survival analysis: the analysis of transition data

Osti, Davide

June 2015

CAMERON AND TRIVEDI: CH. 17 - TRANSITION: SURVIVAL ANALYSIS

The present notes are meant to resume for the sake of our understanding the kind of analysis known as *survival analysis* in the biostatistical literature and *duration analysis* in the applied econometric one[1]. We just intend to briefly survey it, partially inspired by some contributions found among the references contained at the end of each of the three chapters on survival analysis in Cameron and Trivedi (2005)', namely chapters 17, 18, and 19.

17.3.2. DISCRETE DATA $\rightsquigarrow$ define the discrete time cumulative survivor function $S^d(t)$ as the probability of transition from 0 child to 1 child, referred as *starting* behaviour in Knodel '87, given survival up to time $t_i \rightsquigarrow$ recover the discrete hazard rate labeled $\lambda_i = Pr[T = t_i | T \geq t_i] = f^d(t_i)$, namely the discrete density function of duration of each spell, with the property that $S^d(t_i) = lim_{t \rightsquigarrow a} S^d(t_i)$, where a is a given partition of the time interval - which in the case of the Burundian survey data is at a yearly frequency for the time span t = 1993,...,2002.

The discrete time cumulative survivor function $S^d(t)$ is defined as the probability of transition from zero child to one child, referred as *starting* behaviour in Knodel '87, given survival up to time $t_i \rightsquigarrow$ recover the discrete hazard rate labeled $\lambda_i = Pr[T = t_i | T \geq t_i] = f^d(t_i)$, namely the discrete density function of duration of each spell, with the property that $S^d(t_i) = lim_{t \rightsquigarrow a} S^d(t_i)$, where a is a given partition of the time interval - which in the case of the Burundian survey data is at a yearly frequency for the time span t = 1993,...,2002.

$S^d(t) = Pr[T > t]$ and not $Pr[T \geq t]$, i.e. probability of surviving to interval t without a *failure* has verified[2]. Survival function for discrete data is alternatively defined as $S^d(t) = Pr[T \geq t] = \prod_{i|t_i}(1 - \lambda_i) \rightsquigarrow$ the discrete time survivor function is obtained recursively from the hazard function. So, i.e. $Pr[T > t_2]$, namely the probability of no transition at time t_2, for a given sequence of time spells $t_0 < t_1 < ... < t_9$, is $(1 - \underbrace{\lambda_1}_{\text{hazard rate for } \Delta t_1 = t_1 - t_0})(1 - \underbrace{\lambda_2}_{\text{hazard rate for } \Delta t_2 = t_2 - t_1})$[3].

Function $S^d(t)$ is a $\searrow$ step function of time t, with steps at t_j, $j = 0, 1, 2, ..., 9$. We have full pregnancy histories at an annual frequency for nine years, but we do not know when exactly during year the pregnancy occurred. Discrete time cumulative hazard function (chf) is $\Lambda^d(t) = \sum_{i|t_i \leq t} \lambda_i$. Product integral[4] applied to the intervals $\underbrace{[a_0, a_1)}_{\text{'93-'94}}, \underbrace{[a_1, a_2)}_{\text{'94-'95}}, ..., \underbrace{[a_8, a_9)}_{\text{'01-'02}}$. $T = t_i$ denotes a discrete time duration $\rightsquigarrow$ defining the transition in the interval $[a_{j-1}, a_j) \Leftrightarrow$ transition at time a_{j-1} or later. GROUPED DATA $\rightsquigarrow$ underlying data generating process (DGP) is *continuous*, but the data are collected *discretely* $\rightsquigarrow$ transitions in the model are observed discretely $\rightsquigarrow$ and the necessary adjustments are made for grouping[5].

CENSORING

Our data set is *right censored* $\rightsquigarrow$ an individual is observed since $time_0 = 1993$ until the end of the survey period, coinciding with the censoring time $time_c = 2002$, where the *risk set* is $'93 -' 02$ in *calendar year* metric, whereas in *woman-age* metric, it ranges from 12 to 46 years of life[6]. This defines a retrospective history of fertility along the years of civil war in Burundi. Some birth spells between subsequent births are completed or closed, up to $time_c$, some others are instead not and will end in interval $[time_c, age_i)$ for each woman $i = 4783$. The idea is to find a model for *starting*,

[1] Mainly developed at the end of the seventies for applications in labour economics such as in the analysis of the length of unemployment spells, particularly in the UK and the US, see, for example, Lancaster '79.

[2] Where by *failure* we mean a child birth in the statistical demographic literature or the *end of a working contract* in the labour economic literature.

[3] Like a factorization of characteristic equation of an $ARMA$ in Hamilton '94, in the different context of time series analysis

[4] For evaluation of the partial likelihood function.

[5] Which are these necessary adjustments?

[6] Allowing for both right and left censoring

one for *spacing* and one for *stopping* of concievements. A synonym of *right* censoring is *censoring from above* or *top coding*. Also interval censoring exists.

Random or exogenous censoring $\forall i$ has either a T_i^* and a C_i^*, that is, an optimal stopping time and an optimal censoring time, independent between themselves $\rightsquigarrow (t_1, \delta_1), (t_2, \delta_2), ..., (t_9, \delta_9) \equiv$ observed data, realization of $T_i = \min\{T_i^*, C_i^*\}$; $\delta_i = \mathbf{1}[T_i^* < C_i^*] = 1$ if birth occurs and 0 otherwise, $\mathbf{1}$ being an indicator variable for completed observed spells, i.e. transition from zero to one child or from one to two or from two to three, and so on, until a maximum of eight, found in the survey. For the analysis of both *starting* and *spacing*, we need to refer to a multinomial logit or probit, namely to a *discrete choice model*.

For standard survival analysis to be valid, both in the context of analysis of the effects of unemployment benefits on the duration of unemployment spells or war induced displcement on the three main aspects of fertility behaviour (for both cases of micro econometric applications in labour economics or in demography what matters is the occurrence of one of the two in an exogenous manner w.r.t. the other one, to guarantee some causality to the estimates), censoring needs to be an independent (non-informative) one: the parameters of the distribution of C^* are not informative[7] one may treat the censoring indicator δ as exogenous.

NON-PARAMETRIC ESTIMATION

$\rightsquigarrow$ Let $t_1 < t_2 < ... < t_j < ... < t_k \rightsquigarrow$ be the observed discrete failure times of the spells in a sample of size N, $N \geq k$. Define d_j to be the # of spells that ends at time t_j. Data $\equiv$ discrete $\rightsquigarrow$ $d_j > 1$ is possible. $m_j \equiv$ # of spells right censored in $[t_j, t_{j+1})$. Independent censoring is assumed $\rightsquigarrow$ failure time $> t_j$. Spells are *at risk* of failure if they had not yet failed or being censored. $r_j \equiv$ # of spells at risk at time t_j, just before time t_j. $r_j = (d_j + m_j) + ... + (d_k + m_k) = \sum_{l|l \geq j}(d_f + m_l)$. $\lambda_j = Pr[T = t_j | T \geq t_j] \rightsquigarrow$ obvious estimator of the hazard function is the # of spells ending at time t_j divided by the # of spells at risk at time $t_j \rightsquigarrow \hat{\lambda}_j = \frac{d_j}{r_j}$. **Discrete survival time function** $\rightsquigarrow S^d(t) = Pr[T \geq t] = \prod_{j|t_j \leq t}(1 - \lambda_j)$ defines the Kaplan-Meier estimator as the product limit estimator of the survivor function $\Leftrightarrow$ sample analogue of the theoretical survivor function estimated according to the expression proposed by Kaplan-Meier[8] is:

$$\hat{S}(t) = \prod_{j|t_j \leq t}(1 - \hat{\lambda}_j) = \prod_{j|t_j \leq t} \frac{r_j - d_j}{r_j} \tag{1}$$

NOTATION

- $t_j \equiv$ jaar = 1993,...,2002, temporal span for each woman j = 1,...,4,783;
- $r_j \equiv$ # of mothers *at risk*, i.e. age > 14, with no sons/daughters yet;
- $d_j \equiv$ # of pregnancies experience by an individual woman;
- $m_j \equiv$ # of individuals whose record from the survey was censored, (i.e. mothers still in the risk set, with age less than 46[9]), namely with no birth registered at the date the survey was carried on, but who could nonetheless have had a child somewhen after the sample period;
- $\hat{\lambda}_j = \frac{d_j}{r_j} \equiv$ estimated hazard;
- $\hat{\Lambda}(t_j) \equiv$ estimated cumulative hazard;
- $\hat{S}(t_j) \equiv$ estimated survivor function, $\hat{S}(t_j) = \frac{r}{N}$, in case of censoring;

[7]In other words, censoring time is independent across individuals about the parameters of the duration T^*
[8]Or *product limit* estimator.
[9]Assumed by assumption, due to the restrictions imposed on the questionnaire of the ESD, 2002.

- $\underbrace{r_j}_{\text{\# at risk at time } j} - \underbrace{d_j}_{\text{\# of new mothers at time } t_j} = \underbrace{r_{j+1}}_{\text{\# at risk at time } t+1}$, a first order linear difference equation, holding if $\underbrace{m_j}_{\text{this we have to test it!}} = 0$, that is, if no observation is censored.

Say that: $j \equiv$ # at risk at time j, r_j # of mothers at risk, d_j # of pregnancies, m_j # of individual censored, i.e. mothers with no children in the temporal span, $\hat{\lambda}_j = \frac{d_j}{r_j}$, $\hat{\Lambda}(t_j) \equiv$ estimated cumulative hazard; $\hat{S}(t_j) \equiv$ estimated survivor function $\rightsquigarrow = \frac{r}{N}$, in case of no censoring. $r_j - d_j = r_{j+1}$.

$$\hat{S}(t_j) = \prod_{j|t_j \le t} \lambda_j \left[\frac{r_{j+1}}{r_j}\right] \equiv \frac{r_1}{r}, r_1 \equiv N \tag{2}$$

Discrete-time theoretical cumulative hazard function is

$$\Lambda^d(t) = \sum_{j|t_j \le t} \rightsquigarrow \hat{\Lambda}(t) = \sum_{j|t_j \le t} \rightsquigarrow \hat{\Lambda}(t) = \sum_{j|t_j \le t} \hat{\lambda}_j = \sum_{j|t_j \le t} \frac{d_j}{r_j} \rightsquigarrow \tag{3}$$

the latter expression is the Nelson-Aelen estimator of the chf[10].

Again, $\lambda(t) = \frac{f(t)}{S(t)}$, $-f(t) = \frac{dS(t)}{dt} = \frac{d[1-F(t)]}{dt} = -f(t)$, $\lambda(t) = -\frac{d}{dt} \ln S(t)$. $\mu = \mathbf{E}[t] = \int_0^\infty t f(t) dt$, is the usual statistical definition of the mean of continuous proper random variable. Notice that it is linear in $\ln t$, and w/slope α.

SOME PARAMETRIC MODELS

$\rightsquigarrow$ popular choices $\rightsquigarrow$ exponential, Gompertz, Weibull, log-Normal, log-logistic, Gamma. For example $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt \equiv$ Gamma function, $I(a, x) = \int_0^x e^{-t} t^{\alpha-1} / \Gamma(\alpha)$, with $0 < I(a, x) < 1 \equiv$ incomplete Gamma function. Generalized Weibull $\rightsquigarrow$ proposed by Mudholkar, Srivastava and Kollia in 1996 $\rightsquigarrow$ additional shape parameter $\mu \rightsquigarrow$ allows hazard to have a more flexible shape[11]. The $\lim_{\mu \to \infty}$ yields, from the previous expression, the Weibull model of the following form:

$$\ln \lambda(t) = \ln(\gamma\alpha) + (\alpha - 1) \ln t - \underbrace{\mu}_{\mu > 0} \underbrace{\ln(-S(t))}_{\frac{\partial \ln S(t)}{\partial t} < 0}$$

the right hand side is increasing in t. if $\alpha \le 1$ and $\mu < 0$ then ln hazard is a monotonically decreasing function of t which implies that the two together can generate a U shaped unimodal hazard function meaning that the generalized Weibull is a potentially very flexible functional form. Gompertz is also a fairly flexible functional form whose hazard function can be made monotonically decrease depending on $\alpha \lesseqgtr 0$, with the exponential a a special case, constituting a good model for mortality data, used in biostatistics more than in econometrics. Log-normal and log-logistic have a boothub, first increasing with t and then decreasing. Other models are the Rayleigh and Makeham distribution and the generalized Gamma model (Lawless, '82), nesting Gamma and Weibull as special cases.

Distributions are often two parameter distributions where regressors are introduced by making the parameter $\gamma = \exp\{\mathbf{x}'\beta\}$ with α being a constant, but for the log-normal $\mu = \mathbf{x}'\beta$ and σ^2 is left as a constant not to be estimated.

Main issues are the dependence on correct model specification for consistent parameter estimates and the wide range of parametric models available. Classification either as proportional hazards (PH) models or as accelerated failure time models (AFT). Weibull model belongs to both categories. Also widely used in applied micro-econometrics is the piecewise constant hazard model.

[10] Cumulative hazard function
[11] Where by *flexible* is meant

MAXIMUM LIKELIHOOD ESTIMATION

Fully parametric model with independent censoring is estimable via maximum likelihood or via least squares. Regressors are assumed to be time invariant. Notation:

- $T^* \equiv$ observations whose duration is censored;
- $f(t|\mathbf{x};\theta) \equiv$ conditional density;
- $\theta \equiv q \times 1$ vector of parameters;
- $\mathbf{x} \equiv q \times 1$ vector of regressors;

censoring complicates estimation.

Treatment similar to Tobit models[12].

Likelihood contribution.

$$Pr[T > t] = \int_t^\infty f(u|\mathbf{x},\theta)du = 1 - F(t|\mathbf{x},\mathbf{u}) = S(t|\mathbf{x},\theta) \tag{4}$$

Density of the i^{th} observation:

$$f(t_i|\mathbf{x_i},\theta)^{\delta_i} S(t_i|\mathbf{x}_i,\mathbf{u})^{1-\delta_i} = S(t|\mathbf{x},\theta) \tag{5}$$

where $\delta_i = 1$ if no censoring occurs and $\delta_i = 0$ if right-censoring occurs.

Take logs and sum up:

$$\ln L(\theta) = \sum_{i=1}^{N} [\underbrace{\delta_i \ln f(t_i|\mathbf{x_i},\theta)}_{\text{(a)}} + \underbrace{(1-\delta_i)\ln S(t_i|\mathbf{x_i},\theta)}_{\text{(b)}}] \tag{6}$$

we get the log-likelihood function $\rightsquigarrow \hat{\theta}_{\mathbf{MLE}} \equiv$ value of the parameters maximizing the log-likelihood! $(a) \rightsquigarrow$ closed intervals, $(b) \rightsquigarrow$ open intervals.

$\ln S(t) = \Lambda(t)$ and $\ln f(t) = \ln\{\lambda(t)S(t)\} = \ln\{\lambda(t)\} + \ln\{S(t)\} \rightsquigarrow$ rewrite the log-likelihood function in terms of the conditional hazard and integrated hazard functions:

$$\ln L = \sum_{i=1}^{N} \{\delta_i \ln \lambda(t_i|\mathbf{x_i},\theta) + \Lambda(t_i|\mathbf{x_i},\theta)\} \tag{7}$$

useful if the parametric model is defined by specifying the hazard rate $\rightsquigarrow$ usual estimation theory applies, that is $\hat{\theta}_{\mathbf{MLE}} \rightsquigarrow N(\theta, (\mathbb{E}[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}])^{-1})$[13]. If density is correctly specified $\rightsquigarrow$ the $\hat{\theta}_{\mathbf{MLE}}$ is consistent[14]. Exponential duration model $\rightsquigarrow$ in absence of censoring $\rightsquigarrow$ requires correct specification

[12] Allowing consistent estimation of censored data models, through OLS estimation, i.e. estimating equations of the type $y^* = X\beta + u$, where y^* is the complete variable, not censored nor truncated, with the usual OLS assumptions of $E[u] = 0$ (zero mean disturbance term) and $E[X'u] = 0$ (error term and regressors' vector.) $\rightsquigarrow$ with these assumptions, we have that if y^* could be observed, the OLS estimator of β would be consistent. Unfortunately, y^* is not observed. With a censored version of y^*, and considering the simplest version of a regression model, simply linear, we have $E[y^*|X] = X\beta \Leftrightarrow$ we assume the conditional mean of X on β is simply linear. Define y, the *latent variable*, as $y = y^*$ if $y^* < b$ and $y = b$ if $y^* \geq b$, as happens in the case of *top coding* or censoring from above (ref. Pellizari's notes "Maximum Likelihood and Limited Dependent Variable Models", May 24, 2010, largely inspired by the Woolridge for the course in econometric analysis DES LM course at Bocconi University). Then the conditional mean of y given X can be computed as $E[y|X] = Pr(y = b|X) \times E(y|X, y > b) + Pr(y > 0|X)E(y|X, y > 0) = Pr(y > 0|X)Pr(y|X, y > 0)$ THE FORMULA IS FOR THE DERIVATION OF THE LIKELIHOOD CONTRIBUTION OF A DOWN CODED DATASET.

[13] Asymptotically, namely as N, the sample size, grows large, normally this should hold for samples with more than 30 obsevations, as in the time series anlysis literature.

[14] Assuming that our objective is efficient and unbiased estimation of the model parameters.

only of conditional mean function. However, in presence of censoring, inconsistency may occur even for the exponential model and arises for other parametric models even without censoring, which constitutes a major weakness of the parametric approach (**lack of robustness**).

with *left-censoring* $\rightsquigarrow$ spell known to be of length at most $t \rightsquigarrow Pr[T^* < t] = \int_0^t f(s|,\mathbf{x},\theta)ds = F[s|\mathbf{x},\theta]$.

with *interval-censoring* $\rightsquigarrow$ spell known to lie in $[t_a, t_b)$, $\rightsquigarrow$ likelihood contribution: $Pr[t_{\leq} T < t_b] = \int_{t_a}^{t_b} f(s)dsf_j$. Go back to chapter **17.5.2. confidence bands for nonparametrics estimates** $\rightsquigarrow$ as those used in **STATA**. $\lambda_j = \frac{d_j}{r_j}$, estimate of the hazard $\rightsquigarrow$ very discontinuous survivor and cumulative density functions are much more smoother...it is standard practice to plot them against time along with confidence bands. Kaplan-Meier estimate of the survivor function $\rightsquigarrow$

$\rightsquigarrow \hat{V}[S(t)] = \hat{S}^2(t)\sum_{j|t_j\leq t}\frac{d_j}{r_j(r_j-d_j)}$ is the *Greenwood estimate of the variance.*

$\rightsquigarrow \ln(-\ln\hat{S}(t)) \rightsquigarrow$ base for *confidence bands.*

$$S^d(t) \in \left(\hat{S}(t)\exp\{-z_{\frac{\alpha}{2}}\hat{\sigma}(t), z_{\frac{\alpha}{2}}\hat{\sigma}(t)\}\right),$$

where the estimated standard deviation of the regressors, $\ln(-\ln\hat{S}(t))$, is defined as:

$$\hat{\sigma}_g^2 = \frac{\sum_{j|t_j\leq t} d_j/(r_j(r_j-d_j))}{\{\sum_{j|t_j\leq t}\ln\left[\frac{r_j-d_j}{d_j}\right]\}^2}$$

Comparison between Exponential and Weibull:

FUNCTION	EXPONENTIAL	WEIBULL
$f(t)$	$\gamma\exp\{-\gamma t\}$	$\gamma\alpha t^{\alpha-1}\exp\{-\gamma t^{\alpha-1}\}$
$F(t)$	$1-\exp\{-\gamma t\}$	$1-\exp\{-\gamma t^{\alpha}\}$
$S(t)$	$\exp\{-\gamma t\}$	$\exp\{-\gamma t^{\alpha-1}\}$
$\lambda(t)$	γ	$\gamma\alpha t^{\alpha-1}$
$\Lambda(t)$	γt	γt^{α}
$\mathbb{E}[t]$	$\frac{1}{\gamma}$	$\gamma^{-\frac{1}{2}}\Gamma[\alpha^{-1}+1]$
$Var[t]$	$\frac{1}{\gamma^2}$	$\gamma^{-\frac{2}{\alpha}}[\Gamma(2\alpha^{-1}+1)]-[\Gamma(\alpha^{-1}+1)]$
γ,α	$\gamma>0$	$\gamma>0,\alpha>0$

[15]

Nelson-Aelen estimator of cumulative hazard rate[16]'s variance $\rightsquigarrow \hat{V}[\Lambda(t)] = \sum_{j|t_j\leq t}\frac{d_j}{r_j^2} \rightsquigarrow$ transformation $\ln\hat{\Lambda}(t)$ yields the percentage point $(1-\alpha)$ confidence interval for $\Lambda \in \{\hat{\Lambda}(t)\exp\{-z_{\frac{\alpha}{2}}\hat{\sigma}_{\Lambda(t)}\}, \hat{\Lambda}(t)\exp\{+z_{\frac{\alpha}{2}}\hat{\sigma}_{\Lambda(t)}\}\}$ where $\hat{\sigma}_{\Lambda(t)} \equiv$ st. dev. of $\ln\hat{\Lambda}(t)$, estimated via

$$\hat{\sigma}^2_{\Lambda(t)} = \hat{V}[\hat{\Lambda}(t)]/\hat{\Lambda}(t)^2.$$

PARAMETRIC REGRESSION MODELS

EXPONENTIAL AND WEIBULL DISTRIBUTIONS

The natural parametric starting point is an exponential model, a pure Poisson process has durations that are exponentially distributed (see Lancaster, 1990). The exponential duration distribution has a constant hazard rate γ that does not vary with t, the memoryless property of the exponential,

[15] $\Gamma(\alpha) = \int_0^{\alpha} e^{-x}x^{\alpha-1}dx$ is the Gamma function.
[16] cumulative hazard rate.

whose survivor function is written as $S(t) = \exp\{-\int_0^1 \gamma dt\} = \exp\{-\gamma t\}$. The density is $f(t) = -\frac{dS(t)}{dt} = \gamma \exp\{-\gamma t\}$ and the cumulative hazard is $\Lambda(t) = -\ln S(t) = \gamma t$, linear in t.

Nevertheless, the exponential is a one parameter distribution which might be a bit too restrictive in applications. Hence it may be appropiate to use the Weibull distribution, which, with $\alpha = 1$, collapses into an exponential.

Weibull's hazard is $\gamma\alpha t^{\alpha-1}$ which is monotonically $\searrow$ if $\alpha < 1$ or monotonically $\nearrow$ if $\alpha > 1 \rightsquigarrow$ special case of PH family. Parameters (α, γ) should be estimated from the data.

Recall some useful definitions:

$$\lambda(t) = \lim_{dt \to 0} Pr\{t \leq T \leq t + dt | T \geq t\}/dt = f(t) \rightsquigarrow \text{hazard function}$$

$$F(t) = Pr\{T \leq t\} = \int_0^t f(s)ds \rightsquigarrow \text{cumulative hazard function}$$

$$S(t) = 1 - F(t) = Pr\{T \geq t\} \rightsquigarrow \text{survivor function}$$

$$f(t) = \frac{dF(t)}{dt} \rightsquigarrow \text{reltionship between relative and cumulative hazards}$$

$$T_u = \frac{1}{\lambda[1 - H(x)]} \rightsquigarrow \text{average duration of pregnancies}$$

$S(t_a|\mathbf{x}, \theta) - S(t_b|\mathbf{x}, \theta)$, all vectors are of dimension $q \times 1 \rightsquigarrow$ duration data in economics are often interval censored, where in the present example t_a and t_b are two subsequent point on the temporal line, with $t_a < t_b$.

Can we use continuous distributions for analyzing discrete data?

COMPONENTS OF THE LIKELIHOOD FUNCTION

This section discusses the same concepts as from Lancaster (1990). There exist different likelihood specifications depending on which type of data is used, whether complete, truncated, or censored, especially for employment durations.

- Complete duration $\rightsquigarrow f(t)$
- Left-truncation at $t_L \leq t \rightsquigarrow \frac{f(t)}{S(t)}$
- Right-censoring at $t_{c_R} \rightsquigarrow S(t_{c_R})$
- Right-truncation at $t_R \rightsquigarrow f(t_R)/(1 - S(t_R))$
- Interval censoring at t_{c_L} and t_{c_R}.

WEIBULL EXAMPLE

$\lambda(t) = \gamma\alpha t^{\alpha-1}$, where $\alpha > 0$ and $\gamma > 0$. Regressors are introducable as $\gamma = \exp\{\mathbf{x}'\beta\}$ always strictly positive, recalling that as an exponential function with base greater than one behaves asymptotically towards zero on the left and exponentially exploding on the right side. $\alpha \equiv$ independent on the regressors.

a. $\ln f(t|\mathbf{x}, \beta, \alpha) = \ln[\exp(\mathbf{x}', \beta)\alpha t^{\alpha-1}\exp\{-\exp\{\mathbf{x}'\beta\}t^{\alpha-1}\}] =$
$= \mathbf{x}'\beta + \ln\alpha + (\alpha-1)\ln t - \exp\{\mathbf{x}'\beta t^{\alpha}\}$.
b. $\ln S(t|\mathbf{x}', \alpha, \beta) = \ln[\exp\{-\exp\{\mathbf{x}'\beta t^{\alpha}\}] = -\exp\{\mathbf{x}'\beta\}t^{\alpha}$
$\sum_{i=1}^{N}\{\delta_i, 1-\delta_i\} \rightsquigarrow \ln L = \sum_i \left[\delta_i\{\mathbf{x}'\beta + \ln\alpha + (\alpha-1)\ln t_i - \exp\{\mathbf{x_i}'\beta\}t_i^{\alpha}\}\right] -$
$-(1-\delta_i)\{\exp\{\mathbf{x_i}'\beta\}\}t_i^{\alpha}$.

$$\ln L = \sum_{i=1}^{N}\Big[\underbrace{\delta_i}_{\text{dummy for censoring}} \underbrace{\{\mathbf{x}'\beta + \ln\alpha + (\alpha-1)\ln t_i - \exp\{\mathbf{x}'\beta\}t^{\alpha}\}}_{\text{conditional log density}} +$$
$$+ \underbrace{(1-\delta_i)}_{\text{dummy for censoring}} \underbrace{\{\mathbf{x_i}'\beta\}t_i^{\alpha}\Big]}_{\text{conditional log survivor}}$$

First order conditions are:
$\frac{\partial \ln L}{\partial \beta} = \sum_i \left[\left(\delta_i - \exp\{\mathbf{x_i}'\beta\}t^{\alpha}\right)\mathbf{x}_i'\right] = \mathbf{0}$
$\frac{\partial \ln L}{\partial \alpha} = \sum_i \delta_i(\frac{1}{\alpha} + \ln t_i) - \ln t_i \exp\{\mathbf{x}_i'\beta\}t_i^{\alpha} = 0$
from the first f.o.c.,
$\sum_{i=1}^{N}\delta_i\mathbf{x}_i' \sum_{i=1}^{N}\exp\{\mathbf{x_i}'\beta\}t_i^{\alpha}\mathbf{x}_i'$
$\sum_{i=1}^{N}\left[\frac{\delta_i}{t_i^{\alpha}}\right] = \sum_{i=1}^{N}\exp\{\mathbf{x_i}'\beta\}$
$\sum_{i=1}^{N}\ln\delta_i - \sum_{i=1}^{N}\alpha\ln t_i = \mathbf{x_i}'\beta$
$\sum_{i=1}^{N}\left[\frac{\ln\delta_i - \alpha\ln t_i}{\mathbf{x}_i'}\right] = \hat{\beta}_{\mathbf{MLE}}$
$\rightsquigarrow$ maximum likelihood estimator for slopes of the regressors.
For consistency, with no censoring, $\mathbb{E}[\frac{\partial \ln L}{\partial \beta}] = 0$ requires $\mathbb{E}[T^{\alpha}|\mathbf{x}] = \exp\{-\mathbf{x}_i'\beta\}$.

USE OF MODEL ESTIMATES

Usual way $\equiv$ effects of regressors on the conditional mean in linear models. If $\gamma = \exp\{\mathbf{x_i}'\beta\} \rightsquigarrow \mathbb{E}[T^*|\mathbf{x}] = \exp\{-\mathbf{x_i}'\beta/\alpha\}\Gamma(\alpha^{-1}+1) = \exp\{-\mathbf{x_i}'\beta/\alpha\}\Gamma(\alpha^{-1})/\alpha$. Calculate expected length of completed spells at various values of $\mathbf{x}$. The econometrics of duration models is certainly concerned with the role of covariates but is especially involved with the slope of the hazard functions, bacause the economic theories make explicit predictions about the shape of the hazard function.

Interpretation of estimates of parametric duration models focuses on the Weibull hazard rate $\lambda(t) = \gamma\alpha t^{\alpha-1}$ and how it changes over time and with changes in the regressors. One sided test of $\alpha = 1$ are obviously of interest,

$$\frac{d\lambda(t)}{d\mathbf{x}} = \exp\{\mathbf{x_i}'\beta\}\alpha t^{\alpha-1}\beta = \lambda(t)\beta$$

$\Delta x \rightsquigarrow$ influences multiplicatively the $\lambda(t)$.
$\beta_j > 0 \rightsquigarrow +\Delta\lambda(t)$ as a component of $\mathbf{x} \nearrow$. If $\beta_j \nearrow \rightsquigarrow$ the hazard rate $\nearrow$
If $\Delta x \nearrow$.

LEAST SQUARES ESTIMATION

Less efficient than MLE and relying on correct specification of the density. Similar to Tobit model. Exponential duration $\mathbb{E}[T|\mathbf{x}] = 1/\gamma = \exp\{\mathbf{x_i}'\beta\}$, NLS[17] regression of t_i on $\exp\{-\mathbf{x_i}\beta\} \rightsquigarrow$ consistent though inefficient estimator for β. Alternatively $\rightsquigarrow \ln t = \mathbf{x_i}'\beta + \mathbf{u}$, $u \sim$ extreme value of type I

[17] Nonlinear least squares.

(**check type I and II distribution on michele pellizzari's notes**) $\mathbb{E}[\ln T|\mathbf{x}] = \mathbf{x_i}'\beta - c$, $c \equiv$ Euler's constant $\approx .5722$. β consistently estimable via linear regression of $\ln t_i$ on $\mathbf{x}_i$. Analytical censored moments for censored observations.

$$\begin{aligned}
\phi(\mathbf{x},\beta) &= \exp\{\mathbf{x}'\beta\} \\
\int_0^t \lambda(s|\mathbf{x})ds &= \int_0^t \lambda_0(s,\alpha)\exp\{\mathbf{x}'\beta\}ds \\
\Lambda(t|\mathbf{x}) &= \Lambda_0(t,\alpha)\exp\{\mathbf{x}'\beta\} \\
\ln\Lambda(t|\mathbf{x}) &= \ln\Lambda_0(t,\alpha) + \mathbf{x}'\beta \\
-\ln\Lambda_0(t,\alpha) &= \mathbf{x}'\beta \underbrace{-\ln\Lambda(t|\mathbf{x})} \\
&= \mathbf{x}'\beta + \qquad \underbrace{u}_{\text{type I extreme value distrib.}}
\end{aligned}$$

$\mathbb{E}[\ln\Lambda_0(T,\alpha)|T > t^*] \rightsquigarrow$ in censored samples, using a Heckman two steps procedure[18].

SOME IMPORTANT DURATION MODELS

Proportional hazards versus accelerated failure time metric is central in that it changes the metric in which the duration to failure time is measured.

PROPORTIONAL HAZARDS

$\lambda\{t|\mathbf{x}\} = \underbrace{\lambda_0(t,\alpha)}_{\text{basel. haz.,}} \; \underbrace{\phi(\mathbf{x},\beta)}_{\text{scaling factor}}$ $(*)$ $\rightsquigarrow$ the hazard function is factorizable into separate functions.

Polynomial baseline hazards are popular in the literature...usually $\phi(\mathbf{x},\beta) = \exp\{\mathbf{x}'\beta\}$

All hazard functions of the form $(*)$ are proportional to the β baseline hazard with scale factor $\phi(\mathbf{x},\beta) \rightsquigarrow$ widely used (as parameters' vector β can be estimated without specification of the functional form of $\lambda_0(.)$). Exponential, Weibull, Gompertz regression models are all PH models since their hazards are $\exp\{\mathbf{x}'\beta\}$, $\exp\{\mathbf{x}'\beta\}t^{\alpha-1}$ and $\exp\{\mathbf{x}'\beta\}\exp\{\alpha t\}$.

Piecewise constant hazard model $\rightsquigarrow \lambda_0(t,\alpha)$ is a step function with k segments $\rightsquigarrow \lambda_0(t,\alpha) = e^{\alpha_j}$, $c_{j-1} \le t < c_j$, $\forall j = 1,...,k$, $c_0 = 0$, $c_\infty = +\infty \rightsquigarrow$ other breakpoints $c_1,...,c_{k-1}$ are specified and the parameters $\alpha_1,...,\alpha_k$ are to be estimated. These are exponentiated to ensure that $\lambda_0(t,\alpha) > 0$. This model has more baseline parameters to estimate than models such as the Weibull which has only one baseline hazard parameter.

ACCELERATED FAILURE TIME

Similar to PH model but, rather than t, it admits $\ln t = \mathbf{x}'\beta + u \rightsquigarrow \neq$ distributions of $u \rightsquigarrow \neq AFT$ models. If $t \in (-\infty,\infty) \rightsquigarrow$ distribution for u can take any *continuous* distribution on $(-\infty,+\infty)$. The denomination AFT arises because $t = \exp\{\mathbf{x}'\beta\}v$, where $v = e^u$, has hazard rate $\lambda(t|\mathbf{x}) = \lambda_0(v|\exp\{\mathbf{x}'\beta\})$, where the baselinine hazard doesn't depend on t. Substitute $v = t\exp\{-\mathbf{x}\beta\}$ and get:

$$\lambda(t|\mathbf{x}) = \lambda_0(t\exp\{-\mathbf{x}'\beta\})\exp\{x'\beta\}$$

$\rightsquigarrow$ an acceleration of the baseline hazard $\lambda_0(t)$ if $\exp\{-\mathbf{x}\beta\} > 1$ and a deceleration if $\exp\{-\mathbf{x}'\beta\} < 1$. Log-normal model for t results if $u \sim N(0,\sigma^2)$; log-logistic model obtained if $u \sim$ *logistic*; Gamma model reached if $f(u) = \exp\{\alpha u - e^u\}/\Gamma(\alpha)$. $g(t) = \mathbf{x}'\beta + u$.

[18]??

FLEXIBLE HAZARDS MODELS

In flexible or flexibly parametric models or semi-parametric models, the researcher specifies first the hazard rate rather than the probability density function, i.e. $\lambda(t) = \mathbf{x}'\beta + a_1 t + a_2 t^2$ leading to U-shaped hazard functions. $\Lambda(t) = (\mathbf{x}'\beta)t + \frac{a_1}{2}t^2 + \frac{a_2}{3}t^3$[19]. Given $\lambda(t)$ and $\Lambda(t) \rightsquigarrow$ directly from the log-likelihood.

COX PH MODEL

Here, choose parametric functional forms that are flexible and hence provid some protection against mis-specification [...]. Fortunately there exists a semiparametric model that requires less complete distributional assumptions.

PROPORTIONAL HAZARDS MODEL

Their starting point is to propose a particular functional form of the hazard rate, the proportional hazard, conveniently factored as:

$$\lambda(t|\mathbf{x}'\beta) = \underbrace{\lambda_0(t)}_{\text{baseline hazard function of } t \text{ alone}} \times \underbrace{\phi(\mathbf{x},\beta),}_{\text{function of } \mathbf{x} \text{ alone, originally indep. on } t}$$

where $\phi(\mathbf{x},\beta) = \exp\{\mathbf{x}'\beta\}$, most commonly. The independence on time of the second term in the above expression is subsequently relaxed.

Semiparametric model is considered:

$\lambda_0(t)$'s functional form unspecified; $\phi(\mathbf{x},\beta) = \exp\{\mathbf{x}'\beta\}$.

If j^{th} regressor x_j is raised by one unit (i.e. an additional year of education) and the other regressors are kept constant, then,

$$\begin{aligned}\lambda(t|\mathbf{x}_{new},\beta) &= \lambda_0(t)\exp\{\mathbf{x}'\beta + \beta_j\} \\ &= \exp\{\beta_j\}\lambda_0(t)\exp\{\mathbf{x}'\beta_j\} \\ &= \underbrace{\exp\{\beta_j\}}_{\text{new hazard}}\ \underbrace{\lambda\{t|\mathbf{x}',\beta\}}_{\text{original hazard}}\end{aligned}$$

Change in hazard is $1 - \exp\{\beta_j\}$ times the original hazard.

If one uses instead calculus methods,

$$\tfrac{\partial\lambda(t|\mathbf{x},\beta)}{\partial x_j} = \lambda_0(t)\exp\{\mathbf{x}'\beta\}\beta_j = \beta_j\lambda(t|\mathbf{x},\beta),$$

the resulting change in the hazard is β_j times the original hazard.

Consistent with $\exp\{\beta_j\} \approx 1+\beta_j$, a non-calculus result. Statistical packages offer often estimates of β_j and $\exp\{\beta_j\}$ along with confidence intervals.

For more general forms of $\phi(\mathbf{x},\beta)$, changes in the regressor can be interpreted as having a multiplicative effect on the original hazard, since

$$\tfrac{\partial\lambda(t|\mathbf{x},\beta)}{\partial\mathbf{x}} = \lambda_0(t)\tfrac{\partial\phi(\mathbf{x},\beta)}{\partial x_j} = \Big[\lambda(t|\mathbf{x},\beta) \times \tfrac{\partial\phi(\mathbf{x},\beta)}{\partial x_j}\Big]/\phi(\mathbf{x},\beta)$$

$\rightsquigarrow$ requires knowledge of β but not of the baseline hazard function $\lambda_0(t)$.

[19] since $\Lambda(t) = \int \lambda(t)dt$

IDENTIFICATION OF THE COX REGRESSION MODEL

Identification in PH models is discussed in Cox ('72, '79) refers to the problem of how to estimate β in the PH model. In fact that does not require simultanueous estimation of λ_{0t}, the baseline hazard function. Setup $\rightsquigarrow t_1 < t_2 < ... < t_j < ... < t_k \rightsquigarrow$ observed discrete failure times of the spells in a sample of N $(\geq k)$. The risk set $R(t_j)$ is defined to be the set of individuals who're at risk of failing just before the j^{th} ordered failure $\rightsquigarrow D(t_j) \equiv$ set of subjects having births occurring at time t_j, $d_j \equiv$ # of mothers that give birth at time t_j. To sum up:

$$\begin{aligned} R(t_j) &= \{l : t_l \geq t_j\} \rightsquigarrow \textit{set of spells}^{20} \textit{ at risk at time } t \\ D(t_j) &= \{l : t_l = t_j\} \rightsquigarrow \textit{set of spells completed at } t \\ d_j &= \textstyle\sum_l \mathbf{1}\{t_l = t_j\} \rightsquigarrow \#\textit{ of spells completed at } t_j \end{aligned}$$

where $\mathbf{1}$ stands for a dummy variable taking values 1 if the spell was completed before the end of the survey and 0 otherwise.

Tied data are possible at $d_j > 1$. The former of the last three expressions includes spells that are neither completed nor censored.

$$\begin{aligned} Pr[T_j = t_j | R(t_j)] &= \frac{Pr[T_j = t_j | T_j \geq t_j]}{\sum_{l \in R(t_j)} Pr[T_l = t_l | T_l \geq t_j]} \\ &= \frac{\lambda_j(t_j | \mathbf{x_j}, \beta)}{\sum_{l \in R(t_j)} \lambda_l(t_j | \mathbf{x_l}, \beta)} \\ &= \frac{\phi(x_j, \beta)}{\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta)} \end{aligned}$$

In the first line, the numerator indicates the probability that a particular spell at risk ending at time $t_j \equiv$ probability of failure for spell j. The denominator instead expresses the cond. prob. that a spell of any individual in the risk set $R(t_j)$ fails, i.e. has an additional child.

In the third line, λ_0 has dropped out due to the PH assumption[21] $\rightsquigarrow$ intercept not identified $\rightsquigarrow$ basis for estimating β. We must control for tied durations, more likely when we have grouped observations (durations).

Suppose there are two tied values at time t_j, for individuals j_1 and j_2 with regressors $\mathbf{x_{j1}}$ and $\mathbf{x_{j2}}$. If j_1 fails[22] and x_{j_2} then the probability of occurrence of a failure of them is:

$$\frac{\phi(x_{j1}, \beta)}{\sum_{l \in Rt_j} \phi(x_l, \beta)} + \frac{\phi(x_{j2}, \beta)}{\sum_l R(t_j) \phi(x_l, \beta)}$$

A similar term arises if j_2 fails before j_1 and the likelihood contribution's the sum of the two possibilities. Exact likelihood becomes complicated with many tied values.

Cox and Oakes '84 $\rightsquigarrow Pr[T_j = t_j | j \in R(t_j)] \approx \frac{\prod_{m \in D(t_j)} \phi x_m, \beta}{[\sum_{l \in R(t_j)} \phi(x_l, \beta)] d_j}$, approximation working well if the # of failures is small relative to the number of individuals (mothers) at risk (not our case) in the population.

Cox $\rightsquigarrow$ partial likelihood function $\rightsquigarrow$ joint product of $Pr[T_j = t_j | j \in R(t_j)]$ over the k ordered failures. Then,

[20] Spacing between subsequent births, i.e. this setup may be useful for *spacing* not for *starting* fertility behaviour.

[21] PH assumption refers to the hazard model being composed by proportional hazards rate rather than by accelerated failure time rate, referred is Stata as AFT, and proportional hazard as `PH` in the options of the command `streg`.

[22] If a mother remains pregnant for starting or if the mother has an additional child for spacing behaviours, let's say. The present exposition is taken from the Cameron and Trivedi (2005)'s manual, specifically in chapter 17.8.2. Identification of the PH model.

$$L_p(\beta) = \prod_{j=1}^{k} \frac{\prod_{m \in D(t_j)} \phi(\mathbf{x_m}, \beta)}{\sum_{l \in R(t_j)} \phi(\mathbf{x_l}, \beta) d_j}$$

Cox proposed estimation of β by minimizing the log partial likelihood function

$$L_p = \sum_{j=1}^{k} \Big\{ \sum_{m \in D(t_j)} \ln \phi(\mathbf{x_m}, \beta) - \underbrace{d_j \ln \Big(\sum_{l \in R(t_j)} \phi(\mathbf{x_l}, \beta) \Big) \Big\}}.$$

only in the underlined term there appear censored spells, because they do not contribute to the observed births, but until they're censored, they affect the size of the risk set.

Here we change indexation of mothers form j to i,

$$\ln L_p(\beta) = \sum_{i=1}^{k} \delta_i [\sum_{m \in D(t_j)} \ln \phi(x_m, \beta) - \ln(\sum_{l \in R(t_i)} \phi(\mathbf{x_l}, \beta))].$$

where the binary variable δ_i is defined as:

$$\delta_i = \begin{cases} 1 & \text{for censored obs.;} \\ 0 & \text{else;} \end{cases}$$

and

$$\begin{gathered} \phi(\mathbf{x}, \beta) = \exp\{\mathbf{x}'\beta\} \Leftrightarrow \\ \Leftrightarrow \ln \phi(\mathbf{x}, \beta) = \mathbf{x}'\beta, \text{ with F.O.C.:} \\ \frac{\partial \ln L_p(\beta)}{\partial \beta} = \sum_{i=1}^{N} \delta_i \big[\mathbf{x}_i - \mathbf{x}_i^*(\beta) \big] = 0, \\ \mathbf{x}_i^* = \frac{\sum_{l \in R(t_i)} \mathbf{x_l} \exp\{\mathbf{x}_l'\beta\}}{\sum_{l \in R(t_i)} \exp\{\mathbf{x}_l'\beta\}} \end{gathered}$$

which is the weighted average of the regressors $\mathbf{x}_l$ for subjects at risk of failure at time t_i.

The partial likelihood $\equiv$ limited information likelihood, as $\lambda_0(t)$ has dropped out. But it is neither a conditional likelihood nor a marginal likelihood. Then is $L_p(\beta)$ a valid likelihood function? Andersen et al. 1993 show that even $\ln L_p(\beta)$ yields a consistent estimator for β. See Lancaster (1990), ch. 9,

$$\mathbf{A}(\beta) = -\mathbf{B}(\beta), \text{ and } \hat{\beta}_{ML} \rightsquigarrow_a N\big[\beta, (\mathbb{E}\big[\tfrac{\partial \ln L_(\beta)}{\partial \beta \partial \beta}\big])^{-1}\big]$$

The indexation p under the likelihood term stands for *partial*.

Estimator is inefficient (WHY?).

SURVIVOR FUNCTION FOR THE COX *PH* MODEL

Many studies stop at the estimate of β. Other studies focus on the shape of the baseline hazard function. For the PH model, it is possible to estimate nonparametrically the baseline hazard functionor survivor, once β is obtained after having maximized the partial likelihood. Estimates analogous to those of Kaplan-Meier.

PH hazard function's $\rightsquigarrow S(t|\mathbf{x}, \beta) = S_0(t)^{\phi(\mathbf{x}, \beta)}$, using $S(t|\mathbf{x}, \beta) = \exp\{-\int_0^t \lambda_0(s)\phi(\mathbf{x}, \beta)ds\}$ and defining $S_0(t) = \exp\{-\int_0^t \lambda_0(s)ds\}$.

Assume a discrete time formulation with baseline hazard $1 - \alpha_j$, at discrete failure time t_j, $j = 1, ..., k$. $\hat{\alpha}_j$ is the solution to

$$\sum_{l \in D(t_j)} \frac{\phi(\mathbf{x_l}, \hat{\beta})}{1 - \hat{\alpha}_j^{\phi(\mathbf{x}_l, \beta)}} = \sum_{m \in R(t_j)} \phi(\mathbf{x}_m \hat{\beta}), \; j = 1, ..., k.$$

$\hat{\beta}_{PML} \equiv$ partial likelihood estimator of β, $\rightsquigarrow S_0(t) = \prod_{j|t_j \leq t} \alpha_j$, the cumulative product of the instantaneous conditional survivor probabilities. Estimated survival function (baseline) is,

with no regressors, $\hat{S}_0(t) = \prod_{j|t_j \leq t}$

$\hat{S}_0(t)$ reduces to the Kaplan - Meier estimator,

normalize $\phi(\mathbf{x}_l, \beta) = 1$ and the expression yields hazard rate $1 - \hat{\alpha}_j = \frac{d_j}{r_j}$.

With regressors but without ties, the baseline hazard rate,

$$1 - \hat{\alpha}_j = \phi(\mathbf{x}_j, \hat{\beta}) / \sum_{m \in R(t_j)} \phi(\mathbf{x}_j, \hat{\beta}).$$

Survivor function for individuals with regressors $\mathbf{x} = \mathbf{x}^*$ can be estimated via $\hat{S}(t|\mathbf{x}^*, \beta) = \hat{S}_0(t)^{\phi(\mathbf{x}^*, \hat{\beta})}$.

Linear transformation of regressors don't change the estimates of β, but they do change the hazard function.

$$\begin{aligned} \lambda(t|\mathbf{x}, \beta) &= \lambda_0(t) \exp\{\mathbf{x}\beta\} \\ = \lambda_0(t) \exp\{\mathbf{x}'\beta\} & \underbrace{\exp\{(x - \bar{x})'\beta\}}_{\text{deviation from the mean of x's}} \\ &= \lambda_0^*(t) \exp\{(x - \bar{x})'\beta\} \end{aligned}$$

new baseline hazard $\rightsquigarrow$ demeaning each regressor'll change the baseline hazard + care in interpretation.

DERIVATION OF THE SURVIVOR FUNCTION

Following Kalbfleisch and Prentice (2002), we derive α_j.

$$\begin{aligned} S(t_j|\mathbf{x}, \beta) - S(t_{j+t}|\mathbf{x}, \beta) &= S_0(t_j)^{\phi(\mathbf{x}, \beta)} - S_0(t_{j+1})^{\phi(\mathbf{x}, \beta)} \\ &= [\alpha_j^{-1} S_0(t_{j+1})]^{\phi(\mathbf{x}, \beta)} - S_0(t_{j+1})^{\phi(\mathbf{x}, \beta)} \\ &= [\alpha_j^{-\phi(\mathbf{x}, \beta)} - 1] S_0(t_{j+1})^{\phi(\mathbf{x}, \beta)} \end{aligned}$$

since $S_0(t_{j+1}) = \prod_{l=1}^{j} \alpha_l = \alpha_j S_0(t_j)$, and the first term means that subject with duration time t_j has likelihood contribution equal to the probability of survival time $t > t_j$.

For these subjects who are censored at time t_j, the likelihood contribution is the propbability of survival $t > t_j$ or $S_0(t_{t+1})^{\phi(\mathbf{x}, \beta)}$. So subjects that either die or are censored at $[t_j, t_{j-1}]$ contribute probability $S_0(t_{j+1})^{\phi(\mathbf{x}, \beta)} = \prod_{l=1}^{j} \alpha^{\phi(\mathbf{x}, \beta)}$ with an additional multiplier $\{\alpha_j^{-\phi(\mathbf{x}\beta)} - 1\}$ for subjects that deliver before the end of the survey time. The over all failure times the likelihood $\rightsquigarrow L(\alpha, \beta) = \prod_{j=1}^{k} \big[\prod_{l \in D(t_j)} \big(\alpha_j^{-\phi(\mathbf{x}_l, \beta)} - 1 \big) \prod_{m \in R(t_j)} \alpha^{\phi(\mathbf{x}, \beta)} \big]$.

$\ln L(\alpha, \beta) = \sum_{j=1}^{k} \big\{ \sum_{l \in D(t_j)} \ln(\alpha_j^{\phi(\mathbf{x}, \beta)} - 1) + \sum_{m \in R(t-j)} -\phi(\mathbf{x}, \beta) \ln \alpha_j \big\}$.

$\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha_j} = 0 \Leftrightarrow$ as before.

TIME VARYING REGRESSORS

Beyond gender, confession, educational attainment, age, marital status, number of children even concieved, tending towards age at marriage break, *residence, location*, and so on.

DISCRETE-TIME PROPORTIONAL HAZARDS MODEL

Grouped duration models are appropiate when failure times are aggregated and observed/recorded at aggregate time intervals like a *week*, a *month*, a *year*. A simple method is to form a panel and estimate a stacked logit or probit model of the probability of an individual failure in each period, with separate intercept for period (fixed effects). Discrete time variant of a continuous time PH model considered by several authors such as Kalbfleisch and Prentice (1980), [...][23].

Grouped data with grouping point t_a, where a stands for annum in our case, $a = 1, ..., A$, the discrete hazard function is defined by:

$$\lambda^d(t_a|\mathbf{x}) = Pr[t_{a-1} \leq T < t_a|T \geq t_{a-1}, \mathbf{x}(t_{a-1})],\ a = 1, ..., A$$

Time varying regressors are permitted. The associated discrete time survivor function is:

$$S^d(t_a|\mathbf{x}) = Pr[T \geq t_{a-1}|\mathbf{x}] = \prod_{s=1}^{a-1}(1 - \lambda^d(t_s|\mathbf{x}(t_s))).$$

We obtain the first general relation between the discrete and continuous time hazards. Discrete time hazards $\rightsquigarrow$ probability of failure in $[t_{a-1}, t_a)$ divided by the probability of surviving until at least time t_{a-1}:

$$\lambda^d(t_a|\mathbf{x}) = \frac{S(t_a|\mathbf{x}) - S(t_a|\mathbf{x})}{S(t_{a-1}|\mathbf{x})}.$$

$S(t|\mathbf{x}) \equiv$ survivor function.
$S(t|\mathbf{x}) = \exp\{-\int_0^t \lambda(s)ds\}$, and, after some algebra,

$$\lambda^d(t_a|\mathbf{x}) = 1 - \exp\{-\int_{t_{a-1}}^{t_a} \lambda(s)ds\}.$$

Now, specialize to the discrete - time hazard model associated with the continuous PH model $\rightsquigarrow$

$$\lambda(t) = \lambda_0(t)\exp\{\mathbf{x}(t_{a-1})'\beta\},\text{ for } t \in [t_{a-1}, t_a).$$

Regressors are constant within the interval, but can vary across intervals. $\lambda_0(t)$ instead can vary w/in interval.

$$\begin{aligned}\lambda^d(t_a|\mathbf{x}) &= 1 - \exp(-\exp\{\mathbf{x}(t_{a-1})'\beta\}) \int_{t_{a-1}}^{a_t} \lambda_0(s)ds \\ &= 1 - \exp\{-\lambda_0(t_a)\exp\{\mathbf{x}(t_{a-1})'\beta\}\}.\end{aligned}$$

Associated discrete $t_i = \int_{t_{a-1}}^{t_a} \lambda(s)ds$.
The survivor function:

$$S^d(t_a|\mathbf{x}) = \prod_{s=1}^{a-1} \exp\{-\exp\{\ln\lambda_{0s}\mathbf{x}(t_{s-1})'\beta\}\}.$$

The density for the i^{th} subject is th eproduct of the survivor function in eachh period that the subject survives (i.e. the mother does not deliver an additional child) times the hazard at the time of failure.

$$\begin{aligned}L(\beta, \lambda_{01}, ..., \lambda_{0A}) = \prod_{i=1}^{N}\big[\prod_{s=1}^{a_i - 1}\exp\{-\exp\{\ln\lambda_{0s} + \mathbf{x_i}(t_{a-1})'\beta\}\}\big] \times \\ \times\big[1 - \exp\{\exp\{\ln\lambda_{0a_i} + \mathbf{x}_i(t_{a-1})'\beta\big],\end{aligned}$$

when censoring is ignored for simplicity and failure is assumed to occur at time t_{a_i} for the i^{th} observation. At least one failure is assumed to occur in each interval $[t_{a-1}, t_a)$. The MLE maximizes the latter likelihood function w.r.t. βa dn $\lambda_{01}, ..., \lambda_{0A}$, such as a polynomial in time. $\lambda_0(s) = \int_{t_{a-1}}^{t_a} \alpha s^{\alpha-1}ds$, as in a fully parametric Weibull model.

[23] following Blake, Lunde and Timmerman ('99)

HAN AND HAUSMAN APPROACH

Flexibly estimate the baseline hazard $\lambda_0^d(t)$, while maintaining a parametric form for the function of covariates. $\lambda_i(\tau) \equiv$ hazard rate for obs. i, cond. prob. that a spell terminates in $(\tau, \tau+\Delta)$, as a PH form:

$$\lambda_i(\tau) = \lambda_0(\tau)\exp\{-\mathbf{x}'\beta\}.$$

Taking log after integration and rearrangin yields:

$$\Lambda_0(t) - \mathbf{x}_i'\beta = \epsilon_i, \text{ where } \Lambda_0(t) = \ln\int_0^t \lambda_0(\tau)d\tau \equiv \text{log of cumulative baseline hazard, and}$$
$$\epsilon_i \equiv \ln\int_0^t \lambda_i(\tau)d\tau.$$

$\rightsquigarrow Pr[\text{failure in t}] = \int_{\Lambda_0(t-1)-\mathbf{x}'\beta}^{\Lambda_0(t)-\mathbf{x}'\beta} f(\epsilon)d\epsilon.$

Let $y_{it} = 1$ if the i^{th} mother experiences a baby - birth in period t, and $y_{it} = 0$ otherwise. Then the joint-likelihood of N observations [24] is given by:

$$\ln L(\underbrace{\underbrace{\beta}_{\text{slope of the reg.}}, \underbrace{\Lambda_{0(1)}, ..., \Lambda_{0(T)}}_{\text{basel. haz. param.}}}_{\text{estim. without imposing a functional form}}) = \sum i = 1^N \sum_{t=1}^T y_{it} \ln\left[\int_{\Lambda_0(t-1)}^{\Lambda_0(t)} f(\epsilon)d\epsilon\right].$$

The integral in the log-likelihood is of course the difference in the cdf $\left[\Lambda_0(t-1) - \mathbf{x}_i'\beta, \Lambda_0(t) - \mathbf{x}'\beta\right]$. The precise form of this expression depends on the functional form of the cdf. If $\epsilon_i \sim N(\mu, \sigma^2) \rightsquigarrow$ loglik $\equiv$ probit form; if $\epsilon_i \sim$ extreme value distribution $\rightsquigarrow$ loglik $\equiv$ logit model. Under normality $\rightsquigarrow$ $Pr\left[\Lambda_0(t+1) < \mathbf{x}_i'\beta + \epsilon_i \leq \lambda_0(t+1)\right] = \Phi\left[\Lambda_0(t+1) - \mathbf{x}_i'\beta\right] - \Phi\left[\Lambda_0(t) - \mathbf{x}'\beta\right]$.

Contrarily to the partial likelihood method, treating the baseline hazard as a nuisance and then eliminating it, the Han and Hausman, 1990 approach estimates all unknown parameters simultaneously at a modest computational cost. Monte Carlo study...

DISCRETE TIME BINARY CHOICE

An alternative approach is binary choice modeling for transitions leading to a discrete time transition model of the type:

$$Pr\left[t_{a-1} \leq T < t_a | T \geq t_{a-1} | \mathbf{x}\right] = F[\lambda_a + \mathbf{x}'(t_{a-1})\beta],\ a = 1, ..., A.$$

Regressors' coefficient allowed to be constant over time. Obvious choice of F are the standard normal and the logistic cumulative density function. Then λ_a and β can be estimated by a stacked logit or a stacked probit model where a separate intercept is permitted at each duration interval. Resulting likelihood:

$$L(\beta, \lambda_1, ..., \lambda_N) = \prod_{i=1}^N \left[\prod_{s=1}^{a_i-1} 1 - F(\lambda_s + \mathbf{x}_i(t_{s-1})\beta)\right] \times F(\lambda_{ai} + \mathbf{x}'(t_{a_i-1})\beta),$$

similar to the loglik. of the discrete PH model.

[24] The whole sample. In our case $N \approx 11,620$ women interviewed on their retrospective fertility path of $T = 9$ years, for $t = 1993, ..., 2002$.

MIXTURE MODELS AND UNOBSERVED HETEROGENEITY

FINITE MIXTURE MODEL

The sample is a described as a probabilistic mixture from two subpopultions with $f_1(t|\mu_1(\mathbf{x}))$ and $f_2(\mu_2(t|\mathbf{x}))$, where the vector of independent variable is the same for both subpopulations $\rightsquigarrow \pi f_1(.)+(1-\pi)f_2(.)$, $\pi \in [0,1] \rightsquigarrow$ two components mixture $\rightsquigarrow$ parameters to be estimated $\rightsquigarrow \mu_1, \mu_2, \pi$. $\pi \equiv$ const. or $\pi = \frac{e^\lambda}{1+e^\lambda}$ (logistic), and λ (funct. of other covariates). Some i came from $f_1(.)$ and some other from $f_2(.) \rightsquigarrow$ due to the latent partitioning of the sample (unobserved heterogeneity), or simply linear approximation of densities $f_1(.)$ and $f_2(.)$ make better fit. Each subpopulation $\equiv$ a "type", say there exist m homogeneous subpopulations called <u>components</u>. j^{th} component $\equiv$ fraction π_j of the total population, $\sum_{j=1}^m \pi_j = 1$.

$$\pi_j(t_i|\mathbf{x}_i, \pi_j, \beta) = \sum_{j=1}^m f(t_i|\mathbf{x}_j, \underbrace{\nu_j}_{\text{estimated support point}}, \beta) \underbrace{\pi}_{\text{associated probability}} (\nu_j),\ \forall i = 1, ..., N.$$

Semiparametric approach $\rightsquigarrow$ á la Heckman and Singer ('84) $\rightsquigarrow$ as if unobserved heterogeneity observed. Happening if π_j not subject to any parametric assumption $\rightsquigarrow$ semiparametric mixture model for t. Estimable under known or unknown # of components. Usually π_j, $j = 1, ..., m$ are known $\rightsquigarrow$ maximum likelihood estimates $\rightsquigarrow$ Non Parametric Maximum Likelihood Estimator. # of components unknown $\rightsquigarrow$ inference issues. Good to think at a small # of latent classes, rather than a continuum of types.

LATENT CLASS INTERPRETATION

$d_{i1}, ..., d_{im}$ is a dummy for identifying who's taken from which class of individuals drawn, say from j^{th} class for $i = 1, ..., N \Leftrightarrow$ each observation $\Leftrightarrow$ sample from each of the m class (here $m = 1, 2, 3$, i.e. women exposed to conflict and women not expose to conflict, on the base of the residence variable, called `commune2` in the dataset `ind_92_02.dta`). Assume the model is identified:

$$(t_i|d_i, \mu, \pi) \sim i.i.d., \text{ with densities:}$$
$$\sum_{j=1}^m d_{ij} f(\underbrace{t_i}_{\text{duration of ind. } i} | \underbrace{\mu_j}_{\text{means duration of ind. in class } j}) = \prod_{j=1}^m f(t_i|\mu_j) d_{ij}$$

$\mu_j = \mu(\mathbf{x}_j, \beta_j)$, where the dimensions of $\mathbf{x}$ are $m \times 1$, the same of β and $\mu = (\mu_1, ..., \mu_m)$, and $(\mathbf{d}_i|\mu_{m\times 1}, \pi_{m\times 1}) \sim i.i.d.$ with a multinomial distribution (??).

$$\prod_{j=1}^m \prod_{j=1}^{d_{ij}},\ 0 < \pi_j < 1,\ \sum_{j=1}^m \pi_j = 1.$$
$$(t_i|\mu, \pi) \sim_{iid} \underbrace{\sum_{j=1}^m \underbrace{\prod_{j=1}^{d_{ij}}}_{\text{probability weight}} \underbrace{f_j(t|\mu_j)^{d_{ij}}}_{\text{density function}}}_{\text{likelihood contribution}}.$$
$$\underbrace{\rightsquigarrow L(\beta_{m\times 1}, \pi_{N\times 1}|\mathbf{t}_{N\times 1}) = \prod_{i=1}^N \sum_{j=1}^m \prod_j^{d_{ij}} f_j(t_j, \mu_j)^{d_ij}}_{\text{likelihood function}}.$$

the dimension of β is that of the individuals, not that of the classes.

MAXIMIZATION OF THE LIKELIHOOD

via EM algorithm where $\mathbf{d} = (d_1, ..., d_n)$ are treated as missing data, if d were observable:

$$\ln L(\mu, \pi|\mathbf{t}, \mathbf{d}) = \sum_{i=1}^{N}\sum_{j=1}^{m} d_{ij} \ln f_{ij}(\mathbf{t}_{ij}, \mu) + \sum_{i=1}^{N}\sum_{j=1}^{m} d_{ij} \ln \pi_j$$

if π_j, $j = 1, ..., m$ are given, the posterior distribution of obs. t_i belonging to j, say, z_{ij} is given by:

$$z_{i,j} \equiv Pr[y_i \in \text{pop.}j] = \frac{\pi_j \cdot f_j(y_i|\mathbf{x}_i, \beta_j)}{\sum_{j=1}^{m} \pi_j \cdot f_j(y_i|\mathbf{x}_i, \beta_j)}$$

where the numerator is the relative density and the denominator the cumulative one.

$\underbrace{\mathbf{x}_{N\times 1}, \mathbf{y}_{N\times 1}}_{\text{categorical}}$

$\underbrace{\beta_{m\times 1}}_{\text{categorical}}$

average of z_{ij} is, over i, the prob. that a randomly drawn individual belongs to the j^{th} subpopulation.

$$\mathbb{E}[z_{ij}] = \pi_j$$
$$\hat{z}_{ij} \rightsquigarrow \text{estimates of } \mathbb{E}[d_{ij}].$$
$$\mathbb{E}[L(\beta_1, ..., \beta_m, \pi_{m\times 1}|\mathbf{t}, \hat{\mathbf{z}}, \mathbf{x}_1, ..., \mathbf{x}_m)] = \sum_{i=1}^{N}\sum_{j=1}^{m} \hat{z}_{ij} \ln f_j(t_j, \mu(\mathbf{x}, \beta_j)) + \sum_{i=1}^{N}\sum_{j=1}^{m} \hat{z}_{ij} \ln \pi_j$$

the E step pf the algorithm.

the M-step of the algortithm $\rightsquigarrow$ maximisez $\mathbb{E}[L(.)]$ by solving for the first order conditions w.r.t. β and π:

$$\frac{\partial \mathbb{E}[L]}{\partial \beta} = 0 \rightsquigarrow \hat{\pi}_j - \frac{1}{N}\sum_{i=1}^{m} \hat{z}_{ij} = 0, \forall i = 1, ..., m;$$
$$\frac{\partial \mathbb{E}[L]}{\partial \pi} = 0 \rightsquigarrow \sum_{i=1}^{N}\sum_{j=1}^{m} \hat{z}_{ij} \frac{\partial \ln f_j(t_i|\beta_j)}{\partial \beta_j} = 0.$$

To get new values of $\hat{z}_{ij}$, iterate through E and M steps.

How to choose m? No guiding prior theory.

$m \cdot \dim[\beta] + m - 1 \equiv$ dimension of parameters to be estimated, start with $m = 2$ and then use diagnostick checks. Penalized likelihood criterion $\rightsquigarrow AIC, BIC$. LR test not good here $\rightsquigarrow$ check across candidate latent class models. Overparametrization $\rightsquigarrow$ misspecification.

Selecting from competing models $\rightsquigarrow$ goodness of fit, model diagnostics.

Stock or flow sampling? $\rightsquigarrow$ A/D (average interrupted duration) versus expected elapsed duration, completed duration. Renewal theory...stationary Poisson process w/stationary parameters. [...] $\underbrace{\text{backwards recurrence time and forwards recurrence time}}_{\text{duration from current state to transition}}$.

$\mathbb{E}[N|t] \equiv$ expected # of events in the time interval $[0, t)$. $\lim_{dt\rightsquigarrow 0} \mathbb{E}[N(t)] \equiv$ renewal intensity.

Salant, '77 $\rightsquigarrow$ flow versus stock sampling $\rightsquigarrow AID \equiv$ average interrupted duration $ACD \equiv$ average completed duration $\rightsquigarrow \neq$ between them $\rightsquigarrow$ hazard function being heterogeneous.

STOCK SAMPLING

Sampling in the survey period from the stock of individuals who are in a given state.

FLOW SAMPLING

Sample those who enter the state during a particular interval. Survey is more likely to capture longer spells rather than shorter spells $\rightsquigarrow$ upwards bias $\rightsquigarrow$ length bias sampling.

$AID > AIC \rightsquigarrow$ interruption bias.

Density of interrupted spells:

$$f(u) = \frac{\bar{G}(u)}{\int \bar{G}(u)du} = \frac{\bar{G}(u)}{\mathbb{E}[t]}.$$

$g(t) \equiv$ density of completed spells.

$\bar{G}(u) = \int g(x)dx \rightsquigarrow$ survivor function corresponding to density $g(u)$. $\mathbb{E}[t] \equiv$ mean of the distribution of durations.

$g(t)$ is exponential $\rightsquigarrow$ stochastic process for the *event* is a Poisson process. $f(u)$ exponential as well.

$$\mathbb{E}[u] = \tfrac{1}{2}\{\mathbb{E}[t] + \tfrac{V[t]}{\mathbb{E}[t]}\}.$$

$$\underbrace{\mathbb{E}[t^{(s)}]}_{\text{mean of duration for the constant stock}} = \mathbb{E}[t] + \tfrac{\mathbb{E}[t]}{V[t]} > \mathbb{E}[t]$$

holding both is $V[t] \geq \mathbb{E}[t]$ or $\leq \mathbb{E}[t]$.

$\mathbb{E}[u] < \mathbb{E}[t]$ if haz. rate is $\nearrow$ in t,

$\mathbb{E}[u] > \mathbb{E}[t]$ if haz. rate is $\searrow$ in t.

SPECIFICATION TESTING

$\longrightarrow$ 4 types:

1. inclusion and exclusion of covariates;
2. tests of functional forms of the survival function j;
3. tests of unobserved heterogeneity;
4. join tests of state dependence and unobserved heterogeneity.

For 1., just use Wald type tests; for 2., they're the same as tests of unobserved heterogeneity if the restriction is on the absence of unobserved heterogeneity $\rightsquigarrow$ test whether heterogeneoty (variance) parameter is zero.

HYPOTHESES TESTING

$\rightsquigarrow$ score test based on the exponential null model. Heterogeneity $\Leftrightarrow$ duration dependence $\rightsquigarrow$ do a joint test rather than a separate one $\rightsquigarrow$ locally heterogeneous Weibull models, see Lancaster, '85.

A local heterogeneous density is generated by considering a Taylor expansion of arbitrary density around ν, yielding: $S(t|\nu) = e^{-\mu t^\alpha \nu} = e^{-\mu\epsilon} = e^{-\epsilon[1+(-\epsilon)(\nu-1)+(\epsilon^2/2)(\nu-1)^2]+o(\epsilon^3)}$, $\epsilon = \mu t^\alpha$. From second line:

$$\mathbb{E}[e^{-\epsilon\nu}] = e^{-\epsilon}[1 + (\epsilon^2 \underbrace{\sigma^2}_{\text{variance of the heterogeneity distribution, first order taylor expansion}} /2)] \equiv S_m(t)$$

where $S_m(t)$ is the sample variance grouped for sub-populations, say $m = 1, 2, 3 \equiv r, u, c$,

$$\rightsquigarrow f_m(t) = -\frac{\partial S_m(t)}{\partial t} = -\frac{\epsilon}{\nu-1} + \frac{\epsilon^2}{2} \cdot (\nu-1)^2 + o(\epsilon^3) =$$

$$= \underbrace{-\frac{\mu t^\alpha}{\nu-1} + \frac{\mu^2 t^{2\alpha}}{2}(\nu-1)^2 + o(\mu^3 t^{3\alpha})}_{\text{A}}$$

$$\mathbb{E}[A] = \tfrac{\epsilon^2\sigma^2}{2}, \text{ this means that}$$

$$\mathbb{E}[-\tfrac{\epsilon}{(\nu-1)}] = 0!!$$

Hence,

$$f_m(t) = -\frac{\partial S_m(t)}{\partial t} = \alpha\mu t^{\alpha-1}\cdot e^{-\epsilon}\big[1+\big(\tfrac{\epsilon^2\sigma^2}{2}\big)\big] - e^{-\epsilon}\big[2\epsilon(\alpha\mu t^{\alpha-1})\tfrac{\sigma^2}{2}\big]$$
$$= \alpha\mu t^{\alpha-1}e^{-\epsilon}\big[1+\sigma^2(\epsilon^2-2\epsilon)/2\big].$$

Write the log-likelihood as follows:

$$\ln L(\alpha,\beta,\sigma^2) = \textstyle\sum_{i=1}^N \ln\big\{\big[f_m(t)\big]^{\delta_i}\cdot\big[S_m(t)\big]^{1-\delta_i}\big\}$$
$$= \textstyle\sum_{i=1}^N \delta_i\big[\ln\alpha + (\alpha-1)\ln t_i + \ln\mu_i + \ln(1+\sigma^2(\epsilon_i^2-2\epsilon_i)/2) -$$
$$-\epsilon_i + (1-\delta_i)\ln(1+\sigma^2\epsilon_i^2/2)\big],$$

$\delta_i \equiv$ censoring indicator.

$$\ln\mu_i = \beta_0 + \mathbf{x}_i'\beta_1 \text{ and } \epsilon_i = \mu_i t_i^\alpha \rightsquigarrow \text{generalized error.}$$

Tests for heterogeneity:

$H_0 : \sigma^2 = 0$ and $\alpha = 1 \rightsquigarrow$ no heterogeneity in the population about the covariates + exponential distribution specification, recall that w/$\alpha = 1$, Weibull collapses into an exponential.

$$\underbrace{\theta}_{4\times1} = \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} \text{ and } \underbrace{\theta}_{2\times1} = \begin{bmatrix}\sigma^2\\ \alpha\end{bmatrix};\ \underbrace{\theta_2}_{2\times1} = \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} \text{ and } \underbrace{\theta_0}_{4\times1} = \begin{bmatrix}0\\ 1\\ \beta_0\\ \beta_1\end{bmatrix} \rightsquigarrow \text{restricted vectors}$$

connected with the null hypothesis of the test for heterogeneity.

Consider unrestricted data: joint score test statistic (or Lagrangian multiplier test) is:

$$LM\ \underbrace{H}_{\text{heterogeneity}}\ \underbrace{D}_{\text{duration}} = \frac{1}{d}\cdot\underbrace{\begin{bmatrix}\psi'(1) & 1\\ 1 & 1\end{bmatrix}}_{\text{not diagonal}}\cdot s,$$ [25]

$\mathbf{s}' = [1/2\sum_i(\epsilon_i^2-2\epsilon_i), \sum_i(1+(1-\epsilon_i)\ln t_i)]$ and $\psi'(r) \rightsquigarrow$ first derivation of the digamma function $\frac{d\ln\Gamma(r)}{dr}$ and $d = \frac{1}{N(\psi'(1)-1)} \rightsquigarrow$ to implement the test, LM_{HD} is evaluated at the null $\rightsquigarrow$ test stat $\sim \chi_2^2$, a Chi-Squared with two degrees of freedom.

$$LM_{HD} = \frac{1}{4N}\cdot\big(\textstyle\sum_i(\epsilon_i^2-2\epsilon_i)\big)^2 \sim \chi_1^2, \text{ asymptotically under } H_0;$$
$$LM_D = \frac{1}{d}\cdot\big(\textstyle\sum_i(1+(1-\epsilon_i)\ln t_i)\big)^2 \sim \chi_1^2, \text{ asymptotically under } H_0,$$

inferring the direction of the misspecification on the basis of a separate test can be misleading. Jaggia and Trivedi, '94; difficult to discern between unobserved heterogeneity and state dependence.

Weibull, Weibull Gamma or proportional hazards model $\rightsquigarrow$ carry out test using integrated hazards functions because in absence of heterogeneity, Bera and Yoon, '93 $\rightsquigarrow$ more general issue of hypotheses' testing when a model is misspecified.

GENERALIZED RESIDUALS

$\rightsquigarrow$ non parametric graphical test of fit of duration model.

$$S(t|u) = \exp\{-\Lambda(t|\mu)\},$$
$$f(t|\mu) = \lambda(t\mu)\exp\{-\Lambda(t|\mu)\}.$$

Generalized residual $\rightsquigarrow \epsilon = \Lambda(t|\mu) = -\ln(S(t|\mu))$. Jacobian of this transformation is:

[25] The two components of the two tests are correlated.

$$J = dt/d\epsilon = \frac{1}{d\Lambda(t|\mu)/dt} = \frac{1}{d\Lambda(t|\mu)}$$

given $f(t|\mu)$, density of ϵ is given

$$\lambda(t|\mu)\exp\{-\epsilon\}\frac{1}{\lambda(t|\mu)} = \exp\{-\epsilon\},$$

which doesn't depend on $\mu \rightsquigarrow$ density is the unit exponential distribution.

DIAGNOSTIC TESTS BASED ON INTEGRATED HAZARD

These tests exploit the unit exponential property of the generalized residual ϵ under the null of correct specification. $S(\epsilon) = \exp\{-\epsilon\}$, $-\ln S(\epsilon) = \Lambda(\epsilon) = \epsilon$.

Estimated integrated hazard for the Weibull model is $\hat{\epsilon} = \hat{\mu}\hat{t}$. $\hat{S}(\hat{\epsilon}) = N^{-1} \rightsquigarrow \#$ of sample observations $\hat{\epsilon}$. Regress $-\ln S(\hat{\epsilon})$ on $\hat{\epsilon}$ and intercept and test whether intercept is zero and the slope is equal to 1. Generalized error model for the Weibull - gamma mixture, exponential- gamma mixture, by setting $\alpha = 1$. $\epsilon = k\ln(k + \mu t^{\alpha})/k \rightsquigarrow$ compute $\hat{\epsilon}$ given estimates (μ, α, k) and then plot $\hat{\epsilon}$ against $-\ln S(\hat{\epsilon})$.

CENSORED DATA

Observed duration is $t = \min[T, L]$ where L denotes the right censoring at $L \rightsquigarrow \epsilon(t)$ is not unit exponentially distributed.

$$\begin{aligned} \mathbb{E}[\epsilon(T)|T \geq L] &= \int_{\epsilon(L)}^{\infty} \frac{\epsilon f(\epsilon)}{S(\epsilon(L))} d\epsilon \\ &= \frac{1}{e^{-\epsilon(L)}} \cdot \left[\int_{\epsilon(L)}^{\infty} \epsilon e^{\epsilon} d\epsilon\right] \\ \frac{1}{e^{-\epsilon(L)}} &\cdot \left[1 + \epsilon(L) \cdot e^{-\epsilon(L)} + e^{-\epsilon(t)} - 1\right] \\ &= 1 + \epsilon(L), \text{ upon integration by parts.} \end{aligned}$$

$\tilde{\epsilon}(t) = \hat{\epsilon}(t)$ if data are not censored (not our case).

REFERENCES

- Cameron and Trivedi, 2005, Microeconometrics: Methods and Applications, Cambridge University Press...;
- Han and Hausman, 1990, Flexible Parametric Estimation of Duration and Competing Risk Models, Journal of Applied Econometrics ...;
- Heckman and Singer, 1994, The Identifiability of Proportional Hazards Models, Review of Economic Studies ...;