional distribution $f(t|\text{elapsed spell})$
- To estimate an effect of x, it is essential that x varies within the stratum (check this e.g. with a stratum of 2 spells)
- A group only contributes to the partial likelihood function if an individual leaves the state and at least one other individual is there
    - E.g Child mortality within families: single child families do not contribute nor do families where children do not die

- Censoring issues become more difficult in this SPL context.
  (See Ridder & Tunali, Jectrics, 1999)

- When baseline hazards do not differ between clusters, normal (unstratified) partial likelihood is more efficient
  $\Rightarrow$ Hausman-test for presence of cluster specific effects possible by comparing PL with SPL

Multivariate MPH models

- Consider two durations *T* and *S* these may follow each subsequently, or they may run simultaneously

<u>Example 1:</u> *i*) The labor market history as a sequence of transitions between work and non-work

<u>Example 2:</u> *ii*) Time spend in a labor market state and survey participation

iii) Life time of an individual and time spend in marriage and non-marriage

We will focus on i), but ii) & iii) similarly, see e.g. van den Berg, 2001

Competing risk: the simplest case
(single spell, no unobserved heterogeneity)

- Suppose we model work durations and that there are two possible ways to leave work: Unemployment and Disability

  - θ^{U} : the transition rate out of work to unemployment

  - θ^{D} : the transition rate out of work to Disabled

  And:

$$\theta = \theta^{U} + \theta^{D}$$

- Then for a completed spell that ends in *U*(nemployment):

- Then Pr(*T*=*t* & exit to *U*) equals:

$$\theta(t)e^{-\int_0^t \theta(s)ds}\,\frac{\theta^U(t)}{\theta^U(t)+\theta^D(t)}$$

$$= \theta^U(t)e^{-(\int_0^t \theta^U(s)+\theta^D(s)ds)}$$

$$= \theta^U(t)e^{-\int_0^t \theta^U(s)ds}\; e^{-\int_0^t \theta^D(s)ds}$$

- The second term is the survivor function of T associated with a transition to *D*(isability).
  - You need to survive a transition to *D* if you want to make an exit to *U*
  - An exit to *D* has a similar expression

- Suppose we have a sample of spells, some are censored, some completed with an exit to U, some completed with and exit to D.
- If we define d_U and d_D as indicators for exits to *U* and *D*, respectively
- Then a contribution to the likelihood function equals:

$$\theta^U(t)^{d_U}\,\theta^D(t)^{d_D} e^{-\left(\int_0^t \theta^U(s)+\theta^D(s)ds\right)}$$

$$= \theta^U(t)^{d_U} e^{-\int_0^t \theta^U(s)ds}\,\theta^D(t)^{d_D} e^{-\int_0^t \theta^D(s)ds}$$

- Part associated with *U*

Part associated with *D*

- So the likelihood function separates in two parts:

  L^U and L^D, each associated with one risk

  => Separate estimation possible with standard software like STATA

  $\rightarrow$ Treat exits to other state as censored!

- One can generalize this into more states (e.g. *U*, *D*, *O*ut of the labor force)
- With no unobserved heterogeneity multi-state duration models factorize into separate and independent parts

  => standard software like STATA can handle estimation

- Multiple spells do not complicate matters, actually they improve efficiency

- Estimation is much more complicated when there are shared parameters and in particular in presence of random effects (unobservables, say v_1 and v_2)

$$\mathrm{L} = \Pi \iint \theta^U(t|v_U)^{d_U} e^{-\int_0^t \theta^U(s|v_U)ds} \cdot$$

$$\theta^D(t|v_U)^{d_D} e^{-\int_0^t \theta^D(s|v_D)ds} g(v_U, v_D) dv_U dv_D$$

- Expression does not factorize in two parts unless $v_1 \perp v_2$

  i.e. $g(v_1,v_2)=g(v_1)g(v_2)$

- Otherwise all pars have to be estimated jointly ....

- Moreover, identifications in these correlated competing risks models is more difficult than single spell - single state models
- After all, in the single spell competing risks model we only observe one duration and two censoring indicators.

  T_i censored by T_j while both are dependent

  (so less ‘independent variation’)

- With multiple spells this is less of a problem