The Right-Censored Proportional Hazard Model and its Implementation in EasyReg

Herman J. Bierens

May 10, 2005

1 The right-censored mixed proportional hazard model

1.1 The proportional hazard model

Let T be a duration, and let X be a vector of covariates. As is well-known, the conditional hazard function is defined as

$$\frac{f(t|X)}{1-F(t|X)} = \lambda(t,X),$$

where $F(t|X) = P[T \leq t|X]$, $f(t|X)$ is the corresponding conditional density function, and $\int_0^\infty \lambda(\tau,X)d\tau = \infty$. Then the conditional survival function is

$$S(t|X) = 1 - F(t|X) = \exp\left(-\int_0^t \lambda(\tau,X)d\tau\right).$$

The proportional hazard model assumes that the conditional survival function takes the form

$$\begin{aligned} S(t|X,\alpha,\beta) &= S(t|X) = \exp\left(-\exp(\beta' X)\int_0^t \lambda(\tau|\alpha)d\tau\right) \\ &= \exp\left(-\exp(\beta' X)\,\Lambda(t|\alpha)\right) \end{aligned}$$

where $\lambda(t|\alpha)$ is the baseline hazard function depending on a parameter (vector) α, $\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau$ is the integrated baseline hazard, and $\exp(\beta' X)$ is the systematic hazard function.

1.2 Censoring

Usually the duration T is only observed up to an upper bound $\overline{T}$, which may vary per individual. This is called (right) censoring. Indicate censoring by the dummy variable

$$C = I\left(T > \overline{T}\right),$$

where $I(.)$ is the indicator function, i.e., $I(true) = 1$, $I(false) = 0$. Assuming that $\overline{T}$ is exogeneous and does not depend on the covariates, it follows from (**??**) that

$$P[C = 1|X] = S\left(\overline{T}|X, \alpha, \beta\right) = \exp\left(-\exp(\beta' X)\Lambda\left(\overline{T}|\alpha\right)\right) \tag{1}$$

and

$$\begin{aligned} P[T &\leq t|X, C = 0] = \frac{P[T \leq t, T \leq \overline{T}|X]}{P[C = 0|X]} \\ &= \frac{P[T \leq \min\left(t, \overline{T}\right)|X]}{P[C = 0|X]} = \frac{1 - \exp\left(-\exp(\beta' X)\Lambda\left(\min\left(t, \overline{T}\right)|\alpha\right)\right)}{1 - \exp\left(-\exp(\beta' X)\Lambda\left(\overline{T}|\alpha\right)\right)}. \end{aligned}$$

Taking the derivative to t yields the density $f\left(t|X, C = 0, \alpha, \beta\right)$ of T conditional on the covariates and absence of censoring:

$$\begin{aligned} & f\left(t|C = 0, X, \alpha, \beta\right) \\ = & \frac{\exp\left(-\exp(\beta' X)\Lambda(t|\alpha)\right)}{1 - \exp\left(-\exp(\beta' X)\Lambda\left(\overline{T}|\alpha\right)\right)}\exp(\beta' X)\lambda(t|\alpha) \text{ if } t \leq \overline{T}, \\ = & \ 0 \text{ if } t > \overline{T}. \end{aligned} \tag{2}$$

2 The log-likelihood function

Given i.i.d. observations $\{T_j, D_j, C_j\}_{j=1}^{N}$ on (T, D, C), and assuming that $T_j = \overline{T}_j$ if $C_j = 1$, it follows from (1) and (2) that the log-likelihood function takes the form

$$\begin{aligned} \ln\left(L_N\left(\alpha, \beta, h\right)\right) = -\sum_{j=1}^{N}\exp\left(\beta' X_j\right)\Lambda\left(T_j|\alpha\right) \\ + \sum_{j=1}^{N}\left(1 - C_j\right)\beta' X_j + \sum_{j=1}^{N}\left(1 - C_j\right)\ln\left(\lambda\left(T_j|\alpha\right)\right) \end{aligned}$$

3 Implementation in EasyReg

EasyReg module SURVIVAL1 will ask you to select T, the right-censoring dummy variable C, and the covariates X.

The following options for the baseline and integrated hazard of T are available.

3.1 Weibull hazard

$$\lambda(t|\alpha) = \alpha_1\alpha_2 t^{\alpha_2 - 1},$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'.$$
Integrated hazard:
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1 t^{\alpha_2}.$$

3.2 Generalized Weibull hazard

If in the Weibull case $\alpha_2 < 1$ then $\lambda(0|\alpha) = \infty$, whereas if $\alpha_2 > 1$ then $\lambda(0|\alpha) = 0$. This may be too restrictive. The following generalized Weibull hazard specification satisfies $0 < \lambda(0|\alpha) < \infty$:

$$\lambda(t|\alpha) = \alpha_1\alpha_2 (\alpha_3 + t)^{\alpha_2 - 1},$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha_3 > 0,\ \alpha = (\alpha_1, \alpha_2, \alpha_3)'$$
Integrated hazard:
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1 \left((\alpha_3 + t)^{\alpha_2} - \alpha_3^{\alpha_2}\right).$$

3.3 Unimodal hazard

$$\lambda(t|\alpha) = \frac{2\alpha_1 t}{\alpha_2^2 + t^2},\ \alpha_2 = \arg\max_{t\geq 0} \lambda(t|\alpha),$$
$$\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'.$$
Integrated hazard:
$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1 . \ln\left(\frac{\alpha_2^2 + t^2}{\alpha_2^2}\right).$$

3.4 Generalized unimodal hazard

In the unimodal hazard case, $\lambda(0|\alpha) = 0$. Again, this may be too restrictive. The following generalized unimodal hazard specification allows $\lambda(0|\alpha) > 0$:

$$\lambda(t|\alpha) = \frac{2\alpha_1 (\alpha_3 + t)}{(\alpha_2 + \alpha_3)^2 + (\alpha_3 + t)^2}, \ \alpha_2 = \arg\max_{t \geq 0} \lambda(t|\alpha),$$
$$\alpha_1 > 0, \ \alpha_2 > 0, \ \alpha_3 > 0, \ \alpha = (\alpha_1, \alpha_2, \alpha_3)'.$$

Integrated hazard:

$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1.\ln\left(\frac{(\alpha_2 + \alpha_3)^2 + (\alpha_3 + t)^2}{(\alpha_2 + \alpha_3)^2 + \alpha_3^2}\right).$$

In all four cases the parameter α_1 acts as a scale factor. Therefore, $\ln(\alpha_1)$ acts as a constant term. Consequently, the vector X of covariates should **not** contain a constant.

4 Two-step ML estimation

The parameters are estimated by maximum likelihood in two rounds. First, the log-likelihood is maximized with the α_i's fixed to 1. The resulting maximum likelihood estimator $\widetilde{\beta}$ together with the values $\alpha_i = 1$ are the starting values for the second round, yielding the actual maximum likelihood estimators $\widehat{\beta}$ and $\widehat{\alpha}$.

If you check "Batch mode" these two rounds are conducted automatically, where in each round the iteration is automatically restarted until the log-likelihood does not change anymore. This option is recommended for big jobs.