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.

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, fitting proportional hazard models in the presence of interval censoring benefits from hazard function estimation (e.g. Betensky *et al.* 2001).

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-BASED HAZARD ESTIMATION

Suppose we observe data (T_i, δ_i), $1 \le i \le n$, 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 is

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

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 kth knot to approximately correspond to the kth sample quantile of the unique T_i values and sets $K = \min(\lfloor n/4 \rfloor, 30)$. The answers are quite insensitive to the placement of the knots and their number (e.g. Ruppert 2001; French, Kammann and Wand 2001)

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

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

Then define the cumulative hazard to be $\Lambda(t) = \int_0^t \lambda(u)du$ and the sum of cumulative hazards evaluated at the T_i to be

$$\Lambda(\boldsymbol{\beta}, \mathbf{b}) \equiv \sum_{i=1}^{n} \Lambda(T_i) = e^{\beta_0}\{\mathcal{M}(\beta_1, \mathbf{b}) + \mathcal{N}(\beta_1, \mathbf{b})\}\,,$$

where

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

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

$\kappa_0 \equiv 0$ 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 then

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

The right-hand side of (3) involves an intractable K-dimensional integral. Analogous to the Penalized Quasi-Likelihood approach (PQL) (e.g. Breslow and Clayton 1993; Wolfinger and O'Connell 1993), we can approximate (3) with

$$\ell(\boldsymbol{\beta}, \sigma_b) \simeq \boldsymbol{\delta}^\mathsf{T}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\widehat{\mathbf{b}}) - \Lambda(\boldsymbol{\beta}, \widehat{\mathbf{b}}) - \tfrac{1}{2\sigma_b^2}\widehat{\mathbf{b}}^\mathsf{T}\widehat{\mathbf{b}}\,, \qquad (4)$$

where

$$\widehat{\mathbf{b}} = \underset{\mathbf{b}}{\text{argmax}}\, S(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)\,,$$

and

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

Note that the joint log-likelihood $S(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ is the same as the h-likelihood (Lee and Nelder 1996; Ha, Lee and Song 2001; Lee and Nelder 2001) and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}$ are maximum h-likelihood estimators.

The implicit assumption of PQL is that the integrand is approximated well by a multivariate normal density. Often, in the usual mixed model context, the quality of this approximation is a concern since inference about the parameters $\boldsymbol{\beta}$ and σ_b^2 is a primary goal. However in the current context random effects are simply used as a device for flexible curve estimation and the accuracy of PQL is not crucial. For fixed $\sigma_b^2 = 1/\tau$ this mixed model approach with PQL approximation is equivalent to the penalized spline fit

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

with

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}} \end{bmatrix} = \underset{\boldsymbol{\beta},\mathbf{b}}{\operatorname{argmax}} \left\{ \boldsymbol{\delta}^{\mathsf{T}}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \Lambda(\boldsymbol{\beta}, \mathbf{b}) - \tfrac{1}{2}\tau\mathbf{b}^{\mathsf{T}}\mathbf{b} \right\} \qquad (5)$$

and, through transformation from the truncated line basis to the linear B-spline basis, can be equated with one of the class of estimators proposed by 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.

Note that (2) does not constitute a bona fide log-hazard function since the survival probability does not approach zero as $t \to \infty$. We are not concerned by this since we can only estimate the hazard function over the range of the data. A referee has also pointed out that (2) might benefit from some modification in the right hand tail, such as the imposition of exponential drop-off as used in the HEFT (Hazard Estimation with Flexible Tails) algorithm of Kooperberg, Stone and Truong (1995).

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

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

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

$$\ell_{\text{marg}}(\sigma_b) = -K \log \sigma_b + S(\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b), \sigma_b) - \tfrac{1}{2} \log \left| -S''(\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b), \sigma_b) \right| \,, \qquad (6)$$

where $(\widehat{\boldsymbol{\beta}}(\sigma_b), \widehat{\mathbf{b}}(\sigma_b))$ denotes the solution to $S'(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}; \sigma_b) = \mathbf{0}$. This is analogous to the REML estimate of Breslow and Clayton (1993) and (6) corresponds to the adjusted profile h-likelihood (Lee and Nelder 1996). The REML log-likelihood in the Gaussian model is also the marginal log-likelihood for the data when the regression parameters are integrated out against a flat prior (Harville 1974).

In the Appendix we give exact, readily computable, formulae for $S(\boldsymbol{\beta}, \mathbf{b}, \sigma_b)$ 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 hazard

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

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

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

where

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

and

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

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

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

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 $o_1, \ldots, o_n$ 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`. This macro uses restricted pseudo likelihood to find the parameter estimates of the generalized linear mixed model.

When there are ties among $\{T_1, \ldots, T_n\}$, we have to modify the above method to assure that $\mathbf{o}$ is well defined. Suppose $T_{n_1} < T_{n_2} < \ldots < T_{n_m}$ are all the unique values of $\{T_1, \ldots, T_n\}$ 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

$$\begin{aligned}
\tilde{\mathbf{o}} &= \log\Big(\tfrac{1}{2}[2c_1 T_{n_1} + \sum_{j=2}^{m} c_j(T_{n_2} + T_{n_1}), c_2(T_{n_2} - T_{n_1}) + \sum_{j=3}^{m} c_j(T_{n_3} - T_{n_1}), \ldots, \\
&\qquad c_{m-1}(T_{n_{m-1}} - T_{n_{m-2}}) + c_m(T_{n_m} - T_{n_{m-2}}),\ c_m(T_{n_m} - T_{n_{m-1}})]^{\mathsf{T}}\Big)\,.
\end{aligned}$$

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

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

The same general approach can be extended to higher degree truncated polynomial bases. Linear transformations of these bases, such as B-spline or Demmler-Reinsch bases, could also be used for increased numerical stability although we have so far had good experience with truncated lines. Mixed model routines transform the basis functions as part of their fitting algorithms (e.g. Pinheiro and Bates 2000, Section 2.2) so the choice of basis functions is not so crucial for the quadrature approach described in this section.

5. STANDARD ERRORS

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

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

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

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

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:

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

We used sample sizes $n = 200$ and $n = 500$. The number of replications in the simulation was 300. Under this set up, there is about 26% of censoring.

Figure 2 and 3 give a graphical summary of the results. Figure 2 shows the true hazard function and the estimated hazard function with three different methods: the marginal likelihood approach, the trapezoidal approach with `SAS GLIMMIX` and HEFT. 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 3 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 square root mean square

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

and using the sample which is near the 10th, 50th and 90th percentiles of the distances based on the 300 realizations. Shown also on Figure 3 are the 90% empirical pointwise confidence interval for the estimated hazard function (shaded regions).

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 are shown in Figure 1. A bias due to quadrature is apparent, but is negligible in this case.

The results shown above are for the censoring percentage of 26%. We also ran some simulations with various lower and upper bounds for the uniform distribution of the censoring variable such that we have 15%, 25%, 35% and 50% of censoring. We compared the biases and empirical standard errors for the estimated $\lambda(t)$ and found that the estimation procedure seems to be quite stable even though with the increase of censoring percentage, the estimated hazard function does show more variation. Table 1 shows the biases and the empirical standard errors of the estimated $\lambda(t)$ at $t = 0.739,\ 1.065,\ 2.316$ (corresponding to 25th, 50th and 75th quantile of the Weibull mixture distribution of u_i) at different censoring percentages and sample size $n = 200$.

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. For comparison we also plot the estimate obtained using the HEFT (Hazard Estimation with Flexible Tails) algorithm of Kooperberg, Stone and Truong (1995). Approximate 95% pointwise confidence intervals based on the standard error estimation described in Section 5. are indicated by the shading in Figure 4. The existence of features such as the dip between 100 and 150 cannot be deduced from this analysis alone. However, the technology of significant zero crossings of derivatives ('SiZer') (e.g. Chaudhuri and Marron 1999) could be adapted to address this issue.

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 covariate 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 2001).

Acknowledgments

We are grateful to the editor, associate editor and three referees for their comments on the original submission. 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

Here we give readily computable formulae for the estimators (5) and the marginal likelihood given by (6).

First we introduce some notation. Let $\mathbf{x}$, $\mathbf{v}$ be $m \times 1$ vectors and $\mathbf{c}$ be an $n \times 1$ vector consisting of elements of the set $\{1, \ldots, m\}$. Let $\mathbf{X}$ be an $m \times p$ matrix and $\mathbb{X}$ be an $m \times p \times q$ array. Then define

$\mathbf{x}_{\mathbf{c}} = m \times 1$ vector with ith entry equal to x_{c_i} ,

$\mathbf{X}_{\mathbf{c},} = m \times p$ matrix with (i,k)th entry equal to $x_{c_i,k}$,

$\mathbb{X}_{\mathbf{c},,} = m \times p \times q$ array with (i,k,l)th entry equal to $x_{c_i,k,l}$,

$\overline{\mathbf{x}} = (m+1) \times 1$ vector with the ith entry equal to $0\ I(i \geq 2) x_{i-1}$,

$\overline{\mathbf{X}} = (m+1) \times p$ matrix with the (i,j)th entry equal to $I(i \geq 2) x_{i-1,j}$,

$\overline{\mathbb{X}} = (m+1) \times p \times q$ array with the (i,j,k)th entry equal to $I(i \geq 2) x_{i-1,jk}$,

$\mathbf{x}_{\text{-}} = (m-1) \times 1$ 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 $\sum_{l=1}^{i} x_{ljk}$.

where $I(\cdot)$ is the indicator function.

Then to compute (5) and (6) we use the following results:

$$\begin{aligned}
\Lambda(\boldsymbol{\beta},\mathbf{b}) &= e^{\beta_0}(\mathcal{M}+\mathcal{N}),\\
S'(\boldsymbol{\beta},\mathbf{b};\sigma_b) &= \begin{bmatrix} \mathbf{1}^\mathsf{T}\boldsymbol{\delta} - e^{\beta_0}\{\mathcal{M}+\mathcal{N}\} \\ \boldsymbol{\delta}^\mathsf{T}\mathbf{T} - e^{\beta_0}\{\mathcal{M}_\beta+\mathcal{N}_\beta\} \\ \sum_{i=1}^n \delta_i (T_i-\boldsymbol{\kappa})_+ - e^{\beta_0}\{\mathcal{M}_\mathbf{b}+\mathcal{N}_\mathbf{b}\} - \frac{1}{\sigma_b^2}\mathbf{b} \end{bmatrix},\\
\text{and}\qquad S''(\boldsymbol{\beta},\mathbf{b};\sigma_b) &= -e^{\beta_0}\begin{bmatrix} \mathcal{M}+\mathcal{N} & \mathcal{M}_\beta+\mathcal{N}_\beta & \mathcal{M}'_\mathbf{b}+\mathcal{N}'_\mathbf{b} \\ \mathcal{M}_\beta+\mathcal{N}_\beta & \mathcal{M}_{\beta^2}+\mathcal{N}_{\beta^2} & \mathcal{M}'_{\beta\mathbf{b}}+\mathcal{M}'_{\beta\mathbf{b}} \\ \mathcal{M}_\mathbf{b}+\mathcal{N}_\mathbf{b} & \mathcal{M}_{\beta\mathbf{b}}+\mathcal{N}_{\beta\mathbf{b}} & \frac{\mathbf{I}}{e^{\beta_0}\sigma_b^2}+\mathcal{M}_{\mathbf{bb}'}+\mathcal{N}_{\mathbf{bb}'} \end{bmatrix}.
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{M}=\mathcal{M}(\beta_1,\mathbf{b}) &= \mathbf{1}^\mathsf{T}\{\text{cumsum}(\bar{\boldsymbol{\nu}}^{(0)})\}_{\mathbf{k}^*}, \qquad \mathcal{N}=\mathcal{N}(\beta_1,\mathbf{b}) = \mathbf{1}^\mathsf{T}\boldsymbol{\zeta}^{(0)},\\
\mathcal{M}'_\mathbf{b} = \frac{\partial\mathcal{M}(\beta_1,\mathbf{b})}{\partial\mathbf{b}'} &= \mathbf{1}^\mathsf{T}\{\text{cumsum}(\overline{\mathbf{B}})_{\mathbf{k}^*,}\},\\
\mathcal{N}'_\mathbf{b} = \frac{\partial\mathcal{N}(\beta_1,\mathbf{b})}{\partial\mathbf{b}'} &= \mathbf{1}^\mathsf{T}\{(\boldsymbol{\zeta}^{(1)} - \frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}})\otimes\mathbf{D}^{(0)}_{\mathbf{k}^*,} - \boldsymbol{\zeta}^{(0)}\otimes\mathbf{D}^{(0)}_{\mathbf{k}^*,}\},\\
\mathcal{M}_\beta = \frac{\partial\mathcal{M}(\beta_1,\mathbf{b})}{\partial\beta_1} &= \mathbf{1}^\mathsf{T}\{\text{cumsum}(\overline{\boldsymbol{a}})_{\mathbf{k}^*}\}, \qquad \mathcal{N}_\beta = \frac{\partial\mathcal{N}(\beta_1,\mathbf{b})}{\partial\beta_1} = \mathbf{1}^\mathsf{T}\{-\frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(1)}\},\\
\mathcal{M}'_{\beta\mathbf{b}} = \frac{\partial^2\mathcal{M}(\beta_1,\mathbf{b})}{\partial\beta_1\partial\mathbf{b}'} &= \mathbf{1}^\mathsf{T}\{\text{cumsum}(\overline{\mathbf{B}_2})_{\mathbf{k}^*,}\},\\
\mathcal{N}'_{\beta\mathbf{b}} = \frac{\partial^2\mathcal{N}(\beta_1,\mathbf{b})}{\partial\beta_1\partial\mathbf{b}'} &= \mathbf{1}^\mathsf{T}\{(\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}^2_{\mathbf{k}^*}} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)})\otimes\mathbf{D}^{(0)}_{\mathbf{k}^*,} + (\frac{\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} - \boldsymbol{\zeta}^{(1)})\otimes\mathbf{D}^{(1)}_{\mathbf{k}^*,}\},\\
\mathcal{M}_{\beta^2} = \frac{\partial^2\mathcal{M}(\beta_1,\mathbf{b})}{\partial\beta_1^2} &= \mathbf{1}^\mathsf{T}\{\text{cumsum}(\overline{\boldsymbol{a}_2})_{\mathbf{k}^*}\},\\
\mathcal{N}_{\beta^2} = \frac{\partial^2\mathcal{N}(\beta_1,\mathbf{b})}{\partial\beta_1^2} &= \mathbf{1}^\mathsf{T}\{\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}^2_{\mathbf{k}^*}} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)}\},\\
\mathcal{M}_{\mathbf{bb}'} = \frac{\partial^2\mathcal{M}(\beta_1,\mathbf{b})}{\partial\mathbf{b}\partial\mathbf{b}'} &= \mathbf{1}\odot\{\text{cumsum}(\overline{\mathbb{B}})_{\mathbf{k}^*,,}\},\\
\mathcal{N}_{\mathbf{bb}'} = \frac{\partial^2\mathcal{N}(\beta_1,\mathbf{b})}{\partial\mathbf{b}\partial\mathbf{b}'} &= \mathbf{1}\odot\{(\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}^2_{\mathbf{k}^*}} - \frac{2\boldsymbol{\zeta}^{(1)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} + \boldsymbol{\zeta}^{(2)})\otimes\mathbb{V}^{(00)}_{\mathbf{k}^*,,} + (\frac{2\boldsymbol{\zeta}^{(0)}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} - 2\boldsymbol{\zeta}^{(1)})\otimes\mathbb{V}^{(01)}_{\mathbf{k}^*,,}\\
&\qquad + \boldsymbol{\zeta}^{(0)}\otimes\mathbb{V}^{(11)}_{\mathbf{k}^*,,}\}.
\end{aligned}$$

Here,

$$
\begin{aligned}
&\boldsymbol{\nu}^{(j)} = \frac{\boldsymbol{\chi}_{-}}{\boldsymbol{\varepsilon}_{-}} \otimes \boldsymbol{\Delta}^{(j)}\,, &&\boldsymbol{a}_2 = \frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{-}^2} - \frac{2\boldsymbol{\nu}^{(1)}}{\boldsymbol{\varepsilon}_{-}} + \boldsymbol{\nu}^{(2)}\,,\\
&\boldsymbol{\chi} = e^{-\mathrm{cumsum}(\overline{\boldsymbol{\kappa}}\otimes\overline{\mathbf{b}})}\,, &&\boldsymbol{\Delta}^{(j)} = \boldsymbol{\kappa}^j \otimes e^{\boldsymbol{\kappa}^j\otimes\boldsymbol{\varepsilon}_{-}} - \overline{\boldsymbol{\kappa}}^j_{-} \otimes e^{\overline{\boldsymbol{\kappa}}^j_{-}\otimes\boldsymbol{\varepsilon}_{-}}\,,\\
&\boldsymbol{\varepsilon} = \beta_1 + \mathrm{cumsum}(\overline{\mathbf{b}})\,, &&\boldsymbol{\Delta}^{(j)}_{\mathbf{T}} = \mathbf{T}^j \otimes e^{\mathbf{T}^j\otimes\boldsymbol{\varepsilon}_{\mathbf{k}^*}} - \overline{\boldsymbol{\kappa}}^j_{\mathbf{k}^*} \otimes e^{\overline{\boldsymbol{\kappa}}^j_{\mathbf{k}^*}\otimes\boldsymbol{\varepsilon}_{\mathbf{k}^*}}\,,\\
&\boldsymbol{\zeta}^{(j)} = \frac{\boldsymbol{\chi}_{\mathbf{k}^*}}{\boldsymbol{\varepsilon}_{\mathbf{k}^*}} \otimes \boldsymbol{\Delta}^{(j)}_{\mathbf{T}}\,, &&\mathbf{B} = (\boldsymbol{\nu}^{(1)} - \frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{-}}) \otimes \mathbf{D}^{(0)}_{-} - \boldsymbol{\nu}^{(0)} \otimes D^{(1)}_{-}\,,\\
&\boldsymbol{a} = -\frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{-}} + \boldsymbol{\nu}^{(1)}\,, &&\mathbf{B}_2 = (\frac{2\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{-}^2} - \frac{2\boldsymbol{\nu}^{(1)}}{\boldsymbol{\varepsilon}_{-}} + \boldsymbol{\nu}^{(2)}) \odot \mathbf{D}^{(0)}_{-} + (\frac{\boldsymbol{\nu}^{(0)}}{\boldsymbol{\varepsilon}_{-}} - \boldsymbol{\nu}^{(1)}) \odot \mathbf{D}^{(1)}_{-}\,,
\end{aligned}
$$

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

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

and $\mathbb{V}^{(ij)}$ is the $(K+1)\times K\times K$ array with (k,l,m)th entry equal to $\mathbf{V}^{(ij)}(l,m)$ if $l \le k-1$ and $m \le k-1$, and zero otherwise.

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. *Statistics in Medicine*, to appear.

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.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94**, 807–823.

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*, **11**, 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.

French, J.L., Kammann, E.E. and Wand, M.P. (2001). Comment on paper by Ke and Wang. *Journal of the American Statistical Association*, **96**, in press.

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.

Ha, I.D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233–244.

Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

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.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B*, **58**, 619–656.

Lee, Y. and Nelder, J.A. (2001). Hierarchical generalised linear model: synthesis of generalised linear models, random effect models and structured dispersions. *Biometrika*, **88**, in press.

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.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*, New York: Springer.

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

Ruppert, D. (2001). Selecting the number of knots for penalized splines. Unpublished manuscript.

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.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233–243.

Table 1: The biases and empirical standard errors of estimated hazard rates at different censoring percentages (p_c).

t	$\lambda(t)$	Relative Bias (Standard Error)			
		$p_c = 15\%$	$p_c = 25\%$	$p_c = 35\%$	$p_c = 50\%$
0.739	0.997	0.029 (0.147)	0.034 (0.152)	0.023 (0.157)	0.030 (0.155)
1.065	1.399	0.063 (0.199)	0.066 (0.203)	0.067 (0.228)	0.113 (0.223)
2.316	0.436	0.038 (0.102)	0.014 (0.119)	0.022 (0.138)	0.058 (0.218)

Figure 1: Estimated smoothing parameter using trapezoidal approach with `SAS GLIMMIX` versus marginal likelihood approach

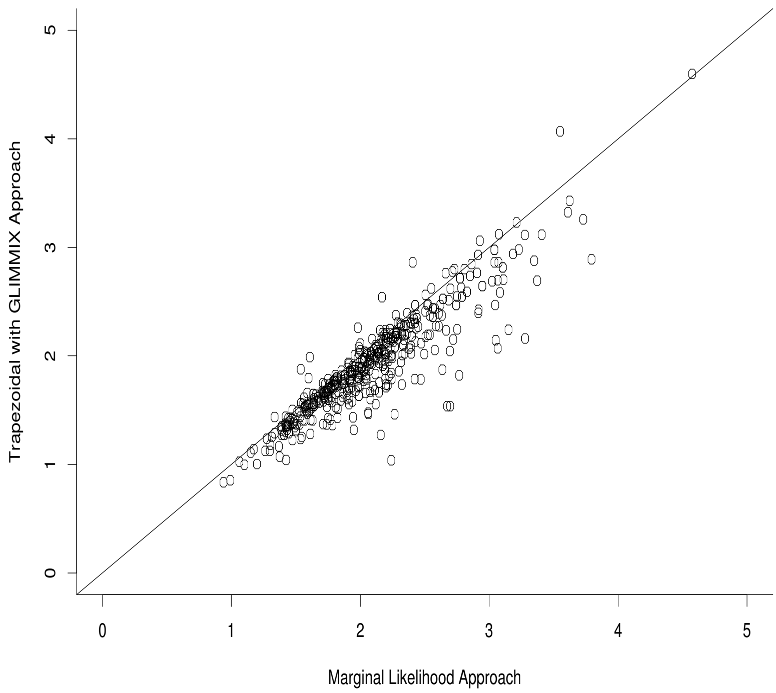


Figure 2: Estimated and the true underlying hazard function with $n = 200$

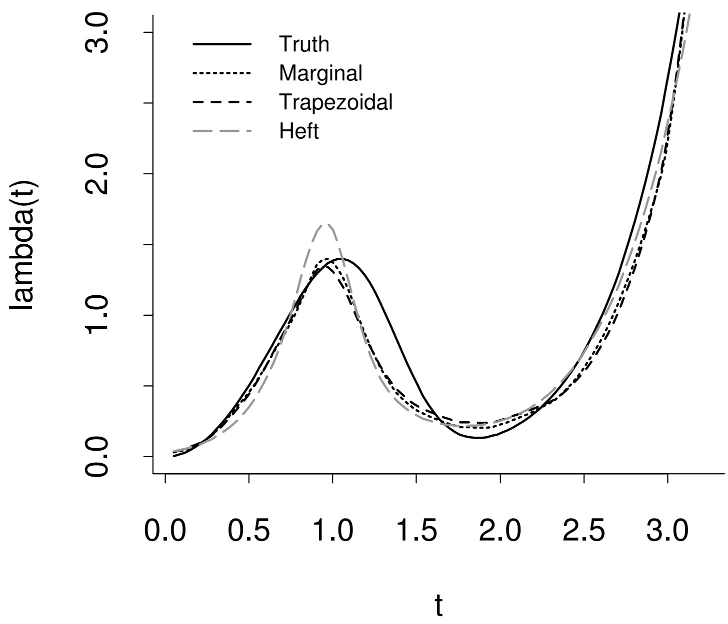
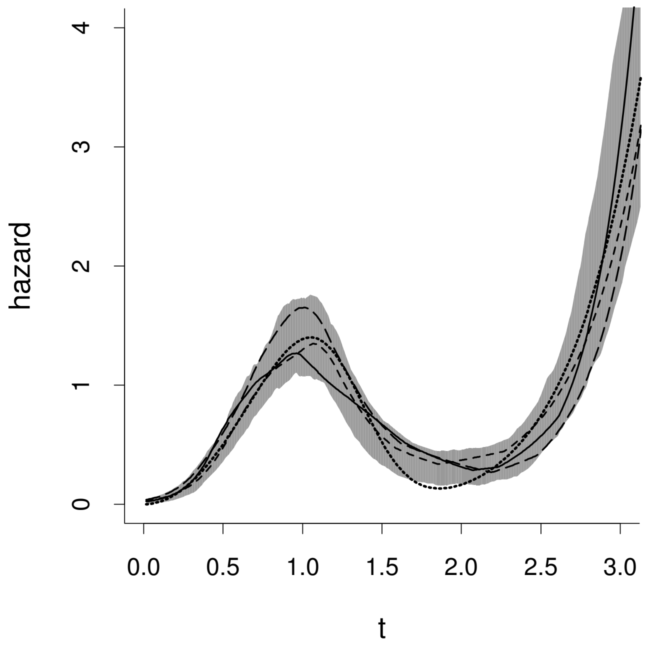
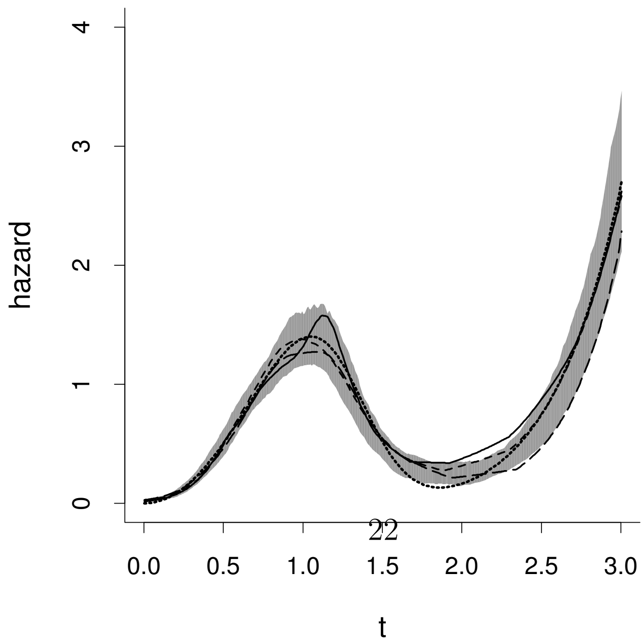


Figure 3: Estimated and the true underlying hazard function: sample estimates near the median of the $\hat{D}$'s (solid curve), 10th percentile (short-dashed curve) and 90th percentile (long-dashed curve). The dotted curve is the true hazard function. The shaded regions correspond to empirical 90% confidence intervals based on the simulation.



(a) $n = 200$



(b) $n = 500$

Figure 4: Estimated hazard function for the test cricket scores of S.R. Waugh (December 1985 – August 1997). The solid curve is the mixed model-based estimate and the dashed curve is result obtained from HEFT. The shaded region corresponds to approximate 95% pointwise confidence intervals based on the standard error estimation described in Section 5..