ISSN 1440-771X
ISBN 0 7326 1080 X

Mixed Model-Based Hazard Estimation

T. Cai, Rob J. Hyndman and M.P. Wand

Working Paper 11/2000

December 2000

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

MONASH UNIVERSITY

AUSTRALIA

Mixed model-based hazard estimation

BY T. CAI
Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

R.J. HYNDMAN
Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia.

AND M. P. WAND
Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.

21st November, 2000

ABSTRACT

We propose a new method for estimation of the hazard function from a set of censored failure time data, with a view to extending the general approach to more complicated models. The approach is based on a mixed model representation of penalized spline hazard estimators. One payoff is the automation of the smoothing parameter choice through restricted maximum likelihood. Another is the option to use standard mixed model software for automatic hazard estimation.

Key words: Non-parametric regression; Restricted maximum likelihood; Variance component; Survival analysis.

1 Introduction

The hazard function is prominent in the field of survival analysis and is useful in many other contexts, such as reliability and actuarial science. While common survival models, in particular the Cox proportional hazards model (Cox, 1972), do not require explicit estimation of the hazard function, there are numerous situations where a good hazard estimate is useful. For example, proportional hazard models in the presence of interval censoring benefit from hazard function estimation (e.g. Betensky *et al.* 2000).

Nonparametric hazard estimation has an established literature, with the proposal of several kernel-based estimators (e.g. Tanner and Wong, 1983; Hjort, 1993) and spline-based estimators (e.g. Bloxom, 1985; Etezadi-Amoli and Ciampi, 1987; Senthilselvan, 1987; Rosenberg, 1995; Joly, Commenges and Letenneur, 1998; Eilers, 2000; O'Sullivan, 1988; Kooperberg et al, 1995). In this paper we take a mixed model approach to spline estimation of the hazard function. Operationally, our estimate is equivalent to a penalized spline fit with a quadratic penalty on the knot coefficients (e.g. Eilers and Marx, 1996; Eilers, 2000). However, the mixed model approach has the following advantages:

(1) a data-driven rule for choosing the amount of smoothing is easily formulated using maximum likelihood.

(2) the penalized spline hazard estimate can be approximated by a Poisson mixed model, with an offset. This allows hazard function estimation to be done using standard software such as the `SAS` macro `GLIMMIX`.

(3) it allows for easier extension to more complex models and censoring types. Examples include additive models, geostatistical models, hazard regression and interval censoring.

The mixed model/penalized spline approach to hazard estimation is described in Section 2. In Section 3 we formulate an automatic smoothing parameter rule based on restricted maximum likelihood. Section 4 describes a Poisson mixed model approximation, Section 5 describes standard error estimation and Section 6 demonstrates practical efficacy. We conclude with some discussion of possible extensions in Section 7.

2 Mixed model hazard estimation

Suppose we observe data (T_i, δ_i), $1 \leq i \leq n$, where T_i is the time to an event, and δ_i is an indicator of non-censoring. Let $\lambda(t)$ be the hazard function of the T_i and

$\eta = \ln\,\lambda$. Then the log-likelihood of the data i

$$\ell = \sum_{i=1}^{n} \left\{ \delta_i \eta(T_i) - \int_0^{T_i} e^{\eta(u)}\,du \right\}. \tag{1}$$

A *linear spline model* for η is

$$\eta(t) = \beta_0 + \beta_1\,t + \sum_{k=1}^{K} b_k (t - \kappa_k)_+ \tag{2}$$

where $x_+ \equiv \max(0, x)$ and corresponds to η being a piecewise linear function with knots at $\kappa_1, \ldots, \kappa_K$. The knots should be relatively dense to allow for detailed structure in η to be estimated. Our implementation chooses the knots to be equally spaced with respect to the quantiles of the unique T_i values and sets $K = \min(\lfloor n/4 \rfloor, 35)$. The answers are quite insensitive to the placement of the knots and this choice represents a trade-off between computational complexity and the ability to handle fine detail.

If the b_k are treated as ordinary parameters and estimated via maximization of (1) then the resulting estimate of η will be a somewhat wiggly piecewise linear function. A remedy is to treat them as *random effects*:

$$b_1, \ldots, b_K \sim N(0, \sigma_b^2).$$

The amount of smoothing is controlled by σ_b^2 and its reciprocal acts as a smoothing parameter.

Let $\boldsymbol{\beta} = [\beta_0\ \beta_1]^\mathsf{T}$, $\mathbf{b} = [b_1, \ldots, b_K]^\mathsf{T}$,

$$\mathbf{X} = [1\ \ T_i]_{1 \le i \le n}, \quad \text{and} \quad \mathbf{Z} = [\ (T_i - \kappa_k)_+]_{\substack{1 \le i \le n \\ 1 \le k \le K}}.$$

Then define

$$\Lambda(\boldsymbol{\beta}, \mathbf{b}) = e^{\beta_0}\{M(\beta_1, \mathbf{b}) + N(\beta_1, \mathbf{b})\}$$

where

$$M(\beta_1, \mathbf{b}) = \sum_{i=1}^{n} \int_0^{\kappa_{k_i^*-1}} e^{\beta_1 u + \sum_{k=1}^{K} b_k (u - \kappa_k)_+} du\,,$$

$$N(\beta_1, \mathbf{b}) = \sum_{i=1}^{n} \int_{\kappa_{k_i^*-1}}^{T_i} e^{\beta_1 u + \sum_{k=1}^{K} b_k (u - \kappa_k)_+} du\,,$$

$\kappa_0 = 0$ and, for each $1 \le i \le n$,

$$k_i^* = \text{smallest } 1 \le k \le K \text{ such that } T_i \le \kappa_k\,.$$

The log-likelihood is the

$$\ell(\boldsymbol{\beta},\sigma_b) = \log \int \exp\{\boldsymbol{\delta}^{\mathsf{T}}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta},\mathbf{b}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^{\mathsf{T}}\mathbf{b}\}\, d\mathbf{b} - K\log(\sigma_b). \quad (3)$$

The right-hand side of (3) involves an intractable K-dimensional integral. A common approach to handling this integral is Laplace approximation (e.g. Breslow and Clayton, 1993) which, when applied to (3), results i

$$\ell(\boldsymbol{\beta},\sigma_b) \simeq \boldsymbol{\delta}^{\mathsf{T}}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\widehat{\mathbf{b}}) - \Lambda(\boldsymbol{\beta},\widehat{\mathbf{b}}) - \tfrac{1}{2\sigma_b^2}\widehat{\mathbf{b}}^{\mathsf{T}}\widehat{\mathbf{b}}$$

where

$$\widehat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmax}}\left\{\boldsymbol{\delta}^{\mathsf{T}}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta},\mathbf{b}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^{\mathsf{T}}\mathbf{b}\right\}.$$

For fixed $\sigma_b^2 = 1/\tau$ this mixed model approach with Laplace approximation is equivalent to the penalized spline fit

$$\widehat{\boldsymbol{\eta}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\mathbf{b}}$$

with

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\mathbf{b}}\end{bmatrix} = \underset{\boldsymbol{\beta},\mathbf{b}}{\operatorname{argmax}}\left\{\boldsymbol{\delta}^{\mathsf{T}}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta},\mathbf{b}) - \tfrac{1}{2}\tau\mathbf{b}^{\mathsf{T}}\mathbf{b}\right\} \quad (4)$$

and has similarities with the B-spline estimator of Eilers (2000). However, as we mentioned in the introduction, the mixed model framework has some compelling advantages: it has a natural automatic smoothing parameter choice (Section 3) and, with some modification, can be implemented using standard software (Section 4).

The final log-hazard is a piecewise linear function. However, with a dense set of knots the final curve estimate will be, visually, quite smooth. Higher degree splines will give a mathematically smoother result, but linear splines have the advantage of admitting exact expressions for $\Lambda(\boldsymbol{\beta},\mathbf{b})$. Computing formulae are given in the Appendix.

3 Choice of amount of smoothing

The reciprocal of σ_b^2 acts as a smoothing parameter, and its choice has a profound influence on the fit. Therefore it is important to have the option of having the data choose the amount of smoothing.

An obvious solution is to replace σ_b^2 by its maximum likelihood estimate. However, restricted maximum likelihood (REML) is slightly more attractive for variance component estimation. REML is well-defined for the Gaussian mixed model (see e.g.

Searle, Casella and McCulloch, 1992) but is less clear-cut for non-Gaussian models. One way around this is to maximize the marginal likelihood, defined by

$$\mathcal{L}_{\text{marg}}(\sigma_b) = \int \mathcal{L}(\boldsymbol{\beta}, \sigma_b) d\boldsymbol{\beta}$$

where $\mathcal{L}(\boldsymbol{\beta}, \sigma_b)$ is the anti-logarithm of (3). The marginal log-likelihood is then

$$\begin{aligned}\ell_{\text{marg}}(\sigma_b) &= \log \int\int \exp\{\boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta}, \mathbf{b}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b}\}\, d\mathbf{b}\, d\boldsymbol{\beta} - K\log(\sigma_b) \\ &= \log \int\int \exp\{\Xi(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)\}\, d\mathbf{b}\, d\boldsymbol{\beta} - K\log(\sigma_b)\end{aligned}$$

where

$$\Xi(\boldsymbol{\beta}, \mathbf{b}, \sigma_b) = \boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta}, \mathbf{b}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b}.$$

We apply Laplace's method to approximate $\ell_{\text{marg}}(\sigma_b)$. Let Ξ' and Ξ'' denote the $(K+2)$ vector and $(K+2)\times(K+2)$ dimensional matrix of first- and second-order partial derivatives of Ξ with respect to $(\boldsymbol{\beta}, \mathbf{b})$. The approximation yields

$$\ell_{\text{marg}}(\sigma_b) = -K\log\sigma_b + \Xi(\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b), \sigma_b) + \tfrac{1}{2}\log|\Xi''(\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b), \sigma_b)|$$

where $\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b)$ denotes the solution to $\Xi'(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}; \sigma_b) = \mathbf{0}$. This is analogous to the Penalized Quasi-Likelihood (PQL) approach of Breslow and Clayton (1993).

In the Appendix we give exact, readily computable, formulae for $\Xi(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ and its first two derivatives with respect to $(\boldsymbol{\beta}, \mathbf{b})$. This allows straightforward estimation of σ_b, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}$.

4 A simpler alternative

The hazard estimators, and data-driven smoothing parameter described in the previous two sections use exact calculation of the cumulative hazard function. However, the formulas are quite involved and specialist software is required for its implementation. In this section we show that a mixed model-based hazard estimate may be obtained using standard software. The key is to approximate the cumulative hazard function via quadrature. For simplicity we will present the formulae for trapezoidal integration. Other quadrature schemes could be used instead.

We first treat the case of no ties: $T_1 < T_2 < \ldots < T_n$. Recall that the likelihood depends on the cumulative hazar

$$\boldsymbol{\Lambda} = [\Lambda(T_1), \ldots, \Lambda(T_n)]^\mathsf{T} = \left[\int_0^{T_1} \lambda(u)du, \ldots, \int_0^{T_n} \lambda(u)du\right]^\mathsf{T}.$$

Instead of computing the integrals exactly, we can approximate $\boldsymbol{\Lambda}$ by numerical integration using the trapezoidal rule. That i

$$\boldsymbol{\Lambda} \simeq \mathbf{Q}\,\boldsymbol{\lambda}$$

where

$$\mathbf{Q} = \frac{1}{2}\begin{bmatrix} T_1 + T_1 & 0 & 0 & \cdots & 0 \\ T_2 + T_1 & T_2 - T_1 & 0 & \cdots & 0 \\ T_2 + T_1 & T_3 - T_1 & T_3 - T_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ T_2 + T_1 & T_3 - T_1 & T_4 - T_2 & \cdots & T_n - T_{n-1} \end{bmatrix}$$

and

$$\boldsymbol{\lambda} = [\lambda(T_1), \dots, \lambda(T_n)]^\mathsf{T}.$$

Then the log-likelihood for $(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ is

$$\begin{aligned} \ell(\boldsymbol{\beta}, \mathbf{b}, \sigma_b) &= \boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \mathbf{1}^\mathsf{T}\boldsymbol{\Lambda} - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b} - K\log\sigma_b \\ &\simeq \boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \mathbf{1}^\mathsf{T}\mathbf{Q}\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b} - K\log\sigma_b \\ &= \boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \mathbf{1}^\mathsf{T}\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{o}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b} - K\log\sigma_b \end{aligned}$$

where $\exp(\boldsymbol{a}) \equiv [\exp(a_1), \dots, \exp(a_n)]^\mathsf{T}$ and

$$\begin{aligned} \mathbf{o} &= \log(\mathbf{Q}^\mathsf{T}\mathbf{1}) \\ &= \log\Big(\tfrac{1}{2}[2T_1 + (n-1)(T_2 + T_1), (T_2 - T_1) + (n-2)(T_3 - T_1), \dots, \\ &\qquad (T_{n-1} - T_{n-2}) + (T_n - T_{n-2}), (T_n - T_{n-1})]^\mathsf{T}\Big). \end{aligned}$$

This shows that $\ell(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ is approximately the log-likelihood corresponding to a Poisson mixed model

$$\delta_i|\mathbf{b} \sim \text{Poisson}[\exp\{(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})_i + o_i\}], \quad \mathbf{b} \sim N(0, \sigma_b^2\mathbf{I})$$

where $o_1, \dots, o_n$ are known offset values. Thus we can estimate the hazard function using mixed Poisson regression. More specifically, we can obtain a REML estimate of σ_b by fitting a mixed Poisson regression with logarithmic link and offset $\mathbf{o} = \log(\mathbf{Q}^\mathsf{T}\mathbf{1})$ using the SAS macro `GLIMMIX`.

When there are ties among $\{T_1, \dots, T_n\}$, we have to modify the above method to assure that $\mathbf{o}$ is well defined. Suppose $T_{n_1} < T_{n_2} < \dots < T_{n_m}$ are all the unique values of $\{T_1, \dots, T_n\}$ and for $1 \le j \le m$, let $c_j \equiv \sum_{i=1}^n I(T_i = T_{n_j})$, and $\tilde{\delta}_j \equiv \sum_{i=1}^n \delta_i I(T_i = T_{n_j})$, where $I(\cdot)$ is the indicator function. It follows that

$$\ell(\boldsymbol{\beta}, \mathbf{b}, \sigma_b) \simeq \tilde{\boldsymbol{\delta}}^\mathsf{T}(\tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b}) - \mathbf{1}^\mathsf{T}\exp(\tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b} + \tilde{\mathbf{o}}) - \tfrac{1}{2\sigma_b^2}\mathbf{b}^\mathsf{T}\mathbf{b} - K\log\sigma_b$$

where $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$ are obtained from $\mathbf{X}$ and $\mathbf{Z}$ by deleting all the duplicated rows, $\tilde{\boldsymbol{\delta}} = (\tilde{\delta}_1, ..., \tilde{\delta}_m)^\mathsf{T}$, and

$$\tilde{\mathbf{o}} = \log\Big(\tfrac{1}{2}[2c_1T_{n_1} + \sum_{j=2}^{m} c_j(T_{n_2} + T_{n_1}), c_2(T_{n_2} - T_{n_1}) + \sum_{j=3}^{m} c_j(T_{n_3} - T_{n_1}), \ldots,$$
$$c_{m-1}(T_{n_{m-1}} - T_{n_{m-2}}) + c_m(T_{n_m} - T_{n_{m-2}}),\ c_m(T_{n_m} - T_{n_{m-1}})]^\mathsf{T}\Big).$$

Therefore $\ell(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ can be approximated by the log-likelihood corresponding to a Poisson mixed model

$$\tilde{\delta}_i|\mathbf{b} \sim \text{Poisson}[\exp\{(\tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b})_i + \tilde{o}_i\}], \quad \mathbf{b} \sim N(0, \sigma_b^2\mathbf{I}).$$

5 Standard errors

The covariance matrix of the estimated coefficients given the smoothing paramete σ_b is approximately

$$\text{cov}\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\mathbf{b}}\end{bmatrix}\right) \simeq \{\Xi''(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)\}^{-1}\text{cov}(\Xi'(\boldsymbol{\beta}, \mathbf{b}, \sigma_b))\{\Xi''(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)\}^{-1}$$

It follows from likelihood theory that the covariance can be approximated b

$$\widehat{\text{cov}}\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\mathbf{b}}\end{bmatrix}\right) = \{\Xi''(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \sigma_b)\}^{-1}.$$

For the quadrature approach of Section 4 the standard errors are given by `SAS GLIMMIX`.

6 Practical performance

6.1 Simulations

In order to assess the performance of these mixed model-based hazard estimates, with REML smoothing parameter choice, we simulated data from the following model:

$$T_i = \min(u_i, c_i), \qquad \delta_i = I(u_i \geq c_i)$$
$$u_i \sim d\text{Weibull}(1,3) + (1-d)\text{Weibull}(3,8), \qquad d \sim \text{Binomial}(0.7)$$
$$c_i \sim \text{Uniform}(0,6).$$

We used sample sizes $n = 200$ and $n = 500$. The number of replications in the simulation was 300.

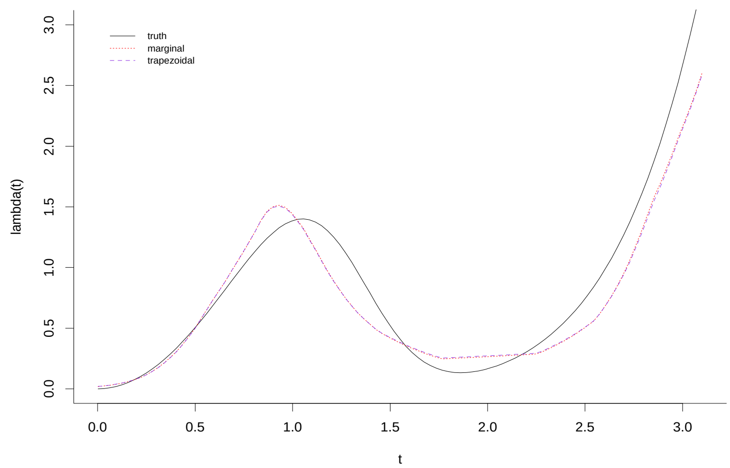


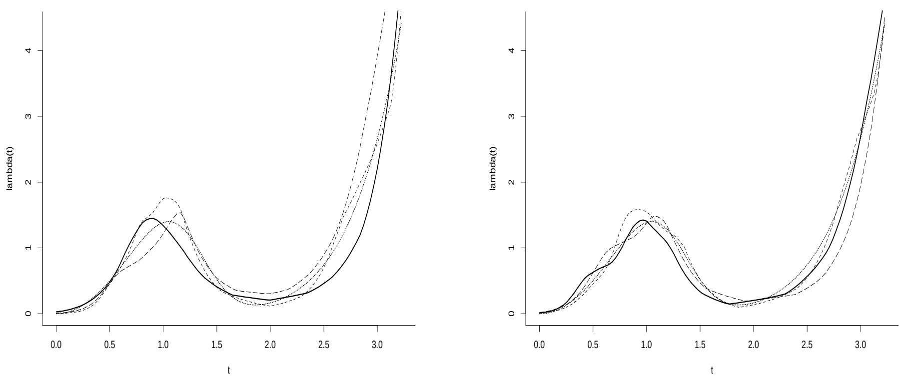
Figure 1: Estimated and the true underlying hazard function with $n = 200$

Figure 1 and 2 give a graphical summary of the results. Figure 1 shows the true hazard function and the estimated hazard function with two different methods obtained by marginal likelihood approach and the trapezoidal approach with `SAS GLIMMIX`. Here we chose 30 knots at $\{1/31, 2/31, \ldots, 30/31\}$ quantiles of $T_1, \ldots, T_n$. For this particular data set, the estimated smoothing parameter $\hat{\sigma}_b$ is 1.70 using the marginal likelihood method, and 1.64 by the trapezoidal approach with `SAS GLIMMIX`. As we can see from the graph, the estimated hazard functions by two approaches are very close. Figure 2 shows the performance of the hazard estimator based on the smoothing parameter chosen by the marginal likelihood approach for sample sizes $n = 200$ and $n = 500$. They are obtained by defining the distance between the estimated hazard function and the true hazard function to be the root mean square

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\{\widehat{\lambda}(T_i) - \lambda(T_i)\}^2}$$

and using the sample which is near the 10th, 50th and 90th percentiles of the distances based on the 300 realizations.

To compare the performance of the trapezoidal approach with the marginal likelihood approach, 300 sets of such data were simulated and for each realization, we computed the corresponding smoothing parameter $\hat{\sigma}_b$ using both methods. The resulting estimates of σ_b were shown in Figure 3.



(a) $n = 200$

(b) $n = 500$

Figure 2: Estimated and the true underlying hazard function: sample estimates near the median of the $\hat{D}$'s (solid curve), 10th percentile (dot-dashed curve) and 90th percentile (dashed curve). The dotted curve is the true hazard function

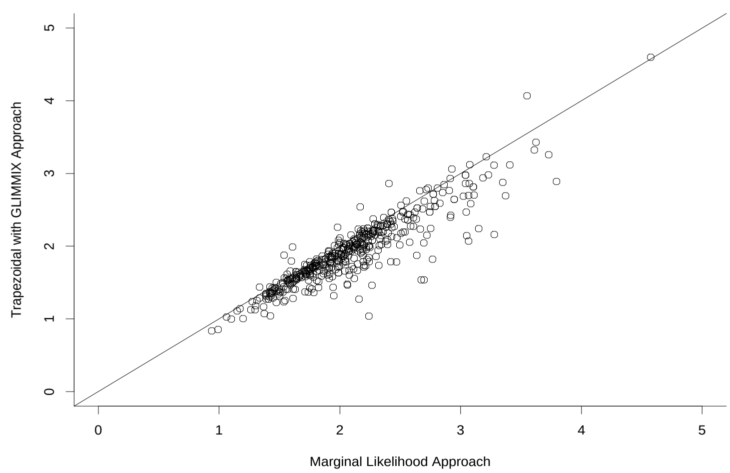


Figure 3: Estimated smoothing parameter using trapezoidal approach with `SAS GLIMMIX` versus marginal likelihood approach

6.2 Example

An application of our hazard estimator to sports statistics is illustrated in Figure 4. The data correspond to runs scored in test cricket innings by Australian player S.R. Waugh over the period December 1985 to August 1997. Censoring corresponds to the player being 'not out' at the completion of the innings. The estimate shows the player's high vulnerability early in the innings and when nearing 200. He also exhibits some slight vulnerability after reaching 50 and after reaching 150. A remarkable feature of S.R. Waugh's record is the ability to continue beyond the landmark score of 100 to a score higher than 150 and this is apparent in the dip in the hazard estimate between 100 and 150. Approximate 95% pointwise confidence intervals based on the standard error estimation described in Section 5 are indicated by the shading.

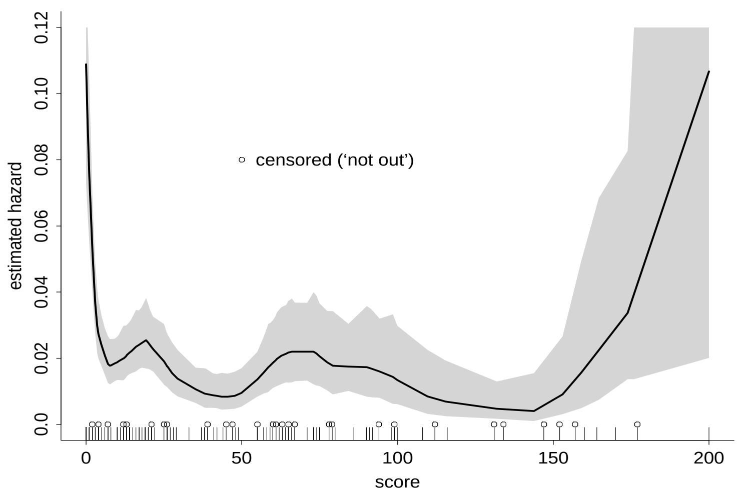


Figure 4: Estimated hazard function for the test cricket scores of S.R. Waugh (December, 1985 – August, 1997). The shaded region corresponds to approximate 95% pointwise confidence intervals based on the standard error estimation described in Section 5.

7 Extension

We have demonstrated that the mixed model approach to hazard estimation performs well and provides an attractive alternative to other methods. However, the biggest advantage, in our view, is the straightforward extension to more complex models such as hazard regression models with time-varying effects (Kooperberg, Stone and Truong, 1995; Fahrmeir and Wagenpfeil, 1996). Finally, this approach should also be beneficial in the interval censoring context where hazard estimation plays a crucial role (e.g. Betensky *et al.* 1999, 2000).

Acknowledgments

T. Cai was partially supported by U.S. National Institutes of Health grant NIH/NIAID 5 R37 AI29168-11 M.P. Wand was partially supported by U.S. Environmental Protection Agency grant R 824757.

Appendix: Computing Formulae

Let $\mathbf{x}$, $\mathbf{v}$ be $m \times 1$ vectors and $\mathbf{b}$ be an $n \times 1$ vector consisting of elements of the set $\{1, \dots, m\}$. Let $\mathbf{X}$ be an $m \times p$ matrix and $\mathbb{X}$ be an $m \times p \times q$ array. Then define

$\mathbf{x_b} = m \times 1$ vector with ith entry equal to x_{b_i},
$\mathbf{X}_{\mathbf{b},} = m \times p$ matrix with (i,k)th entry equal to $x_{b_i,k}$,
$\mathbb{X}_{\mathbf{b},,} = m \times p \times q$ array with (i,k,l)th entry equal to $x_{b_i,k,l}$,
$\overline{\mathbf{x}} = (m+1) \times 1$ vector with the ith entry equal to 0 if $i = 1$, x_{i-1} if $i \geq 2$;
$\overline{\mathbf{X}} = (m+1) \times p$ matrix with the (i,j)th entry equal to 0 if $i = 1$, $x_{i-1,j}$ if $i \geq 2$;
$\overline{\mathbb{X}} = (m+1) \times p \times q$ array with the (i,j,k)th entry equal to 0 if $i = 1$, $x_{i-1,jk}$ if $i \geq 2$;
$\mathbf{x}_{\text{-}} = (m-1) \times 1$ vector with the ith element equal to x_i .
$\mathbf{X}_{\text{-}} = (m-1) \times p$ matrix with the (i,j)th element equal to x_{ij} .
$\mathbb{X}_{\text{-}} = (m-1) \times p \times q$ array with the (i,j,k)th element equal to x_{ijk} .
$\mathbf{v} \otimes \mathbf{x} = m \times 1$ vector with ith entry equal to $v_i x_i$,
$\mathbf{v} \otimes \mathbf{X} = m \times p$ matrix with (i,j)th entry equal to $v_i x_{ij}$,
$\mathbf{v} \otimes \mathbb{X} = m \times p \times q$ array with (i,j,k)th entry equal to $v_i x_{ijk}$,
$\mathbf{v} \odot \mathbb{X} = m \times q$ matrix with (j,k)th entry equal to $\sum_{i=1}^{n} v_i x_{ijk}$.
$\text{cumsum}(\mathbf{x}) = m \times 1$ vector with ith entry equal to $\sum_{j=1}^{i} x_j$,
$\text{cumsum}(\mathbf{X}) = m \times p$ matrix with (i,j)th entry equal to $\sum_{l=1}^{i} x_{lj}$
$\text{cumsum}(\mathbb{X}) = m \times p \times q$ array with (i,j,k)th entry equal to $\sum_{l=1}^{i} x_{ijk}$,

Then

$$M(\beta_1, \mathbf{b}) = \mathbf{1}^\mathsf{T} \{\text{cumsum}(\bar{\boldsymbol{\nu}}^{(0)})\}_{\mathbf{k}^*}, \qquad N(\beta_1, \mathbf{b}) = \mathbf{1}^\mathsf{T} \boldsymbol{\zeta}^{(0)}$$

$$\boldsymbol{\nu}^{(j)} = \frac{\boldsymbol{\chi}_{\text{-}}}{\boldsymbol{\varepsilon}_{\text{-}}} \otimes \boldsymbol{\Delta}^{(j)}, \qquad \boldsymbol{\zeta}^{(j)} = \frac{\boldsymbol{\chi}_{\mathbf{k}^*}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} \otimes \boldsymbol{\Delta}_{\mathbf{T}}^{(j)}$$

$$\boldsymbol{\chi} = e^{-\text{cumsum}(\overline{\boldsymbol{\kappa}} \otimes \overline{\mathbf{b}})}, \qquad \boldsymbol{\varepsilon} = \beta_1 + \text{cumsum}(\overline{\mathbf{b}})$$

$$\boldsymbol{\Delta}^{(j)} = \boldsymbol{\kappa}^j \otimes e^{\boldsymbol{\kappa}^j \otimes \boldsymbol{\varepsilon}_{\text{-}}} - \overline{\boldsymbol{\kappa}}^j_{\text{-}} \otimes e^{\overline{\boldsymbol{\kappa}}^j_{\text{-}} \otimes \boldsymbol{\varepsilon}_{\text{-}}}, \qquad \boldsymbol{\Delta}_{\mathbf{T}}^{(j)} = \mathbf{T}^j \otimes e^{\mathbf{T}^j \otimes \boldsymbol{\varepsilon}_{\mathbf{k}^*}} - \overline{\boldsymbol{\kappa}}^j_{\mathbf{k}^*} \otimes e^{\overline{\boldsymbol{\kappa}}^j_{\mathbf{k}^*} \otimes \boldsymbol{\varepsilon}_{\mathbf{k}^*}}$$

To calculate first and second derivatives for $M(\beta_1, \mathbf{b})$ and $N(\beta_1, \mathbf{b})$, first, Define

$$\mathbf{D}^{(j)} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \kappa_1^j & 0 & 0 & \cdots & 0 \\ \kappa_1^j & \kappa_2^j & 0 & \cdots & 0 \\ \kappa_1^j & \kappa_2^j & \kappa_3^j & \cdots & \kappa_K^j \end{bmatrix}, \qquad \mathbf{V}^{(ij)} = \tfrac{1}{2}\{\boldsymbol{\kappa}^i (\boldsymbol{\kappa}^j)^\mathsf{T} + \boldsymbol{\kappa}^j (\boldsymbol{\kappa}^i)^\mathsf{T}\}$$

$\mathbb{V}^{(ij)} = (K+1) \times K \times K$ array with the (k,l,m)th entry equal to
$I(l \leq k-1; m \leq k-1) \mathbf{V}^{(ij)}(l,m)$

and

$$\boldsymbol{a} = -\frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}} + \boldsymbol{\nu}^{(1)}, \qquad \mathbf{B} = (\boldsymbol{\nu}^{(1)} - \frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}}) \otimes \mathbf{D}_{\text{-}}^{(0)} - \boldsymbol{\nu}^{(0)} \otimes D_{\text{-}}^{(1)},$$

$$\boldsymbol{a}_2 = \frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}^2} - \frac{2\boldsymbol{\nu}^{(1)}}{\boldsymbol{\varepsilon}_{\text{-}}} + \boldsymbol{\nu}^{(2)}, \quad \mathbf{B}_2 = (\frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}^2} - \frac{2\boldsymbol{\nu}^{(1)}}{\boldsymbol{\varepsilon}_{\text{-}}} + \boldsymbol{\nu}^{(2)}) \odot \mathbf{D}_{\text{-}}^{(0)} + (\frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}} - \boldsymbol{\nu}^{(1)}) \odot \mathbf{D}_{\text{-}}^{(1)}$$

$$\mathbb{B} = (\frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}^2} - \frac{2\boldsymbol{\nu}^{(1)}}{\boldsymbol{\varepsilon}_{\text{-}}} + \boldsymbol{\nu}^{(2)}) \otimes \mathbb{V}_{\text{-}}^{(00)} + (\frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{\text{-}}} - 2\boldsymbol{\nu}^{(1)}) \otimes \mathbb{V}_{\text{-}}^{(01)} + \boldsymbol{\nu}^{(0)} \otimes \mathbb{V}_{\text{-}}^{(11)}$$

Then

$$\frac{\partial M(\beta_1, \mathbf{b})}{\partial \beta_1} = \mathbf{1}^\top \{\text{cumsum}(\overline{\boldsymbol{a}})_{\mathbf{k}^*}\}, \quad \frac{\partial N(\beta_1, \mathbf{b})}{\partial \beta_1} = \mathbf{1}^\top \{-\frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(1)}\},$$

$$\frac{\partial M(\beta_1, \mathbf{b})}{\partial \mathbf{b}^\top} = \mathbf{1}^\top \{\text{cumsum}(\overline{\mathbf{B}})_{\mathbf{k}^*,}\}, \quad \frac{\partial N(\beta_1, \mathbf{b})}{\partial \mathbf{b}^\top} = \mathbf{1}^\top \{(\boldsymbol{\zeta}^{(1)} - \frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}}) \otimes \mathbf{D}_{\mathbf{k}^*,}^{(0)} - \boldsymbol{\zeta}^{(0)} \otimes \mathbf{D}_{\mathbf{k}^*,}^{(0)}\}$$

$$\frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \beta_1^2} = \mathbf{1}^\top \{\text{cumsum}(\overline{\boldsymbol{a}_2})_{\mathbf{k}^*}\}, \quad \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \beta_1^2} = \mathbf{1}^\top \{\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}^2} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)}\},$$

$$\frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}^\top} = \mathbf{1}^\top \{\text{cumsum}(\overline{\mathbf{B}_2})_{\mathbf{k}^*,}\}, \quad \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}^\top} = \mathbf{1}^\top \{(\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}^2} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)}) \otimes \mathbf{D}_{\mathbf{k}^*,}^{(0)} + (\frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} - \boldsymbol{\zeta}^{(1)}) \otimes \mathbf{D}_{\mathbf{k}^*,}^{(1)}\}$$

$$\frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}^\top} = \mathbf{1} \odot \{\text{cumsum}(\overline{\mathbb{B}})_{\mathbf{k}^*,,}\}, \quad \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}^\top} = \mathbf{1} \odot \{(\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}^2} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)}) \otimes \mathbb{V}_{\mathbf{k}^*,,}^{(00)} + (\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} - 2\boldsymbol{\zeta}^{(1)}) \otimes \mathbb{V}_{\mathbf{k}^*,,}^{(01)} + \boldsymbol{\zeta}^{(0)} \otimes \mathbb{V}_{\mathbf{k}^*,,}^{(11)}\}$$

$$\Xi'(\boldsymbol{\beta}, \mathbf{b}; \sigma_b) = \begin{bmatrix} \mathbf{1}^\top \boldsymbol{\delta} - e^{\beta_0}\{M(\beta_1, \mathbf{b}) + N(\beta_1, \mathbf{b})\} \\ \boldsymbol{\delta}^\top \mathbf{T} - e^{\beta_0}\{\frac{\partial M(\beta_1, \mathbf{b})}{\partial \beta_1} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \beta_1}\} \\ \sum_{i=1}^n \delta_i (T_i - \boldsymbol{\kappa})_+ - e^{\beta_0}\{\frac{\partial M(\beta_1, \mathbf{b})}{\partial \mathbf{b}} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \mathbf{b}}\} - \frac{1}{\sigma_b^2}\mathbf{b} \end{bmatrix}$$

and

$$\Xi''(\boldsymbol{\beta}, \mathbf{b}; \sigma_b) = -e^{\beta_0} \begin{bmatrix} M(\beta_1, \mathbf{b}) + N(\beta_1, \mathbf{b}) & \frac{\partial M(\beta_1, \mathbf{b})}{\partial \beta_1} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \beta_1} & \frac{\partial M(\beta_1, \mathbf{b})}{\partial \mathbf{b}'} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \mathbf{b}'} \\ \frac{\partial M(\beta_1, \mathbf{b})}{\partial \beta_1} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \beta_1} & \frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \beta_1^2} + \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \beta_1^2} & \frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}'} + \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}'} \\ \frac{\partial M(\beta_1, \mathbf{b})}{\partial \mathbf{b}} + \frac{\partial N(\beta_1, \mathbf{b})}{\partial \mathbf{b}} & \frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}} + \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \beta_1 \partial \mathbf{b}} & \frac{\mathbf{I}}{e^{\beta_0}\sigma_b^2} + \frac{\partial^2 M(\beta_1, \mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}'} + \frac{\partial^2 N(\beta_1, \mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}'} \end{bmatrix}.$$

References

Betensky, R.A, Lindsey, J.C., Ryan, L.M. and Wand, M.P. (1999). Local EM estimation of the hazard function for interval censored data *Biometrics*, **55**, 238–245.

Betensky, R.A, Lindsey, J.C., Ryan, L.M. and Wand, M.P. (2000). A local likelihood proportional hazards model for interval censored data.

Bloxom, B. (1985). A constrained spline estimator of a hazard function. *Psychometrika*, **50**, 301–321.

Breslow, N.E. and Clayton, D.G. (1993). Approximated inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistics Society, Series B* , **34**, 187–220.

Eilers, P.H.C. (2000). Hazard estimation with B-splines. Unpublished manuscript.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, **89**, 89–121.

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the baseline hazard function. *Biometrics*, **43**, 181–192.

Joly, P., Commenges, D. and Letenneur, L. (1998) A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics*, **54**, 185–194.

Hjort, N.L. (1993). Dynamic likelihood hazard rate estimation. Technical Report 4, Institute of Mathematics, University of Oslo.

Kooperberg, C. and Stone, C.J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, 301–328.

Kooperberg, C., Stone, C.J. and Truong, Y.K. (1995). Hazard regression. *Journal of the American Statistical Association*, **90**, 78–94.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators *SIAM Journal on Scientific and Statistical Computing*, **9**, 363–379.

Rosenberg, P.S. (1995). Hazard function estimation using B-splines *Biometrics*, **51**, 874–887.

Senthilselvan, A. (1987). Penalized likelihood estimation of hazard and intensity functions. *Journal of the Royal Statistical Society, Series B*, **49**, 170–174.

Tanner, M.A. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *The Annals of Statistics*, **11**, 989–993.