Survival Analysis / Modèles de Durée

Chapitre 2 : Hazards Model

Gilles de Truchis

Master 2 ESA

Plan du chapitre

1. Partial Likelihood Estimation
2. Covariates
3. Model Diagnostics
4. Time Dependent Covariates

Plan

1. Partial Likelihood Estimation
2. Covariates
3. Model Diagnostics
4. Time Dependent Covariates

Non parametric models

- As discussed in Chapter 1, Lehman-type alternatives are defined as

$$H_1 : S_1(t) = \big(S_0(t)\big)^{\psi}$$

where $\psi \neq 1$ unless under

$$H_0 : S_1(t) = \big(S_0(t)\big)^{1}$$

⇒ theses hypotheses can be formulated in terms of proportional hazards

$$h_1(t) = \psi h_0(t)$$

- The latter Eq. is the key to quantify the difference between two hazard functions by means of the so-called proportional hazards model
- We can extend the model to include covariate information x as follows

$$\psi = e^{x\beta}$$

- Other functional are possible albeit this is the most common in practice

Note The estimation is complicated in absence of parametric form for

$$h_0(t),$$

and require the concept of partial likelihood developed by Cox

Introduction to the partial likelihood

- Let j denotes the j'th failure time (sorted from lowest to highest)
- Let $h_i(t_j)$ be the hazard function for subject i at failure time t_j

⇒ The Cox proportional hazards (semi-parametric) model is

$$h_i(t_j) = \psi_i h_0(t_j), \quad \psi_i = e^{x_i'\beta}$$

Note ψ_i characterize the hazard ratio $h_i(t_j)/h_0(t_j)$

- In the simplest case where we compare two groups (dummy variable)

$$x_i = \{0, 1\}$$

- In the particular case of control vs treatment group we expect

$$\beta < 0$$

as the experimental group is less likely than control patients to fail

⇒ Hence, $\psi_i < 1$ ($\psi_i = 1$) is expected in the treatment (control) group

The partial likelihood

- Consider the first failure time t_1 and let

$$R_1$$

be the set of all subjects at risk for failure at this time (the risk set)

- The probability that the subject i fails is its hazard divided $\sum h_k(t_1)$

$$\mathbb{P}_1 = \frac{h_i(t_1)}{\sum_{k \in R_1} h_k(t_1)} = \frac{\psi_i h_0(t_1)}{\sum_{k \in R_1} \psi_k h_0(t_1)} = \frac{\psi_i}{\sum_{k \in R_1} \psi_k}$$

where $h_0(t_1)$ is the hazard for a subject from the control group

- At failure time t_2 a new (smaller) risk set R_2 is considered

⇒ We repeat this calculation to obtain p_2 and so on up to t_n

- The partial likelihood is the product

$$\mathcal{L}(\psi) = \mathbb{P}_1 \mathbb{P}_2 \dots \mathbb{P}_n$$

Example of partial likelihood computation

- Consider the following (artificial) data (see also Chapter 1)

Table – Survival data

Patient	Survtime	Censor	Group
1	6	1	$C(x_1 = 0)$
2	7	0	$C(x_2 = 0)$
3	10	1	$T(x_3 = 1)$
4	15	1	$C(x_4 = 0)$
5	19	0	$T(x_5 = 1)$
6	25	1	$T(x_6 = 1)$

- Consider the following (artificial) data (see also Chapter 1)

⇒ the first failure time is at $t = 6$ and for each patient we have either

$$\psi_1 = \psi_2 = \psi_4 = 1 \text{ or } \psi_3 = \psi_5 = \psi_6 = \psi$$

i.e. we have 6 patients at risk (3 in the “C” group for which $\psi = 1$) and

$$\mathbb{P}_1 = \frac{\psi_1 h_0(t_1)}{3\psi h_0(t_1) + 3h_0(t_1)} = \frac{1}{3 \times \psi + 3}$$

Example of partial likelihood computation

- The second failure time is at $t = 10$ because at $t = 7$ there is no failure

Note At $t = 7$ we have a “C” patient that dropped out due to censoring

⇒ Of the 6 patients at risk at the first time, only 4 remains in R_2 and

$$\mathbb{P}_2 = \frac{\psi}{3\psi + 1}$$

where ψ appears in the numerator as the patient 3 was in the “T” group

- The third failure time (t_3) is at $t = 15$ with 3 patients in R_3 and

$$\mathbb{P}_3 = \frac{1}{2\psi + 1}$$

- The last failure time (t_4) is at $t = 25$ with 1 patient in R_4 and

$$\mathbb{P}_4 = \frac{\psi}{\psi} = 1$$

as she is in the “T” group

Example of partial likelihood computation

- Now we are ready to compute the partial likelihood

$$\mathcal{L}(\psi) = \mathbb{P}_1\mathbb{P}_2\mathbb{P}_3\mathbb{P}_4 = \frac{\psi}{(3\psi+3)(3\psi+1)(2\psi+1)}$$

- In the case of a Cox model the log partial likelihood is

$$\ell(\beta) = \beta - \log(3\exp(\beta)+3) - \log(3\exp(\beta)+1) - \log(2\exp(\beta)+1)$$

as ψ is assumed to be of exponential form : $\psi = e^{\beta}$

⇒ The maximum partial likelihood estimate is

$$\widehat{\beta}$$

the value of β that maximizes this function

Note 1 As discussed above, it is nonparametric because the hazard function

$$h_0(t)$$

does not enter the partial likelihood and hence requires no specification

Note 2 Unlike traditional likelihood, $\mathcal{L}(\psi)$ is not a probability but allows to estimate β

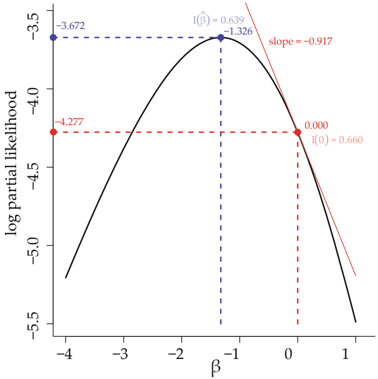
Example of partial likelihood computation

- $\widehat{\beta} = -1.3261$ is obtain by numerical optimization
- We anticipate on the next slide and report some test statistics

Note 1 The null hypothesis ($\beta = 0$) is reported for comparison

Note 2 The slope of the tangent is given by the LM statistic $S(\beta) = \ell'(\beta)$

Note 3 $I(\beta) = -S'(\beta) = -\ell''(\beta)$ denotes the fisher information



Partial likelihood hypothesis tests

- As in standard likelihood one can derive 3 types of test for $H_0 : \beta = 0$
  - The Wald test
  - The LM test
  - The LR test
- The limit theory of theses tests can differ and is often more difficult to derive
- In view of presenting them, define
  - $S(\beta) = \ell'(\beta)$, the score function
  - $I(\beta) = -S'(\beta) = -\ell''(\beta)$, the fisher information
  - $I(\widehat{\beta})$, the observed information

The Wald test

- The Wald test is of form

$$Z_W = \frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}}$$

where $\sigma^2_{\widehat{\beta}}$ is obtained numerically from the negative inverse of the Hessian

$$I(\widehat{\beta})^{-1} = -\ell''(\widehat{\beta})^{-1}$$

Note As the second derivative reflects the curvature of the likelihood, a sharper curve (i.e. more information) leads to lower variance

- Under the null hypothesis $H_0 : \beta = 0$, this normalized statistic if Gaussian

⇒ We reject H_0 if $|Z_W| > z_{\alpha/2}$ or $Z_W^2 > \chi^2_{\alpha,1}$

- The asymptotic normality can be used to construct confidence intervals

$$\widehat{\beta} \pm z_{\alpha/2} \times \sigma_{\widehat{\beta}}$$

The Lagrange Multiplier (score) test

- The LM test is based on the score of the partial log-likelihood

⇒ The variance of this test is hence directly $I(\beta)$

- The test is computed under the null hypothesis as follows

$$Z_{LM} = \frac{S(\beta = 0)}{\sqrt{I(\beta = 0)}}$$

⇒ We reject $H_0 : \beta = 0$ if $|Z_{LM}| > z_{\alpha/2}$ or $Z_W^2 > \chi^2_{\alpha,1}$

Note 1 This test can be computed without finding the MPLE

Note 2 This test is equivalent to the log-rank test statistic U_0 discussed in Chapter 1

⇒ With the same artificial data of Table 1, U_0 was equal to $0.917 \equiv -S(0)$

The Likelihood Ratio test

- The LR test is based on the asymptotic behavior of

$$Z_{LR} = 2\big(\ell(\beta = \widehat{\beta}) - \ell(\beta = 0)\big) \sim \chi^2_1$$

- Z_{LR} is invariant to monotonic transformations of β (unlike the LM and Wald tests)

$\Rightarrow$ Whether the test is computed in terms of β or $\psi = \exp(\beta)$ has no effect on the p-value

$\Rightarrow$ We reject H_0 if $Z^2_{LR} > \chi^2_{\alpha,1}$

Exercise : computation of partial likelihood hypothesis tests

- Consider the MPLE results plotted on S10

⇒ All elements needed to compute Z_W, Z_{LM}, Z_{LR} are there

Exercise : computation of partial likelihood hypothesis tests

- Consider the MPLE results plotted on S10

⇒ All elements needed to compute Z_W, Z_{LM}, Z_{LR} are there

- For Z_{LM} we have

$$Z_{LM}^2 = \left(\frac{S(\beta=0)}{\sqrt{I(\beta=0)}}\right)^2 = \frac{(-0.917)^2}{0.660} = 1.274$$

Any software can compute the p-value which is $p = 0.2591$

- For Z_W we have

$$Z_W^2 = \left(\frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}}\right)^2 = \left(\frac{-1.326129}{\sqrt{1/0.639}}\right)^2 = 1.124$$

Any software can compute the p-value which is $p = 0.2891$

- Finally, for Z_{LR} we have

$$Z_{LR} = 2\big(\ell(\beta=\widehat{\beta}) - \ell(\beta=0)\big) = 2(-3.672 + 4.277) = 1.209$$

Any software can compute the p-value which is $p = 0.2715$

Pseudo-R^2 statistic

- At this stage one can also use

$$\ell(\beta = \widehat{\beta}) \text{ and } \ell(\beta = 0)$$

  to compute an adaptation of the R^2 statistic to survival analysis

- The R^2_{CS} statistic (Cox and Snell) is defined as follows

$$R^2_{CS} = 1 - \left(\frac{\ell(0)}{\ell(\beta)} \right)^{2/n}$$

⇒ R^2_{CS} reflects the improvement in the fit of the model with the covariate compared to $\beta = 0$

Note R^2_{CS} has a major drawback as it is capped to 0.75 but alternatives are not consensual

The partial likelihood with multiple covariates

- To achieve greater generality we now consider the case where

$$x_i = (x_{i,1}, \cdots, x_{i,p})'$$

is a vector of p dummy covariates for each individual i

- To save place we use ψ_i in place of $\psi_i(x_i, \beta)$, where β is now a vector of p coefficients

- In the particular case of the Cox model, the hazard ratio is $\exp(x_i'\beta)$

- As in S6, before the first failure time, all of the subjects are said to be at risk

⇒ Among them one will fail at time t_1 in the risk set R_1

- More generally, at time t_j, the risk set is R_j leading to

$$\mathcal{L}(\beta) = \prod_{j=1}^{D} \frac{h_i(t_j)}{\sum_{k \in R_j} h_k(t_j)} = \prod_{j=1}^{D} \frac{\psi_j h_0(t_j)}{\sum_{k \in R_j} \psi_k h_0(t_j)} = \prod_{j=1}^{D} \frac{\psi_j}{\sum_{k \in R_j} \psi_k}$$

for the Cox proportional hazard model, with D the number of failures

The log partial likelihood with multiple covariates

- The log partial likelihood is simply given by

$$\ell(\beta) = \sum_{j=1}^{D} \left(\log(\psi_j - \log\left(\sum_{k\in R_j} \psi_k\right)\right) = \sum_{j=1}^{D} x_j'\beta - \sum_{j=1}^{D} \log\left(\sum_{k\in R_j} \exp(x_k'\beta)\right)$$

- The score function has p components, one for each of the p covariates

⇒ For the l'th component the score is given by

$$S_l(\beta) = \frac{\partial \ell(\beta)}{\partial \beta_l} = \sum_{j=1}^{D} \left(x_{jl} - \frac{\sum_{k\in R_j} x_{jk} \exp(x_j'\beta)}{\sum_{k\in R_j} \exp(x_j'\beta)} \right)$$

Note We may view the score function as the sum of “residuals”

⇒ The observed value x_{jl} of the covariate l minus an “expected” value

Recall When x_j is a single binary covariate, $S(\beta = 0)$ is the log-rank statistic

Note The Fisher information matrix is now a matrix

$$I(\beta; x) = -\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} = -\frac{S(\beta)}{\partial \beta}$$

Wald, LR and LM tests with multiple covariates

- In presence of multiple covariates the usual tests are as follows
- The Wald test under $H_0 : \beta = 0$ is

$$Z_W^2 = \widehat{\beta}' I(\widehat{\beta}; x)\widehat{\beta}$$

- The LM test :

$$Z_{LR}^2 = S'(\beta = 0; x) I(\beta = 0; x)^{-1} S(\beta = 0; x)$$

- The LR test :

$$Z_{LM}^2 = 2\big(\ell(\beta = \widehat{\beta}) - \ell(\beta = 0)\big)$$

- Under H_0, all 3 statistics are asymptotically χ^2_{k-1}

Exercise with multiple covariates

- Consider the exponential survival data simulated in Chapter 1

$\Rightarrow$ A confounding binary genotype factor was introduced :

$$g = 1 \text{ (wild type) or } g = 2 \text{ (mutant type)}$$

- When estimating the Cox model to compare trivially the “T” and “C” group we obtain

$$\widehat{\beta} = 0.464(\sigma_{\widehat{\beta}} = 0.117) \text{ with } LR = 15.5(p = 0.00000)$$

$\Rightarrow$ How to interpret those results ?

Exercise with multiple covariates

- Consider the exponential survival data simulated in Chapter 1

⇒ A confounding binary genotype factor was introduced :

$$g = 1 \text{ (wild type) or } g = 2 \text{ (mutant type)}$$

- When estimating the Cox model to compare trivially the "T" and "C" group we obtain

$$\widehat{\beta} = 0.464(\sigma_{\widehat{\beta}} = 0.117) \text{ with } LR = 15.5(p = 0.00000)$$

⇒ How to interpret those results ?

Note 1 It suggests higher hazards for the "T" group ($\widehat{\beta} > 0$) with a significant difference with the "C" group

Note 2 Also, $\exp(\widehat{\beta}) = 1.59$ indicates that the "T" group is associated with a 59% additional risk of death over the "C" group

Exercise with multiple covariates

- As for the log-rank test, it is possible to stratified the data
- When estimating the stratified Cox model to compare the "T" and "C" group we obtain

$$\widehat{\beta} = -0.453(\sigma_{\widehat{\beta}} = 0.164) \text{ with } LR = 7.66(p = 0.00566)$$

⇒ How to interpret those results ?

Exercise with multiple covariates

- As for the log-rank test, it is possible to stratified the data
- When estimating the stratified Cox model to compare the “T” and “C” group we obtain

$$\widehat{\beta} = -0.453(\sigma_{\widehat{\beta}} = 0.164) \text{ with } LR = 7.66(p = 0.00566)$$

⇒ How to interpret those results ?

Note 1 It suggests higher hazards for the “C” ($\widehat{\beta} < 0$) group with a significant difference with the “T” group

Note 2 Also, $\exp(\widehat{\beta}) = 0.636$ indicates that the “T” group is associated with

$$1 - 0.636 = 36\%$$

less risk of death over the “C” group

Exercise with multiple covariates

- Finally, we introduce the genotype as a covariate
- When estimating the Cox model with the two covariates we obtain

$$\widehat{\beta}_{grp} = -0.453(\sigma_{\widehat{\beta}_{grp}} = 0.163)$$

and

$$\widehat{\beta}_{gen} = -1.568(\sigma_{\widehat{\beta}_{gen}} = 0.183)$$

with

$$LR = 93.4(p = 0.00000)$$

⇒ How to interpret those results ?

Exercise with multiple covariates

- Finally, we introduce the genotype as a covariate
- When estimating the Cox model with the two covariates we obtain

$$\widehat{\beta}_{grp} = -0.453(\sigma_{\widehat{\beta}_{grp}} = 0.163)$$

and

$$\widehat{\beta}_{gen} = -1.568(\sigma_{\widehat{\beta}_{gen}} = 0.183)$$

with

$$LR = 93.4(p = 0.00000)$$

⇒ How to interpret those results ?

Note 1 As for the stratified Cox model, the correct treatment effect is identified

Note 2 Indeed, we see higher hazards for the "C" ($\widehat{\beta} < 0$) group with a significant difference with the "T" group

Tied survival times

- Tied survival time are failure that occurs simultaneously

Note 1 In continuous time data this is likely to arise due to rounding

Note 2 In discrete time data this can genuinely appear

Note 3 If censoring times are tied with failure times, the convention is to consider the failures to precede the censoring

Example Consider a continuous time process and the following reports

Table – Survival data with tied survival times

Patient	Survtime	Censor	Group
1	1	1	T
2	1	1	T
3	2	1	C
4	3	0	T
5	4	1	T
6	4	1	C
7	5	0	C
8	6	1	C
9	6	0	C
10	7	0	C

Tied survival times and partial likelihood

- As the underlying times are actually continuous we use the Cox model

$$h(t;x) = e^{x\beta}h_0(t)$$

where $x = 1$ or 0 for the treatment or control group, respectively

- As in the regular case, the likelihood is the product of probabilities

$\mathbb{P}_1$ At $t = 1$, all 10 patients are at risk and two of them fail, both from the "T" group, and either of those two patients may have failed first

$\Rightarrow$ We account for those two possibilities when constructing $\mathbb{P}_1$

$$\mathbb{P}_1 = \frac{\exp(\beta)}{4\exp(\beta)+6}\frac{\exp(\beta)}{3\exp(\beta)+6}+\frac{\exp(\beta)}{4\exp(\beta)+6}\frac{\exp(\beta)}{3\exp(\beta)+6} = A\times B+C\times D$$

- The first (second) product assumes that patient 1 (2) fails first

Note 1 In B, 4 becomes 3 as patient 1 has failed

Note 2 In D, 4 becomes 3 as patient 2 has failed

Note 2 As both patients are in the "T" group the $A\times B$ and $C\times D$ are symmetric

Exercise : tied survival times and partial likelihood

- We want to derived the remaining terms of the partial likelihood

Exercise : tied survival times and partial likelihood

- We want to derived the remaining terms of the partial likelihood

$\mathbb{P}_2$ At $t = 2$, 8 patients are at risk (2 and 6 in the "T" and "C" group resp.)

$\Rightarrow$ As there is only 1 failure in the "C" group we have

$$\mathbb{P}_2 = \frac{1}{2\exp(\beta) + 6}$$

Exercise : tied survival times and partial likelihood

- We want to derived the remaining terms of the partial likelihood

$\mathbb{P}_2$ At $t = 2$, 8 patients are at risk (2 and 6 in the “T” and “C” group resp.)

⇒ As there is only 1 failure in the “C” group we have

$$\mathbb{P}_2 = \frac{1}{2\exp(\beta) + 6}$$

$\mathbb{P}_3$ At $t = 4$, 6 patients are at risk (as at $t = 3$ patient 4 is censored)

⇒ We have two failures, one in each group, and

$$\mathbb{P}_3 = \frac{1}{\exp(\beta) + 5} \times \frac{\exp(\beta)}{\exp(\beta) + 4} + \frac{\exp(\beta)}{\exp(\beta) + 5} \times \frac{1}{5}$$

to account for all scenarios of failure (patient 5 first or patient 6 first)

Exercise : tied survival times and partial likelihood

- We want to derived the remaining terms of the partial likelihood

$\mathbb{P}_2$ At $t = 2$, 8 patients are at risk (2 and 6 in the "T" and "C" group resp.)

⇒ As there is only 1 failure in the "C" group we have

$$\mathbb{P}_2 = \frac{1}{2\exp(\beta) + 6}$$

$\mathbb{P}_3$ At $t = 4$, 6 patients are at risk (as at $t = 3$ patient 4 is censored)

⇒ We have two failures, one in each group, and

$$\mathbb{P}_3 = \frac{1}{\exp(\beta) + 5} \times \frac{\exp(\beta)}{\exp(\beta) + 4} + \frac{\exp(\beta)}{\exp(\beta) + 5} \times \frac{1}{5}$$

to account for all scenarios of failure (patient 5 first or patient 6 first)

- Only 1 constant factor remains as patients 7 and 10 are censored and

$$\mathbb{P}_4 = \frac{1}{3}$$

as at $t = 6$, by convention, the censored patient 9 failed after patient 8

⇒ One may express the partial likelihood as $\mathcal{L}(\beta) = \mathbb{P}_1\mathbb{P}_2\mathbb{P}_3$ or $\mathbb{P}_1\mathbb{P}_2\mathbb{P}_3\mathbb{P}_4$

Discrete tied survival times

- Consider now that times are in fact discrete in the table below

⇒ In such a case, the Cox model is transformed to a discrete logistic model

$$\frac{h(t;x)}{1-h(t;x)} = e^{x\beta}\frac{h_0(t)}{1-h_0(t)}$$

Table – Survival data with tied survival times

Patient	Survtime	Censor	Group
1	1	1	T
2	1	1	T
3	2	1	C
4	3	0	T
5	4	1	T
6	4	1	C
7	5	0	C
8	6	1	C
9	6	0	C
10	7	0	C

Discrete tied survival times and partial likelihood

- At $t = 1$, as 2 patients fail among the 10 patients at risk we now have

$$\binom{10}{2} = \frac{10!}{2!(n-k)!} = 45$$

pairs that could represent the two failures

- All factors are summarized in the matrix below and lead to

$$\mathbb{P}_1 = \frac{e^{2\beta}}{6e^{2\beta} + 24e^{\beta} + 15}$$

Table – Pairs that could represent two failures among 10 patients

	e^{β}	e^{β}	e^{β}	e^{β}	1	1	1	1	1	1
e^{β}	●									
e^{β}	$e^{2\beta}$	●								
e^{β}	$e^{2\beta}$	$e^{2\beta}$	●							
e^{β}	$e^{2\beta}$	$e^{2\beta}$	$e^{2\beta}$	●						
1	e^{β}	e^{β}	e^{β}	e^{β}	●					
1	e^{β}	e^{β}	e^{β}	e^{β}	1	●				
1	e^{β}	e^{β}	e^{β}	e^{β}	1	1	●			
1	e^{β}	e^{β}	e^{β}	e^{β}	1	1	1	●		
1	e^{β}	e^{β}	e^{β}	e^{β}	1	1	1	1	●	
1	e^{β}	e^{β}	e^{β}	e^{β}	1	1	1	1	1	●

Exercise : discrete tied survival times and partial likelihood

- We want to compute the remaining factors

Exercise : discrete tied survival times and partial likelihood

- We want to compute the remaining factors
- At $t = 2$, there is only 1 failure in the "C" group $\Rightarrow \mathbb{P}_2 = 1/(2e^{\beta} + 6)$

Exercise : discrete tied survival times and partial likelihood

- We want to compute the remaining factors
- At $t = 2$, there is only 1 failure in the “C” group $\Rightarrow \mathbb{P}_2 = 1/(2e^\beta + 6)$
- At $t = 4$, there are 2 failures and 6 patients are at risk such that we have

$$\binom{6}{2} = 15$$

possible pairs, of which 1 is from the “T” group and 1 from the “C” group

$$\mathbb{P}_3 = \frac{\exp(\beta) \times 1}{5\exp(\beta) + 10}$$

$\Rightarrow$ Again, one may simply express the partial likelihood as $\mathcal{L}(\beta) = \mathbb{P}_1\mathbb{P}_2\mathbb{P}_3$

Table – Pairs that could represent two failures among 6 patients

	e^β	1	1	1	1	1
e^β	●					
1	e^β	●				
1	e^β	1	●			
1	e^β	1	1	●		
1	e^β	1	1	1	●	
1	e^β	1	1	1	1	●

Approximation in presence of tied survival times

- With many ties, the discrete and continuous methods are cumbersome

⇒ Two approximation methods can be implemented

Breslow It adjusts the denominator to simply reflect all patients at risk

⇒ In the previous example, $\mathbb{P}_1$ and $\mathbb{P}_3$ becomes

$$\mathbb{P}_1 = \frac{2e^{2\beta}}{(6e^{\beta}+4)^2} \text{ and } \mathbb{P}_3 = \frac{2(e^{\beta}\times 1)}{(e^{\beta}+5)^2}$$

Efron It is better as it reflects all patients at risk before and after the failure

⇒ In the previous example, $\mathbb{P}_1$ and $\mathbb{P}_3$ becomes

$$\mathbb{P}_1 = \frac{e^{\beta}}{(6e^{\beta}+4)} \frac{e^{\beta}}{(0.5e^{\beta}+0.5e^{\beta}+4e^{\beta}+4)}$$

and

$$\mathbb{P}_3 = \frac{e^{\beta}}{(e^{\beta}+5)} \frac{1}{(0.5+0.5e^{\beta}+3)}$$

with the weight 0.5 reflecting that each of the 2 patients has a chance of 1/2 of being in the second denominator since 1 of them would have been the first failure

Left truncated data

- Consider the data of Table 1 with left truncation information

e.g. A patient can be diagnosed before entering a trial (i.e. backwards recurrence times is $\neq 0$)

Note 1 The standard way to compare the 2 groups is to ignore “back times”

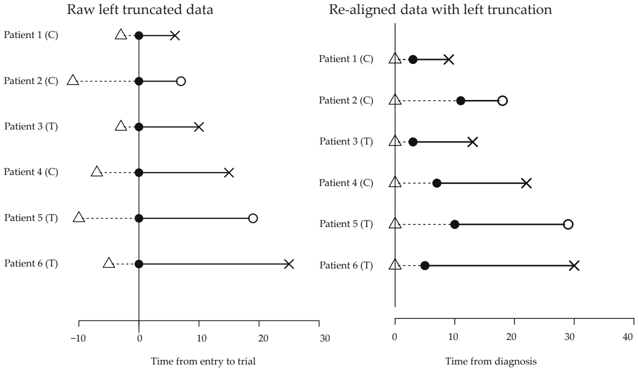
⇒ Nothing wrong (i.e. no bias) in that way to proceed but starting from diagnosis could be of interest

Note 2 To account for backwards recurrence times, one can re-configure the data so that they start at 0

Table – Survival left truncated data

Patient	Survtime	Censor	Group	Back time
1	6	1	C	-3
2	7	0	C	-11
3	10	1	T	-3
4	15	1	C	-7
5	19	0	T	-10
6	25	1	T	-5

Left truncation and re-configured data



- In that case, estimation results are similar for the two data sets

⇒ No statistical difference between "C" and "T" (but n is too small)

  - Raw data :

$$\widehat{\beta} = -1.33(\sigma_{\widehat{\beta}} = 1.25) \text{ with } LR = 1.21(p = 0.271)$$

  - Re-configured data :

$$\widehat{\beta} = -1.07(\sigma_{\widehat{\beta}} = 1.24) \text{ with } LR = 0.81(p = 0.368)$$

Plan

1. Partial Likelihood Estimation
2. Covariates
3. Model Diagnostics
4. Time Dependent Covariates

Categorical and Continuous Covariates

- All covariates considered until now are dummy variables

Note An exception is the confounder “genotype” that is categorical

$$g \in \{1, 2\}$$

but can easily be transformed to $\{0, 1\}$ as it is dichotomous

- More generally one can encode categorical variables with dummies

e.g. If we have a 3-level variable we need : “Ba (x_1), Ma (x_2), no-diploma (x_3)”

⇒ If “Ba” is the reference, then $x_1 = 1$, $x_2 = x_3 = 0$

⇒ An individual without any diploma implies $x_1 = x_2 = 0$ and $x_3 = 1$

- Continuous variables are also frequent and have to be considered

e.g. income, age, etc.

The Cox model with categorical and continuous covariates

- For a set of k covariates (categorical or/and continuous) the model is

$$\log(\psi_i) = x_{1i}\beta_1 + x_{2i}\beta_2 + \ldots + x_{ki}\beta_k = x_i'\beta$$

- For the covariate x_j, β_j is the log hazard ratio for the effect of that parameter on survival, adjusting for the other covariates
- For continuous covariates, it represents the effect of a unit change in the covariate
- For dummy covariates, it represents the effect of the corresponding level as compared to the reference

Note 1 As for logistic regression, a variable can enter non-linearly the model

Note 2 Interaction terms can be introduced

Note 3 At this stage, all covariate are assumed to be fixed in time

Note 4 This model differs from the logistic model as there is no intercept term : if there were one, it would cancel out just as $h_0(t)$ canceled out

Example of Cox model estimation with categorical and continuous covariates

- Consider artificial survival data with two covariates : age and diploma

⇒ individual at risk can loose their job

- Ages are between 40 and 80 at random
- We set the diploma variable so that there are 20 of each 3 categories
- We assume an exponential distribution with parameter as follows
  - We set the log-rate parameter to have baseline -4.5
  - The diploma variable take the values 1 and 2 for “Ba” and “No diploma” when compared to “Ma”
  - We let “age” decrease the log rate by 0.05 per year
- We do not introduce censoring in the data set and $n = 60$

Example of Cox model estimation with categorical and continuous covariates

- When applying the Cox model we obtain the following estimates

$$\widehat{\beta}_{Ba} = 1.151,\ (\sigma_{\widehat{\beta}_{Ba}} = 0.368), \quad z = 3.113\ (p = 0.00173)$$

and

$$\widehat{\beta}_{No} = 2.499,\ (\sigma_{\widehat{\beta}_{No}} = 0.429), \quad z = 5.820\ (p = 0.00000)$$

and

$$\widehat{\beta}_{age} = -0.078,\ (\sigma_{\widehat{\beta}_{age}} = 0.014), \quad z = 5.385\ (p = 0.00000)$$

⇒ Estimates of log hazard ratios are close to the true values (1, 2 and 0.05)

- When looking at exponential coefficient, $\exp(\beta)$, we conclude that
  - Individuals with Bachelor degree have $\exp(\beta_{Ba}) = 3.16$ times the risk of being fired as do subject with Ma degree
  - Individuals without diploma have $\exp(\beta_{No}) = 12.17$ times the risk of being fired as do subject with Ma degree

Note The z statistics is a generalizations of the 2-group comparison Wald tests

Nested models

- When comparing models we have to determine whether that are nested
- Here is an illustration of nested models in terms of covariates
  - Model A : “Age”
  - Model B : “Employment”
  - Model C : “Age” + “Employment”

⇒ Model A is nested in Model C as well as model B

- To test for the presence of nested models we can compute LR tests

Note Models A and B are not nested and requires specific testing procedures

Example of nested models

- Consider the data on therapies to aid smokers to quit (Chapter 1)
- In this study, “Age” and “Employment” have 4 and 3 levels
  - Age : “21-34”, “35-49”, “50-64” and “65+”
  - Employment : “ft” (full-time), “other” and “pt” (part-time)

⇒ By default we choose the first level as the reference level

- Estimation of the Cox model on model A, B and C

	coef	exp(coef)	se(coef)	z	p
LR : 12.2 ($p = 0.006$)			Model A		
age35-49	0.0293	1.030	0.309	0.0947	0.920
age50-64	-0.7914	0.453	0.336	-2.3551	0.019
age65+	-0.3173	0.728	0.444	-0.7153	0.470
LR : 2.06 ($p = 0.357$)			Model B		
other	0.198	1.22	0.237	0.836	0.40
pt	0.450	1.57	0.323	1.394	0.16
LR : 16.8 ($p = 0.005$)			Model C		
age35-49	-0.130	0.878	0.321	-0.404	0.6900
age50-64	-1.024	0.359	0.359	-2.856	0.0043
age65+	-0.782	0.457	0.505	-1.551	0.1200
other	0.526	1.692	0.275	1.913	0.0560
pt	0.500	1.649	0.332	1.508	0.1300

Example of nested models

- From the Wald test (z) for Model C we see that some levels are significant

e.g. The "50-64" age group has a lower hazard when compared to the reference "21-34" with $\widehat{\beta} = -1.024$

e.g. The "other" employment group has higher hazard when compared to the reference "ft" with $\widehat{\beta} = 0.526$

- However, we cannot easily see whether "Age" or "Employment" should be part of the model

⇒ We assess this issue using (partial) likelihood ratio tests based on

$\ell(\widehat{\beta})$ Model A : -380.043, Model B : -385.123, Model C : -377.759

LR : A|C $2(\ell(\widehat{\beta}_C) - \ell(\widehat{\beta}_A) = 4.567$ compare to $\chi^2_{\nu=5-3}$ which leads to $p = 0.1019$

⇒ "Age" is not significant when "Employment" is included in the model

LR : B|C $2(\ell(\widehat{\beta}_C) - \ell(\widehat{\beta}_B) = 14.727$ compare to $\chi^2_{\nu=5-2}$ which leads to $p = 0.0020$

⇒ "Employment" is significant when "Age" is included in the model

Example of nested models

- These results raise the question of including "Age" in model A

$\Rightarrow$ To test this hypothesis we consider the null model N

$$\ell(\widehat{\beta}_N) = -386.153$$

free of any covariate

LR : N|A $2(\ell(\widehat{\beta}_A) - \ell(\widehat{\beta}_N) = 12.220$ compare to $\chi^2_{\nu=3-0}$ which leads to $p = 0.0066$

$\Rightarrow$ "Age" is significant when included in the model N

When a large number of potential factors can enter the model

⇒ The forward stepwise model selection

Step 1 fit univariate models (1 for each covariate) and retain the one with the smallest p-value

Step 2 apply Step 1 again but with the selected covariate included

Step 3 continue until no additional covariate has a p-value less than a pre-defined threshold (e.g. 5%)

⇒ The backward stepwise model selection

Step 1 fit a model with all covariates

Step 2 remove one by one the covariates, each time removing the one with the largest p-value

Step 3 continue the procedure until the p-values are all below a pre-defined threshold (e.g. 5%)

- The stepwise approach can be automatized but has 2 main drawbacks
    - Due to multiple comparisons, the p-values produced from one stage to the next are misleading

Note Corrections like the one of Bonferroni exist

- - Also, p-values are only valid for nested models and hence this approach is not recommended for non-nested models

Non-nested models and criterion based selection

- Information criteria apply to partial log likelihood
- We discuss some examples based on the so-called AIC

$$AIC = -2\ell(\widehat{\beta}) + 2k$$

  where k is the number of parameters in the model

- One can view the AIC as balancing two quantities
  - The goodness of fit $-2\ell(\widehat{\beta})$ (smaller for models that fit the data well)
  - The complexity measure that enter the criterion as a penalty term $2k$
- Applying the AIC to the previous model selection issue we obtain

$\ell(\widehat{\beta})$ Model A : 766.086, Model B : 774.246, Model C : 765.519

$\Rightarrow$ The model C is the one that minimizes the AIC and offers the best fit

Note The BIC (or SIC) also applies to survival analysis

$$BIC = -2\ell(\widehat{\beta}) + k\log(n)$$

and as it penalizes by a factor of $\log(n)$, it will tend to select models with fewer parameters as compared to AIC

Information criterion and the stepwise approach

- We can implement the backward stepwise procedure with the AIC
- Let consider additional covariates for the smokers therapies
  - "yearsSmoking"+"levelSmoking"+"priorAttempts"+"longestNoSmoke"
  + "gender"+ "morphotype"+ "age"+ "employment"

Note 1 (+) & (-) show the effect on AIC of adding or removing the covariate

Note 2 Covariates are listed in order from the one which, when removed, yields the greatest AIC reduction to the smallest reduction

Information criterion and the stepwise approach

- When starting the procedure, all covariates are there (AIC = 770.2)
  - ⇒ "(-) morpho" is at the top of the list and will be removed first
- Intermediate results are unreported but proceed in the same way
- At final step (AIC = 758.42) and all per-covariate are above 758.42
  - ⇒ The sign (-) remains for employment & age and reveal that removing them would be detrimental
  - ⇒ At the opposite, variables for which a "(+)" appears indicate that adding would deteriorate the fit of the model

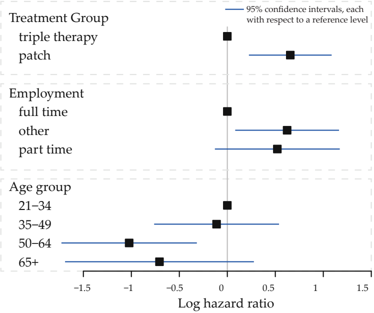
Sign	Covariate	Level	AIC	Sign	Covariate	Level	AIC
Step 1			770.2	Final Step			758.42
-	morpho	3	766.98		<none>		758.42
-	years	1	768.20	+	longest	1	759.10
-	gender	1	768.20	-	employment	2	760.31
-	prior	1	768.24	+	years	1	760.34
-	level	1	768.47	+	gender	1	760.39
-	longest	1	769.04	+	prior	1	760.40
	none		770.20	+	level	1	760.41
-	employment	2	772.45	+	morpho	3	761.53
-	age	3	774.11	-	age	3	767.24

Forest plot

Final model	coef	exp(coef)	se(coef)	z	p
grppatchOnly	0.656	1.928	0.220	2.986	0.0028
employmentother	0.623	1.865	0.276	2.254	0.0240
employmentpt	0.521	1.684	0.332	1.570	0.1200
ageGroup435-49	-0.112	0.894	0.322	-0.348	0.7300
ageGroup450-64	-1.023	0.359	0.360	-2.845	0.0044
ageGroup465+	-0.707	0.493	0.502	-1.410	0.1600

- The Forest plot offers an alternative representation :

e.g. 1 triple therapy is better than the patch alone
e.g. 2 subjects with full-time work have a better success rate than others
e.g. 3 the upper age groups have better results than younger patients



Smooth estimates of continuous covariates

- For continuous covariates, the relationship with the log-hazard can be

... linear, quadratic, or of any other nonlinear nature

e.g. in the previous study, the age has been split into 4 groups and

... the forest plot reveals different effects and hence nonlinearities

⇒ An alternative way to capture this nonlinearity is via pieces of

... polynomial functions (Splines) that are stitched to form a smooth curve

- The points where these pieces are joined are called “knots”

... and a crucial issue is to determined their locations

⇒ The Splines enter the penalized partial likelihood via a penalty term

$$\mathcal{P}(\beta, \omega) = \ell(\beta, \omega) - g(\omega, \theta)$$

with ω the set of constrained parameters and θ some tuning parameters

Penalized Cox model and Spline fit

- Splines with many knots increase the complexity of the likelihood

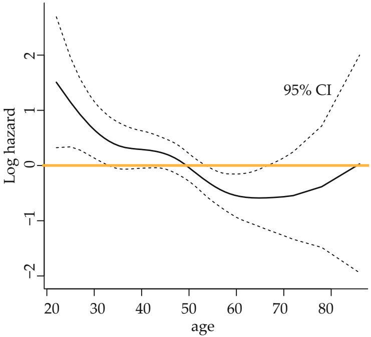
... but also improve the goodness of fit

⇒ $\mathcal{P}(\beta, \omega)$, when maximized, balances goodness of fit against complexity

e.g. When plotting the penalized spline fit from the Cox model we observe

  - a decreasing relationship with age with a slight upward turn after age 65
  - but for most of the part, the effect seems not significant

Figure – Splines



Penalized Cox model and Spline fit

- The penalized Cox model estimation results are reported below

	coef	exp(coef)	se(coef)	χ^2	ν	p
grppatchOnly	0.651	0.221	0.219	8.67	1.00	0.0032
employmentother	0.633	0.277	0.275	5.21	1.00	0.0220
employmentpt	0.570	0.340	0.333	2.81	1.00	0.0940
pspline(age,linear)	-0.034	0.010	0.010	11.07	1.00	0.0009
pspline(age,nonlinear)				4.20	3.08	0.2500

- For the 3 first factors the coefficient are stable as compared to S45
- The Splines are decomposed in two parts : linear and nonlinear
  - the linear one is highly significant
  - the nonlinear one is not significant (probably because the data set is sparse)

Plan

1. Partial Likelihood Estimation
2. Covariates
3. Model Diagnostics
4. Time Dependent Covariates

Martingale residuals

- Assessing goodness of fit using residuals also applies to survival analysis

- Residual analysis essentially relies on graphical analysis

⇒ Typically, residuals are plotted versus some quantity

- To construct the residuals sequence, we compare the censoring indicator

$$\delta_i$$

to the expected value of the indicator under the Cox model

⇒ In absence of time dependent covariates and for right-censored data

$$\widehat{m}_i = \delta_i - \widehat{H}_0(t_i)\exp(x_i'\widehat{\beta})$$

- These Martingale residuals range in value from $-\infty$ to 1 and $\mathbb{E}(\widehat{m}_i) = 0$

- However these residuals can be asymmetric and hence cannot be used as a measure of goodness of fit

Deviance residuals

- An alternative is the so-called deviance residual defined as

$$\widehat{d}_i = \text{sign}(\widehat{m}_i)\Big(-2\big(\widehat{m}_i + \delta_i \log(\delta_i - \widehat{m}_i)\big)\Big)^{1/2}$$

- d_i residuals are symmetrically distributed with $\mathbb{E}(\widehat{d}_i)$

Note 1 The sum of squares of $\widehat{d}_i$ is the value of the partial likelihood ratio test

- While their properties might seem preferable to those of $\widehat{m}_i$, only $\widehat{m}_i$ have the property of showing us the functional form of a covariate

⇒ In practice, the martingale residuals are more useful

Note 2 Other types of residuals will be discussed later

Example : Martingale versus deviance residuals

- Consider again the Cox model for smoking therapies data
- As discussed earlier, the null model (N) is the one without covariates

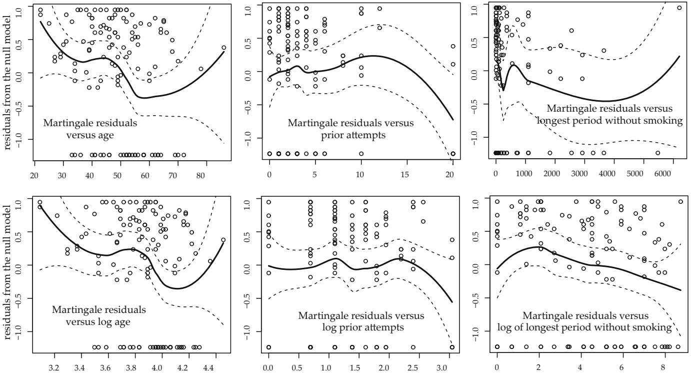
⇒ We may plot $\widehat{m}_i$ against continuous covariates to get a preliminary assessment of which of them should be in the model

Note 1 We also include the log of covariates and use a LOESS curve to identify patterns

Note 2 LOESS (LOcally Estimated Scatterplot Smoothing) is a nonparametric regression based on the nearest neighbor method

Note 3 The 95% confidence intervals for the LOESS curve are also reported

Example : Martingale versus deviance residuals



- For the raw covariates we observe strong non-linearities

  e.g. For "age", we find something similar to Figure 1 (Spline fit)

  ⇒ This null model residual based approach is an alternative way to identify nonlinearity

- For the log-transformed covariates we observe less non-linearities

  e.g. For "LongestNoSmoke", the log seems sufficient to remove the non-linearity

Example : Martingale versus deviance residuals

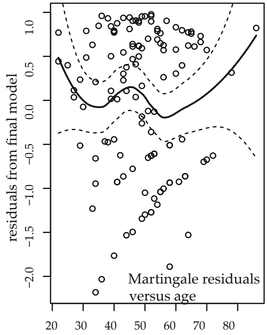
- We apply the stepwise approach with the log of “LongestNoSmoke”

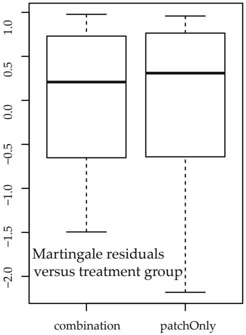
⇒ The results are unchanged (only “age” and “employment” are retained)

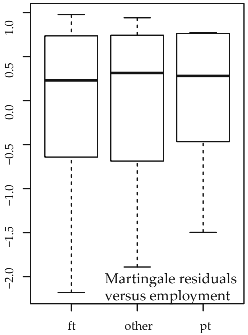
- We compute the final model residuals and obtain the following plots

⇒ Some non-linearity remains for “age” albeit less than for the null model

- The residual distributions of both “group” and “employ” are reasonably comparable, indicating that these variables are modeled successfully







Jackknife residuals

- Some subject may have a huge influence on the parameter estimates

⇒ As this may indicate a problem with the data

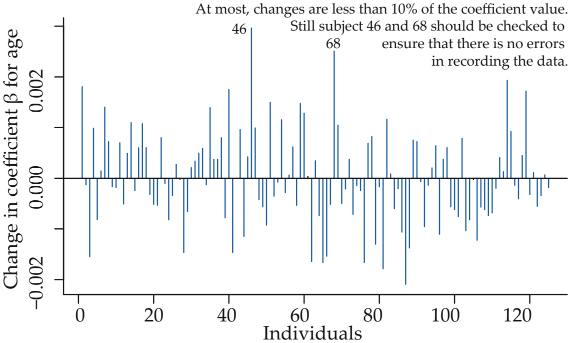
… we need tools that can identify those individuals

- The Jackknife residuals are computed as the difference in the value of

$$\widehat{\beta}$$

when all data are used and when an individual is deleted from the data

⇒ Then, we can plot the change in coefficients for each subject



Log cumulative hazard plots

- When comparing survival times between two groups

… the proportional hazards assumption is of importance

$$S_1(t) = \big(S_0(t)\big)^{\exp(\beta)}$$

with $\exp(\beta)$ the proportional hazards constant

⇒ This is the foundation of Lehman alternatives and the Cox model

- The log-transformation gives

$$\log(S_1(t)) = \exp(\beta)\log(S_0(t))$$

with all logs being negative as survival functions are less than 1

- $g(u) = \log(-\log(u))$ changes the range of u from $(0, 1)$ to $(-\infty, \infty)$

⇒ The so-called log cumulative hazard plot, that is a plot of

$$g(S_1(t)) \text{ and } g(S_0(t)) \text{ versus } \log(t)$$

should lead to parallel curves separated by β if the assumption is correct

Example of log cumulative hazard plots

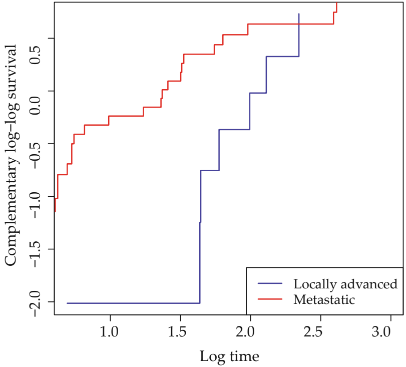
- Consider the pancreatic cancer data (see in Chapter 1)

Recall We performed the Prentice-modification test and found a stronger group difference than did the log-rank test

⇒ As this test places higher weight on earlier survival times it suggests non-proportional hazards

- This is confirmed by the log cumulative hazard plot

Note However, statistical inference is unavailable and this approach is limited



Schoenfeld residuals

- Schoenfeld residuals can assess the proportional hazards assumption more rigorously

- To compute them, let start from the partial log-likelihood

$$\ell(\beta) = \sum_{i\in D}\left(\log(\psi_i) - \log\Big(\sum_{k\in R_i}\psi_k\Big)\right) = \sum_{i\in D}\left(x_i\beta - \log\Big(\sum_{k\in R_i}\exp(x_k\beta)\Big)\right)$$

and its derivative (the score function)

$$\ell(\beta)' = \sum_{i\in D}\Big(x_i - \sum_{k\in R_i} x_k p(\beta, x_k)\Big),\;\; p(\beta, x_k) = \exp(x_k\beta)\Big(\sum_{j\in R_k}\exp(x_j\beta)\Big)^{-1}$$

where the second term can be viewed as the weighted expected value $\mathbb{E}(X_i) = \bar{x}(t_i)$

- The Schoenfeld residuals are the individual terms of the score

$$\hat{r}_i = x_i - \sum_{k\in R_i} x_k p(\beta, x_k) = x_i - \bar{x}(t_i)$$

- A plot of $\hat{r}_i$ versus x_i will yield a pattern of points

⇒ They are centered on 0 if the proportional hazards assumption is correct

Example of Schoenfeld residuals

- Consider the artificial data of S1 ($\hat{\beta} = -1.32$) and compute the weights

t_i	n_{0i}	n_{1i}	$p(\beta, x_k = 0)$	$p(\beta, x_k = 1)$	Grp
6	3	3	$1/(3 + 3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(3 + 3e^{\hat{\beta}})$	C
10	1	3	$1/(1 + 3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1 + 3e^{\hat{\beta}})$	T
15	1	2	$1/(1 + 2e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1 + 2e^{\hat{\beta}})$	C
25	0	1	$1/e^{\hat{\beta}}$	$e^{\hat{\beta}}/e^{\hat{\beta}} = 1$	T

- It remains to compute $\mathbb{E}(X_i)$ and $\hat{r}_i$ which for $t_i = 6$ gives

Example of Schoenfeld residuals

- Consider the artificial data of S1 ($\hat{\beta} = -1.32$) and compute the weights

t_i	n_{0i}	n_{1i}	$p(\beta, x_k = 0)$	$p(\beta, x_k = 1)$	Grp
6	3	3	$1/(3+3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(3+3e^{\hat{\beta}})$	C
10	1	3	$1/(1+3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1+3e^{\hat{\beta}})$	T
15	1	2	$1/(1+2e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1+2e^{\hat{\beta}})$	C
25	0	1	$1/e^{\hat{\beta}}$	$e^{\hat{\beta}}/e^{\hat{\beta}} = 1$	T

- It remains to compute $\mathbb{E}(X_i)$ and $\hat{r}_i$ which for $t_i = 6$ gives

$$\mathbb{E}(X_i) = 3\times 0\times 1/(3+3e^{\hat{\beta}}) + 3\times 1\times e^{\hat{\beta}}/(3+3e^{\hat{\beta}}) = 0.2098 \Rightarrow \hat{r}_i = 0-0.2098$$

Example of Schoenfeld residuals

- Consider the artificial data of S1 ($\hat{\beta} = -1.32$) and compute the weights

t_i	n_{0i}	n_{1i}	$p(\beta, x_k = 0)$	$p(\beta, x_k = 1)$	Grp
6	3	3	$1/(3+3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(3+3e^{\hat{\beta}})$	C
10	1	3	$1/(1+3e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1+3e^{\hat{\beta}})$	T
15	1	2	$1/(1+2e^{\hat{\beta}})$	$e^{\hat{\beta}}/(1+2e^{\hat{\beta}})$	C
25	0	1	$1/e^{\hat{\beta}}$	$e^{\hat{\beta}}/e^{\hat{\beta}} = 1$	T

- It remains to compute $\mathbb{E}(X_i)$ and $\hat{r}_i$ which for $t_i = 6$ gives

$$\mathbb{E}(X_i) = 3\times 0\times 1/(3+3e^{\hat{\beta}})+3\times 1\times e^{\hat{\beta}}/(3+3e^{\hat{\beta}}) = 0.2098 \Rightarrow \hat{r}_i = 0-0.2098$$

- For $t_i = 10 : \mathbb{E}(X_i) = 1\times 0\times 1/(1+3e^{\hat{\beta}})+3\times 1\times e^{\hat{\beta}}/(1+3e^{\hat{\beta}}) = 0.4434$

$$\Rightarrow \hat{r}_i = 1 - 0.4434 = 0.5566$$

- For $t_i = 15 : \mathbb{E}(X_i) = 1\times 0\times 1/(1+2e^{\hat{\beta}})+2\times 1\times e^{\hat{\beta}}/(1+2e^{\hat{\beta}}) = 0.3468$

$$\Rightarrow \hat{r}_i = 0 - 0.3468 = -0.3468$$

- For $t_i = 25$ we have $\mathbb{E}(X_i) = 1$

$$\Rightarrow \hat{r}_i = 1 - 1 = 0$$

Grambsch and Therneau residuals

- They propose to scale each residual by an estimate of its variance

$$\widehat{r}_i^* = \widehat{r}_i \times d \times \mathbb{V}(\widehat{\beta})$$

  where d is the total number of death

- Then, Grambsch and Therneau show that if hazards are non proportional

$$\mathbb{E}(r_i^*) \approx \beta + \beta(t)$$

  i.e. a survival time dependent β (unknown) enter the $\mathbb{E}(\hat{r}_i^*)$ whereas

$$\mathbb{E}(r_i^*) = \beta$$

  in presence of proportional hazards

$\Rightarrow$ $\beta(t)$ can be approximated by

$$\widehat{\beta}(t) \approx \widehat{r}_i^* - \widehat{\beta}$$

where $\widehat{\beta}$ is estimated from the Cox model

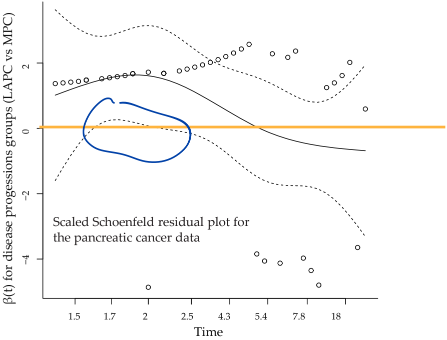
Note Statistical inference is now possible to test $H_0 : \beta(t) = 0$

Example for Grambsch and Therneau residuals

- We compute $\widehat{\beta}(t)$ for the pancreatic cancer data and plot it versus time

Note 1 we also compute the LOESS curve and its 95% confidence intervals

Note 2 the time axis is scaled to match the Kaplan-Meier-transformed time



- The curve reveals a slight increase, followed by a steady decline

Note 3 Zero seems to be almost always in the confidence intervals

Example for Grambsch and Therneau residuals

- A more formal test can be obtained by fitting as straight line to $\widehat{r}_i^*$
- This score-type test statistic, denoted $\widehat{\rho} \sim \chi_1^2$, gives

$$\widehat{\rho} = -0.328, \;\; p = 0.0496$$

⇒ we reject the null of a constant β (i.e. we reject the proportional hazards)

- The way we defined the time axis matters (Kaplan-Meier-transformed time here)

e.g. If we consider time ordered by the ranks survival times we obtain

$$\widehat{\rho} = -0.330, \;\; p = 0.0486$$

⇒ very similar results

e.g. If we consider the untransformed time line we obtain

$$\widehat{\rho} = -0.197, \;\; p = 0.2390$$

⇒ here we cannot reject the null of proportional hazards

Note This latter approach is not to be preferred when the failure times are sparse and not uniformly spaced over time

Plan

1. Partial Likelihood Estimation
2. Covariates
3. Model Diagnostics
4. Time Dependent Covariates

What are time dependent covariates ?

- The partial likelihood theory assumes that covariates are time invariant

⇒ The value of z at $t = 0$ is the same at any $t_i > 0$

- In some cases this assumption is unrealistic

e.g. In credit scoring analysis, the employment status is likely to change

e.g. In job market analysis, the skills are likely to evolve

⇒ Time dependent covariates require special measures to obtain valid parameter estimates

Impact of time dependent covariates

- Unfortunately, we cannot predict survival using future covariate values

- This deceptively principle can ensnare even experienced research

⇒ We illustrate this with the following example :

e.g. consider data on patients enrolled in a transplant program

- Here are the results of the survival study :

	coef	exp(coef)	se(coef)	z-test	p
transplant	-1.71711	0.17958	0.27853	-6.165	7.05e-10
age	0.05889	1.06065	0.01505	3.913	9.12e-05
surgery	-0.41902	0.65769	0.37118	-1.129	0.259

⇒ It seems that heart transplanted patients live longer than others

- The covariate “transplant” equals 1 for transplanted patients

⇒ The issue is that “transplant” is time dependent as patients in a transplant program have to live long enough to be transplanted

⇒ It only shows that patients who live long enough to receive a transplant have longer lives than patients who do not live as long (tautology)

Landmark time

- In that particular case, a simple fix is to define a landmark time τ
- It divide patients into two groups : intervention and comparison groups

Intervention those who received the intervention prior to τ

Comparison those who did not received the intervention prior to τ

- If only patients who survive up to τ are included

and all patients remain in their assigned group, this method is valid

Note Hence, patients transplanted after τ remain in the comparison group

$\Rightarrow$ the comparison group could be renamed “no transplant within τ days”

Example of landmark time

- If we set $\tau = 30$ days, 79 of the 103 patients lived this long
- Of these 79 patients, 33 had a transplant before τ and 46 did not
- Of these 46 patients, 30 subsequently had a transplant

Note we still count them in the comparison group

⇒ we have hence created a new variable "transplant30" which has a fixed value for all patients in the set of 30-day survivors

- Here are the valid results of survival study :

	coef	exp(coef)	se(coef)	z-test	p
transplant30	-0.04214	0.95874	0.28377	-0.148	0.8820
age	0.03720	1.03790	0.01714	2.170	0.0300
surgery	-0.81966	0.44058	0.41297	-1.985	0.0472

- The "transplant" covariate is no longer significant

Note However, one could discuss the choice of the landmark τ

Beyond the landmark approach

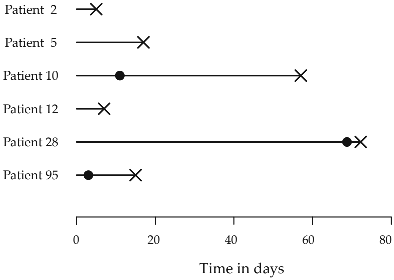
- Unfortunately there is no clear way to select the landmark τ

$\Rightarrow$ we prefer another approach based on adjustments of the Cox model

- Let consider a subset of 6 patients to illustrate this approach
- 3 of them received a transplant and 3 of them did not

id	wait.time	futime	fustat	transplant
2	–	5	1	0
5	–	17	1	0
10	11	57	1	1
12	–	7	1	0
28	70	71	1	1
95	1	15	1	1

futime : following-up (failure) time
fustat : 0 if censored, 1 otherwise
● and waiting time : time of transplant



Modified partial likelihood

- We first model incorrectly the data

	coef	exp(coef)	se(coef)	z-test	p
transplant	-1.6878	0.1849	1.1718	-1.44	0.150

- To correct the model we allow the contributions of each subject to change from one failure time to the next

⇒ The hazard function is now given by

$$h(t) = h_0(t)e^{x_k(t_i)\beta}$$

with $x_k(t_i)$ the time-varying covariate for the kth subject at time t_i

- This leads to the modified partial likelihood

$$\mathcal{L}(\beta) = \prod_{i=1}^{D} \psi_{ii} \Big(\sum_{k \in R_i} \psi_{ki} \Big)^{-1}$$

with $\psi_{ki} = e^{x_k(t_i)\beta}$

- In the fixed-time case we were able, as time passes, to successively delete ψ_i for subject that failed at that time
- We here have to recalculate the entire denominator at each failure time

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

$\mathcal{L}_2(\beta)$ P12 fails at $t = 7$, 5 being at risk, still 1 patient being transplanted

$$\mathcal{L}_2(\beta) = \frac{1}{4 + e^{\beta}}$$

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

$\mathcal{L}_2(\beta)$ P12 fails at $t = 7$, 5 being at risk, still 1 patient being transplanted

$$\mathcal{L}_2(\beta) = \frac{1}{4 + e^{\beta}}$$

$\mathcal{L}_3(\beta)$ P95 fails at $t = 15$, 4 being at risk, but the P10 “transplant” status has switched to 1

$$\mathcal{L}_3(\beta) = \frac{e^{\beta}}{2 + 2e^{\beta}}$$

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

$\mathcal{L}_2(\beta)$ P12 fails at $t = 7$, 5 being at risk, still 1 patient being transplanted

$$\mathcal{L}_2(\beta) = \frac{1}{4 + e^{\beta}}$$

$\mathcal{L}_3(\beta)$ P95 fails at $t = 15$, 4 being at risk, but the P10 “transplant” status has switched to 1

$$\mathcal{L}_3(\beta) = \frac{e^{\beta}}{2 + 2e^{\beta}}$$

$\mathcal{L}_4(\beta)$ P5 fails at $t = 17$, 3 being at risk, still 2 patients being transplanted

$$\mathcal{L}_4(\beta) = \frac{1}{2 + e^{\beta}}$$

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

$\mathcal{L}_2(\beta)$ P12 fails at $t = 7$, 5 being at risk, still 1 patient being transplanted

$$\mathcal{L}_2(\beta) = \frac{1}{4 + e^{\beta}}$$

$\mathcal{L}_3(\beta)$ P95 fails at $t = 15$, 4 being at risk, but the P10 “transplant” status has switched to 1

$$\mathcal{L}_3(\beta) = \frac{e^{\beta}}{2 + 2e^{\beta}}$$

$\mathcal{L}_4(\beta)$ P5 fails at $t = 17$, 3 being at risk, still 2 patients being transplanted

$$\mathcal{L}_4(\beta) = \frac{1}{2 + e^{\beta}}$$

$\mathcal{L}_5(\beta)$ P10 fails at $t = 57$, 2 being at risk, still 2 patients being transplanted

$$\mathcal{L}_5(\beta) = \frac{e^{\beta}}{1 + e^{\beta}}$$

Example of modified partial likelihood computation

- Let compute $\mathcal{L}(\beta)$ for the six patients (labeled 2, 5, 10, 12, 28, 95)

$\mathcal{L}_1(\beta)$ P2 fails at $t = 5$, all 6 being at risk, the P95 being the only 1 transplanted

$$\mathcal{L}_1(\beta) = \frac{1}{5 + e^{\beta}}$$

$\mathcal{L}_2(\beta)$ P12 fails at $t = 7$, 5 being at risk, still 1 patient being transplanted

$$\mathcal{L}_2(\beta) = \frac{1}{4 + e^{\beta}}$$

$\mathcal{L}_3(\beta)$ P95 fails at $t = 15$, 4 being at risk, but the P10 “transplant” status has switched to 1

$$\mathcal{L}_3(\beta) = \frac{e^{\beta}}{2 + 2e^{\beta}}$$

$\mathcal{L}_4(\beta)$ P5 fails at $t = 17$, 3 being at risk, still 2 patients being transplanted

$$\mathcal{L}_4(\beta) = \frac{1}{2 + e^{\beta}}$$

$\mathcal{L}_5(\beta)$ P10 fails at $t = 57$, 2 being at risk, still 2 patients being transplanted

$$\mathcal{L}_5(\beta) = \frac{e^{\beta}}{1 + e^{\beta}}$$

$\mathcal{L}_6(\beta)$ P28 is the last to fail ($t = 71$), just after having been transplanted

$$\mathcal{L}_6(\beta) = \frac{e^{\beta}}{e^{\beta}} = 1$$

Example of modified partial likelihood computation

- Overall, the modified partial likelihood is

$$\mathcal{L}(\beta) = \frac{1}{5+e^\beta} \times \frac{1}{4+e^\beta} \times \frac{e^\beta}{2+2e^\beta} \times \frac{1}{2+e^\beta} \times \frac{e^\beta}{1+e^\beta} \times 1$$

- On the numerical side, $\mathcal{L}(\beta)$ is based on the start-stop format
  - It divides the time data for patients with a time-varying covariate

e.g. As P10 was a non-transplant patient until day 11, its future as a non-transplant patient is unknown

⇒ we censor that portion of the patient's life experience at $t = 11$:

$$\text{start} : t = 0, \ \text{stop} : t = 11$$

⇒ we start a new record of P10 (which is left-truncated at $t = 11$)

$$\text{start} : t = 11, \ \text{stop} : t = 57$$

- For our subset of 6 patient it results in new lines in the database

P#	start	stop	death	transpl
2	0	5	1	0
5	0	17	1	0
10	0	11	0	0
10	11	57	1	1
12	0	7	1	0
28	0	70	0	0
28	70	71	1	1
95	0	1	0	0
95	1	15	1	1

Example of modified partial likelihood computation

- Once the data are in this start-stop format the Cox model applies
- For our subset of 6 patient the conclusions remain unchanged

	coef	exp(coef)	se(coef)	z-test	p
transplant	0.2846	1.3292	0.9609	0.296	0.767

- When considering the whole data set and all covariates we obtain

	coef	exp(coef)	se(coef)	z-test	p
transplant	0.01405	1.01415	0.30822	0.046	0.9636
surgery	-0.77326	0.46150	0.35966	-2.150	0.0316
age	0.03055	1.03103	0.01389	2.199	0.0279

- As with the landmark analysis we confirm that there is no evidence that receiving a heart transplant increases survival

Predictable time dependent variables

- An alternative way to model non-proportional hazard is to allows for

$$\beta = \beta(t)$$

  for a particular covariate

- If there is only one covariate we have

$$h(t) = h_0 e^{x_k \beta(t)}$$

- Characterizing the functional form of $\beta(t)$ is challenging

⇒ A way to proceed is to define a new time dependent variable with fixed coefficients

Note As this variable is defined by the econometrician, it is referred as predictable variable

- The pattern of the Schoenfeld residuals are helpful to identify an appropriate time dependent function

Time transfer function

- Consider again the pancreatic cancer data as in S61
- A simple estimation of the Cox model gives

	coef	exp(coef)	se(coef)	z-test	p
stage of progress	0.593	1.81	0.401	1.48	0.14

Recall the Schoenfeld plot revealed that the hazard ratio might vary

- An alternative way is to define a time dependent covariate as

$$g(t) = \theta_0 + \theta_1 \times \log(t)$$

where θ_0 denotes the usual time-invariant group indicator

⇒ Plugging $g(t)$ in the Cox model, the fitted time transfer function is

$$\beta(t) = 6.01 - 1.09 \log(t)$$

	coef	exp(coef)	se(coef)	z-test	p
l(stage)	6.01	407.339	3.060	1.96	0.050
nl(stage)	-1.09	0.338	0.589	-1.84	0.065

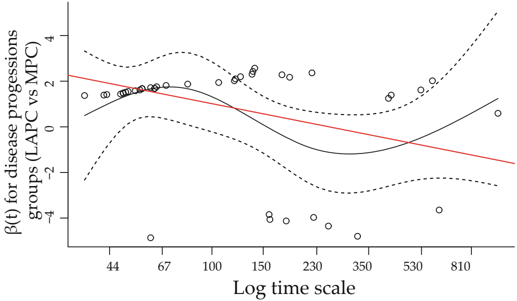
- The LR test that compares the two groups accounting $\beta(t)$ gives

$$LR = 6.33 \ \ (p = 0.0423)$$

⇒ As θ_0 and θ_1 are significant, this suggests that the group indicator combined with a time-varying hazard ratio yields evidence of group difference

Visualization of the time transfer function

- We can use the Schoenfeld residuals plot of S61 to visualize $\theta_1 \times \log(t)$



- The red curve, $-1.09\log(t)$, is linear as the time axis is in log

⇒ It indicates that overall, the log hazard ratio decreases over time

Note The results are dependent of the functional and e.g. no longer old for

$$g(t) = \theta_0 + \theta_1 \times t$$

```
stage.n        1.27810    3.590      0.66103   1.93    0.053
tt(stage.n)    -0.00366   0.996      0.00253   -1.44   0.150
LR test        4.56       p=0.102
```

Variables that linearly increase with time

- A common source of confusion is whether the age variable is time dependent
- Indeed, the age increases with time itself

⇒ the age is definitely a time dependent variable

But it has no effect on the model if one includes it as time varying covariate

- To see why this happens defined the current age of a subject by

$$x(t) = x(0) + t$$

where $x(0)$ denotes the age at entry into the study

⇒ Then, the hazard function is given by

$$h(t) = h_0(t)e^{\beta x(t)} = \big(h_0(t)e^{\beta t}\big)e^{\beta x(0)}$$

such that once we insert $h(t)$ in the partial likelihood,

$$e^{\beta t}$$

appears in both the numerator and the denominator of each factor

⇒ Hence, it cancels out as does the baseline hazard

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. Biometrika, 66(1), 59-65.