A semi-parametric generalization of the Cox proportional hazards regression model: Inference and Applications

Karthik Devarajan[a],[*] and **Nader Ebrahimi**[b]

[a]Division of Population Science, Fox Chase Cancer Center, Philadelphia, PA 19111

[b]Division of Statistics, Northern Illinois University, DeKalb, IL 60115

Abstract

The assumption of proportional hazards (PH) fundamental to the Cox PH model sometimes may not hold in practice. In this paper, we propose a generalization of the Cox PH model in terms of the cumulative hazard function taking a form similar to the Cox PH model, with the extension that the baseline cumulative hazard function is raised to a power function. Our model allows for interaction between covariates and the baseline hazard and it also includes, for the two sample problem, the case of two Weibull distributions and two extreme value distributions differing in both scale and shape parameters. The partial likelihood approach can not be applied here to estimate the model parameters. We use the full likelihood approach via a cubic *B*-spline approximation for the baseline hazard to estimate the model parameters. A semi-automatic procedure for knot selection based on Akaike's Information Criterion is developed. We illustrate the applicability of our approach using real-life data.

Keywords

censored survival data analysis; crossing hazards; Frailty model; maximum likelihood; regression; spline function; Akaike information criterion; Weibull distribution; extreme value distribution

1 Introduction

The modeling and analysis of data in which the principal endpoint is the time until an event occurs is often of prime interest in medical and engineering studies. Typically, such an event is the onset of a disease or death itself as seen in clinical trials or failure of an item or a system as seen in industrial life testing. The time to an event is normally referred to as survival or failure time.

The primary goal in analyzing censored survival data is to assess the dependence of survival time on covariates. The secondary goal is the estimation of the underlying distribution of survival time. The Cox Proportional Hazards (PH) model (Cox, 1972) is a standard tool for exploring the association of covariates with survival time. An interesting feature of this model is that it is semi-parametric in the sense that it can be factored into a parametric part consisting

[*]Corresponding author. karthik.devarajan@fccc.edu (Karthik Devarajan), nader@math.niu.edu (Nader Ebrahimi).

of a regression parameter vector associated with the covariates and a non-parametric part that can be left completely unspecified.

In the Cox PH model, given a vector of possibly time-dependent covariates $\mathbf{z}$, the hazard function at time t is assumed to be of the form

$$\lambda(t|\mathbf{z}) = \lambda_0(t) g(\mathbf{z}). \tag{1.1}$$

where $\lambda_0(t)$ is the baseline hazard function, denoting the hazard under no covariate effect and $g(\mathbf{z})$ is a non-negative function of the covariate vector $\mathbf{z}$, referred to as the risk function, such that $g(\mathbf{0}) = 1$. The most commonly used form of the Cox PH model is

$$\lambda(t|\mathbf{z}) = \lambda_0(t) e^{\beta' \mathbf{z}} \tag{1.2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a p vector of regression coefficients. The focus is on inference for $\boldsymbol{\beta}$, with the baseline hazard function, $\lambda_0(t)$, the non-parametric part, left completely unspecified.

In spite of its semi-parametric feature, the Cox PH model implicitly assumes that the hazard and survival curves corresponding to two different values of the covariates do not cross. Although this assumption may be valid in many experimental settings, it has been found to be suspect in others. For example, if the treatment effect decreases with time, then one might expect the hazard curves corresponding to the treatment and control groups to converge. Other examples that indicate the presence of non-proportional hazards are also given in Gore et al. (1984), and Tonak et al. (1979), among others.

In this paper, we describe a semi-parametric generalization of the Cox PH model which allows crossing of hazards as well as survival functions. In Section 2, we discuss its unique properties and place it within the context of censored survival data analysis. In Section 3, we describe an estimation procedure for this model using cubic B-spline approximations for the baseline hazard. We illustrate our method with real-life examples in Section 4, and in Section 5 we provide some concluding remarks.

2 A Semi-Parametric Generalization of the Cox PH Model

We describe a semi-parametric generalization of the Cox PH model in which the hazard functions corresponding to different values of the covariates can cross. The special case of this model was originally introduced by Quantin et al. (1996) for the purpose of goodness of fit testing of the Cox PH model. Devarajan (2000) outlined the unique properties of this non-proportional hazards regression model as well as inference for this model using maximum penalized likelihood estimation, and provided a theoretical justification for using spline approximations for the baseline hazard. In addition, Devarajan and Ebrahimi (2002, p.237) used this model and developed a goodness of fit procedure for testing the Cox PH model. In independent work, Hsieh (2001) and Wu & Hsieh (2009) discussed an estimating equations approach for this model by approximating the baseline hazard using piecewise constants.

In our model, the survival function corresponding to a covariate vector $\mathbf{z}$ is assumed to be of the form

$$S(t|\mathbf{z}) = g(\mathbf{z}) \{S_0(t)\}^{h(\mathbf{z})}, \tag{2.1}$$

where $S_0(t)$ is an arbitrary baseline survival function, and $g(\mathbf{z})$ and $h(\mathbf{z})$ are nonnegative functions of the covariate vector $\mathbf{z}$ such that $g(\mathbf{0}) = h(\mathbf{0}) = 1$. This model includes, for the two sample problem, the case of two Weibull and two extreme value distributions differing in both scale and shape parameters. The Cox PH model is obtained as a special case of model (2.1) by setting $h(\mathbf{z}) = 1$. In this paper, we will consider only the exponential function for g and h, i.e., $g(\mathbf{z}) = \exp(\boldsymbol{\beta}'\mathbf{z})$ and $h(\mathbf{z}) = \exp(\boldsymbol{\gamma}'\mathbf{z})$ where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown p vectors of parameters. In terms of cumulative hazard functions, our non-proportional hazards regression model takes the specific form

$$\Lambda(t|\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}} \{\Lambda_0(t)\}^{\exp\left(\boldsymbol{\gamma}'\mathbf{z}\right)}. \tag{2.2}$$

The Cox PH model is obtained as a special case of model (2.2) by setting $\gamma = 0$. The conditional survival function is

$$S(t|\mathbf{z}) = \exp\{-e^{\boldsymbol{\beta}'\mathbf{z}}[-\log S_0(t)]^{\exp\left(\boldsymbol{\gamma}'\mathbf{z}\right)}\} \tag{2.3}$$

and the conditional hazard function is

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp[(\beta+\gamma)'\mathbf{z} + \{e^{\boldsymbol{\gamma}'\mathbf{z}} - 1\} \log\{\Lambda_0(t)\}]. \tag{2.4}$$

Applying a complementary $\log(-\log)$ transformation in (2.3), we get

$$\log\{-\log\{S(t|\mathbf{z})\}\} = \boldsymbol{\beta}'\mathbf{z} + \exp(\boldsymbol{\gamma}'\mathbf{z}) \log\{-\log\{S_0(t)\}\}$$

The above equation can be expressed as

$$\psi\{S(t|\mathbf{z})\} = \exp(\boldsymbol{\gamma}'\mathbf{z}) \cdot h(t) + \boldsymbol{\beta}'\mathbf{z} \tag{2.5}$$

where $\psi(x) = \log(-\log(x))$ and $h(t) = \log\{-\log\{S_0(t)\}\}$. This can be shown to be a member of the family of models described in Cheng et al. (1997) (see Devarajan, 2000, pp.45-48 for details). An equivalent version of (2.5) is

$$\exp(\boldsymbol{\gamma}'\mathbf{z}) \cdot h(T) = -\boldsymbol{\beta}'\mathbf{z} + \varepsilon \tag{2.6}$$

where the error ε has distribution function $F = 1 - \psi^{-1}$. Since the distribution of the baseline $h(T) = \log\{-\log(S_0(T)\}$ is unit extreme value, equation (2.6) results in a scale and shape transformation of this unit extreme value distribution. The generalized model (2.3) can be interpreted as a transformation of the unit extreme value distribution in terms of reparametrizations of the scale and shape parameters. Similarly, applying a log transformation in (2.3) (with $\psi(x) = \log(x)$ and $h(t) = \log\{S_0(t)\}$) and using similar arguments as above, one can interpret our generalized model as a transformation of the unit exponential distribution in terms of reparametrizations of the scale and shape parameters.

2.1 Some Useful Features of the Generalization

We describe several useful features of our generalized model and highlight its relationship to other survival models. These properties provide the framework for a generalized approach to censored survival data analysis.

2.1.1 Crossing hazards over time—The hazards ratio corresponding to two different covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ is

$$\frac{\lambda(t|\mathbf{z}_1)}{\lambda(t|\mathbf{z}_2)} = \exp\{(\beta+\gamma)'(\mathbf{z}_1 - \mathbf{z}_2) + [e^{\gamma'\mathbf{z}_1} - e^{\gamma'\mathbf{z}_2}]\log[\Lambda_0(t)]\}. \tag{2.7}$$

Since this ratio is a monotone function of t, the model allows the hazards ratio to invert over time. In other words, it allows crossing of hazard curves. For example, when treatment effect decreases or increases over time, model (2.2) can be applied.

2.1.2 Relation to the Time-Dependent Coefficient Cox PH Model—If γ is assumed to be close to zero, we can approximate the right hand side of equation (2.2) as follows:

$$\Lambda(t|\mathbf{z}) = e^{\beta'\mathbf{z}}\{\Lambda_0(t)\}^{1+\gamma'\mathbf{z}}.$$

Thus, for two different covariate vectors $\mathbf{z_1}$ and $\mathbf{z_2}$, we have,

$$\begin{aligned}
\log\{\tfrac{\Lambda(t|\mathbf{z}_1)}{\Lambda(t|\mathbf{z}_2)}\} &= [\beta+\log\{\Lambda_0(t)\}\cdot\gamma]'(\mathbf{z_1} - \mathbf{z_2}) \\
&= \eta'(t)(\mathbf{z}_1 - \mathbf{z}_2),
\end{aligned} \tag{2.8}$$

where $\boldsymbol{\eta}(t) = \boldsymbol{\beta} + g(t)\cdot\boldsymbol{\gamma}$ with $g(t) = \log\{\Lambda_0(t)\}$. This is a special case of the Cox PH model with time-dependent coefficients $\boldsymbol{\eta}(t)$ which allows for crossing of hazard curves. When the deviation from proportional hazards is small, our proposed model approximates the Cox PH model with time dependent coeffcients.

In terms of the hazard functions,

$$\frac{\lambda(t|\mathbf{z_1})}{\lambda(t|\mathbf{z_2})} = \exp\{\beta+h(t)\cdot\gamma\}'(\mathbf{z_1} - \mathbf{z_2}), \tag{2.9}$$

where $h(t) = 1+\log\{\Lambda_0(t)\}$. Cox (1972) considered the case where $h(t) = t$, a dummy time-dependent variable for goodness of fit testing of the Cox PH model. Therneau and Grambsch (1994) have considered the model (2.9) with an assumed form for $h(t)$.

2.1.3 Proportionality of the Hazard-Cumulative Hazard Ratios—Using equations (2.2) and (2.4), we see that,

$$\frac{\lambda(t|\mathbf{z})}{\Lambda(t|\mathbf{z})} = \frac{\lambda_0(t)}{\Lambda_0(t)}\exp(\gamma'\mathbf{z}). \tag{2.10}$$

Let $k(t|\mathbf{z})$ denote the conditional growth rate of the logarithm of the cumulative hazard function. Then $k(t|\mathbf{z})=\dfrac{d}{dt}\log \Lambda(t|\mathbf{z})=\dfrac{\lambda(t|\mathbf{z})}{\Lambda(t|\mathbf{z})}.$ Using (2.10), we see that

$$k(t|\mathbf{z})=k_0(t)\exp(\gamma'\mathbf{z}), \tag{2.11}$$

where $k_0(t)=\dfrac{\lambda_0(t)}{\Lambda_0(t)}$ is the growth rate of the logarithm of the baseline cumulative hazard function.

Equation (2.10) has a form similar to that of (1.2), the Cox PH model. The difference between the two is that (1.2) models the growth rate of the cumulative hazard function while (2.10) models the growth rate of the logarithm of the cumulative hazard function. Equation (2.10) implies that the ratio of the hazard function to the cumulative hazard function, given a covariate vector $\mathbf{z}$, is proportional to the ratio of the hazard function to the cumulative hazard function at baseline. It is worth mentioning here that this property is related to the constancy of the ratio of the hazard function to the cumulative hazard function of the extreme value distribution. Note that if we set $\gamma = 0$ in (2.10), the ratios of hazard to cumulative hazard in (2.10) are equal and it reduces to the Cox PH model.

2.1.4 Proportionality of the Logarithm of Cumulative Hazards—For $\beta = 0$, model (2.2) reduces to

$$\Lambda(t|\mathbf{z})=\{\Lambda_0(t)\}^{\exp(\gamma'\mathbf{z})}. \tag{2.12}$$

Model (2.12) is similar to (2.2) and has all its features with the exception that it does not include the Cox PH model. By taking logarithm on both sides of (2.12), it is easy to see that the logarithms of the cumulative hazards are proportional.

2.1.5 Relation to the Frailty Model—In survival analysis, it has been found that differences between hazards due to covariate effects tend to converge as follow-up time elapses. Such an effect can be accounted for by postulating the existence of unobserved random effects or frailties with prognostic value. In terms of hazard functions, we can write

$$\lambda(t|\mathbf{z}, u)=\lambda_0(t)e^{\beta'\mathbf{z}}u, \tag{2.13}$$

where u is an unobserved frailty or random effect (see Hougaard (2000) pp.215-245 for more details). Assuming that u has a positive stable distribution with parameter θ, one can easily show that

$$\Lambda(t|\mathbf{z})=\{\Lambda_0(t)\}^{\exp(\theta)}e^{u\theta\beta'\mathbf{z}}. \tag{2.14}$$

It is clear from (2.14) that if we define $\gamma = (\theta, 0, \cdots, 0)$, then our proposed model and (2.14) will be the same. When frailty interacts with treatment, we can re-write (2.14) as

$$\lambda(t|\mathbf{z}, u)=\lambda_0(t)e^{u\beta'\mathbf{z}}. \tag{2.15}$$

This also leads to crossing or diverging hazards (Cuzick & Trejo, 1992, p.65; Hastie & Tibshirani, 1993) and can be shown to be equivalent to (2.14), and hence to our proposed model.

2.1.6 Interaction between covariates and baseline hazard—We can rewrite the conditional survival function given by (2.3) as

$$
\begin{aligned}
S(t|\mathbf{z}) &= \exp\{-e^{\beta' \mathbf{z}}[-\log S_0(t)]^{\exp(\gamma' \mathbf{z})}\}\\
&= g_0\{\beta' \mathbf{z} + e^{\gamma' \mathbf{z}} h_0(t)\}
\end{aligned}
\tag{2.16}
$$

where $g_0(x) = \exp\{-exp(x)\}$ and $h_0(t) = \log \Lambda_0(t)$. The generalized model (2.3) allows for interaction between covariates and the baseline hazard.

2.2 Interpretation of the model

For the two-sample problem, the non-proportional hazards model (2.2) can be written as

$$
\Lambda(t|z=1) = \{\Lambda_0(t)\}^{\exp(\gamma)} e^{\beta}
\tag{2.17}
$$

where $\Lambda_0(t) = \Lambda(t|z = 0)$ and $\Lambda(t|z = 1)$ are the conditional cumulative hazard functions for the control and treatment groups respectively. From (2.17), we see that as $\gamma \to 0$, the model becomes proportional in hazards. As $|\gamma| \to \infty$, the model deviates from the proportionality assumption, i.e., large values of $|\gamma|$ may be indicative of diverging or converging hazards over time. One way to interpret the parameter γ in this model is to consider the following equation obtained from (2.17) using (2.10),

$$
\frac{\lambda(t|z=1)}{\Lambda(t|z=1)} = \frac{\lambda_0(t)}{\Lambda_0(t)} \exp(\gamma),
\tag{2.18}
$$

which implies that

$$
\log\left\{ \frac{\frac{\lambda(t|z=1)}{\Lambda(t|z=1)}}{\frac{\lambda_0(t)}{\Lambda_0(t)}} \right\} = \gamma,
$$

a result analogous to that obtained (as the logarithm of the relative risk) for the regression parameter β in the Cox PH model in the two-sample setting.

3 Estimation for the Non-Proportional Hazards Model

The observed data consist of independent observations on the triple $(X, \delta, \mathbf{z})$, where X is the minimum of a failure and censoring time pair (T, C), $\delta = I(T \leq C)$ is the indicator of the event that a failure has been observed and $\mathbf{z} = (z_1, \dots , z_p)'$ is a p vector of covariates. The random variables T and C denote the survival and censoring times respectively which are assumed to be independent.

The fundamental assumption of proportionality of hazards in the Cox PH model (1.2) requires that the hazards ratio remains constant over time. The partial likelihood approach of Cox (1972) is the standard inferential method for this model. The baseline hazard, $\lambda_0(t)$, drops out of this partial likelihood and can be left completely unspecified (Cox, 1972). On the other hand,

it is evident from equations (2.4) and (2.7) that the generalized model (2.2) allows crossing of hazard curves over time. Due to the non-constant hazards ratio, a factorization of the likelihood into a "partial" likelihood and a nuisance part is not possible. The baseline hazard must be specified in the likelihood and needs to be explicitly estimated. There are several methods available for estimation in this model. For example, Bordes & Breuils (2006) discussed sequential estimation for semi-parametric models with application to the PH model, while Wu & Hsieh (2009) adopted a piecewise constant approach to estimating the baseline hazard in model (2.2). One could use a set of parametric basis functions, such as splines, to approximate the baseline hazard and then proceed with maximum likelihood (ML) methods to estimate the parameters of the model. Splines have been used to model the baseline hazard in the Cox PH model by Durrelman & Simon (1989), Gray (1992) and Rosenberg (1995).

Models based on spline approximations are intermediate in structure between completely parametric models and non-parametric methods, and provide flexibility in incorporating a variety of shapes. Under certain regularity conditions, Devarajan (2000, pp.53-58) showed that the maximum penalized likelihood estimate of the baseline hazard and baseline cumulative hazard in model (2.2) are exponential splines with knots at the unique failure times. This provides us with a theoretical justification for using spline functions to model the baseline hazard and cumulative hazard functions. de Boor (2001, Chapter IX, p.87) describes in detail the computational properties of *B-spline basis functions*. The *B*-spline basis function has the properties of a probability density function and a linear combination of such functions is like a mixture of densities that allows for various distributions. In this paper, we use a set of cubic *B*-spline basis functions to model the baseline hazard.

3.1 Modeling the Baseline Hazard using B-Splines

We assume that the baseline hazard function, $\lambda_0(t)$, and its first two derivatives are continuous and model it as a linear combination of cubic *B*-spline basis functions. This model consists of $K + 1$ segments determined by K knots $t_0 < t_1 < t_2 < \ldots < t_K < t_{K+1}$ over the interval $[t_0, t_{K+1}]$ subject to the constraints that adjacent functions and their first two derivatives agree at the knots. It requires total of $4(K + 1)$ parameters and $K + 4$ basis functions (Devarajan (2000, p. 72); de Boor (2001)).

We also define six additional 'slack' knots (Rosenberg (1995)) given by $t_{-3} = t_0 - 3$, $t_{-2} = t_0 - 2$, $t_{-1} = t_0 - 1$, $t_{K+2} = t_{K+1} + 1$, $t_{K+3} = t_{K+1} + 2$ and $t_{K+4} = t_{K+1} + 3$ where $t_0 = \min(x_i)$, $t_{K+1} = \max(x_i)$ and (x_i, δ_i), $i = 1, \ldots, n$ are right-censored observations. Using the result from Schumaker (1981, p.46), the *cubic B-Spline basis functions* are given by

$$B_k(t) = (t_{k+4} - t_k) \sum_{l=k}^{k+4} \left\{ \frac{(t_\ell - t)_+^3}{\prod_{\substack{m=k \\ m \neq l}}^{k+4} t_m - t_l} \right\} \tag{3.1}$$

where $x_+ = x$ if $x \geq 0$ and zero otherwise. Using (2.2) and (2.4), the log-likelihood for right-censored observations (x_i, δ_i), $i = 1, \ldots, n$ where δ_i is the censoring indicator, is given by

$$\ell = \sum_{i=1}^{n} \left\{ \delta_i \left[\log \lambda_0(x_i) + (\beta+\gamma)' \mathbf{z}_i + \left(e^{\gamma' \mathbf{z}_i} - 1 \right) \cdot \log \Lambda_0(x_i) \right] - \left[\Lambda_0(x_i) - \Lambda_0(t_0) \right]^{\exp\left(\gamma' \mathbf{z}_i\right)} e^{\beta' \mathbf{z}_i} \right\}.$$

(3.2)

Using a linear combination of *B*-spline basis functions, the baseline hazard is given by

$\lambda_0(t|\alpha) = \sum_{k=-3}^{K} \eta_k B_k(t)$ with $B_k(t)$ defined as in (3.1). With the number of knots fixed at K, the order of approximation attainable is $O(K^{-4})$ (Devarajan (2000, p.73); de Boor (2001)). We take $\eta_k = e^{\alpha k}$ to ensure positivity of the hazard function $\lambda_0(t|\alpha)$. The cumulative baseline hazard function is $\Lambda_0(t|\alpha) = \sum_{k=-3}^{K} e^{\alpha k} I_k(t)$ where

$$I_k(t) = \frac{(t_{k+4} - t_k)}{4} \left\{ \sum_{l=k}^{k+4} \frac{t_{l+}^4}{\prod_{\substack{m=k \\ m \neq l}}^{k+4} (t_m - t_l)} \right\} - \frac{(t_{k+4} - t_k)}{4} \left\{ \sum_{l=k}^{k+4} \frac{(t_l - t)_+^4}{\prod_{\substack{m=k \\ m \neq l}}^{k+4} (t_m - t_l)} \right\}$$

is a *quartic B-spline basis function*. Since $I_k(t)$ is monotone increasing and positive and $e^{\alpha k}$ is positive for each k, monotonicity and positivity of the cumulative hazard function is ensured. Now, the log-likelihood given by (3.2) becomes,

$$\ell = \sum_{i=1}^{n} \delta_i \left\{ \log \left(\sum_{k=-3}^{K} e^{\alpha_k} B_k(x_i) \right) + (\beta+\gamma)' \mathbf{z}_i + (e^{\gamma' \mathbf{z}_i} - 1) \cdot \log \left(\sum_{k=-3}^{K} e^{\alpha_k} I_k(x_i) \right) \right\} - \sum_{i=1}^{n} \left\{ \sum_{k=-3}^{K} e^{\alpha_k} \left[I_k(x_i) - I_k(t_0) \right] \right\}^{\exp\left(\gamma' \mathbf{z}_i\right)} e^{\beta' \mathbf{z}_i}.$$

(3.3)

This equation is maximized with respect to the model parameters α, β and γ to obtain the maximum likelihood (ML) estimates. Using the second partial derivatives of the likelihood function with respect to the model parameters, we obtain the Fisher information matrix and hence the variance-covariance matrix of the parameter estimates. If $n \to \infty$ at a rate faster than the rate at which $K \to \infty$, then one can show that our proposed estimators are consistent estimators and are normally distributed (Devarajan, 2000, pp.73-74).

3.2 Implementation

ML estimation of model parameters was implemented using the Broyden-Fletch-Goldfarb-Shanno (BFGS) method for the unconstrained optimization of a nonlinear multivariate function in double precision. Routines from the IMSL Fortran Library (DU2ING and DU2POL) were used for maximization of the likelihood function with respect to model parameters α, β and γ to obtain the ML estimates. Multiple restarts of the optimization procedure were performed with various starting values to check the consistency and validity of the results.

3.3 Knot Selection

An integral part of the estimation procedure using any spline approximations is the choice of the number and location of the knots. In general, it is desirable to have the knots reasonably spread out with roughly equal number of data points between successive knots. Knot selection has been discussed by many authors in the context of hazard estimation. We develop a semi-automatic knot selection procedure that involves initial knot placement based on quantiles equally spaced in percentile ranks of the uncensored observations followed by stepwise knot deletion and final model selection based on *Akaike's Information Criterion (AIC)*. During the stepwise deletion process, a sequence of models indexed by v is obtained such that the vth model has n_v parameters. In each step, only the significant parameters are retained. The significance of any parameter associated with a covariate or a knot, say β, is determined by its Wald z-score. All non-significant knots are removed from the likelihood function at each stage as they do not provide a good approximation to the baseline hazard. If l_v denotes the log-likelihood of the vth model, then the *AIC* for this model, with penalty parameter $a = 2$ is given by

$$AIC(a, v, K) = -2\ell_v + an_v = -2\ell(\alpha, \beta, \gamma; K) + 2n_v.$$

The model v corresponding to the minimum value of $AIC(a = 2, v, K)$ is chosen as the final model where n_v is the number of significant parameters in the final model. This method provides a balance between computational complexity and model flexibility.

4 Illustration of our Methods

We illustrate estimation in the non-proportional hazards model (2.2) using real-life examples. All figures presented were created using the R statistical language and environment (R Development Core Team (2009), www.R-project.org).

4.1 Example 1: Leukemia Data

This dataset consists of times to remission (in weeks) of 42 leukemia patients (Kalbfleisch and Prentice, 1980, pp.205-206). Patients are randomized into two groups of 21 each - a control and a treated group. Group membership is the only covariate of interest and is specified by z where $z = 0, 1$ denote the treatment and control groups, respectively. Figure 1(a) presents the Kaplan-Meier survival curves for these groups.

First, we modeled this data using the non-proportional hazards regression model (2.2) with *B*-spline approximations for the baseline hazard. Slack knots were chosen as described in Section 3.1. The remaining knots were chosen based on quantiles equally spaced in percentile ranks of the uncensored observations. This model consists of two parameters, β and γ. The knot selection procedure described in Section 3.3 was implemented for this data set. The model with three knots based on quantiles of uncensored observations provided the best fit using the *AIC* criterion. Table 4.1 presents a summary of the knot selection procedure where K denotes the number of knots. Results of the final model fit are provided in Table 4.2. The log-likelihood corresponding to this final model is -106.084. The parameter estimates corresponding to the 4 slack knots are $\hat{\alpha}_{-3} = -19.93$, $\hat{\alpha}_{-2} = -4.15$, $\hat{\alpha}_{-1} = -3.97$ and $\hat{\alpha}_0 = -3:45$; and those corresponding to the 3 knots based on quantiles are $\hat{\alpha}_1 = -3.68$, $\hat{\alpha}_2 = -2.88$ and $\hat{\alpha}_3 = -16.92$. The parameter γ provides information about the non-proportionality in the hazards and its estimate is close to zero. From the results presented in Table 4.2, we observe that this parameter is highly non-significant while the parameter β is highly significant at the 5% level. Figure 1(b) presents the fitted survival curves for the two groups based on model (3.1) computed using the parameter estimates specified above.

Next, we fit the Cox PH model to this data. We obtained an estimate of 1.57 (with standard error 0.412) for the regression coefficient β and a highly significant p-value of 0.00014. Our results for the fit based on the non-proportional hazards model (2.2) corroborate those based on a fit of the Cox PH model. In addition, the fitted parametric survival curves presented in Figure 1(b) lend evidence to the same conclusion as their non-parametric counterparts shown in Figure 1(a). The Cox PH model thus provides a good fit for this example and the estimated treatment effect is indeed 1.57 under both models.

Example 2: Mice Data

As a general rule, we recommend fitting the generalized model (2.2) to a dataset since the Cox PH model is embedded within this model. If the assumption of proportionality is indeed true, we expect it to be reflected in the model fit as seen in this example. On the other hand, any evidence of crossing hazards will likely be indicated by the fit of the more general model. We illustrate this in our next example.

In this example, we analyze data from an animal experiment presented in Wei (1984) in which a control group of 22 male mice are subject to 300 rads of radiation and followed for cancer incidence. The treatment group consists of 29 mice placed in a germ-free environment. The purpose of the study was to determine the effect of a germ-free environment in developing thymic lymphoma. All 51 observations in this dataset were uncensored. Figure 2(a) presents the survival curves for the control and treatment groups. The survival curves cross at various points until about 250 days after which the treatment group has higher survival probability than the control group.

First, we modeled this data using the non-proportional hazards regression model (2.2) with B-spline approximations for the baseline hazard. Slack knots were chosen as described in Section 3.1, and the remaining knots were chosen based on quantiles equally spaced in percentile ranks of the uncensored observations. As in the previous example, this model contains two parameters, β and γ. Once again, the knot selection procedure described in Section 3.3 was implemented for this data. The best model, chosen using the *AIC* criterion, consisted of four knots based on quantiles of uncensored observations. Table 4.3 presents a summary of the knot selection procedure for this data. Results of the final model fit are provided in Table 4.4. The log-likelihood corresponding to this final model is -301.369. The parameter estimates corresponding to the 4 slack knots are $\hat{\alpha}_{-3} = -5.93$, $\hat{\alpha}_{-2} = -169.71$, $\hat{\alpha}_{-1} = -4.73$ and $\hat{\alpha}_0 = -4.18$; and those corresponding to the 4 knots based on quantiles are $\hat{\alpha}_1 = -34{:}21$, $\hat{\alpha}_2 = -3.20$, $\hat{\alpha}_3 = -12.79$ and $\hat{\alpha}_4 = -2.54$.

From the results of the final model fit presented in Table 4.4, we observe that the parameter β is marginally significant at the 5% level while the parameter γ is highly significant. This indicates that the hazards for the two treatment groups are non-proportional and corroborates the graphical evidence from Figure 2(a). We computed the fitted survival curves for the two groups based on (3.1) using the estimates of the model parameters specified above. These parametric curves, displayed in Figure 2(b), cross in a similar fashion to that seen in their non-parametric counterparts shown in Figure 2(a).

For exploratory purposes, we also analyzed this data set using the Cox PH model. The regression coefficient in this model, β, was only marginally significant at the 5% level (p-value = 0.0732). This suggests a possible non-proportionality of the hazards corresponding to the treatment and placebo groups. Using (2.18), we obtain

$$\log\left\{\frac{\frac{\widehat{\lambda(t|z=1)}}{\Lambda(t|z=1)}}{\frac{\lambda_0(t)}{\Lambda_0(t)}}\right\}=\widehat{\gamma}=-0.348$$

for this example and thus we expect

$$\frac{\lambda(t|z=1)}{\Lambda(t|z=1)}<\frac{\lambda_0(t)}{\Lambda_0(t)}.$$

This indicates an interaction between the baseline hazard and treatment. Note that proportionality of hazards amounts to equality of the hazard-cumulative hazard ratios in (2.18). Using (2.17), we also obtain

$$\frac{\log\widehat{\Lambda}(t|z=1)}{\log\widehat{\Lambda}_0(t)}=\exp(-0.348)=0.71.$$

That is, mice in the treatment group have significantly better survival rate than the control group.

These examples suggest that a few knots based on equally spaced quantiles of uncensored observations provide a reasonably good approximation to the unknown baseline hazard. Also, a fit of this generalized model provides a method for testing the assumption of proportional hazards.

Acknowledgments

References

1. Bordes L, Breuils C. Sequential Estimation for Semiparametric Models with Application to the Proportional Hazards Model. Journal of Statistical Planning and Inference 2006;136:3735–3759.

2. de Boor, C. A Practical Guide to Splines. Springer-Verlag; New York: 2001.

3. Cheng SC, Wei LJ, Ying Z. Predicting Survival Probabilities with Semiparametric Transformation Models. Journal of the American Statistical Association 1997;92:227–235.

4. Cox DR. Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society-Series B 1972;34:187–202.

5. Cuzick, J.; Trejo, B. Analysis of Trials with Treatment-Individual Interactions. In: Klein, J.; Goel, P., editors. Survival Analysis: State of the Art. Kluwer, Dordrecht; The Netherlands: 1992. p. 65-71.

6. Devarajan, K. PhD Dissertation. Division of Statistics, Northern Illinois University; 2000. Inference for a Non-proportional Hazards Regression Model and Applications.

7. Devarajan, K.; Ebrahimi, N. Goodness-of-Fit Testing for the Cox Proportional Hazards Model. In: Huber, C.; Balakrishnan, N.; Nikulin, MS.; Mesbah, M., editors. Goodness-of-Fit Tests and Model Validity. Birkhauser; Boston: 2002. p. 237-254.

8. Devarajan K, Ebrahimi N. Testing for Covariate Effect in the Cox Proportional Hazards Regression Model. Communications in Statistics - Theory & Methods 2009;38(14):2333–2347. [PubMed: 20054448]

9. Durrelman S, Simon R. Flexible Regression Models with Cubic Splines. Statistics in Medicine 1989;8:551–561. [PubMed: 2657958]

10. Gore SM, Pocock SJ, Kerr GR. Regression Models and Non-proportional Hazards in the Analysis of Breast Cancer Survival. Applied Statistics 1984;33:176–195.

11. Gray RJ. Spline-based Tests in Survival Analysis. Biometrics 1994;50:640–652. [PubMed: 7981391]

12. Hastie T, Tibshirani RJ. Varying-Coefficient Models. Journal of the Royal Statistical Society-Series B 1993;55:757–796.

13. Hougaard, P. Analysis of Multivariate Survival Data. Springer-Verlag; New York: 2000.

14. Hsieh F. On Heteroscedastic Hazards Regression Models: Theory and Application. Journal of the Royal Statistical Society-Series B 2001;63(1):63–80.

15. IMSL Fortran Math Library. Fortran Subroutines for Mathematical Applications 1991;3(version 2)

16. Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. John Wiley; New York: 1980.

17. Quantin C, Moureau T, Asselain B, Maccario J, Lellouch J. A Regression Model for Testing the Proportional Hazards Hypotheses. Biometrics 1996;52:874–885. [PubMed: 8924576]

18. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2009. URL http://www.R-project.org.

19. Rosenberg PS. Hazard Function Estimation Using B-Splines. Biometrics 1995;51:874–887. [PubMed: 7548706]

20. Schumaker, LL. Spline Functions Basic Theory. John Wiley & Sons; New York: 1981.

21. Therneau TR, Grambsch PM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika 1994;81:515–526.

22. Tonak V, Hermanek P, Banz H, Groitl H. Cytotoxic and Hyperthermic Perfusion: A Preliminary Study. Cancer Treatment Review 1979;6:135–141.

23. Wei LJ. Testing Goodness-of-Fit for Proportional Hazards Model with Censored Observations. Journal of the American Statistical Association 1984;79:649–652.

24. Wu H-DI, Hsieh F. Heterogeneity and Varying Effect in Hazards Regression. Journal of Statistical Planning and Inference 2009;139:4213–4222.
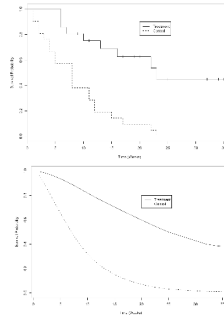
Figure 1.
(a): Kaplan-Meier Survival Curves (Leukemia Data)
(b) Fitted Survival Curves (Leukemia Data)
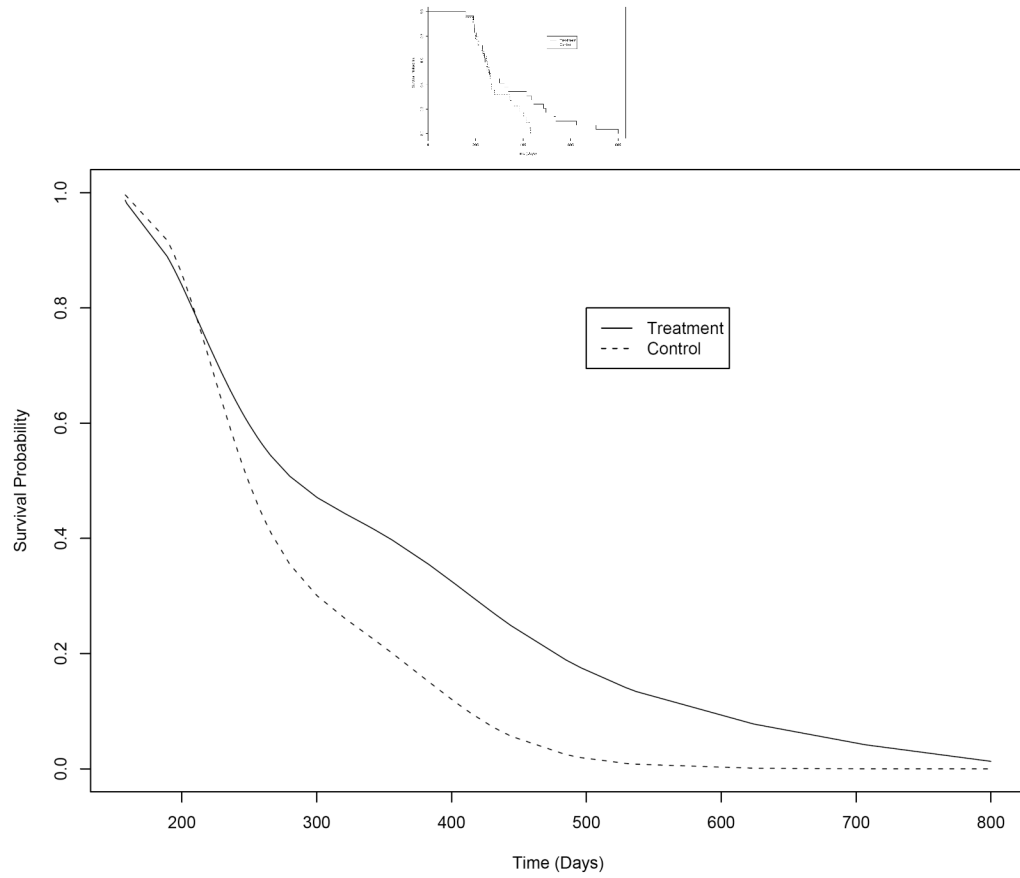
Figure 2.
(a) Kaplan-Meier Survival Curves (Mice Data)
(b) Fitted Survival Curves (Mice Data)

Table 4.1

Summary of Knot Selection Procedure for Leukemia Data

K	$\hat{\beta}$	$\hat{\gamma}$	log-likelihood	AIC
10	1.890	0.099	-104.772	239.544
9	1.616	-0.076	-104.556	223.112
8	1.692	0.070	-104.927	231.854
7	1.719	0.058	-105.735	227.470
6	1.672	0.007	-105.683	223.366
5	1.578	-0.066	-105.522	223.044
4	1.564	-0.088	-105.909	223.818
3	1.573	-0.111	-106.084	222.168
2	1.549	-0.129	-106.106	222.212
1	1.538	-0.138	-106.120	222.240

Table 4.2

Leukemia Data - Knot Sequence: 5,8,12.75 Quantiles of Uncensored Observations based on Final Model

Parameter	Estimate	Std. Error	Wald z-score	*p*-value
β	1.573	0.5044	3.1187	0.0018
γ	-0.111	0.3113	-0.3568	0.7213

Table 4.3

Summary of Knot Selection Procedure for Mice Data

K	$\hat{\beta}$	$\hat{\gamma}$	log-likelihood	AIC
6	-0.453	-0.411	-300.516	625.032
5	-0.434	-0.393	-301.369	624.154
4	-0.414	-0.348	-301.369	622.738
3	-0.438	-0.270	-302.743	623.486
2	-0.364	-0.587	-304.204	624.408
1	-0.438	-0.312	-305.732	625.464

Table 4.4

Mice Data - Knot Sequence: 202, 244, 300, 428 Quantiles of Uncensored Observations based on Final Model

Parameter	Estimate	Std. Error	Wald z-score	*p*-value
β	-0.414	0.222	-1.865	0.0622
γ	-0.348	0.144	-2.417	0.0157