An Adaptive Approach for Hazard Regression Modeling

George J. Knafl

School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, USA
Email: gknafl@unc.edu

Abstract

Regression models for survival time data involve estimation of the hazard rate as a function of predictor variables and associated slope parameters. An adaptive approach is formulated for such hazard regression modeling. The hazard rate is modeled using fractional polynomials, that is, linear combinations of products of power transforms of time together with other available predictors. These fractional polynomial models are restricted to generating positive-valued hazard rates and decreasing survival times. Exponentially distributed survival times are a special case. Parameters are estimated using maximum likelihood estimation allowing for right censored survival times. Models are evaluated and compared using likelihood cross-validation (LCV) scores. LCV scores and tolerance parameters are used to control an adaptive search through alternative fractional polynomial hazard rate models to identify effective models for the underlying survival time data. These methods are demonstrated using two different survival time data sets including survival times for lung cancer patients and for multiple myeloma patients. For the lung cancer data, the hazard rate depends distinctly on time. However, controlling for cell type provides a distinct improvement while the hazard rate depends only on cell type and no longer on time. Furthermore, Cox regression is unable to identify a cell type effect. For the multiple myeloma data, the hazard rate also depends distinctly on time. Moreover, consideration of hemoglobin at diagnosis provides a distinct improvement, the hazard rate still depends distinctly on time, and hemoglobin distinctly moderates the effect of time on the hazard rate. These results indicate that adaptive hazard rate modeling can provide unique insights into survival time data.

Keywords

Adaptive Regression, Fractional Polynomials, Hazard Rate, Likelihood Cross-Validation, Survival Times

1. Introduction

Cox regression, also called proportional hazard regression, is the commonly used approach for modeling survival time or time-to-event data [1]. The hazard function is assumed to equal a product of an unspecified function of time and the exponent of a linear function of available predictors with estimation addressing only the slope parameters for the predictors while ignoring the underlying dependence of the hazard function on time. An important extension would be an approach that estimates the dependence on time as well as on the predictors. Kooperberg, Stone, and Truong [2] provide such an extension that generalizes standard proportional hazard regression by using splines to estimate the dependence on time. More recently, Qian and Peng [3] have proposed using quantile regression as an alternative for survival time modeling while Bouaziz and Nuel [4] have proposed the use of models based on piecewise constant hazard rates. See [5] for an overview of hazard rate modeling.

The objectives of this article are to formulate an adaptive approach for hazard regression modeling based on fractional polynomials [6] [7] and then to demonstrate this approach using two survival time data sets, one for lung cancer patients and one for multiple myeloma patients. This is a novel, original approach accounting for the dependence of the hazard rate on time while allowing for non-proportionality of the hazard rate as well as for nonlinearity in time and in other available predictors.

2. Methods

2.1. Survival Times

Let $T > 0$ denote a continuous random survival time variable with conditional cumulative distribution function $F(t;z) = P(T \le t \mid \mathbf{Z} = z)$ for time values $t > 0$ where $\mathbf{Z}$ is a random $r \times 1$ vector with vector value z consisting of predictor variables Z_i with values z_i for $1 \le i \le r$. T has conditional density function

$$f(t;z) = \frac{\mathrm{d}F(t;z)}{\mathrm{d}t},$$

conditional survival function

$$S(t;z) = 1 - F(t;z) = P(T > t \mid \mathbf{Z} = z),$$

and conditional hazard function

$$\lambda(t;z) = \lim_{t \downarrow 0} \frac{P(t < T \le t + \Delta t \mid T > t, \mathbf{Z} = z)}{\Delta t} = \frac{f(t;z)}{S(t;z)}.$$

Integration gives

$$\Lambda(t;z) = \int_0^t \lambda(w;z) \cdot \mathrm{d}w = -\log(S(t;z))$$

so that $S(t;z) = \exp(-\Lambda(t;z))$ and $f(t;z) = \lambda(t;z) \cdot \exp(-\Lambda(t;z))$. The exponential distribution generates the special case with the hazard function constant

in time t so that $\lambda(t;z) = \lambda(z)$ and $S(t;z) = \exp(-\lambda(z) \cdot t)$.

2.2. Fractional Polynomial Hazard Rate Modeling

Let u be a primary predictor and q a real-valued power. As in [7], define the associated general power transform $g(u;q)$ as

$$g(u;q) = \begin{cases} \cos(\pi \cdot q) \cdot |u|^q, & u < 0 \\ 0, & u = 0 \\ u^q, & u > 0 \end{cases}$$

where π is the usual constant. Note that when $u = t$, only the case $u > 0$ can hold. Predictors z_i, for $1 \le i \le r$, are typically nonnegative, and so assume this holds so that the case $u < 0$ is not needed and $g(u;q)$ is always nonnegative. Indicator predictors u with values 0 or 1 are untransformed (*i.e.*, $q = 1$).

Let the hazard function $\lambda(t;z)$ be a general fractional polynomial in the primary predictors t and the coordinates z_i of z, that is, a linear combination of products of power transforms of those primary predictors. Specifically, let $x(t;z)$ be a $p \times 1$ vector of power transforms $x_j(t;z) = g(u_j;q_j)$ for $1 \le j \le p$ where u_j is either t, z_i for $1 \le i \le r$, or a geometric combination, that is, a product of power transforms of some subset of t and z_i for $1 \le i \le r$. Note that geometric combinations generalize standard interactions and so provide for a nonlinear assessment of the concept of moderation [8].

The fractional polynomial hazard rate is defined as

$$\lambda(t;z) = x^{\mathrm{T}}(t;z) \cdot \boldsymbol{\beta}$$

where x^{T} denotes the transpose of x while $\boldsymbol{\beta}$ is a $p \times 1$ vector of slope parameters β_j for $1 \le j \le p$. An intercept parameter can be included in the model using the unit transform as a primary predictor. This definition requires that $\lambda(t;z)$ be constrained to take on positive values as addressed later. This restriction could be resolved by modeling instead the natural log of $\lambda(t;z)$ as a linear function in $\boldsymbol{\beta}$ (as in [2] and [9]), but then the integral $\Lambda(t;z)$ cannot always be expressed in closed form. The advantage of the above definition is that power transforms $g(u_j;q_j)$ are simply integrated with respect to t, which can also speed up computation of estimates for the parameter vector $\boldsymbol{\beta}$. The possible forms for $x_j(t;z)$ are a power transform of t including the unit transform $g(t;0) \equiv 1$, a power transform of z_i for $1 \le i \le r$, a power-transformed geometric combination in t together with one or more z_i for $1 \le i \le r$, and a power-transformed geometric combination in two or more z_i for $1 \le i \le r$ not including t. Consequently, $x_j(t;z)$ can be expressed as the product of a function of t and a function of z, that is,

$$x_j(t;z) = x_{j,1}(t) \cdot x_{j,2}(z) = g(t;q_j) \cdot x_{j,2}(z)$$

where $x_{j,2}(z)$ is either the unit transform or depends on some subset of z_i for $1 \le i \le r$ but not on t. When for all j $q_j = 0$, none of the $x_j(t;z) = x_j(z)$ depend on t, and so the fractional polynomial model is the same as the model

generated by the exponential distribution with $\lambda(t;z) = \lambda(z) = \boldsymbol{x}^{\mathrm{T}}(z) \cdot \boldsymbol{\beta}$. Also, $x_{j,2}(z)$ equals the unit transform when $x_j(t;z) = x_{j,1}(t)$ depends only on t. The power q_j is either the power for a single power transform of t or equals a product of a power q' of t combined with a second power q'' transforming a geometric combination containing the power q' of t. In general, $x_{j,2}(z)$ can contain power transforms of z_i for $1 \leq i \leq r$.

Integration gives

$$\Lambda(t;z) = \boldsymbol{h}^{\mathrm{T}}(t;z) \cdot \boldsymbol{\beta}$$

where $\boldsymbol{h}(t;z)$ is the $p \times 1$ vector with entries

$$h_j(t;z) = \int_0^t x_j(w;z) \cdot \mathrm{d}w = \left(\int_0^t g(w;q_j) \cdot \mathrm{d}w \right) \cdot x_{j,2}(z).$$

Assume that $q_j > -1$, so that

$$h_j(t;z) = \frac{g(t;q_j+1)}{q_j+1} \cdot x_{j,2}(z)$$

with $\dfrac{g(t;q_j+1)}{q_j+1}$ a positive function of time t. The second term $x_{j,2}(z)$ is nonnegative, but not identically 0, for example, when z consists of a single indicator predictor z. By definition, $S(t;z)$ is a decreasing function of t for each unique possible value of the vector z, and so $\Lambda(t;z)$ must be an increasing function of t for functions of z determined by the fractional polynomial model.

Let $J = \{j : 1 \leq j \leq p\}$, $J_1 = \{j : 1 \leq j \leq p, q_j \neq 0\}$, and $J_2 = J \setminus J_1$. Partition $\boldsymbol{x}(t;z)$ into $\boldsymbol{x}_1(t;z)$ with entries $x_j(t;z)$ for $j \in J_1$ and associated slope parameter vector $\boldsymbol{\beta}_1$ as well as $\boldsymbol{x}_2(z)$ with entries $x_j(z)$ for $j \in J_2$ depending only on z and not on t with associated slope parameter vector $\boldsymbol{\beta}_2$. The hazard function $\lambda(t;z)$ then satisfies

$$\lambda(t;z) = \boldsymbol{x}^{\mathrm{T}}(t;z) \cdot \boldsymbol{\beta} = \boldsymbol{x}_1^{\mathrm{T}}(t;z) \cdot \boldsymbol{\beta}_1 + \boldsymbol{x}_2^{\mathrm{T}}(z) \cdot \boldsymbol{\beta}_2$$

so that

$$\lambda(t;z) = \sum_{j \in J_1} g(t;q_j) \cdot x_{j,2}(z) \cdot \beta_j + \sum_{j \in J_2} x_{j,2}(z) \cdot \beta_j.$$

Assume for now that J_2 is empty. For models based on a single transform of t,

$$\lambda(t;z) = \lambda(t) = g(t;q_1) \cdot \beta_1$$

can be guaranteed to be positive by restricting the associated slope β_1 to be positive-valued. For models based on multiple transforms of t but not on z,

$$\lambda(t;z) = \lambda(t) = \sum_{j \in J_1} g(t;q_j) \cdot \beta_j$$

can generate complicated integrals $\Lambda(t;z) = \Lambda(t)$ that are not increasing in t in general. Restricting all the associated slopes β_j for $j \in J_1$ to be positive-valued is a straight forward way to guarantee that $\Lambda(t)$ is increasing in t. More general models with

$$\sum_{j \in J_1} g(t; q_j) \cdot x_{j,2}(z) \cdot \beta_j$$

generate integrals $\Lambda(t; z)$ that need to be increasing in each unique choice for the terms $x_{j,2}(z)$. This can be guaranteed in a straight forward way once again by restricting all the associated slopes β_j for $j \in J_1$ to be positive-valued.

On the other hand, when J_2 is nonempty, the assumption of all positive-valued slope parameters would be too restrictive. For example, consider a model based on two primary predictors: t and an indicator variable z for membership in one of two groups of study participants with additive effects to t and z. The term

$$x_1^{\mathrm{T}}(t; z) \cdot \boldsymbol{\beta}_1 = x_1^{\mathrm{T}}(t) \cdot \boldsymbol{\beta}_1$$

models the dependence of the hazard function on t in terms of $p-1$ parameters while $x_2^{\mathrm{T}}(z) \cdot \boldsymbol{\beta}_2 = z \cdot \beta_p$ provides for a shift in the hazard function for the group coded as a 1 compared to the hazard function for the group coded as a 0. In general, this shift can be in the negative or the positive direction, and so the slope of $\boldsymbol{\beta}_2$ needs to be allowed to be negative. However, this means that $x^{\mathrm{T}}(t; z) \cdot \boldsymbol{\beta}$ and $h^{\mathrm{T}}(t; z) \cdot \boldsymbol{\beta}$ can sometimes take on negative or zero values. See Section 2.3 for how to handle such cases.

Hazard ratios are naturally generated for Cox regression models based on proportional hazard rates. Hazard ratios are not needed to assess a fractional polynomial model but can be generated if desired. For example, consider the simple fractional polynomial model $\lambda(t; z)$ based on time t and a single indicator z for one of two groups. The associated hazard ratio is given by $\lambda(t; 1) / \lambda(t; 0)$, that is, the hazard rate for $z = 1$ divided by the hazard rate for $z = 0$. Note that in general this is a function of time t.

2.3. Maximum Likelihood Estimation

Observed survival times can be of two types: an actual survival time or a right censored survival time with the actual survival time only known to be larger than the observed right censored time. Thus, the observations have the form $O_s = \{t_s, c_s, z_s\}$ for $s \in S = \{s : 1 \le s \le n\}$ where $t_s > 0$ is a time value, c_s is an indicator for whether t_s is right censored (*i.e.*, $c_s = 1$) or an actual, uncensored survival time (*i.e.*, $c_s = 0$), and z_s is a possibly trivial $r \times 1$ predictor vector with values $z_{s,i}$ for $1 \le i \le r$. Note that all observations can be uncensored, but not all observations should be censored.

Let $x(t_s; z_s)$ be the associated observed vectors determined by some fractional polynomial model. For each $s \in S$, the likelihood term $L(O_s; \boldsymbol{\beta})$ for observation O_s satisfies

$$L(O_s; \boldsymbol{\beta}) = \begin{cases} f(t_s; z_s), & c_s = 0 \\ S(t_s; z_s), & c_s = 1 \end{cases}$$

so that the associated log-likelihood term satisfies

$$\ell(O_s; \boldsymbol{\beta}) = \log(L(O_s; \boldsymbol{\beta})) = (1 - c_s) \cdot \log(\lambda(t_s; z_s)) - \Lambda(t_s; z_s).$$

The likelihood $L(S;\boldsymbol{\beta})$ equals the product over $s \in S$ of the likelihood terms $L(O_s;\boldsymbol{\beta})$ with associated log-likelihood

$$\ell(S;\boldsymbol{\beta}) = \log\left(L(S;\boldsymbol{\beta})\right) = \sum_{s \in S} \ell(O_s;\boldsymbol{\beta}).$$

The maximum likelihood estimate $\boldsymbol{\beta}(S)$ of $\boldsymbol{\beta}$ maximizes the log-likelihood $\ell(S;\boldsymbol{\beta})$ over $\boldsymbol{\beta}$, which is achieved by solving the estimating equations

$$\frac{\partial \ell(S;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0.$$

For simplicity of notation, parameter estimates $\boldsymbol{\beta}(S)$ are denoted as functions of the index set S for the observed data used in their computation without hat (^) symbols.

The gradient vector has entries satisfying

$$\frac{\partial \ell(S;\boldsymbol{\beta})}{\partial \beta_j} = \sum_{s \in S} \frac{\partial \ell(O_s;\boldsymbol{\beta})}{\partial \beta_j} = \sum_{s \in S}\left((1-c_s)\cdot\frac{x_j(t_s;z_s)}{\boldsymbol{x}^{\mathrm{T}}(t_s;z_s)\cdot\boldsymbol{\beta}} - h_j(t_s;z_s)\right)$$

for $j \in J$. The Hessian matrix has entries satisfying

$$\frac{\partial^2 \ell(S;\boldsymbol{\beta})}{\partial \beta_{j'}\partial \beta_j} = \sum_{s \in S} \frac{\partial^2 \ell(O_s;\boldsymbol{\beta})}{\partial \beta_{j'}\partial \beta_j} = -\sum_{s \in S}\left((1-c_s)\cdot\frac{x_{j'}(t_s;z_s)\cdot x_j(t_s;z_s)}{\left(\boldsymbol{x}^{\mathrm{T}}(t_s;z_s)\cdot\boldsymbol{\beta}\right)^2}\right)$$

for $j, j' \in J$. The gradient vector and Hessian matrix can be used in a Newton-Raphson algorithm to compute the maximum likelihood estimate $\boldsymbol{\beta}(S)$.

When $\boldsymbol{x}^{\mathrm{T}}(t_s;z_s)\cdot\boldsymbol{\beta} \leq 0$ or $\boldsymbol{h}^{\mathrm{T}}(t_s;z_s)\cdot\boldsymbol{\beta} \leq 0$ for some s, set $\lambda(t_s;z_s) = \delta$ for some small positive value δ like 0.00001. Also, set $\Lambda(t_s;z_s) = \delta \cdot t$. However, leave the gradient vector and Hessian matrix unchanged so that maximum likelihood estimation uses actual derivatives with respect to the parameter vector $\boldsymbol{\beta}$.

The estimated covariance matrix for the parameter estimate vector $\boldsymbol{\beta}(S)$ equals the matrix with entries $-\left(\frac{\partial^2 \ell(S;\boldsymbol{\beta}(S))}{\partial \beta_{j'}\partial \beta_j}\right)^{-1}$. Variances for the model parameters are given by the diagonal entries of this covariance matrix. These can be used to compute z tests for zero individual model parameters for assessing specific theoretically important models. However, tests for parameters of adaptively generated models are typically significant as a consequence of the adaptive modeling process (as summarized in Section 2.5). Consequently, results for such tests are not reported for models generated in example analyses.

Models based on a single transform with $p = 1$ can be computed directly. In these cases, $\boldsymbol{\beta} = (\beta_1)$ and the gradient is a scalar value satisfying

$$\frac{\partial \ell(S;\boldsymbol{\beta})}{\partial \beta_1} = \sum_{s \in S}\left(\frac{1-c_s}{\beta_1} - h_1(t_s;z_s)\right)$$

so that

$$\beta_1(S) = \frac{n - \sum_{s \in S} c_s}{\sum_{s \in S} h_1(t_s;z_s)}.$$

2.4. Likelihood Cross-Validation

Partition the index set S into $k > 1$ disjoint sets $S(h)$, called folds, for $h \in H = \{h : 1 \le h \le k\}$. The LCV score for a given model is defined as

$$\text{LCV} = \prod_{h \in H} \left(L\left(S(h); \boldsymbol{\beta}\left(S \setminus S(h) \right) \right) \right)^{1/n}.$$

In other words, evaluate the likelihood for each fold $S(h)$ using the parameter vector $\boldsymbol{\beta}\left(S \setminus S(h) \right)$ computed by maximizing the likelihood for the complement $S \setminus S(h)$ of the fold, normalize it by the sample size n, and multiply these normalized deleted fold likelihoods together to get the LCV score.

When one or more slopes $\beta_j \le 0$ for some $j \in J_1$ are generated as part of the estimation computations, reset $\ell(S; \boldsymbol{\beta}) = -\Delta$ for some large value Δ such as 700 unless $\ell(S; \boldsymbol{\beta})$ is already smaller than $-\Delta$. For a model with parameter vector estimate $\boldsymbol{\beta}(S)$ having one or more slopes $\beta_j(S) \le 0$ for some $j \in J_1$, reset the LCV score to a very small value δ' such as 10^{-12}. These adjustments guarantee that the adaptive modeling process of Section 2.5 generates models with acceptable parameter vector estimates $\boldsymbol{\beta}(S)$.

A larger LCV score indicates a better model, but not necessarily a distinctly (substantially, significantly) better model. A χ^2-based LCV ratio test, analogous to a likelihood ratio test, can be used to decide if there is a distinct improvement or not in the LCV score. Following [7], these tests are expressed in terms of the percent decrease (PD) in the LCV score for the model with the smaller score compared to the model with the larger score and a cutoff for a distinct PD, changing with the sample size. A PD larger than the cutoff indicates that the model with the larger score distinctly (substantially, significantly) improves on the model with the smaller score. Otherwise, the model with the smaller score is a competitive alternative and, if also simpler, it is a parsimonious, competitive alternative and so a preferable choice. Examples of LCV ratio tests are provided in Section 3.

2.5. Adaptive Model Selection

An effective choice of a fractional polynomial hazard rate model based on a subset of the primary predictors t and the coordinates of z, is identified adaptively using a heuristic search process controlled by tolerance parameters indicating how much of a change in the LCV score is allowable at each stage of the process. Starting from a base hazard rate model, which is usually the constant model, power transforms and also geometric combinations if requested are systematically added to the model. Then, the expanded model is contracted, removing extraneous transforms if any and adjusting the powers of the remaining transforms. LCV ratio tests are used to decide whether to stop the contraction or continue removing transforms. Only a brief overview of the adaptive modeling process is provided here; details are provided in [10]. Adaptive modeling applies to both independent and correlated outcomes in a variety of regression contexts such as linear regression [7], logistic regression [11], Poisson regression [12]

[13], and discrete regression [14].

Power transforms of general time variables T can cause computational problems for the adaptive modeling process when some time values are large. For this reason, fractional polynomial hazard rate models are computed using the normalized survival time variable T' with observed values $t'_s = t_s / t^*$ where

$$t^* = \max \left\{ t_s : s \in S \right\}.$$

This generates estimates $\lambda'(t';z)$ and $S'(t';z)$ of the hazard rate and survival function, respectively. Then, define the estimated hazard rate and survival function for the original survival time variable T as

$$\lambda(t;z) = \lambda'(t/t^*;z) \quad \text{and} \quad S(t;z) = S'(t/t^*;z)$$

for any t and z. This adjustment guarantees that power transforms are computed using bounded time values within the interval (0, 1). It is not needed when $t^* \leq 1$ to start with, but that is unlikely to hold.

2.6. Modeling Constraints

A variety of constraints are needed so that a fractional polynomial provides an appropriate hazard rate model. Specifically, predictors z_i for $1 \leq i \leq r$ are restricted to be nonnegative. Power transforms $g(t;q)$ have powers restricted to satisfy $q > -1$. Estimates of hazard rates $\lambda(t;z)$ and their integrals $\Lambda(t;z)$ are adjusted to be positive-valued (as addressed in Section 2.3). Slope parameters β_j for $j \in J_1$ corresponding to predictors depending on time t are restricted to be positive-valued (as addressed in Section 2.4). Also, fractional polynomial hazard rate models are computed using the normalized survival times (as addressed in Section 2.5).

3. Example Analyses

Example analyses are presented in this section demonstrating the use and applicability of adaptive hazard regression modeling based on factional polynomials. Two survival time data sets are used in these analyses including survival times for lung cancer patients in Section 3.1 and for multiple myeloma patients in Section 3.2.

All analyses are conducted using SAS® Version 9.4. A SAS macro for conducting adaptive modeling in a variety of regression contexts including hazard regression is available on request from the author.

3.1. Example Analyses of Survival Times for Lung Cancer Patients

Data are provided in [15] consisting of $n = 137$ survival times for lung cancer patients, ranging from 1 to 999 days with 6 (6.6%) right censored. The two largest survival times of 991 and 999 days are much larger than the next largest survival time of 587 days and so have the potential for being highly influential. Note that since $t^* = 999$ is of the order of 10^3, even moderate-sized powers can gen-

erate very large transformed time values. For example, the power $q = 4$ generates some transformed time values of the order of 10^{12}, indicating the need for analyzing normalized time values. There are several available predictors other than time, but the example analyses only address cancer cell type with 27 (19.7%), 27 (19.7%), 48 (35.0%), and 35 (25.6%) patients having adeno, large, small, and squamous cell types, respectively.

The cutoff for a distinct percent decrease (PD) in the LCV score for data with $n = 137$ observations is 1.39%. All analyses use $k = 5$ folds with fold sizes ranging from 19 (13.9%) to 32 (23.4%) patients so that both fold sizes and fold complement sizes are not proportionately sparse.

The adaptively generated hazard rate model in time alone is based on the single time transform $t'^{-0.16}$ without an intercept and with LCV score 2.68420. In contrast, the constant hazard rate model has LCV score 2.61568 with distinct PD 2.55% (*i.e.*, greater than the cutoff of 1.39% for a distinct PD). Consequently, the hazard rate is distinctly nonconstant in time. Figure 1 provides the plot for the estimated hazard rate as a function of time. The hazard rate decreases nonlinearly from 15.1 at 1 day to 5.0 at 999 days.

The impact of cell type on the hazard rate is assessed using the adaptive model based on time, indicators for each of the four cell types, and geometric combinations in these five primary predictors. The generated model is based on indicators for having the squamous and large cell types with an intercept and no transforms of time. The LCV score is 2.92430. In contrast, the LCV score of 2.68420 for the model based on time alone generates a distinct PD of 8.21%. Consequently, consideration of cell type provides a distinct improvement.

Furthermore, the hazard rate is reasonably considered to no longer depend on
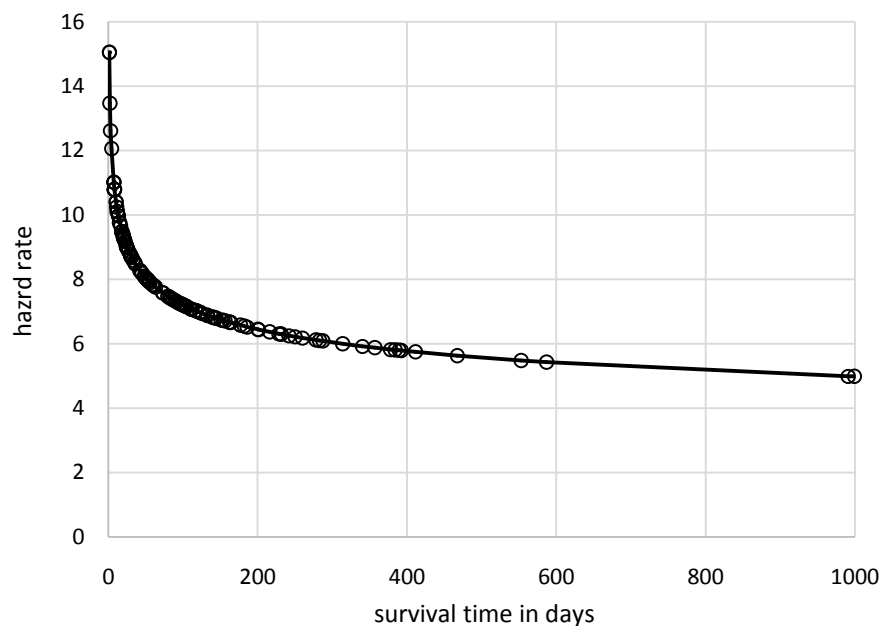

Figure 1. The estimated hazard rate as a function of time for lung cancer patients.

time after controlling for cell type and so supports the use of the exponential distribution for modeling lung cancer survival times. Under the cell type model, the estimated hazard rate is 13.7 for patients with adeno or small cell types, decreases to 5.8 for patients with large cell type, and then decreases further to 4.4 for patients with squamous cell type. **Figure 2** provides the plot of the estimated survival probability as a function of time. The estimated probability of survival decreases over time with the value at any given time smallest (worst) for patients with adeno or small cell type, larger for patient with large cell type, and largest for patients with squamous cell type.

The standard Cox regression model for survival as a one-way analysis of variance factor based on cell type has non-significant F test with $p = 0.57$. Under the reduced model based on effects to large and squamous cell types as generated adaptively, both effects are nonsignificant ($p = 0.17$ for large and $p = 0.71$ for squamous cell types). These results indicate that Cox regression can in some cases be unable to identify effects on survival identifiable with adaptive hazard regression modeling and in this case using the special case based on the exponential distribution.

3.2. Example Analyses of Survival Times for Multiple Myeloma Patients

Data are provided in [16] of $n = 65$ survival times for multiple myeloma patients, ranging from 1.25 to 92 days with 17 (26.2%) right censored. There are several available predictors other than time, but the example analyses only address hemoglobin at diagnosis. Hemoglobin values range from 4.9 to 14.6 g/dL.
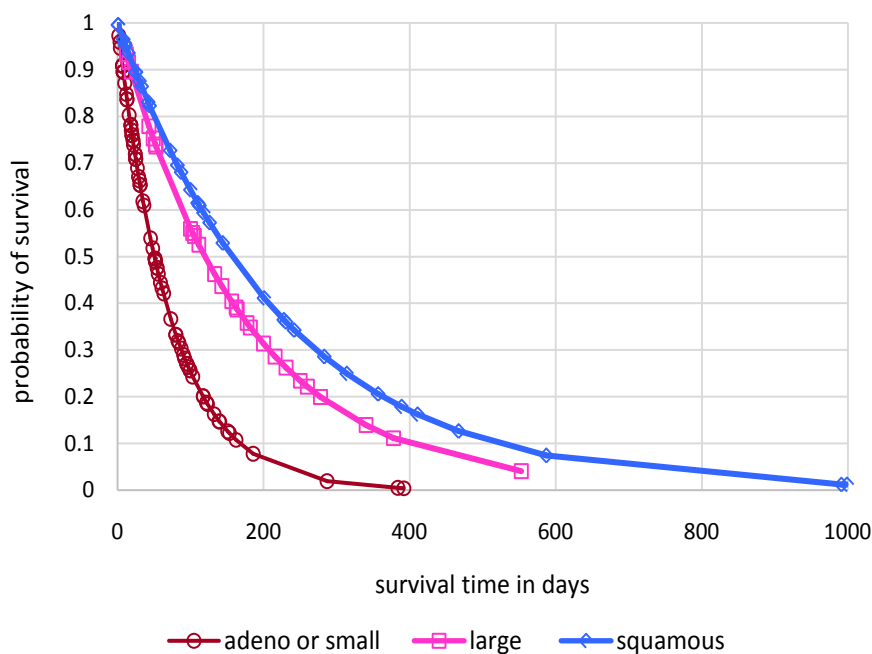
Figure 2. The estimated probability of survival as a function of time by cell type for lung cancer patients.

The cutoff for a distinct percent decrease (PD) in the LCV score for data with $n = 65$ observations is 2.91%. All analyses use $k = 5$ folds with fold sizes ranging from 8 (12.3%) to 16 (24.6%) patients so that both fold sizes and fold complement sizes are not proportionately sparse.
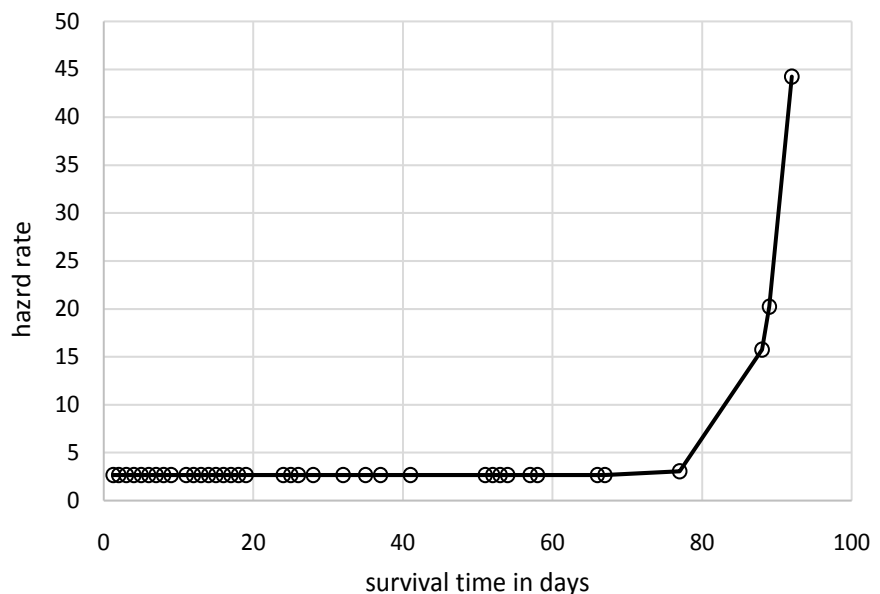
The adaptively generated hazard rate model in time alone is based on the single time transform t'^{26} with an intercept and with LCV score 1.07551. In contrast, the constant hazard rate model has LCV score 1.02430 with distinct PD 4.76% (*i.e.*, greater than the cutoff of 2.91% for a distinct PD). Consequently, the hazard rate is distinctly nonconstant in time. **Figure 3** provides the plot for the estimated hazard rate as a function of time. The hazard rate is essentially constant at 2.7 for 1.25 to 67 days and then increases very quickly to 44.2 at 92 days.

The adaptively generated model additive in time and hemoglobin (*i.e.*, not considering geometric combinations) is based on t'^{26} and hemoglobin$^{-0.9}$ without an intercept. The LCV score is 1.11557. In contrast, the LCV score of 1.07551 for the time alone model generates a distinct PD 3.59%. Consequently, consideration of hemoglobin provides a distinct improvement. Furthermore, the adaptively generated model in time, hemoglobin, and geometric combinations is based on hemoglobin$^{-0.9}$ and

$$\left(t'^{26} \cdot \text{hemoglobin}^3 \right)^{3.4} = t^{88.4} \cdot \text{hemoglobin}^{10.2}$$

without an intercept. The LCV score is 1.14286. In contrast, the LCV score for the additive model generates a distinct PD 2.39%. Consequently, hemoglobin distinctly moderates (or modifies) the effect of time on the hazard rate.

Interpretation of the model based on hemoglobin can be complicated. A more readily interpretable adaptive model is the one with the continuous hemoglobin predictor replaced by the ordinal hemoglobin quintile predictor (with values 1 - 5

Figure 3. The estimated hazard rate as a function of time for multiple myeloma patients.

corresponding to consecutive quintiles). The adaptively generated moderation model using hemoglobin quintile (HG_quintile) instead of hemoglobin is based on HG_quintile$^{2.5}$,

$$\left(t'^{26} \cdot \text{HG\_quintile}^{-7}\right)^{3.2} = t^{83.2} \cdot \text{HG\_quintile}^{-22.4},$$

$$\left(t'^{11} \cdot \text{HG\_quintile}\right)^{6} = t^{66} \cdot \text{HG\_quintile}^{6},$$

and an intercept. The LCV score is 1.21511. In contrast, the LCV score of 1.14286 for the moderation model in unadjusted hemoglobin generates a distinct PD 5.95%. Consequently, consideration of hemoglobin quintiles not only provides a more interpretable model but also provides a distinct improvement.

Table 1 contains ranges for exact quintiles, each consisting of 13 (20%) patients, as well as estimated hazard rates within these quintiles. For larger quintiles, the estimated hazard rate is a decreasing value, constant in time except for the largest time value in four of the five quintiles (all but quintile 2). Figure 4 contains the plot of the survival function by quintile. The estimated probability of survival decreases over time with the value at any given time increasing (better) for patients within quintiles 1 - 5, but more so for quintiles 4 - 5.

Figure 5 contains the plot of the hazard ratio for quintile 5 compared to quintile 1 as a function of time. The hazard ratio is constant at 0.31 up to about day 58 then decreases quickly to 0 by about day 68 and stays there to day 92. However, patients in quintile 1 have observed survival times of at most 66 days (Table 1), and so hazard rates based on quintile 1 after that day and associated hazard ratios may be of questionable value.
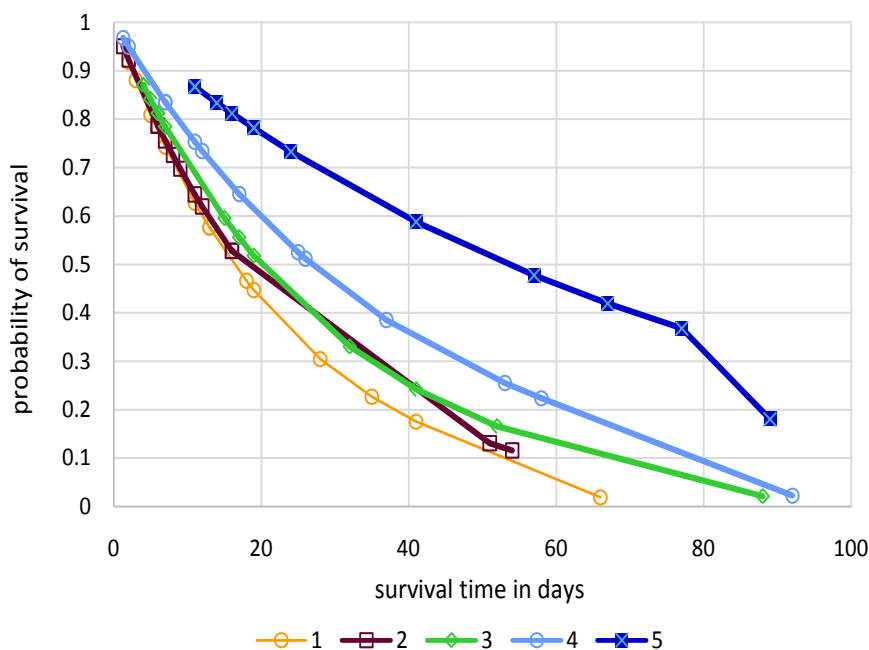

Figure 4. The estimated probability of survival as a function of time by quintiles 1 - 5 for hemoglobin at diagnosis for multiple myeloma patients.

Table 1. Estimated hazard rates over time by hemoglobin quintiles for multiple myeloma patients.

HG quinitile[1]	HG range (g/dL)	Time (days)	Hazard rate	Time (days)	Hazard rate
1	4.9 - 7.5	2 - 41	3.9	66	137.5
2	7.7 - 9.9	1.25 - 54	3.7		
3	10.0 - 10.8	4 - 52	3.2	88	72.1
4	11.0 - 12.4	1.25 - 58	2.4	92	96.8
5	12.5 - 14.6	11 - 77	1.2	89	39.7

HG-hemoglobin; [1]All quintiles contain exactly 13 (20%) of the 65 patients.
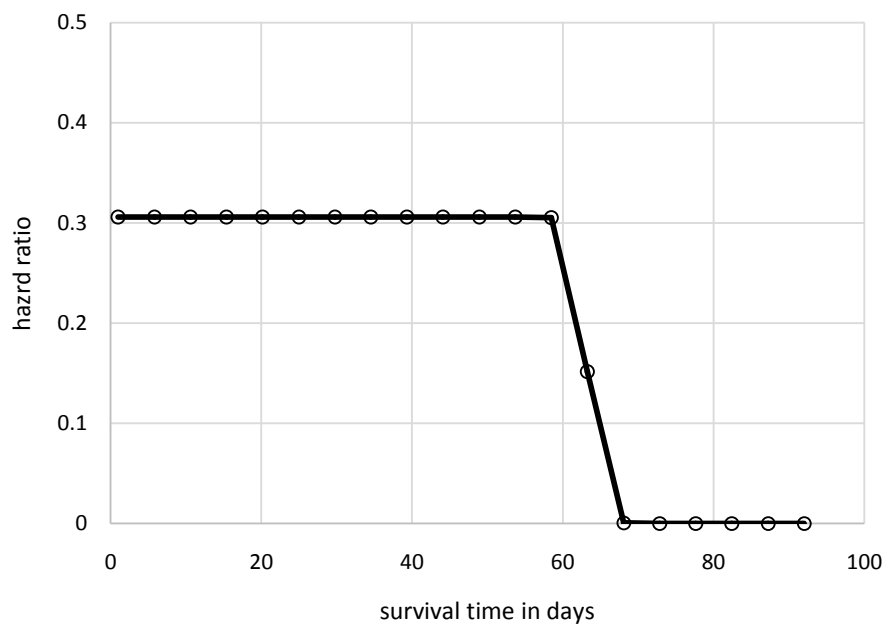
Figure 5. The estimated hazard ratio as a function of time for the fifth quintile of hemoglobin at diagnosis compared to the first quintile for multiple myeloma patients.

4. Discussion

An adaptive approach is formulated for hazard regression modeling. Hazard rates are based on fractional polynomials, that is, linear combinations of products of powers of time and other available predictors. Models are restricted so that hazard rate estimates are positive-valued and so that estimated survival functions are decreasing in time for appropriate functions of the non-time predictors. Models are estimated using maximum likelihood estimation accounting for the possibility of right censored survival times.

Models are compared and evaluated using k-fold likelihood cross-validation (LCV) scores, that is, normalized products of deleted likelihoods for k subsets of the data, called folds, computed using parameters estimated using the complements of those folds. Larger LCV scores indicate better models, but not necessarily distinctly better models. Distinct differences in LCV scores are assessed using LCV ratio tests generalizing standard likelihood ratio tests expressed in

terms of a cutoff for a distinct percent decrease in the LCV score.

Effective models of a specific type (e.g., time alone, additive, or moderation) for analyzing a given survival time data set are generated using adaptive searches through alternate models of that specific type. The search first expands the model in terms of power transforms of time and other available predictors. The expansion can optionally generate geometric combinations of available predictors, that is, products of powers of multiple predictors generalizing standard interactions and providing for a nonlinear extension to standard linear moderation. The expanded model is then contracted by removing extraneous transforms and adjusting the powers for the remaining transforms to improve the LCV score. An LCV ratio test is used to decide whether to stop the contraction. This guarantees that the generated model is effective in the sense that the removal of each of its transforms generates a distinct percent decrease in the LCV score.

Adaptive hazard rate modeling using fractional polynomials is demonstrated using two different survival time data sets. The first data set consists of survival times for patients with lung cancer of differing cell types (adeno, large, small, or squamous). The second data set consists of survival times for multiple myeloma patients with varying values for hemoglobin at diagnosis.

For the lung cancer patients, the hazard rate depends distinctly nonlinearly on time (Figure 1). However, cell type provides a distinct improvement. The hazard rate no longer depends on time after accounting for cell type and so supports the use of the exponentially distributed survival times for lung cancer patients. The estimated hazard rate is a decreasing constant for patients having lung cancer with the following three sets of cell types: adeno or small, large, and squamous. The associated estimated survival function decreases over time but more slowly as cell type changes from adeno or small to large and then to squamous (Figure 2). Moreover, Cox regression is unable to identify a cell type effect.

For the multiple myeloma data, the hazard rate also depends distinctly on time (Figure 3). Consideration of hemoglobin at diagnosis provides a distinct improvement with the hazard rate still depending distinctly on time so that the exponential distribution is not an appropriate choice for modeling these data. Furthermore, hemoglobin at diagnosis distinctly moderates the effect of time on the hazard rate so that an assumption of proportional hazards is not applicable for these data. Consideration of quintiles for hemoglobin at diagnosis provides for a readily interpretable effect of hemoglobin on the hazard rate (Table 1) and on the associated survival function (Figure 4). While the hazard ratio need not be considered to understand adaptive hazard regression models, it can be generated if desired and can be a function of time. For example, Figure 5 displays the hazard ratio for the fifth hemoglobin quintile versus the first quintile as a function of time.

Adaptive hazard rate modeling is not restricted to proportional hazard rates as for Cox regression. It also accounts for the dependence of the hazard rate on time. In contrast to the approach of [2], the dependence on time is based on fractional polynomials rather than on splines and so generates smoother nonli-

near curves. Moreover, it models the hazard rate directly rather than modeling the log of the hazard rate so that the integral of the hazard rate can be computed in closed form resulting in reduced times to compute parameter estimates. In contrast to [3], it addresses the full distribution of survival times as opposed to selected quantiles. In contrast to [4], hazard rate models based on fractional polynomials are more extensive than models based on piecewise constant hazard rates. Also, consideration of geometric combinations accounts for varying dependence of the hazard rate on time for differing sets of predictor values.

In summary, adaptive hazard regression modeling has been formulated to generalize exponential distribution modeling to account for nonlinearity of the hazard rate as well as additive and moderation effects due to other available predictors. It has also been demonstrated using analyses of survival times for lung cancer patients and for multiple myeloma patients. Results of these analyses indicate that adaptive hazard rate modeling can provide unique insights into survival time data. For example, it is possible to assess whether there is a distinct dependence of the hazard rate on time (as held for both the lung cancer data and the multiple myeloma data), whether consideration of another available predictor can provide distinct additive effects (as for lung cancer cell types), and whether another available predictor can moderate the effect of time on the hazard rate (as for hemoglobin at diagnosis for multiple myeloma patients). However, adaptive hazard rate modeling is limited to modeling independent times to the occurrence of a single event, for example, times to death. Future research is needed to address modeling of correlated times to recurrent events, for example, times to multiple hospitalizations for cancer patients. An extension is also needed to address discrete survival times.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Cox, D.R. (1975) Partial Likelihood. *Biometrika*, **62**, 269-276.
https://doi.org/10.1093/biomet/62.2.269

[2] Kooperberg, C., Stone, C.J. and Truong, Y.K. (1995) Hazard Regression. *Journal of the American Statistical Association*, **90**, 78-94.
https://doi.org/10.1080/01621459.1995.10476491

[3] Qian, J. and Peng, L. (2010) Censored Quantile Regression with Partially Functional

Effects. *Biometrika*, **97**, 839-850. https://doi.org/10.1093/biomet/asq050

[4] Bouaziz, O. and Nuel, G. (2017) L0 Regularization for the Estimation of Piecewise Constant Hazard Rates in Survival Analysis. *Applied Mathematics*, **8**, 377-394. https://doi.org/10.4236/am.2017.83031

[5] Turkson, A.J. (2022) Perspectives on Hazard Rate Functions: Concepts; Properties; Theories; Methods; Computations; and Application to Real-Life Data. *Open Access Library Journal*, **9**, e8275. https://doi.org/10.4236/oalib.1108275

[6] Royston, P. and Altman, D.G. (1994) Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, **43**, 429-467. https://doi.org/10.2307/2986270

[7] Knafl, G.J. (2018) Adaptive Fractional Polynomial Modeling. *Open Journal of Statistics*, **8**, 159-186. https://doi.org/10.4236/ojs.2018.81011

[8] Baron, R.M. and Kenny, D.A. (1986) The Moderator-Mediator Variable Distinction in Social Psychology Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality & Social Psychology*, **51**, 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

[9] Royston, P. and Sauerbrei, W. (2008) Multivariable Model-Building: A Practical Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. John Wiley & Sons, Hoboken. https://doi.org/10.1002/9780470770771

[10] Knafl, G.J. and Ding, K. (2016) Adaptive Regression for Modeling Nonlinear Relationships. Springer International Publishing, Switzerland.

[11] Knafl, G.J. (2018) A Reassessment of Birth Defects for Children of Participants of the Air Force Health Study. *Open Journal of Epidemiology*, **8**, 187-200. https://doi.org/10.4236/ojepi.2018.84015

[12] Knafl, G.J. and Meghani, S.H. (2021) Modeling Individual Patient Count/Rate Data over Time with Applications to Cancer Pain Flares and Cancer Pain Medication Usage. *Open Journal of Statistics*, **11**, 633-654. https://doi.org/10.4236/ojs.2021.115038

[13] Knafl, G.J. (2022) Adaptive Regression for Nonlinear Interrupted Time Series Analyses with Application to Birth Defects in Children of Vietnam War Veterans. *Open Journal of Statistics*, **12**, 789-809. https://doi.org/10.4236/ojs.2022.126045

[14] Knafl, G.J. and Meghani, S.H. (2021) Regression Modeling of Individual-Patient Correlated Discrete Outcomes with Applications to Cancer Pain Ratings. *Open Journal of Statistics*, **12**, 456-485. https://doi.org/10.4236/ojs.2022.124029

[15] Kalbfleisch, J.D. and Prentice, R.L. (1980) The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York.

[16] Krull, J.M., Uthoff, V.A. and Harley, J.B. (1975) A Step-Up Procedure for Selecting Variables Associated with Survival. *Biometrics*, **31**, 49-57. https://doi.org/10.2307/2529709