1 Duration models, Hazard rate models, or Event history analysis

What determines the length of different states?

Unemployment spells

Strikes

Lives

Time on different income security programs

Why not use OLS?

Distributional assumptions

Time-varying covariates

Censoring

Meyer, B., K. Viscusi and D. Durbin "Worker's compensation and Injury Duration: Evidence from a Natural experimen" AER 1995. Example where OLS can be used on log duration data.

1.1 Key concepts

Right censored: date of spell termination not observed.

Left censored: date of spell beginning not observed.

$f(t)-$ unconditional probability of an event.

Example in Kiefer: the probability that your favorite team loses in round t.

$\lambda(t)$ - hazard rate. Probability of an event conditional on survival until time t.

Example in Kiefer: Probability that your favorite team loses in round t conditional on survival to that round.

If the hazard is constant, the probability of losing in round 6 is $f(6) = \lambda(1-\lambda)^5$.

Cumulative distribution function: $F(t) = P(T < t)$.

Survival function: $S(t) = 1 - F(t)$.

Hazard function: $\lambda(t) = f(t)/(1 - F(t)) = f(t)/S(t)$.

$\lambda(t) = \frac{d\ln(S(t))}{dt}$

Integrated hazard: $\Lambda(t) = \int_0^t \lambda(s)\,ds = -\ln S(t)$.

Duration dependence

Positive duration dependence: $\frac{d\lambda(t)}{dt} > 0$ at point t^*. Probability that a spell ends increases with time.

Negative duration dependence: $\frac{d\lambda(t)}{dt} > 0$ at point t^*. Discourage worker effect. Scars of unemployment or depreciation of human capital.

Dynamic selection: Most employable workers disappears first from the unemployment pool.

Flow sample: sample of individuals who enter a state and started their spell in a specific time interval, e.g. between August 15 and September 1.

Stock sample: sample of individuals that were in a particular state at some moment of time, e.g., September 1.

1.2 Non-parametric inference

1.2.1 Kaplan-Meier

h_j - number of completed spells of duration t_j.

m_j - number of censored observations between t_j and t_{j+1}.

n_j - number at risk at time t_j.

$$n_j = \sum_{i \geqslant j}^{K} (m_i + h_i)$$

Neither completed nor censored at time t_j.

$$\begin{aligned} \widehat{\lambda}(t_j) &= h_j/n_j \\ \widehat{S}(t_j) &= \prod_{i=1}^{J} (n_i - h_i)/n_i = \prod_{i=1}^{J} (1 - \widehat{\lambda}_i). \end{aligned}$$

1.2.2 Log-rank test

r groups

k different failure times

n_j - number at risk at time t_j.

d_j - number of failures at time t_j.

For each time t_j it is possible to compute the expected number of completed spells in each group under H_0 that all groups are equal:

$$\begin{aligned} E_{ij} &= n_{ij} d_j / n_j \\ \mathbf{u}' &= \sum_{j=1}^{k} d_{1j} - E_{1j}, d_{2j} - E_{2j}, ..., d_{rj} - E_{rj} \end{aligned}$$

Variance:

$$V_{il} = \sum_{j=1}^{k} \frac{W^2\left(t_j\right) n_{ij} d_j \left(n_j - d_j\right)}{n_j \left(n_j - 1\right)} \left(\delta_{il} - \frac{n_{ij}}{n_j}\right)$$

Test statistics:

$$\mathbf{u}'\mathbf{V}^{-1}\mathbf{u} \sim \boldsymbol{\chi}^2(r-1)$$

Wilcoxon test: more weight on early observations, when more are at risk.

1.3 Parametric models

1.3.1 Exponential Distribution

1.3.2

Memory-less - constant hazard function. No duration dependence.

$$
\begin{aligned}
F(t) &= 1 - \exp(-\gamma t) \\
S(t) &= \exp(-\gamma t) \\
f(t) &= \gamma \exp(-\gamma t) \\
\lambda(t) &= \gamma \\
\Lambda(t) &= \gamma t \\
E(t) &= 1/\gamma \\
var(T) &= 1/\gamma
\end{aligned}
$$

1.3.3 Weibull

Generalization of the exponential distribution. Exponential if $\alpha = 1$.

$$\begin{aligned}
\lambda(t) &= \gamma\alpha t^{\alpha-1} \\
F(t) &= 1 - \exp(-\gamma t^{\alpha}) \\
S(t) &= \exp(-\gamma t^{\alpha}) \\
\Lambda(t) &= \gamma t^{\alpha}
\end{aligned}$$

If $\alpha > 1 \rightarrow$ increasing hazard.

If $\alpha < 1 \rightarrow$ decreasing hazard.

1.3.4 Log-logistic distribution

1.3.5

Non-monotonic hazard.

$$\begin{aligned}
\lambda(t) &= \gamma\alpha t^{\alpha-1}/(1+t^{\alpha}\gamma), \text{ where } \gamma > 0 \text{ and } \alpha > 0. \\
F(t) &= 1 - [1/(1+t^{\alpha}\gamma)] \\
S(t) &= 1/(1+t^{\alpha}\gamma)
\end{aligned}$$

If $0 < \alpha \leq 1 \rightarrow$ decreasing hazard.

1.4 Estimation of parametric models

Log-likelihood function when we consider the possibility of censoring, i.e., we also consider the contribution of the likelihood that the observation survived to at least t_k :

$$\ln L^*(\theta) = \sum_{i=1}^{n} d_i \ln f\left(t_i, \theta\right) + \left(1 - d_i\right) \ln S\left(t_i, \theta\right),$$

where $d_i = 1$ if the observation is not censored.

Using $f(t, \theta) = \lambda(t, \theta) S(t, \theta)$ and $\ln S(t, \theta) = -\Lambda(t, \theta)$, this can be rewritten as

$$\ln L^* = \sum_{i=1}^{n} d_i \ln \lambda\left(t_i, \theta\right) + \sum_{i=1}^{n} \Lambda\left(t_i, \theta\right).$$

For the exponential distribution we have ($\lambda(t) = \gamma$ and $\Lambda(t, \gamma) = \gamma t$)

$$\ln L(\gamma) = \sum_{i=1}^{n} d_i \ln \gamma - \gamma \sum_{i=1}^{n} t_i.$$

$$\partial L / \partial \gamma = \gamma^{-1} \sum_{i=1}^{n} d_i - \sum_{i=1}^{n} t_i.$$

Gives a closed form solution

$$\widehat{\gamma} = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i},$$

with the variance-covariance

$$V\left(\widehat{\gamma}\right) = -\left[\partial^2 L / \partial \gamma^2\right]^{-1} = \gamma^2 / \sum_{i=1}^{n} d_i.$$

Example in Kiefer: Strike data. $\widehat{\gamma} = 0.0236$. Almost 2.5 % daily chance of a settlement. Expected duration: $1/\gamma = 1/0.0236 \approx 42$ days.

In general, first order conditions does not lead to a closed form solution. Numerical maximization is needed.

1.5 The proportional hazard model

$$\lambda(t, x, \beta, \lambda_0) = \phi(x, \beta) \lambda_0(t)$$

$\lambda_0(t)$ - baseline hazard.

$\phi(x, \beta)$ - depends on the covariates.

Exponential function: $\phi(x, \beta) = \exp(x'\beta)$

Constant proportional effect from a change in x :

$$\partial \ln \lambda(\cdot) / \partial x_k = \partial \ln \phi(\cdot) / \partial x_k = x'\beta / \partial x_k = \beta_k$$

Transformations of the proportional hazard model shows that

$$-\ln \Lambda_0(t) = t^* = x'\beta + \varepsilon,$$

where the distribution of ε is not normal.

If we use the exponential distribution, we have: -$\ln t = t^* = x'\beta + \varepsilon$.

1.6 Estimation of the proportional hazard model

Previously obtained log-likelihood function:

$$\ln L^* = \sum_{i=1}^{n} d_i \ln \lambda\left(t_i, x, \beta\right) + \sum_{i=1}^{n} \Lambda\left(t_i, x, \beta\right).$$

We use the exponential distribution with and λ_0 normalized to 1, i.e., $\lambda\left(t_i, x, \beta\right) = \exp(x'\beta) \cdot 1$.

$$\begin{aligned}
\log L &= \sum d_i x_i'\beta - \sum t_i \exp(x'\beta) \\
\partial \ln L / \partial \beta_k &= \sum d_i x_k - \sum t_i \exp(x'\beta) x_{ik} \\
\partial^2 \ln L / \partial \beta_j \partial \beta_k &= -\sum t_i \exp(x'\beta) x_{ij} x_{ik}
\end{aligned}$$

Strike data. $\beta_0 + \beta_1 x$, where x is industrial production. $\widehat{\beta}_1 = 10.21$.

1.7 Accelerated Lifetimes

Alternative specification that re-scales time:

$$\lambda\left(t_i, x, \beta\right) = \lambda_0(t\phi\left(x, \beta\right))\phi\left(x, \beta\right).$$

If $\phi\left(x, \beta\right) = \exp(x'\beta)$ we can rewrite this model to

$$-\ln t = x'\beta + \upsilon,$$

where υ has a special form.

In general:

ALM: Restricts transformation of the duration model, but more robust to error term distributions.

PH: Restricts the error term, but is more flexible to transformations.

1.8 The Cox Proportional Hazard Model

Semi-parametric model. Partial likelihood.

Completed durations are ordered: $t_1 < t_2 < ..... < t_n$.The probability of completing spell 1 at $t_1$conditional on that the other spells are still on is

$$\frac{\lambda\left(t_1, x_1, \beta\right)}{\sum_{i=1}^{n} \lambda\left(t_1, x_i, \beta\right)}$$

For the second observation, we have

$$\frac{\lambda\left(t_2, x_2, \beta\right)}{\sum_{i=2}^{n} \lambda\left(t_2, x_i, \beta\right)},$$

and so on. Until the last observation.

If we use the proportional hazard specification, we get

$$\frac{\phi(x_1,\beta)}{\sum_{i=1}^{n} \phi(x_i,\beta)}, \frac{\phi(x_2,\beta)}{\sum_{i=2}^{n} \phi(x_i,\beta)} \ldots\ldots$$

This means that the contribution to the likelihood from each observation consists of the ratio of the hazard of the just completed spells and the sum of the hazards that are still on. The likelihood is formed by the product of each contribution, which gives the following log-likelihood

$$\log L(\beta) = \sum_{i=1}^{n} \{\ln \phi\left(x_i, \beta\right) - \ln[\sum_{j=1}^{n} \phi\left(x_j, \beta\right)]\}.$$

The idea is that we do not need any information on the baseline hazard, only the ordering is important.

Table 1: Discrete-time Cox proportional hazard model estimates (*Est.*) and standard errors (*s.e.*) of the effect of the sickness insurance reform on incidence of work absence.

	Males				Females			
	Est.	*s.e.*	*Est.*	*s.e.*	*Est.*	*s.e.*	*Est.*	*s.e.*
I^R	-0.316	0.005	-0.310	0.007	-0.211	0.005	-0.240	0.007
Unemployment	-		-0.114	0.010	-		-0.077	0.003
Unemployment^2	-		0.022	0.002	-		0.018	0.001
County factor	No		Yes		No		Yes	
Log likelihood	-9668.6		-9646.6		-9929.2		-9892.9	
$\chi^2(25)$; $p-value$	44.0; 0.01				72.6; <0.001			

Note: χ^2 statistics and $p-value$ for likelihood ratio test of joint significance of local labor market unemployment rate and county factor.

1.9 Time-varying Explanatory variables

Example: Johansson and Palme (2005) Journal of Public Economics.

March 1, 1991 reform of the sickpay insurance used as variation in replacement levels.

Pre-reform: 90% replacement level from day 1.

Post reform: 65% day 1-3; 80% day 4-90; 90% day 91-

In work spells

$$\lambda_1(t) = \lambda_0(t)\exp(\delta I^R).$$

In sickpay spells

$$\lambda_1(t) = \lambda_0(t)\exp(I^R(1-3)\beta_1 + I^R(4-7)\beta_2 + I^R(8-90)\beta_3 + I^R(90-)\beta_4)$$

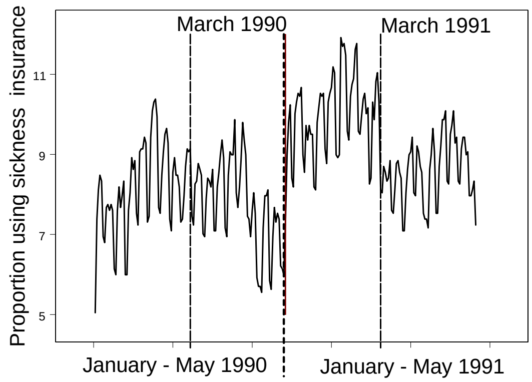

Figure 1: Daily work absence rate January 1 to April 30, 1990 and January 1 to April 30, 1991. Reform date in March 1, 1991 is marked. Males and Females.

Table 2: Discrete-time Cox proportional hazard regression estimates ($Est.$) and standard errors ($s.e.$) of the effect of the reform on the duration in work hazard (hazard of ending a work absence spell).

	Males				Females			
	$Est.$	$s.e.$	$Est.$	$s.e.$	$Est.$	$s.e.$	$Est.$	$s.e.$
$I^R(1-3)$	0.062	0.029	0.099	0.033	0.068	0.028	0.077	0.032
$I^R(4-7)$	0.015	0.300	0.040	0.045	0.050	0.031	0.074	0.046
$I^R(8-90)$	-0.022	0.021	0.006	0.029	-0.101	0.020	-0.095	0.027
$I^R(91-)$	-0.127	0.030	-0.100	0.036	-0.639	0.027	-0.591	0.035
Unemployment			0.150	0.028			-0.013	0.027
Unemployment^2			-0.032	0.004			0.017	0.042
County factor	No		Yes		No		Yes	
Log likelihood	-4362.5		-4350.8		-4256.7		-4241.9	
$\chi^2(25)$; $p-value$	23.32; 0.57				29.6; 0.27			

Note: The baseline hazard is specified as piecewise constant. Indicators for 1-3 days, 4-7 days and 8-90 days in a spell are also included in the specification. χ^2 statistics and $p-value$ for likelihood ratio test for joint significance of local labor market unemployment rate and county factors.

1.10 Functional form misspecification and heterogeneity

Two sources of heterogeneity

Example: Fraction p has $\lambda_1(t) = \gamma_1$ and a fraction $(1-p)$ has $\lambda_2(t) = \gamma_2$.Densities $f_1(t) = \gamma_1 \exp(-\gamma_1 t)$ and $f_2(t) = \gamma_2 \exp(-\gamma_2 t)$.

Suppose we are sampling from this *mixture* distribution. Then we have

$$f(t) = pf_1(t) + (1-p)f_2(t),$$

and we do not know from which group that we are sampling.

Indicator: $x = 0$ if group 1 and $x = 1$ if group 2.

$$\lambda(t, x) = \gamma_1 + x_i(\gamma_2 - \gamma_1)$$

or in exponential form:

$$\lambda(t, x) = \exp(\beta_0 + \beta_1 x),$$

where $\beta_0 = \ln \gamma_1$ and $\beta_1 = \ln \gamma_2 - \ln \gamma_1$.

Hazard for the mixture distribution:

$$\lambda(t) = \frac{p\gamma_1 \exp(-\gamma_1 t) + (1-p)\gamma_2 \exp(-\gamma_2 t)}{p \exp(-\gamma_1 t) + (1-p) \exp(-\gamma_2 t)}$$

If we fit this model to data we will get a negative duration dependence.

Intuition: Initial proportion p and $(1-p)$.If group 2 has a lower hazard, more individuals from group 2 will remain. Will look like duration dependence.

Effect on estimated coefficients

True: $\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$

We estimate: $\exp(\beta_0 + \beta_1 x)$

$$bias = \left(\sum x_{1i} x'_{1i}\right)^{-1} \left(\sum x_{1i} x'_{2i}\right) \beta_2$$

Similar to the linear model, but will only hold locally and is non-linear.

1.10.1 Estimation in the presence of heterogeneity

Two different strategies: either assume a number of different groups with different hazards or a continuous distribution.

Mixed proportional hazard models.

$$S(t \mid \mathbf{x}) = E_\upsilon \left[(St \mid \mathbf{x}, \boldsymbol{\upsilon}) \right] = \int \exp\left(-\upsilon \Lambda_0(t) \phi(\mathbf{x})\right) g(\upsilon)\, d\upsilon.$$

The model is said to be nonparametrically identifiable if, given the data, $\lambda_0, \phi(\cdot)$ and $g(\cdot)$ are uniquely, nonparametrically since no functional form is assumed for $\phi(\cdot)$.

There are several distribution alternatives for $g(\cdot)$.The most common in empirical work is the gamma distribution.

Latent class models. Simplest possible example: two groups with different hazards. We know that p has a higher hazard and $(1-p)$ a lower, i.e.,

$$pf(t_i, \theta_1) + (1-p) f(t_i, \theta_2).$$

Suppose we know that there are two groups.

$$L(\theta_1, \theta_2, p) = \prod \left(p^d f(t_i, \theta_1) + (1-p)^d f(t_i, \theta_2) \right)$$

If d was observable, it would be straight forward to estimate this using maximum likelihood. However, the usual case is that we don know d.Can be treated as "missing data".

EM algorithm:

1. Conditional on the estimates of the parameter vector (including p), calculate the expectation of the likelihood.

2. Maximize and obtain the parameter vector of θ_1, θ_2 and p.

If there is a unknown number of groups: estimate the model with an increasing number of groups. Use AIC or BIC as information criteria for choosing the optimal number of groups.

1.11 Informal methods of specification checking

Generalized residuals:

$$\epsilon = \Lambda(t, \theta) = -\ln S(t, \theta).$$

The density of the generalized residuals is given by the unit exponential, i.e.,

$$\lambda(t, \theta) \exp(-\epsilon) \frac{1}{\lambda(t, \theta)} = \exp(-\epsilon).$$

Exploit the unit exponential distribution of the generalized residuals under H_0.

$$S(\epsilon) = \exp(-\epsilon).$$

For a Weibull model,

$$\widehat{\epsilon} = \widehat{\mu} t^{\widehat{\alpha}}.$$

Given estimates of the parameters in the model, $\widehat{\epsilon}$ can be computed and compared to $-\ln \widehat{S}(\widehat{\epsilon})$ from its survival function:

$$\widehat{S}(\widehat{\epsilon}) = (\text{Number of observations } \geq \widehat{\epsilon})/N.$$

These estimates can be plotted in a diagram. Under H_0 the observations should coincide with the 45° line.

Specification check: regress the $-\ln \widehat{S}(\widehat{\epsilon})$ on the generalized residuals $\widehat{\epsilon}$.

1.12 Competing risk models

In some applications it could be different exit routes.

- Mortality: different causes of death.
- Unemployment: employment (half time/full time) or different forms of labor market programs.

Joint survival function:

$$\Pr\left(\tau > t\right) = \Pr(t_1 > t, ...t_m > t) = S_\tau\left(t\right).$$

If the probabilities are independent, the probability of terminating the spell due to cause j is

$$\Pr\left[r \mid \mathbf{x}_r, \boldsymbol{\beta}_r\right] = \frac{\lambda_r(t, \mathbf{x}_r, \boldsymbol{\beta}_r)}{\sum_{j=1}^{m} \lambda_j(t, \mathbf{x}_j, \boldsymbol{\beta}_j)} = \lambda_r(t, \mathbf{x}_r, \boldsymbol{\beta}_r) S(t, \mathbf{x}_j, \boldsymbol{\beta}_j).$$

Same as analysing the causes separately. Extension of the model where you allow for correlation between exit routes.