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

Herman J. Bierens

May 10, 2005

1 The proportional hazard model

This *EasyReg* module (SURVIVAL) estimates a proportional hazard model for a duration T, conditional on a vector of covariates, **without** unobserved heterogeneity.

Given a vector $X \in \mathbb{R}^k$ of covariates, the conditional hazard function is defined as

$$\widetilde{\lambda}(t|X) = \lim_{\delta \downarrow 0} \frac{P[T \in [t, t+\delta) \mid X, T \geq t]}{\delta}.$$

It can be shown that

$$P[T > t|X] = \exp\left(-\int_0^t \widetilde{\lambda}(\tau|X)d\tau\right).$$

In the case of a proportional hazard model it is assumed that $\widetilde{\lambda}(t|X)$ takes the form

$$\widetilde{\lambda}(t|X) = \exp(\beta'X)\,\lambda(t|\alpha),$$

where $\exp(\beta'X)$ is the systematic hazard and $\lambda(t|\alpha)$ is the baseline hazard function depending on a parameter (vector) α. Thus in this case,

$$P[T > t|X] = \exp(-\exp(\beta'X)\,\Lambda(t|\alpha)),$$

where

$$\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau$$

is the integrated hazard. The conditional probability $P[T > t|X]$ is known as the conditional survival function.

The integrated baseline hazard $\Lambda(t|\alpha)$ will be specified such that there is no need to include a constant in X.

2 Interval censoring

It is assumed that the duration T is not observed directly, but only in the form of dummy variables corresponding to intervals: Let $0 \leq b_0 < b_1 < ... < b_M$. Then the dependent variables are M dummy variables

$$D_i = I\left(T \in (b_{i-1}, b_i]\right), i = 1, ..., M,$$

where $I(.)$ is the indicator function: $I(true) = 1$, $I(false) = 0$. The intervals $(b_{i-1}, b_i]$ may be replaced by $[b_{i-1}, b_i)$. However, these intervals should be disjoint. If the duration T is censored from above as well, the value of b_M should be such that for all observations, T is not censored if $T \in (0, b_M]$.

It is recommended to indicate the bracket values in the name of D_i, for example, use names like "Dummy T in (2,4]" or "I(T in (2,4])". *EasyReg* will then extract the lower and upper bounds of the bracket from the name.

Note that for $i = 1, 2, ..., M$,

$$P[D_i = 1|X] = S\left(b_{i-1}|\alpha, \beta' X, h\right) - S\left(b_i|\alpha, \beta' X, h\right)$$

and

$$P\left[\sum_{i=1}^{M} D_i = 0 \middle| X\right] = S\left(b_M|\beta' X, h_1, h_2\right).$$

3 Hazard function options

3.1 Piecewise linear integrated baseline hazard

Because these probabilities depend on the values of $\Lambda(b_i|\alpha)$ in M bracket points b_i only, we may without loss of generality parametrize $\Lambda(t|\alpha)$ as a piecewise linear function:

$$\Lambda(t|\alpha) = \Lambda(b_{i-1}|\alpha) + \alpha_i\left(t - b_{i-1}\right) \quad (1)$$

$$
\begin{aligned}
&= \sum_{k=1}^{i-1} \alpha_k \left(b_k - b_{k-1}\right) + \alpha_i \left(t - b_{i-1}\right) \text{ for } t \in (b_{i-1}, b_i], \\
\alpha_i &> 0 \text{ for } i = 1, ..., M,\ \alpha = (\alpha_1, ..., \alpha_M)' \in \mathbb{R}^M.
\end{aligned}
$$

There are other equivalent ways to specify $\Lambda(b_i|\alpha)$, but the advantage of the specification (1) is that the null hypothesis $\alpha_1 = ... = \alpha_M$ corresponds to the integrated Weibull hazard $\Lambda(t|\alpha) = \alpha_1 t$. In that case $\exp\left(\beta' X\right) \Lambda(t|\alpha) = \exp\left(\ln(\alpha_1) + \beta' X\right) t$, so that $\ln(\alpha_1)$ acts as a constant term in the systematic hazard.

The specification (1) is adopted in *EasyReg* as the default option, and is called the piecewise linear integrated baseline hazard.

Next to this specification, you have four other options for the baseline hazard:

3.2 Weibull baseline hazard

$$
\begin{aligned}
&\lambda(t|\alpha) = \alpha_1 \alpha_2 t^{\alpha_2 - 1}, \\
&\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'. \\
&\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha) d\tau = \alpha_1 t^{\alpha_2}.
\end{aligned} \tag{2}
$$

The reason for the scale factor α_1 here and below is that X does not contain a constant, hence $\ln(\alpha_1)$ plays the role of constant: $\exp(\beta' X)\alpha_1 t^{\alpha_2} = \exp(\ln(\alpha_1) + \beta' X) t^{\alpha_2}$. Note that if $\alpha_2 > 1$ then $\lambda(t|\alpha)$ is monotonic increasing, starting from $\lambda(0|\alpha) = 0$, and if $\alpha_2 < 1$ then $\lambda(t|\alpha)$ is monotonic decreasing, and $\lambda(0|\alpha) = \infty$.

3.3 Generalized Weibull baseline hazard

$$
\begin{aligned}
&\lambda(t|\alpha) = \alpha_1 \alpha_2 \left(\alpha_3 + t\right)^{\alpha_2 - 1}, \\
&\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha_3 > 0,\ \alpha = (\alpha_1, \alpha_2, \alpha_3)' \\
&\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha) d\tau = \alpha_1 \left(\left(\alpha_3 + t\right)^{\alpha_2} - \alpha_3^{\alpha_2}\right).
\end{aligned} \tag{3}
$$

The reason for the extra parameter α_3 is to allow $0 < \lambda(0|\alpha) < \infty$.

3.4 Unimodal baseline hazard

$$
\begin{aligned}
&\lambda(t|\alpha) = \frac{2\alpha_1 t}{\alpha_2^2 + t^2},\ \alpha_2 = \arg\max_{t\geq 0} \lambda(t|\alpha), \qquad (4)\\
&\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha = (\alpha_1, \alpha_2)'.\\
&\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1.\ln\left(\frac{\alpha_2^2 + t^2}{\alpha_2^2}\right).
\end{aligned}
$$

This hazard function is increasing on $(0, \alpha_2)$ and decreasing on (α_2, ∞), with $\lambda(0|\alpha) = \lambda(\infty|\alpha) = 0$.

3.5 Generalized unimodal baseline hazard

$$
\begin{aligned}
&\lambda(t|\alpha) = \frac{2\alpha_1(\alpha_3 + t)}{(\alpha_2 + \alpha_3)^2 + (\alpha_3 + t)^2},\ \alpha_2 = \arg\max_{t\geq 0} \lambda(t|\alpha), \qquad (5)\\
&\alpha_1 > 0,\ \alpha_2 > 0,\ \alpha_3 > 0,\ \alpha = (\alpha_1, \alpha_2, \alpha_3)'.\\
&\Lambda(t|\alpha) = \int_0^t \lambda(\tau|\alpha)d\tau = \alpha_1.\ln\left(\frac{(\alpha_2 + \alpha_3)^2 + (\alpha_3 + t)^2}{(\alpha_2 + \alpha_3)^2 + \alpha_3^2}\right).
\end{aligned}
$$

Again, the reason for the extra parameter α_3 is to allow $\lambda(0|\alpha) > 0$.

4 Two-step ML estimation

EasyReg estimates the parameter vectors α and β in two steps. In first instance the parameters α_i are fixed to $\alpha_i = 1$, except that in the cases (4) and (5) $\alpha_2 = b_1$.

The (quasi-)maximum likelihood estimator $\widetilde{\beta}_0$ of β in the first step will be used as starting values in the second step, together with the initial values of α_i. This step yields the maximum likelihood estimators $\widehat{\alpha}$ of α and $\widehat{\beta}$ of β.

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.