HAZARD REGRESSION

In survival analysis*, proportional hazards models (*see* PROPORTIONAL HAZARDS MODEL, COX'S) are commonly used to estimate covariate effects. Two advantages of this approach are that the interpretation of the results is similar to that for ordinary linear models, and that the effects are estimated regardless of the baseline hazard function.

Inspired by the success of polynomial splines and their tensor* products in adaptive multiple regression (MARS, Friedman [6]), Kooperberg et al. [11] developed a similar adaptive hazard regression (HARE) methodology for estimating the conditional log-hazard function based on possibly censored survival data with one or more covariates. This methodology circumvents the proportionality used in proportional hazards models while still retaining the usual interpretation of the estimated effects. HARE also provides greater flexibility in modeling these effects through the use of polynomial splines (*see* SPLINE FUNCTIONS) and stepwise addition and deletion of basis functions.

Early attempts to use splines in survival analysis are described in Abrahamowicz et al. [1], Anderson and Senthilselvan [2], Efron [4], Etezadi-Amoli and Ciampi [5], Gray [8], Hastie and Tibshirani [9], O'sullivan [13], Senthilselvan [14], Sleeper and Harrington [15], and Whittemore and Keller [17]. Other nonparametric methods, such as kernel estimates, have been used to test for nonproportionality [7]. Intrator and Kooperberg [10] compare the use of trees and splines in survival analysis.

In this entry, hazard regression refers to the HARE methodology [11]. Some authors use "hazard(s) regression" or "proportional hazard(s) regression" to refer to the Cox proportional hazards model [3].

Let T be a (nonnegative) survival time whose distribution may depend on a vector $\mathbf{x} = (x_1, \ldots, x_M)$ of covariates ranging over a subset $\chi = \chi_1 \times \cdots \times \chi_M$ of R^M. Suppose $f(t|\mathbf{x}), F(t|\mathbf{x}) = \int_0^t f(u|\mathbf{x})du, \lambda(t|\mathbf{x}) = f(t|\mathbf{x})/[1 - F(t|\mathbf{x})]$, and $\alpha(t|\mathbf{x}) = \log \lambda(t|\mathbf{x})$ denote the corresponding conditional density, distribution, hazard, and log-hazard functions, respectively. Let G be a p-dimensional linear space of functions on $[0, \infty) \times \chi$, and let $B_1, \ldots, B_p$ be a basis, i.e., a collection of *basis functions*, of G. The HARE model for the log-hazard function is given by

$$\alpha(t|\mathbf{x}; \boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j B_j(t|\mathbf{x}), \quad t \geqslant 0. \qquad (1)$$

The coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ in (1) is estimated by the maximum likelihood* method. Specifically, consider n randomly selected individuals. For $1 \leqslant i \leqslant n$, let T_i be the survival time, C_i the censoring time, and $\mathbf{x}_i$ the vector of covariates for the ith individual; set $Y_i = \min(T_i, C_i)$ and let δ_i be the indicator function of the event $\{T_i \leqslant C_i\}$. The random variable Y_i is *uncensored* or *censored** according as $\delta_i = 1$ or $\delta_i = 0$. The partial likelihood* corresponding to $Y_i = y_i$, δ_i, $\mathbf{x}_i$, and $\boldsymbol{\beta}$ equals $[f(y_i|\mathbf{x}_i; \boldsymbol{\beta})]^{\delta_i}[1 - F(y_i|\mathbf{x}_i; \boldsymbol{\beta})]^{1-\delta_i}$, so the log likelihood for that individual is given by

$$\phi(y_i, \delta_i|\mathbf{x}_i; \boldsymbol{\beta}) = \delta_i \alpha(y_i|\mathbf{x}_i; \boldsymbol{\beta}) - \int_0^{y_i} \exp \alpha(u|\mathbf{x}_i; \boldsymbol{\beta})du, \quad y_i \geqslant 0 \text{ and } \delta_i \in \{0, 1\},$$

which is a concave function of $\boldsymbol{\beta}$.

The log-likelihood function corresponding to the observed data is given by $l(\boldsymbol{\beta}) = \sum_i \phi(Y_i, \delta_i|\mathbf{x}_i; \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathbf{R}^p$. The maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ can be computed efficiently using the Newton–Raphson* algorithm, and the log likelihood of the fitted model is then given by $\hat{l} = l(\hat{\boldsymbol{\beta}})$.

The basis functions of G that HARE allows are splines (smooth, piecewise polynomial functions) in the covariates, splines in t, and tensor products of two such splines. [The tensor product of the functions $g_1(x_1)$ and $g_2(x_2)$ is the function $g_1(x_1)g_2(x_2)$.] Both the space G and its dimension p are determined adaptively.

Note that if none of the basis functions of G depend on both t and $\mathbf{x}$, then (1) is a proportional hazards model. A particularly interesting feature of HARE is that the model

selection procedure may or may not result in such a model. If any of the basis functions in the selected model is a tensor product of a function in t and a function in one of the covariates, then a proportional hazards model may not be appropriate.

The selection of the dimension and basis functions of G employs stepwise addition and stepwise deletion of basis functions. Initially, the one-dimensional model $\alpha(t|\mathbf{x}; \boldsymbol{\beta}) = \beta_1$ is fitted; then stepwise addition is applied. At each step a candidate basis function for addition to the model is chosen by a heuristic search that is designed approximately to maximize the Rao score statistic* for adding a basis function to the model, which is similar to what is sometimes done in generalized linear modeling. Typically, the candidate basis functions that are considered are linear terms for variables that are not yet in the model, basis functions that introduce one new knot for the spline in a variable already in the model, and tensor products of two basis functions for different variables already in the model.

Upon stopping the stepwise addition stage, a stepwise deletion algorithm is applied. At each step the basis function corresponding to the smallest Wald statistic is deleted (*see* WALD'S *W*-STATISTICS).

During the combination of stepwise addition and stepwise deletion, a sequence of models is obtained. Let p_ν denote the number of parameters and $\hat{l}_\nu$ the log likelihood of the νth model. HARE selects the model that minimizes the Bayes information criterion $\mathrm{BIC} = -2\hat{l}_\nu + p_\nu \log\, n$.

The use of Rao statistics during the stepwise addition and Wald statistics during the stepwise deletion can be motivated by considering a quadratic Taylor approximation of the log-likelihood function. As such, these statistics give an approximation to the change in the log likelihood due to adding or deleting a basis function that does not require finding the new maximum likelihood estimates of the parameters. When all of the basis functions $B_j(t|\mathbf{x})$ in (1) are piecewise linear in time, all integrals needed for the Newton–Raphson algorithm to maximize the log-likelihood function and to compute Rao and Wald statistics can be computed analytically. However, when piecewise cubic splines in time are used, integrals have to be computed numerically.

Kooperberg et al. [11] give more details about the HARE procedure. Under suitable conditions they extended the theoretical framework in Stone [16] to obtain the L_2 rate of convergence for a nonadaptive version of HARE [12].

Example. Data from six breast cancer studies conducted by the Eastern Cooperative Oncology Group have been analyzed [8] using a hybrid of penalized likelihood* and polynomial splines, and using HARE [11]. The data involve 2404 breast cancer patients, 1116 of whose survival times are uncensored and 1288 are censored. The response is the survival time (years) from entry into the study. There are six covariates: estrogen receptor status (ER: 0 is "negative"; 1 is "positive"); the logarithm of the number of positive axilliary lymph nodes at diagnosis; the size of the primary tumor (in centimeters); age at entry; menopause (0 is premenopause, 1 is postmenopause); and body mass index (BMI, defined as weight/height^2 (in kg/m^2).

In the example presented here linear splines (continuous, piecewise linear functions) are used. If quadratic or cubic splines were used, the fitted conditional hazard function would be smoother, but the model description would be more complicated. The use of linear splines has the additional advantage of avoiding the need for numerical integration*.

The HARE fit to the breast-cancer data is summarized in Table 1. The fitted model has knots in time located at 0.441, 1.889, and 7.948 years. Note that this is not a proportional-hazards model because of the presence of the basis functions $(7.948 - t)_+ \times$ ER and $(1.889 - t)_+ \times$ size. Also, the model includes a nonlinear effect (knot) in age and an interaction between log(nodes) and size, and it is nonproportional with respect to ER and size.

From a fitted HARE model, it is easy to plot various functions of interest, such as conditional hazard functions for given sets of covariates. In Fig. 1 the fitted conditional hazard and survival functions for three sets of covariates are shown. As can be seen from

the left side of this figure, ER has a distinctly nonproportional effect: the solid and dotted curves actually cross each other at $t = 3.22$ years. On the other hand, the effect of the number of nodes is proportional.

HARE is very convenient for fitting and comparing linear proportional hazards models, additive proportional hazards models, proportional hazards models with time-varying coefficients, and nonparametric proportional hazards models. For example, when HARE was forced to fit an additive proportional hazards model, there was a decrease of 40.55 in the log likelihood, which is substantial compared to the decrease of three degrees of freedom.

The HARE analysis of these data provides us with a somewhat complicated, nonproportional model for the conditional hazard function, which could not be found by fitting a linear proportional hazards model in the usual manner. Indeed, the HARE analysis suggests that a nonlinear term for age and an interaction between nodes and size should be included in the model. However, the HARE model, which also includes interactions between time and estrogen receptor status and between time and size, is not a proportional hazards model. Thus, even with the appropriate nonlinear effects and interactions between covariates, a proportional hazards model could not be expected to provide a good fit to these data.

Table 1. HARE Analysis of the Breast Cancer Data

Basis Function	Coefficient	Basis Function	Coefficient
1	−3.405	$(0.441 - t)_+$	−5.743
ER	1.060	$(1.889 - t)_+$	−0.966
Log(nodes)	0.688	$(7.948 - t)_+$	0.365
Size	0.159	log(nodes) × size	−0.065
Age	−0.041	$(7.948 - t)_+ \times$ ER	−0.259
$(\text{Age-43})_+$	0.041	$(1.889 - t)_+ \times$ size	0.105
Menopause	0.404		

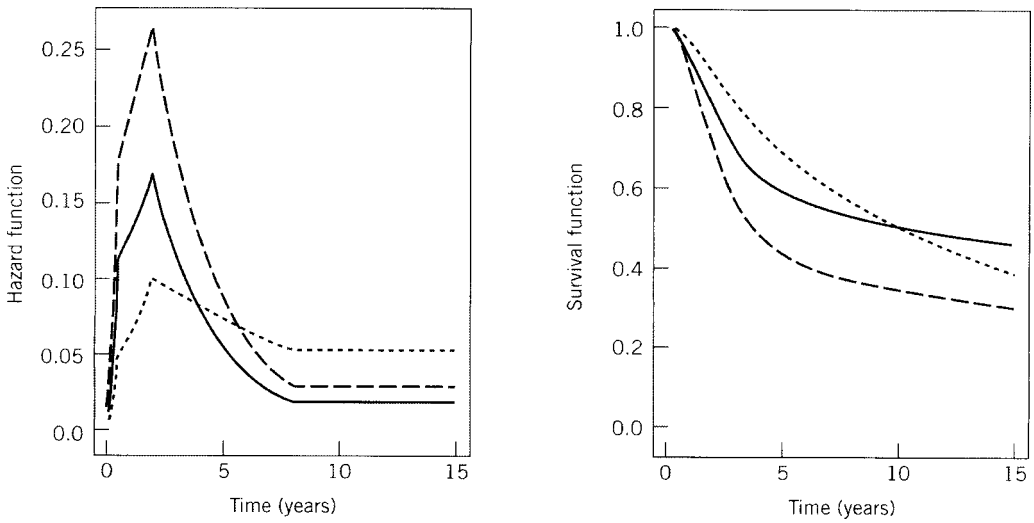

Figure 1. Fitted hazard and survival functions for a premenopausal woman of age 50 with body mass index 25 kg/m^2 and tumor size 3 cm. Solid, 4 nodes and ER negative; dotted, 4 nodes and ER positive; dashed, 10 nodes and ER negative.

REFERENCES

1. Abrahamowicz, M., Ciampi, A., and Ramsay, J. O. (1992). Nonparametric density estimation for censored survival data: regression-spline approach. *Can. J. Statist.,* **20**, 171–185.
2. Anderson, J. A. and Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. R. Statist. Soc. B,* **42**, 322–327.
3. Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B,* **34**, 187–202.
4. Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan–Meier curve. *J. Amer. Statist. Ass.,* **83**, 414–425.
5. Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics,* **43**, 181–192.
6. Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.,* **19**, 1–141.
7. Gray, R. J. (1990). Some diagnostic models for Cox regression models through hazard smoothing. *Biometrics,* **46**, 93–102.
8. Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *J. Amer. Statist. Ass.,* **87**, 942–951.
9. Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
10. Intrator, O. and Kooperberg, C. (1995). Trees and splines in survival analysis. *Statist. Methods Med. Res.,* **4**, 237–261.
11. Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). Hazard regression. *J. Amer. Statist. Ass.,* **90**, 78–94.
12. Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.,* **22**, 143–157.
13. O'sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Sci. Statist. Comput.,* **9**, 363–379.
14. Senthilselvan, A. (1987). Penalized likelihood estimation of hazard and intensity functions. *J. R. Statist. Soc. B,* **49**, 170–174.
15. Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *J. Amer. Statist. Ass.,* **85**, 941–949.
16. Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.,* **22**, 118–184.
17. Whittemore, A. S., and Keller, J. B. (1986). Survival estimation using splines. *Biometrics,* **42**, 495–506.

See also CENSORED DATA; KAPLAN–MEIER ESTIMATOR—I; PROPORTIONAL HAZARDS MODEL, COX'S; SPLINE FUNCTIONS; and SURVIVAL ANALYSIS.

CHARLES KOOPERBERG
CHARLES J. STONE
YOUNG K. TRUONG