Published in final edited form as:

Stat Probab Lett. 2013 July ; 83(7): 1703–1710. doi:10.1016/j.spl.2013.03.007.

On penalized likelihood estimation for a non-proportional hazards regression model

Karthik Devarajan^{a,*} and Nader Ebrahimi^b

^aBiostatistics & Bioinformatics Department, Fox Chase Cancer Center, Philadelphia, PA 19111

^bDivision of Statistics, Northern Illinois University, DeKalb, IL 60115

Abstract

In this paper, a semi-parametric generalization of the Cox model that permits crossing hazard curves is described. A theoretical framework for estimation in this model is developed based on penalized likelihood methods. It is shown that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are hyperbolic splines with knots at the distinct failure times.

Keywords

censored survival data analysis; crossing hazards; proportional hazards; spline; penalized likelihood

1 Introduction

The primary goal in analyzing censored survival data is to assess the dependence of survival time on one or more explanatory variables or covariates. The Cox Proportional Hazards (PH) model has become the standard tool for exploring this association (Cox (1972)). Given a vector of possibly time-dependent covariates $\mathbf{z}$, the hazard function at time t is assumed to be of the form $\lambda(t|\mathbf{z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{z}}$ where $\lambda_0(t)$ is the baseline hazard function, denoting the hazard under no covariate effect and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ 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 attractive semi-parametric feature, the Cox PH model implicitly assumes that the hazard 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-

© 2013 Elsevier B.V. All rights reserved.

*Corresponding author. karthik.devarajan@fccc.edu.
nader@math.niu.edu (Nader Ebrahimi).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

proportional hazards are detailed in Devarajan and Ebrahimi (2002, 2011), Wu and Hsieh (2009) and references therein.

In order to provide modeling flexibility that accounts for non-proportional hazards, several models have been proposed and investigated in the literature. In this paper, we describe one such semi-parametric generalization of the Cox PH model in which the hazard functions corresponding to different values of covariates can cross (see Devarajan and Ebrahimi (2011) and references therein). The survival function corresponding to a covariate vector $\mathbf{z}$ is assumed to be of the form

$$S(t|\mathbf{z}) = \exp(\boldsymbol{\beta}'\mathbf{z}) \{S_0(t)\}^{\exp(\boldsymbol{\gamma}'\mathbf{z})}, \quad (1.1)$$

where $S_0(t)$ is an arbitrary baseline survival function, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown p vectors of parameters. In terms of cumulative hazard functions, this model takes the form

$$\Lambda(t|\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}} \{\Lambda_0(t)\}^{\exp(\boldsymbol{\gamma}'\mathbf{z})}. \quad (1.2)$$

The Cox PH model is obtained as a special case of model (1.2) by setting $\boldsymbol{\gamma} = 0$. The conditional hazard function is

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp\left[(\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z} + \left\{e^{\boldsymbol{\gamma}'\mathbf{z}} - 1\right\} \log\{\Lambda_0(t)\}\right]. \quad (1.3)$$

The hazards ratio corresponding to two different covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ is

$$\frac{\lambda(t|\mathbf{z}_1)}{\lambda(t|\mathbf{z}_2)} = \exp\left\{(\boldsymbol{\beta} + \boldsymbol{\gamma})'(\mathbf{z}_1 - \mathbf{z}_2) + \left[e^{\boldsymbol{\gamma}'\mathbf{z}_1} - e^{\boldsymbol{\gamma}'\mathbf{z}_2}\right] \log[\Lambda_0(t)]\right\}. \quad (1.4)$$

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 (1.2) can be applied. 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. An interesting property of this model is that the ratios of hazard to cumulative hazard functions corresponding to two covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are proportional. Using (1.2) and (1.3), it can be shown that

$$\frac{\lambda(t|\mathbf{z})}{\Lambda(t|\mathbf{z})} = \frac{\lambda_0(t)}{\Lambda_0(t)} \exp(\boldsymbol{\gamma}'\mathbf{z}). \quad (1.5)$$

Devarajan and Ebrahimi (2011) discussed the unique properties of this model and provided an interpretation of it. In addition, they demonstrated its relationship to the time-dependent coefficient Cox PH model. For small $\boldsymbol{\gamma}$, equation (1.2) reduces to

$$\Lambda(t|\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}} \{\Lambda_0(t)\}^{1+\boldsymbol{\gamma}'\mathbf{z}}. \quad (1.6)$$

For two different covariate vectors z_1 and z_2 , $\log \left\{ \frac{\Lambda(t|z_1)}{\Lambda(t|z_2)} \right\} = \boldsymbol{\eta}(t) (z_1 - z_2)$, where $\boldsymbol{\eta}(t) = [\boldsymbol{\beta} + \log\{\Lambda_0(t)\} \cdot \boldsymbol{\gamma}]'$. This is the Cox PH model with time-dependent coefficients $\boldsymbol{\eta}(t)$ that allows crossing hazard curves. Quantin et al. (1996) and Devarajan and Ebrahimi (2002) used model (1.2) for goodness of fit testing of the Cox PH model. Inference for model (1.2) has been discussed in Devarajan (2000), Devarajan and Ebrahimi (2011), Hsieh (2001) and Wu and Hsieh (2009).

In the following section, we develop a theoretical framework for estimation in this non-proportional hazards model based on penalized likelihood methods. We show that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are hyperbolic splines with knots at the distinct failure times. We conclude by briefly discussing the relationship of this approach to prior computational approaches for this model.

2 Maximum Penalized Likelihood Estimation for the Non-Proportional Hazards Regression Model

The partial likelihood approach of Cox (1972) is the standard inferential method for the Cox PH model. Due to the PH assumption, the baseline hazard $\lambda_0(t)$, drops out of this partial likelihood and can be left completely unspecified. On the other hand, it is evident from equations (1.3) and (1.4) that the generalized model (1.2) allows crossing 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 or explicitly estimated. For example, Bordes and Breuils (2006) discussed sequential estimation for semi-parametric models with application to the PH model while Wu and Hsieh (2009) adopted the method of sieves and used piecewise constants for estimating the baseline hazard in model (1.2). Devarajan (2000) and Devarajan and Ebrahimi (2011) harnessed the attractive computational properties of B -splines and used a set of cubic B -spline basis functions to model the baseline hazard. This spline-based approach can be viewed as a generalization of the piecewise constant approach. Here we develop a theoretical framework for estimating the baseline hazard, baseline cumulative hazard and their ratio using penalized likelihood methods. We show that the maximum penalized likelihood estimate of the baseline hazard and the ratio in model (1.2) are hyperbolic splines with knots at the distinct failure times. We briefly discuss the relationship of this approach to prior computational approaches for approximating the baseline hazard in this model.

The maximum penalized likelihood (MPL) approach applies the method of maximum likelihood to curve estimation. Often in survival analysis, one is interested in estimating the hazard function non-parametrically for a group of individuals in a study in the presence of censoring. In many situations, several other concomitant variables are also included in the study, thus giving rise to regression parameters associated with those variables that need to be estimated as well. First, we give a brief introduction to the concept of maximum penalized likelihood estimation in the context of hazard estimation.

The observed data consist of independent observations on the triple $(X, \delta, \mathbf{z})$, where X is the minimum of a failure (or survival) and censoring time pair (T, C) , $\delta = I(T < C)$ 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 failure and censoring times respectively which are assumed to be independent. We are interested in estimating the hazard function in the presence of censoring. For the sake of simplicity, let us assume that there are no covariates. The inclusion of covariates is discussed later. The log-likelihood of a random sample of size n of censored survival data can be expressed in terms of the hazard function as

$$l[\lambda(\cdot)] = \sum_{i=1}^n \left\{ \delta_i \log \lambda(t_i) - \int_0^{t_i} \lambda(u) du \right\} \quad (2.1)$$

where $\lambda(t)$ is the hazard function and δ is as defined above. The subscript refers to the i^{th} individual in the sample for $i = 1, \dots, n$. An application of the maximum likelihood method to the log-likelihood function given above results in a sum of Dirac delta functions with spikes at the uncensored observations as the hazard estimate. This is due to the fact that the log-likelihood $l[\cdot]$ as a functional of $\lambda(\cdot)$ is unbounded from above and hence cannot be maximized with respect to $\lambda(t)$. Good and Gaskins (1971) introduced the idea of non-parametric roughness penalties and showed how to impose a smoothness constraint on $l[\cdot]$ by maximizing a penalized version of the log-likelihood to obtain a bounded estimate. The penalized log-likelihood is defined as

$$\begin{aligned} J[\lambda(t)] &= l[\lambda(t)] - \alpha R[\lambda(t)] \\ &= \sum_{i=1}^n \left\{ \delta_i \log \lambda(t_i) - \int_0^{t_i} \lambda(u) du \right\} - \alpha R[\lambda(t)] \end{aligned} \quad (2.2)$$

where $\alpha > 0$ is the *smoothing parameter* that controls the amount of smoothing and $R[\lambda(t)]$ is the *roughness penalty*. If S denotes the class of all sufficiently smooth functions $\lambda(t)$ such that $R[\lambda(t)]$ is defined and for which $\lambda(t) > 0, \forall t > 0$, then the hazard estimate $\widehat{\lambda}(t)$ is a *maximum penalized likelihood estimate* if $\widehat{\lambda}(t) \in S$ and $J[\widehat{\lambda}(t)] \geq J[\lambda(t)], \forall \lambda(t) \in S$. That is, $\widehat{\lambda}(t)$ maximizes the penalized log-likelihood $J[\lambda(t)]$ over the class of all $\lambda(t)$ satisfying the conditions given above.

An excellent exposition of penalized likelihood methods is given in Tapia and Thompson (1978) and Silverman (1986). Cox and O'Sullivan (1990) discuss the asymptotic properties of penalized likelihood estimates. Penalized likelihood methods have been applied for hazard estimation by Gu (1996), O'Sullivan (1988), Anderson and Senthilselvan (1980), Bartoszynski et al. (1981), among others. The most general form of the penalty functional $R[\lambda(t)]$ is

$$R[\lambda(t)] = \int_0^\infty [\lambda^{(m)}(x)]^2 dx, \quad m > 0$$

where $\lambda^{(m)}(t)$ is the m^{th} derivative of $\lambda(t)$. The commonly used forms for $R[\lambda(t)]$ are

$$\int_0^\infty [\lambda^{(1)}(x)]^2 dx = \int_0^\infty [\lambda'(x)]^2 dx \text{ which penalizes the slope of the curve and}$$

$\int_0^\infty [\lambda^{(2)}(x)]^2 dx = \int_0^\infty [\lambda''(x)]^2 dx$ which penalizes the curvature. Sometimes, it is convenient to work with the logarithm or the square root of the hazard to avoid explicit positivity constraints on the hazard. The functional (2.2) is maximized with respect to $\lambda(t)$ over the class of absolutely continuous functions on $[0, \infty)$ whose m^{th} derivative is square integrable. As noted earlier, the solution corresponding to $\alpha = 0$ is the sum of Dirac delta function spikes at the uncensored observations. As $\alpha \rightarrow \infty$, the solution tends to an $(m - 1)^{th}$ degree polynomial. O'Sullivan (1988) notes that for $m = 2$, the log-hazard estimator is linear in t and has a Weibull form on the logarithmic scale.

Next, we describe estimation of the underlying baseline quantities of interest (such as the hazard, cumulative hazard and their ratio) in models (1.2) and (1.6) using MPL methods.

2.1 MPL Estimation of the Baseline Hazard

We obtain the MPL estimator (MPLE) of the baseline hazard function, $\lambda_0(t)$, in model (1.2) assuming that the regression parameters β and γ are known. In order to avoid explicit positivity constraints, we estimate the square root of the hazard function. That is, given β and γ , we obtain the MPLE of $g(t) = \sqrt{\lambda_0(t)}$ using calculus of variations (Reinsch (1967)).

Let $t_1 < t_2 < \dots < t_n$ denote the distinct, ordered failure and censoring times, and let $C_2(I)$ be the class of continuous functions on the interval I with continuous first derivative and piecewise continuous second derivative. Here $I \equiv [a, b] \subset [0, \infty)$ such that it contains the closed interval $[t_0, t_n]$ where $t_0 = 0$. Using (2.1), (1.2) and (1.3), the log-likelihood for this model is given by

$$\begin{aligned} l[\lambda(\cdot)] &= \sum_{i=1}^n \delta_i \log \lambda(t_i | z_i) - \int_0^{t_i} \lambda(y | z_i) dy \\ &= \sum_{i=1}^n \left\{ \delta_i \left[\log \lambda_0(t_i) + (\beta + \gamma)' z_i + (e^{\gamma' z_i} - 1) \cdot \log \Lambda_0(t_i) \right] \right\} - \quad (2.3) \\ &\quad \sum_{i=1}^n \left\{ [\Lambda_0(t_i)]^{\exp(\gamma' z_i)} e^{\beta' z_i} \right\} \end{aligned}$$

Using the log-likelihood (2.3) above and the penalty functional

$$R[\lambda(t)] = \alpha \int_I [g'(u)]^2 \quad (2.4)$$

where $\alpha > 0$ is the smoothing parameter, the penalized log-likelihood can be written in terms of $g(t) = \sqrt{\lambda_0(t)}$ as

$$l[g(t)] = \sum_{i=1}^n \left\{ 2\delta_i \log g(t_i) + \delta_i (\beta + \gamma)' z_i - e^{\beta' z_i} \left(\int_0^{t_i} g^2(u) du \right)^{\exp(\gamma' z_i)} + \delta_i (e^{\gamma' z_i} - 1) \log \left(\int_0^{t_i} g^2(u) du \right) \right\} - \left\{ \alpha \int_I [g'(u)]^2 du \right\}. \quad (2.5)$$

This penalized log-likelihood is maximized with respect to $g(t)$ to obtain the MPLE of $\lambda_0(t)$. Theorem 3.2 of Cox and O'Sullivan (1985, 1990) and the following theorem, re-stated from O'Sullivan (1988), guarantee existence of the MPLE of $g(t)$ in (2.5).

Theorem 1—If there exists a unique $(m-1)^{th}$ degree polynomial maximizing the likelihood part of (2.5), then the penalized likelihood in (2.5) has a unique maxima over the Sobolev space $W_2^m[a, b]$ of absolutely continuous functions whose m^{th} derivative is square integrable.

Prior to stating our result in Theorem 2 we provide the expression for a hyperbolic spline based on the definition given in Baum (1976), and then state a key result from Hewitt & Stromberg (1975) used in obtaining the MPLE.

Definition: A hyperbolic spline with knots at $t_1 < t_2 < \dots < t_n$ is any function f that can be expressed in the form

$$f(t) = a_i + b_i \cdot t + c_i \cdot e^{\psi_i(t-t_i)} + d_i \cdot e^{\phi_i(t-t_i)}, \quad t_i < t < t_{i+1}, \quad i=1, \dots, n-1,$$

for some real spline coefficients $a_i, b_i, c_i, d_i, \psi_i$ and $\phi_i, i=1, 2, \dots, n-1$.

Lemma 1: For continuously differentiable functions f and g , we have,

$$\int_a^b f(u) g(u) du = f(b-) g(b-) - f(a+) g(a+) - \int_a^b g(u) f'(u) du \quad (2.6)$$

Theorem 2—Given the penalized log-likelihood (2.5) and β and γ , the MPLE of

$g(t) = \sqrt{\lambda_0(t)}$ is a hyperbolic spline with knots at the distinct failure times where $g(t)$ is a member of $C_2(I)$ such that $\int_I [g'(u)]^2 du < \infty$.

Proof: Consider a small increment $\epsilon \cdot \eta(t)$, $\epsilon > 0$ to the function $g(t)$ in the class $C_2(I)$ of functions where $[t_0, t_n] \subset I$ with $t_0 = 0$. A necessary condition for the functional $l[\cdot]$ to have a maximum is that as $\epsilon \rightarrow 0$,

$$\frac{\partial}{\partial \epsilon} l[g(u) + \epsilon \cdot \eta(u)] = 0$$

Now, $\frac{\partial}{\partial \epsilon} l[g(u) + \epsilon \cdot \eta(u)] = 0$ gives

$$\frac{\partial}{\partial \epsilon} \left[\sum_{i=1}^n \left\{ 2\delta_i \log \{g(t_i) + \epsilon \cdot \eta(t_i)\} + \delta_i (\beta + \gamma)' z_i - e^{\beta' z_i} \left(\int_0^{t_i} [g(u) + \epsilon \cdot \eta(u)]^2 du \right)^{\exp(\gamma' z_i)} \right. \right. \\ \left. \left. + \delta_i (e^{\gamma' z_i} - 1) \log \left(\int_0^{t_i} [g(u) + \epsilon \cdot \eta(u)]^2 du \right) \right\} - \left\{ \alpha \int_I [g'(u) + \epsilon \cdot \eta'(u)]^2 du \right\} \right] = 0,$$

which implies that

$$0 = \sum_{i=1}^n \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} - \left\{ e^{(\beta + \gamma)' z_i} \right\} \left(\int_0^{t_i} g^2(u) du \right)^{\exp(\gamma' z_i) - 1} \left(\int_0^{t_i} 2g(u) \cdot \eta(u) du \right) \right. \\ \left. + \delta_i (e^{\gamma' z_i} - 1) \left(\frac{\int_0^{t_i} 2g(u) \cdot \eta(u) du}{\int_0^{t_i} g^2(u) du} \right) \right\} - \left\{ 2\alpha \int_I g'(u) \cdot \eta'(u) du \right\}. \quad (2.7)$$

The last term of equation (2.7) can be rewritten using Lemma 1 and integration by parts as follows,

$$\int_I g'(u) \cdot \eta'(u) du = \int_{t_0}^{t_n} g'(u) \cdot \eta'(u) du + \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du \\ = \sum_{i=1}^n \left\{ \int_{t_{i-1}}^{t_i} g'(u) \cdot \eta'(u) du \right\} + \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du. \quad (2.8)$$

Integrating the first term on the right hand side of equation (2.8) by parts and using Lemma 1, we get,

$$\int_I g'(u) \cdot \eta'(u) du = \sum_{i=1}^n \left[g'(t_i^-) - g'(t_i^+) \right] \eta(t_i) - g'(t_0^+) \eta(t_0) + g'(t_n^+) \eta(t_n) \\ - \sum_{i=1}^n \left\{ \int_{t_{i-1}}^{t_i} g''(u) \cdot \eta(u) du \right\} + \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du. \quad (2.9)$$

Using (2.9), (2.7) becomes

$$\sum_{i=1}^n \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} - \left\{ e^{(\beta + \gamma)' z_i} \right\} \left(\int_0^{t_i} g^2(u) du \right)^{\exp(\gamma' z_i) - 1} \left(\int_0^{t_i} 2g(u) \cdot \eta(u) du \right) \right. \\ \left. + \delta_i (e^{\gamma' z_i} - 1) \left(\frac{\int_0^{t_i} 2g(u) \cdot \eta(u) du}{\int_0^{t_i} g^2(u) du} \right) \right\} - 2\alpha \left\{ \sum_{i=1}^n \left[g'(t_i^-) - g'(t_i^+) \right] \eta(t_i) - g'(t_0^+) \eta(t_0) \right. \\ \left. + g'(t_n^+) \eta(t_n) - \sum_{i=1}^n \left\{ \int_{t_{i-1}}^{t_i} g''(u) \cdot \eta(u) du \right\} + \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du \right\} = 0.$$

Now, collecting terms with $\int a(u)\eta(u)du$, $\eta(u)$ etc. and simplifying, we have,

$$\sum_{i=1}^n \left\{ \int_{t_{i-1}}^{t_i} \left(\sum_{l=i}^n \left(\frac{-e^{(\beta+\gamma)' z_l} \left(\int_0^{t_l} g^2(u) du \right)^{\exp(\gamma' z_l)} + \delta_l (e^{\gamma' z_l} - 1)}{\int_0^{t_l} g^2(u) du} \right) 2g(u) + 2\alpha \cdot g''(u) \right) \eta(u) du \right\} \\ + \sum_{i=1}^n \left\{ \frac{2\delta_i}{g(t_i)} - 2\alpha \left[g'(t_i^-) - g'(t_i^+) \right] \right\} \eta(t_i) - 2\alpha g'(t_n^+) \eta(t_n) + 2\alpha g'(t_0^+) \eta(t_0) \\ - 2\alpha \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du = 0.$$

Since $\eta(\cdot)$ is an arbitrary function in $C_2(I)$, by suitable choice of $\eta(\cdot)$, we can show that each term in the above equation which is a coefficient of $\eta(\cdot)$ or $\eta'(\cdot)$ is identically zero. Hence, we obtain a set of second order ordinary linear differential equations given by

$$0 = \alpha g''(u) + g(u) \left\{ \sum_{l=i}^n \left(\frac{-e^{(\beta+\gamma)' z_l} \left(\int_0^{t_l} g^2(u) du \right)^{\exp(\gamma' z_l)} + \delta_l (e^{\gamma' z_l} - 1)}{\int_0^{t_l} g^2(u) du} \right) \right\}, i=1, 2, \dots, n, \quad (2.10)$$

and the following side conditions

$$\left. \begin{aligned} g'(t_i^-) - g'(t_i^+) &= \frac{\delta_i}{\alpha g(t_i)}, \quad i=1, \dots, n, \\ g'(t_n^+) &= g'(t_0^+) = 0, \\ g'(u) &= 0, \quad u \in I/[t_0, t_n]. \end{aligned} \right\} \quad (2.11)$$

The solution to the differential equation

$$g''(u) - \frac{1}{\alpha} \left\{ \sum_{l=i}^n \left(\frac{e^{(\beta+\gamma)' z_l} \left(\int_0^{t_l} g^2(u) du \right)^{\exp(\gamma' z_l)} - \delta_l (e^{\gamma' z_l} - 1)}{\int_0^{t_l} g^2(u) du} \right) \right\} g(u) = 0 \quad (2.12)$$

is given by

$$g(t) = a_i \cdot \exp(\sqrt{\kappa_i}(t - t_i)) + b_i \cdot \exp(-\sqrt{\kappa_i}(t - t_i)), \quad t_{i-1} < t < t_i, \quad i=1, \dots, n, \quad (2.13)$$

where a_i and b_i are arbitrary constants, κ_i is the coefficient of $g(u)$ in (2.12) and $g(t)$ satisfies the conditions stated in (2.11) (Boyce and DiPrima (1969)). Hence $g(t) = \sqrt{\lambda_0(t)}$ is a hyperbolic spline with knots at the distinct failure times satisfying the conditions in (2.11). From (2.13), it is straightforward to see that,

$$g^2(t) = \lambda_0(t) = 2a_i b_i + a_i^2 \exp(2\sqrt{\kappa_i}(t - t_i)) + b_i^2 \exp(-2\sqrt{\kappa_i}(t - t_i)), \quad t_{i-1} < t < t_i \quad i=1, \dots, n,$$

is also a hyperbolic spline with knots at the distinct failure times. $\square$

Furthermore, it can be easily shown that $\Lambda_0(t) = \int_0^t \lambda_0(x) dx$ is also a hyperbolic spline with knots at the distinct failure times and is given by

$$\Lambda_0(t) = \frac{a_i^2 e^{-2t_i \sqrt{\kappa_i}}}{2 \sqrt{\kappa_i}} (e^{2t \sqrt{\kappa_i}} - 1) + \frac{b_i^2 e^{2t_i \sqrt{\kappa_i}}}{2 \sqrt{\kappa_i}} (1 - e^{-2t \sqrt{\kappa_i}}) + 2a_i b_i t, \quad t_{i-1} < t < t_i, \quad i=1, \dots, n.$$

2.2 MPL Estimation of the Ratio $\frac{\lambda_0(t)}{\Lambda_0(t)}$

For two covariate vectors z_1 and z_2 , the ratios of the hazard to cumulative hazard functions are proportional not only in model (1.2) but also in the time-dependent coefficient Cox

model (1.6). We utilize this property by working with the baseline ratio $\frac{\lambda_0(t)}{\Lambda_0(t)}$ and develop an approach to estimation in these models. In order to avoid explicit positivity constraints,

we work with the square root of this ratio. The function $g(t) = \sqrt{\frac{\lambda_0(t)}{\Lambda_0(t)}}$ is estimated non-parametrically assuming that the regression parameters β and γ are known. That is, given β and γ , we obtain the MPLE of $g(t)$ using similar methods as in §2.1.

Using the log-likelihood (2.3) and penalty functional (2.4), the penalized log-likelihood can

be written in terms of $g(t) = \sqrt{\frac{\lambda_0(t)}{\Lambda_0(t)}}$ as

$$l[g(t)] = \sum_{i=1}^n \left\{ 2\delta_i \log g(t_i) + \delta_i (\beta + \gamma)' z_i - e^{\beta' z_i} \exp \left[e^{\gamma' z_i} G(t_i) \right] + \delta_i e^{\gamma' z_i} G(t_i) \right\} - \left\{ \alpha \int_I [g'(u)]^2 du \right\} \quad (2.14)$$

where $G(t) = \int_0^t g^2(u) du$. This quantity is maximized with respect to the $g(t)$. As noted earlier, Theorem 3.2 of Cox and O'Sullivan (1985, 1990) and Theorem 1 guarantee existence of the MPLE of $g(t)$ in (2.14). We now state our result in the following theorem.

Theorem 3—Given the penalized log-likelihood (2.14) and β and γ , the MPLE of $g(t) =$

$\sqrt{\frac{\lambda_0(t)}{\Lambda_0(t)}}$ is a hyperbolic spline with knots at the distinct failure times where $g(t)$ is a member of $C_2(I)$ such that $\int_I [g'(u)]^2 < \infty$.

Proof: Consider a small increment $\epsilon \cdot \eta(t)$, $\epsilon > 0$ to the function $g(t)$ in the class $C_2(I)$ of functions where $[t_0, t_n] \subset I$ with $t_0 = 0$. A necessary condition for the functional $l[\cdot]$ to have a maximum is that as $\epsilon \rightarrow 0$,

$$\frac{\partial}{\partial \epsilon} l[g(u) + \epsilon \cdot \eta(u)] = 0$$

Now, $\frac{\partial}{\partial \epsilon} l[g(u) + \epsilon \cdot \eta(u)] = 0$ gives

$$\begin{aligned} & \frac{\partial}{\partial \epsilon} \left[\sum_{i=1}^n \left\{ 2\delta_i \log[g(t_i) + \epsilon \cdot \eta(t_i)] + \delta_i (\beta + \gamma)' z_i \right. \right. \\ & \quad \left. \left. - e^{\beta' z_i} \exp \left\{ e^{\gamma' z_i} \int_0^{t_i} [g(u) + \epsilon \eta(u)]^2 du \right\} + \delta_i e^{\gamma' z_i} \int_0^{t_i} [g(u) + \epsilon \cdot \eta(u)]^2 du \right\} \right. \\ & \quad \left. - \left\{ \alpha \int_I [g'(u) + \epsilon \cdot \eta'(u)]^2 du \right\} \right] = 0, \end{aligned}$$

which implies that

$$\begin{aligned} & \sum_{i=1}^n \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} + 2\delta_i e^{\gamma' z_i} \left(\int_0^{t_i} g(u) \eta(u) du \right) - 2e^{(\beta + \gamma)' z_i} \left(\int_0^{t_i} g(u) \eta(u) du \right) \exp \left(e^{\gamma' z_i} \int_0^{t_i} g^2(u) du \right) \right\} \\ & \quad - \left\{ 2\alpha \int_I g'(u) \cdot \eta'(u) du \right\} = 0. \end{aligned}$$

Using Lemma 1 and following the steps outlined in the proof of Theorem 2 (equation (2.8)), the last term of the above equation can be re-written as specified in (2.9). Hence the above equation becomes

$$\begin{aligned} & \sum_{i=1}^n \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} + 2\delta_i e^{\gamma' z_i} \left(\int_0^{t_i} g(u) \eta(u) du \right) \right. \\ & \quad \left. - 2e^{(\beta + \gamma)' z_i} \left(\int_0^{t_i} g(u) \eta(u) du \right) \exp \left(e^{\gamma' z_i} \int_0^{t_i} g^2(u) du \right) \right\} + 2\alpha \sum_{i=1}^n \int_{t_{i-1}}^{t_i} g''(u) \cdot \eta(u) du \\ & - 2\alpha \sum_{i=1}^n \left[g'(t_i^-) - g'(t_i^+) \right] \eta(t_i) + 2\alpha g'(t_0^+) \eta(t_0) - 2\alpha g'(t_n^+) \eta(t_n) - 2\alpha \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du = 0. \end{aligned}$$

Now, collecting terms with $\int a(u)\eta(u)du$, $\eta(u)$ etc. and simplifying, we have,

$$\begin{aligned} & \sum_{i=1}^n \left\{ \int_{t_{i-1}}^{t_i} \left[\sum_{l=i}^n \left(2\delta_l e^{\gamma' z_l} - 2e^{(\beta + \gamma)' z_l} \exp \left(e^{\gamma' z_l} \int_0^{t_l} g^2(u) du \right) \right) \right] g(u) + 2\alpha g''(u) \right\} \eta(u) du \\ & \quad + \sum_{i=1}^n \left\{ \frac{2\delta_i}{g(t_i)} - 2\alpha [g'(t_i^-) - g'(t_i^+)] \right\} \eta(t_i) \\ & \quad + 2\alpha g'(t_0^+) \eta(t_0) - 2\alpha g'(t_n^+) \eta(t_n) - 2\alpha \int_{I/[t_0, t_n]} g'(u) \cdot \eta'(u) du = 0. \end{aligned}$$

Since, $\eta(\cdot)$ is an arbitrary function in $C_2(I)$, by suitable choice of $\eta(\cdot)$, we can show that each term in the above equation which is a coefficient of $\eta(\cdot)$ or $\eta'(\cdot)$ is identically zero. Hence, we obtain a set of second order ordinary linear differential equations given by

$$\alpha \cdot g''(u) + \left\{ \sum_{l=i}^n \left(\delta_l e^{\gamma' z_l} - e^{(\beta + \gamma)' z_l} \exp \left(e^{\gamma' z_l} \int_0^{t_l} g^2(u) du \right) \right) \right\} g(u) = 0, i=1, 2, \dots, n,$$

and the following side conditions

$$\left. \begin{aligned} g'(t_i^-) - g'(t_i^+) &= \frac{\delta_i}{\alpha g(t_i)}, \quad i=1, \dots, n, \\ g'(t_n^+) &= g'(t_0^+) = 0, \\ g'(u) &= 0, \quad u \in I/[t_0, t_n]. \end{aligned} \right\} \quad (2.15)$$

The solution to the differential equation

$$g''(u) - \frac{1}{\alpha} \left\{ \sum_{l=i}^n \left(-\delta_l e^{\gamma' z_l + e^{(\beta+\gamma)' z_l}} \exp \left(e^{\gamma' z_l} \int_0^{t_l} g^2(u) du \right) \right) \right\} g(u) = 0 \quad (2.16)$$

is given by

$$g(t) = c_i \cdot \exp(\sqrt{\tau_i}(t - t_i)) + d_i \exp(-\sqrt{\tau_i}(t - t_i)), \quad t_{i-1} < t < t_i, \quad i=1, \dots, n, \quad (2.17)$$

where c_i and d_i are arbitrary constants, τ_i is the coefficient of $g(u)$ in (2.16) and $g(t)$ satisfies

the conditions stated in (2.15) (Boyce and DiPrima (1969)). Hence $g(t) = \sqrt{\frac{\lambda_0(t)}{\Lambda_0(t)}}$ is a hyperbolic spline with knots at the distinct failure times satisfying the conditions in (2.15).

From (2.17), it is straightforward to see that,

$$g^2(t) = \frac{\lambda_0(t)}{\Lambda_0(t)} = 2c_i d_i + c_i^2 \exp(2\sqrt{\tau_i}(t - t_i)) + d_i^2 \exp(-2\sqrt{\tau_i}(t - t_i)), \quad t_{i-1} < t < t_i, \quad i=1, \dots, n,$$

is also a hyperbolic spline with knots at the distinct failure times. $\square$

Remark 1: Using a linear approximation to the third term in the log-likelihood (2.14) and

penalizing for both slope and curvature of $g(t) = \sqrt{\frac{\lambda_0(t)}{\Lambda_0(t)}}$ (i.e., using $R[g(x)] = \alpha_1 \int_I [g'(x)]^2 + \alpha_2 \int_I [g''(x)]^2$), it can be shown that the MPLE of $g(t)$ is a hyperbolic spline with knots at the distinct failure times where $\int_I [g'(u)]^2 du < \infty$ and $\int_I [g''(u)]^2 du < \infty$ (based on arguments similar to that used in the proof of Theorem 3).

Remark 2: Using the log-likelihood (2.14) and penalizing for both slope and curvature of

$g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$, it can be shown that the MPLE of $g(t)$ is also a hyperbolic spline with knots at the distinct failure times where $\int_I [g'(u)]^2 du < \infty$ and $\int_I [g''(u)]^2 du < \infty$ (based on arguments similar to that used in the proof of Theorem 3). However, explicit positivity constraints have to be imposed on some of the spline coefficients in order to ensure positivity of $g(t)$.

3 Summary and Discussion

In summary, we developed a theoretical framework for maximum penalized likelihood estimation for a semi-parametric generalization of the Cox PH model. We showed that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are

hyperbolic splines with knots at the distinct failure times. From a computational perspective, models based on splines are intermediate in structure between completely parametric models and non-parametric methods, and provide flexibility in incorporating a variety of shapes. It is well-known that any spline function of a given order and pre-specified knot sequence can be approximated by a linear combination of B -splines for that knot sequence. Moreover, an interesting feature of the B -spline basis function is that it has the properties of a probability density function. So a linear combination of B -spline basis functions is like a mixture of densities that allows for various distributions (de Boor (2001)). A full likelihood approach based on a linear combination of B -splines for the baseline hazard has already been outlined in our earlier work (Devarajan and Ebrahimi (2011)) and would aid in the implementation of the proposed method. Simultaneous estimation of the regression parameters β and γ , and the baseline hazard can be achieved via an iterative scheme. The testing and comparison of the performance of this approach to existing methods of estimation for this generalized model using real data are especially important and could form the core of a future, larger investigation.

Acknowledgments

The authors would like to thank the Associate Editor and the reviewer for their helpful comments which significantly improved the presentation of this paper. The work of the first author was supported in part by NIH grant P30 CA 06927.

References

- [1]. Anderson JA, Senthilselvan A. Smooth estimates for the hazard function. *Journal of the Royal Statistical Society - Series B*. 1980; 42:322–327.
- [2]. Bartoszynski R, Brown BW, McBride CM, Thompson JR. Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. *Annals of Statistics*. 1981; 9:1050–1060.
- [3]. Baum AM. An algebraic approach to simple hyperbolic splines on the real line. *Journal of Approximation Theory*. 1976; 17:189–199.
- [4]. de Boor, C. *A Practical Guide to Splines*. Springer-Verlag; New York: 2001.
- [5]. 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.
- [6]. Boyce, W.; DiPrima, RC. *Elementary Differential Equations and Boundary Value Problems*. Second Edition. John Wiley & Sons, Inc.; New York: 1969.
- [7]. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society - Series B*. 1972; 34:187–202.
- [8]. Cox, DD.; O’Sullivan, F. Analysis of penalized likelihood type estimators with application to generalized smoothing in Sobolev spaces. Statistics Department, University of California; Berkeley, California: 1985. Technical Report #51
- [9]. Cox DD, O’Sullivan F. Asymptotic analysis of penalized likelihood and related estimators. *Annals of Statistics*. 1990; 18:1676–1695.
- [10]. Devarajan, K. Ph.D dissertation. Division of Statistics, Northern Illinois University; 2000. Inference for a non-proportional hazards regression model and applications.
- [11]. 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.
- [12]. Devarajan K, Ebrahimi N. A semi-parametric generalization of the Cox proportional hazards regression model: Inference and Applications. *Computational Statistics and Data Analysis*. 2011; 55:667–676. [PubMed: 21076652]

- [13]. Good IJ, Gaskins RA. Non-parametric roughness penalties for probability densities. *Biometrika*. 1971; 58:255–277.
- [14]. Gu C. Penalized likelihood hazard estimation: A general procedure. *Statistica Sinica*. 1996; 6:861–876.
- [15]. Hewitt, A.; Stromberg, P. *Real and Abstract Analysis*. Springer; New York: 1975.
- [16]. Hsieh F. On heteroscedastic hazards regression models: Theory and Application. *Journal of the Royal Statistical Society - Series B*. 2001; 63:63–80.
- [17]. O’Sullivan F. Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal of Scientific Statistical Computing*. 1988; 9:363–379.
- [18]. Quantin C, Moureau T, Asselain B, Maccario J, Lellouch J. A regression model for testing the proportional hazards hypotheses. *Biometrics*. 1996; 5:874–885. [PubMed: 8924576]
- [19]. Reinsch CH. Smoothing by spline functions. *Numerische Mathematik*. 1967; 10:177–183.
- [20]. Silverman, BW. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall; London: 1986.
- [21]. Tapia, RA.; Thompson, JR. *Nonparametric Probability Density Estimation*. Johns Hopkins University Press; Baltimore, Maryland: 1978.
- [22]. Wu H-DI, Hsieh F. Heterogeneity and varying effect in hazards regression. *Journal of Statistical Planning and Inference*. 2009; 139:4213–4222.