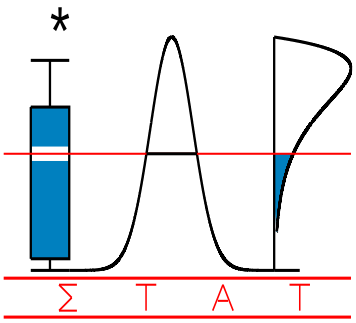
T E C H N I C A L

R E P O R T

0435

BAYESIAN SURVIVAL MODELS WITH SMOOTH TIME-VARYING COEFFICIENTS USING PENALIZED POISSON REGRESSION

LAMBERT, Ph. and P. H. C. EILERS



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

BAYESIAN SURVIVAL MODELS WITH SMOOTH TIME-VARYING COEFFICIENTS USING PENALIZED POISSON REGRESSION

Philippe Lambert[1,2*] & Paul H.C. Eilers[3]

[1] *Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

[2] *Unité d'épidémiologie, biostatistique et méthodes opérationnelles, Faculté de Médecine, Université catholique de Louvain, Belgium.*

[3] *Department of Medical Statistics, Leiden University, Leiden, The Netherlands*

SUMMARY

It is rather well known that one can approach survival problems without covariates in an actuarial way. The time axis is divided into intervals (named *bins*), and in each bin the number of people at risk is counted as well as the number of events. The relationship between time and probability of an event can then be estimated with a parametric or semi-parametric model. Here, we consider a subdivision of the time scale into a large number of bins. The number of events observed in each bin is described using a Poisson distribution with the log mean specified using a flexible penalized B-splines model with knots located at the bins limits. Regression on pertinent covariates can easily be performed using the same log-linear model, leading to the classical proportional hazard model. We propose to extend that model by allowing the regression coefficients to vary in a smooth way with time. Penalized B-splines models will be proposed for each of these coefficients. We show how the regression parameters and the penalty weights can be estimated efficiently using Bayesian inference tools based on the Metropolis-adjusted Langevin algorithm.

KEY WORDS: proportional hazards; P-splines; Markov chain Monte Carlo methods; Metropolis-adjusted Langevin algorithm.

1. INTRODUCTION

The discrete, or life table, approach to survival analysis has many advantages, for modelling as well as computation. The principle of a life table is simple: the time

*Correspondence to: Philippe Lambert, Université catholique de Louvain, Institut de statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve (Belgium). E-mail: lambert@stat.ucl.ac.be Phone: +32-10-47.28.01 Fax: +32-10-47.30.32

axis is divided into relatively small intervals and in each interval we count the number at risk an the number of events. The ratio of the two gives crude estimate of the hazard. For small data sets this estimate generally will be too erratic to be useful. However, if we introduce smoothing, very useful results can be obtained. Efron [1] uses a small set piecewise polynomial function, while Tutz *et al.* [2] and Kauermann [3] use penalized B-splines (see also the references in [3]). The discrete intervals reduce survival analysis to a Poisson log-linear model. We are dealing with proper likelihoods. Multiple events, complicated risk patterns and large data sets are handle with ease.

In this manuscript we use the smoothed life table approach to develop Bayesian models for varying coefficient survival models. In the standard proportional hazards model the (linear) effect of a covariate on the log-hazard is constant over time. The VCM (varying coefficient model) allows it to change gradually. Smoothness is introduced by means of P-splines: regression on a basis of B-splines in combination with a roughness penalty on their coefficients [4]. The Poisson VCM does not allow many useful analytic results, so we have to rely on simulation to estimate parameters. In the Bayesian setting the penalty translates to an (informative, normal) prior on differences of coefficients. The variance of the prior corresponds to the (inverse of the) weight of the penalty. The beauty of the Bayesian approach is that parameters like this are included in the estimation scheme and so all uncertainties are quantified.

Bayesian P-splines were studied by various authors to describe the conditional mean of a normal response. Berry *et al.* [5] use them in normal regression models. A similar approach is proposed by Lang *et al.* [6] to build additive and varying coefficient models, with extensions to be able to deal with spatially-correlated responses. Bayesian P-splines were also considered in non-normal contexts, see e.g. [7], [8] and the very interesting review paper by Fahrmeir [9] on semiparametric Bayesian function estimation. Alternative Bayesian spline approaches propose to estimate the number of knots and their locations, see e.g. Denison *et al.* ([10], Chap. 3), [11] and [6] for a comparison with Bayesian P-splines. It will be not be considered here.

The plan of our paper is as follows. The lifetime approach to survival and the associated VCM model are reviewed in Section 2. Notations are introduced there. The penalized likelihood and the gradients associated to the P-splines approach are derived. The corresponding Bayesian formulation is proposed in Section 3. The Langevin-Hastings algorithm is explained and suggested to update blocks of splines parameters. A re-parametrization strategy is developed to improve the mixing of the generated chain. We conclude our presentation in Section 4 with the illustrative analysis of a clinical study and a discussion in Section 5.

2. THE LIFE TABLE APPROACH

2.1. Basic ideas

Here we review how one can approach survival problems without covariates in an actuarial way. The time axis is divided into intervals (named *bins*) and in each bin the number of people at risk is counted as well as the number of events. The relationship between time and probability of an event can then be estimated with a parametric or semi-parametric model. See [1] for details.

Let $\tilde{r}_j$ be the number of persons at risk in interval j and let $\tilde{y}_j$ be the number of events. Then a basic model for the number of events assumes a binomial distribution for it, $\tilde{Y}_j \sim \text{Bin}(\tilde{r}_j, \pi_j)$. One could further assume that

$$\log \frac{\pi_j}{1-\pi_j} = \eta_j = \sum_k b_{jk}\zeta_{0k},$$

where the elements b_{jk} are values of basis functions at the midpoints of the bins. The basis functions might be global functions of time, or local splines.

It is advantageous to switch to a Poisson model, $(\tilde{Y}_j|\tilde{R}_j > 0) \sim \text{Pois}(\mu_j)$, with

$$\log \mu_j = \eta_j = \sum_k b_{jk}\zeta_{0k} + \log \tilde{r}_j$$

For small probabilities π_j, the two models are nearly equivalent and provide MLEs for the parameters ζ_{0k}'s with negligible numerical differences. But the advantages of additivity on the log scale with the Poisson model are large, as will be become clear.

It is interesting to develop the model in more detail, as a basis for further exploration. Consider subject i in the sample. We can connect a vector $\boldsymbol{r}_i$ to it, with elements r_{ij}, indicating in which bins it was at risk. We can also introduce a vector $\boldsymbol{y}_i$, with elements y_{ij}. It this subject failed in bin j, y_{ij} will be 1, and all other elements of y_i will be zero. If subject i was lost to follow up, there will be no failure and $\boldsymbol{y}_i$ will contain only zeros.

We can combine all vectors $\boldsymbol{r}_i$ into a matrix R with I rows (the number of subjects) and J columns (the number of bins). Then it is clear that $\tilde{r}_j = \sum_i r_{ij}$ and $\tilde{y}_j = \sum_i y_{ij}$, sums over all subjects, per interval.

Note that this framework can be used directly for all kinds of risks patterns, like staggered entry or intermittent risk: for each subject it can be coded in the patterns of ones and zeros in its row of R. Multiple events are no problem either: just put counts of events per subject, per interval in the rows of Y. In the case of multiple events the Poisson model is the natural one; the binomial model is not applicable there.

Thus, we assume that $(Y_{ij}|R_{ij} = 1) \sim \text{Pois}(\mu_{ij})$ with

$$\log \mu_{ij} = \eta_{ij} = \sum_k b_{jk}\zeta_{0k}$$

Of course, Y_{ij} is identically 0 when $r_{ij} = 0$. The corresponding log-likelihood is

$$l = \sum_i \sum_j r_{ij}(y_{ij} \log \mu_{ij} - \mu_{ij}) = \sum_i \sum_j r_{ij}(y_{ij} \sum_k b_{jk}\zeta_{0k} - \mu_{ij})$$

Using the fact that $r_{ij}y_{ij} = y_{ij}$, because there can be no event when $r_{ij} = 0$, we get:

$$l = \sum_i \sum_j (y_{ij} \sum_k b_{jk}\zeta_{0k} - r_{ij}\mu_{ij}).$$

The derivatives w.r.t. the elements of $\boldsymbol{\zeta}_0$ are given by

$$g_{0k} = \partial l / \partial \zeta_{0k} = \sum_i \sum_j b_{jk}(y_{ij} - r_{ij}\mu_{ij}). \tag{1}$$

The sums can be rearranged, to give

$$g_{0k} = \sum_j b_{jk} \sum_i (y_{ij} - r_{ij}\mu_{ij}) = \sum_j b_{jk}(\tilde{y}_j - \tilde{r}_j\tilde{\mu}_j),$$

because $\mu_{ij} = \tilde{\mu}_j$, is the same for all subjects still at risk. This proves that we can fit a model to the aggregated data $\tilde{r}_j$ and $\tilde{y}_j$: there are no advantages to look at the individual subjects. The only thing that counts is the number at risk in each bin and the number of events. This is not true when we introduce covariates.

Assume that Q (non time-varying) covariates are available for each subject, collected in the I by Q matrix X. They can be introduced in a proportional hazards scheme, giving for a subject at risk

$$\log \mu_{ij} = \eta_{ij} = \sum_k b_{jk}\zeta_{0k} + \sum_q x_{iq}\beta_q \tag{2}$$

and the log-likelihood can be adapted accordingly. For the partial derivatives w.r.t. $\boldsymbol{\zeta}_0$, we find the same formula (1), and for the partial derivatives w.r.t. the elements of β we find

$$g_q = \partial l / \partial \beta_q = \sum_i \sum_j x_{iq}(y_{ij} - r_{ij}\mu_{ij}).$$

If we introduce

$$\phi_i = \exp(\sum_q x_{iq}\beta_q) \quad ; \quad \psi_j = \exp(\sum_k b_{jk}\zeta_{0k}),$$

we see this, we can write $\mu_{ij} = \phi_i\psi_j$. This helps us to simplify the partial derivatives to

$$g_{0k} = \sum_j b_{jk}(\sum_i y_{ij} - \psi_j \sum_i r_{ij}\phi_i) \quad ; \quad g_q = \sum_i x_{iq}(\sum_j y_{ij} - \phi_i \sum_j r_{ij}\psi_j).$$

Let $\bar{\phi}_j = \sum_i r_{ij}\phi_i$. It is the mean relative hazard for interval j, averaged over the subjects being at risk. Let $\bar{\psi}_i = \sum_j r_{ij}\psi_j$. It is the mean relative hazard of subject I, averaged over the bins in which it was at risk. Now we can write, with $y_{i+} = \sum_j y_{ij}$ and $y_{+j} = \sum_i y_{ij}$,

$$g_{0k} = \sum_j b_{jk}(y_{+j} - \psi_j\bar{\phi}_j) \quad ; \quad g_q = \sum_i x_{iq}(y_{i+} - \phi_i\bar{\psi}_i).$$

The simplification of computing prior sums is still applicable to Y, but no longer to R, because we are dealing with sums of $r_{ij}\phi_i\psi_j$: both columns and rows of R are weighted (by ϕ and ψ respectively). Still there are possibilities to reduce the computational work, as will become clear later.

2.2. Varying coefficient model

One arguable hypothesis in the proportional hazard model is the constant ratio of the hazards across strata. This might not be a reasonable hypothesis as, among other reasons, the characteristics of the patients at risk could be changing over time.

We allow the regression coefficients to vary in a smooth way over time. Again, a B-splines representation is used with the same knots as in the description of the baseline hazard. This is not a restriction when a large number of knots is considered. Therefore, β_q in Equation (2) becomes

$$\beta_{qj} = \sum_k b_{jk}\zeta_{qk},$$

yielding

$$\log\mu_{ij} = \eta_{ij} = \sum_k b_{jk}\{\zeta_{0k} + \sum_{q=1}^{Q} x_{iq}\zeta_{qk}\} = \sum_k b_{jk}\sum_{q=0}^{Q} x_{iq}\zeta_{qk},$$

with $x_{i0} = 1$ for all i. Thus, the partial derivative of the log-likelihood w.r.t. the baseline hazard and regression splines parameters is

$$g_{qk} = \partial l/\partial\zeta_{qk} = \sum_i x_{iq}\sum_j b_{jk}(y_{ij} - r_{ij}\mu_{ij})$$

2.3. Penalized likelihood

We follow Eilers and Marx [4] to force the fitted baseline hazard and the time varying regression coefficients to be smooth functions in time. Therefore, we shall work with the penalized likelihood

$$\begin{aligned} l_{\text{pen}} &= l - \sum_q \frac{\lambda_q}{2} \sum_k (\Delta^r\zeta_{qk})^2 \\ &= l - \sum_q \frac{\lambda_q}{2} \boldsymbol{\zeta_q}^T P \boldsymbol{\zeta_q} \end{aligned}$$

where Δ^r denotes the difference operator of order r, $\boldsymbol{\zeta_q}^T = (\zeta_{q1}, \ldots, \zeta_{qK})$ and $\{\lambda_q : q = 0, \ldots, Q\}$ is the set of roughness penalty parameters.

The partial derivatives of the penalized likelihood can easily be obtained using the expressions from the previous section:

$$g_{qk} \;=\; \partial l_{\text{pen}}/\partial\zeta_{qk} = \sum_i x_{iq} \sum_j b_{jk}(y_{ij} - r_{ij}\mu_{ij}) - \lambda_q \sum_s P_{sk}\zeta_{qs} \qquad (3)$$

3. BAYESIAN INFERENCE USING THE LANGEVIN ALGORITHM

A Bayesian simulation framework, based on Monte-Carlo Markov chains (MCMC) methods, is very convenient to estimate spline parameters and credibility envelopes for the associated time-varying regression parameters. In addition, the penalty parameters can be estimated jointly as well as the uncertainties of all estimated quantities. This is a major advantage over the traditional (penalized likelihood) estimation procedure which implicitly assumes that the penalty parameters are fixed known quantities, thereby providing too narrow confidence intervals for the splines parameter estimates.

3.1. Bayesian P-splines

In Bayesian terms, each added penalty can be translated by introducing a prior distribution on the rth order differences of the corresponding B-splines coefficients, i.e.

$$\Delta^r \zeta_{qk} \sim N(0, \tau_q^{-1})$$

(see [5] for a Bayesian implementation of P-splines in normal regression models and [6] for a similar exercise in additive models). Consequently, we propose to multiply the improper prior for the B-splines coefficients by

$$\prod_q \tau_q^{\rho(P)/2} \exp\left\{-\frac{1}{2}\,\tau_q\,\boldsymbol{\zeta}_q^T P \boldsymbol{\zeta}_q\right\}.$$

The inverse variance τ_q $(q = 0, \ldots, Q)$ plays the role of λ_q in the penalized likelihood. A vague prior distribution can be chosen for each, say, a gamma distribution $G(a, b)$, i.e.

$$\pi(\tau_q) = \frac{b^a}{\Gamma(a)} \tau_q^{a-1} \exp(-b\,\tau_q).$$

Taking e.g. $a = b = .0001$ is suitable and clearly expresses our ignorance, while yielding a proper posterior distribution for τ_q.

Therefore, the posterior distribution for the the penalty parameters, $\boldsymbol{\tau} = (\tau_0, \ldots, \tau_q)$, and the splines parameters, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0, \ldots, \boldsymbol{\zeta}_Q)$, is

$$\pi(\boldsymbol{\zeta}, \boldsymbol{\tau}|\boldsymbol{y}) \;\propto\; L(\boldsymbol{\zeta}; \boldsymbol{y}) \prod_q \tau_q^{a+\rho(P)/2-1} \exp\left\{-[b + 0.5\,\boldsymbol{\zeta}_q^T P \boldsymbol{\zeta}_q]\,\tau_q\right\}$$

with the log-likelihood

$$\log L(\boldsymbol{\zeta}; \boldsymbol{y}) = \sum_i \sum_j (y_{ij} \log \mu_{ij} - r_{ij}\mu_{ij})$$

The conditional posterior distributions of the penalty parameters, given the other parameters can easily be shown to be

$$\pi(\tau_q|\boldsymbol{\zeta}_0, \boldsymbol{\zeta}, \boldsymbol{\tau}_{-q}; \boldsymbol{y}) = \pi(\tau_q|\boldsymbol{\zeta}_q) \equiv G\left(a + 0.5\ \rho(P), b + 0.5\ \boldsymbol{\zeta}_q^T P \boldsymbol{\zeta}_q\right). \tag{4}$$

3.2. Sampling using the Langevin-Hastings algorithm

We propose to explore the high-dimensional posterior distribution using Monte-Carlo Markov chains (MCMC), see [12] for an excellent introduction. More precisely, we propose to combine the Metropolis-adjusted Langevin algorithm (MALA, [13]) to sample blocks of transformed B-splines coefficients and the Gibbs algorithm to sample the inverse variance (penalty) parameters.

The MALA algorithm enables to sample from potentially high dimensional posterior distribution using Monte Carlo Markov where each proposal is made using the gradient of the log posterior distribution at the current state. More precisely, if $\pi(\boldsymbol{\theta})$ is the posterior distribution and $\boldsymbol{\theta}^t \in \mathbb{R}^K$ the state of the chain at iteration t, then the proposal $\boldsymbol{\theta}$ for the next state is obtained by a random generation from the K-variate normal distribution $N_K(\boldsymbol{\theta}^t + 0.5\ \delta\ \nabla \log \pi(\boldsymbol{\theta}^t), \delta I_K)$ where I_K is the K dimensional identity matrix and δ a carefully chosen variance parameter. This proposal is accepted with probability

$$\min\left\{1, \frac{\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}^t)} \frac{q(\boldsymbol{\theta}, \boldsymbol{\theta}^t)}{q(\boldsymbol{\theta}^t, \boldsymbol{\theta})}\right\}$$

where

$$q(x, y) = (2\pi\delta)^{-K/2}\ \exp\left[-\frac{1}{2\delta}\left\|\boldsymbol{y} - \boldsymbol{x} - 0.5\delta\ \nabla \log \pi(\boldsymbol{x})\right\|_2^2\right]$$

i.e. $\boldsymbol{\theta}^{t+1}$ is set equal to $\boldsymbol{\theta}$ if accepted and to $\boldsymbol{\theta}^t$ otherwise.

Roberts and Rosenthal [14] have shown that the relative efficiency of the algorithm can be characterized by its overall acceptance rate, independently of the target distribution. The asymptotic optimal value for that last quantity is 0.57 with acceptance probabilities in the range $(0.40, 0.80)$ still reasonable. That information can be used to choose the above δ parameter by increasing (decreasing) δ when the observed acceptance rate is larger (smaller) than the targeted value.

3.3. Block use of the MALA algorithm

We prefer the MALA algorithm over the usual random-walk Metropolis-Hastings algorithm as its use of the local properties of the log posterior distribution was

found to provide more quickly converging and better mixing chains while remaining simple to implement.

Let us detail our proposal to build a sample from the posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\zeta}_0, \ldots, \boldsymbol{\zeta}_Q, \tau_0, \ldots, \tau_Q)$. At iteration $(t+1)$:

1. *Langevin-Hastings steps.* `For q from 0 to Q`: in block q, generate $\boldsymbol{\zeta}_q$ using the MALA algorithm with $\boldsymbol{\zeta}_{r<q} = \boldsymbol{\zeta}_r^{t+1}$, $\boldsymbol{\zeta}_{r>q} = \boldsymbol{\zeta}_r^t$ and $\boldsymbol{\tau} = \boldsymbol{\tau}^t$. An expression for the components of the gradient is given in Equation (3). The value of the variance parameter δ_q in the normal proposal density associated to block q, see Section 3.2, must be tuned to achieve an acceptance rate in the above metioned range. These $(Q+1)$ tuning parameters are chosen during the burn-in iterations.

2. *Gibbs steps.* `For q from 0 to Q`: generate τ_q^{t+1} from $\pi(\tau_q|\boldsymbol{\zeta}_q^{t+1})$, cf. Equation (4).

The theory ensures that after a sufficiently large number of iterations, say M, $\{\boldsymbol{\theta}^{M+1}, \boldsymbol{\theta}^{M+2}, \ldots\}$ can be considered as a random sample from the posterior distribution.

In practice, updating the B-splines parameters by blocks in the Langevin-Hastings steps is essential to get convergence of the chains after a few thousands iterations. Convergence speed was much larger than with a block random-walk Metropolis-Hastings algorithm.

We also found that a careful re-parametrization of the posterior distribution within blocks of B-splines parameters substantially boosts the convergence of the corresponding chain. Indeed, P-splines parameters associated to close knots are forced to take similar values leading to large cross-correlations between successive components of the chain: it limits the mixing of the chain and, hence, slows the convergence of the Markov chain to the target posterior distribution. Consequently, after running the above algorithm for R (say, one thousand) iterations and having managed to have the acceptance rates within the above prescribed intervals, we propose to use the following strategy to re-parameterize the posterior distribution within blocks:

Block reparametrization: `For q from 0 to Q`, compute the mean $\bar{\boldsymbol{\zeta}}_q$ and the variance-covariance matrix S_q of $\{\boldsymbol{\zeta}_q^1, \ldots, \boldsymbol{\zeta}_q^R\}$. Reparametrize the posterior distribution using $\boldsymbol{\zeta}_q'$ where

$$\boldsymbol{\zeta}_q = S_q^{1/2}\boldsymbol{\zeta}_q' + \bar{\boldsymbol{\zeta}}_q$$

The initial algorithm can be used again to sample from the re-parameterized posterior distribution. If required, that re-parameterization can be repeated until the cross-correlations between the different components of the chain are smaller than some threshold. The usual diagnostics can be used to assess convergence.

4. ILLUSTRATION ON CLINICAL DATA

We illustrate our the method with the analysis of a clinical study involving 358 patients with advanced ovarian cancer, as described in [15]. Survival/censoring times vary between 9 and 2729 days, with about 25% censoring. The covariates are

- KARN: the Karnovsky performance status measuring the ability of the patient to lead her daily life at the start of therapy, coded as an integer from 0 (100% functional ; 38% of the patients), 1 (30%), 2 (13%), 3 (13%) to 4 (< 60% functional ; 6%).
- DIAM: diameter of residual tumor after surgery, coded as 0 (8% of the patients), 1 (19%), 2 (14%), 3 (19%), 4 (41%) (small to large).
- FIGO: stage of the cancer, coded as 0 (FIGO = III ; 73% of the patients) and 1 (FIGO = IV ; 27%).

All these covariates were centered and scaled to zero mean and unit standard deviation. The block Langevin-Hastings algorithm described in Section 3 was used with four blocks of parameters, the first block corresponding to the baseline hazard and the other three to the standardized covariates. Five equidistant knots were considered to span the observed range for the survival data, yielding eight parameters per block with cubic B-splines.

The Langevin variance parameters $\delta_0, \ldots, \delta_3$ of the proposal distributions associated to each of the four blocks were increased or decreased several times after each of a few sets of hundreds iterations to yield, finally, acceptance rates in the above suggested $(0.40, 0.80)$ range, see [14].

Then, a few thousands extra burn-in iterations were considered to estimate the correlation structure within each block. These structures were used to reparameterize the posterior distribution, as suggested at the end of Section 3.3. That strategy was repeated two or three times until the cross-correlations of the chain and the autocorrelations of the chain for the transformed parameters were found reasonably low.

A final chain of length 20,000 corresponding to the back-transformed B-splines parameters for the baseline hazard and for DIAM is plotted in Figures 1 and 2 respectively.

The baseline hazard, the baseline survival curve and the time-varying regression coefficients evaluated at the estimated mean of the posterior distribution are plotted in Figure 3. The 90% credibility envelopes for the time-varying regression coefficients are in Figure 4. It suggests that the patient performance status before surgery is just a very short term survival predictor. The diameter of the residual tumor after surgery has a significant negative effect on survival for more than 3 years, while a FIGO stage IV is associated to patients with a significantly worse prognosis during most of the observation period.

5. DISCUSSION

The Bayesian formulation of P-splines, implemented with modern simulation techniques, allows the implementation of sophisticated semi-parametric models while accounting for the uncertainty involved in the choice of smoothing parameters. The large number of parameters is not a problem in non-normal settings provided that the local properties of the log posterior density are used to derive proposals in Metropolis-Hastings steps. Here, the gradient of that function was used to update blocks of parameters. Re-parametrization suggested during the burn-in period was essential to have a well mixing chain. Alternatively, one could use MCMC simulation techniques based on iteratively weighted least squares (IWLS) proposals in the Metropolis-Hastings steps, see [16].

Our MCMC approach relies on three elements to achieve efficient computation: 1) use of the Langevin-Hastings algorithm to get good proposals in the steps of the Markov chain, 2) block-wise updating of P-spline coefficients and 3) proper rotation and scaling of parameters (after a burn-in period) to improve mixing. The result is an R program interfacing a routine in C that can handle a data set of 400 observations in about 5 minutes (20,000 simulations after burn-in) with a Pentium IV 3.0 Ghz. This leads to quite acceptable computation times for moderately sized data sets.

The Poisson approach to survival has multiple advantages, among which the easy modelling of sophisticated risk patterns and multiple events, the possibility to include time-varying frailty components.

We are currently working on a similar project with accelerated failure time models: these results will be reported elsewhere.

Finally, note that the Bayesian setting allows an easy specification and estimation of frailty components in the model. This is a useful addition to deal with multicentre studies or with recurrent events.

ACKNOWLEDGMENTS

We thank Jean-Luc Marrion (UCL) for his technical help at an hectic stage of this project. Financial support from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

REFERENCES

1. B, E. 'Logistic regression, survival analysis, and the kaplan-meier curve', *Journal of the American Statistical Association*, **83**, 414–425 (1988).

2. Tutz, G. and Binder, H. 'Flexible modelling of discrete failure time including time-varying smooth effects', *Statistics in Medicine*, **23**, 2445–2461 (2004).

3. Kauermann, G. 'Penalised spline smoothing in multivariable survival models with varying coefficients', *Computational Statistics and Data Analysis* (2004, in press).

4. Eilers, P. H. C. and Marx, B. D. 'Flexible smoothing with B-splines and penalties (with discussion)', *Statistical Science*, **11**, 89–121 (1996).

5. Berry, S. M., Carroll, R. J. and Ruppert, D. 'Bayesian smoothing and regression splines for measurement error problems', *Journal of the American Statistical Association*, **97**, 160–169 (2002).

6. Lang, S. and Brezger, A. 'Bayesian P-splines', *Journal of Computational and Graphical Statistics*, **13**, 183–212 (2004).

7. Lambert, P. Archimedean copula estimation using Bayesian splines smoothing techniques, Discussion paper, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 2004.

8. Fahrmeir, L., Kneib, T. and Lang, S. 'Penalized structured additive regression for space-time data: a Bayesian perspective', *Statistica Sinica*, **14**, 715–745 (2004).

9. Fahrmeir, L. Recent advances in semiparametric Bayesian function estimation, Discussion Paper 137, Ludwig Maximilians University Munich, 1998. http://www.stat.uni-muenchen.de/sfb386/papers/dp1998.html.

10. Denison, D., SMith, A. F. M. and Mallick, B. K. *Bayesian methods for nonlinear classification and regression*, Wiley, 2002.

11. Biller, C. 'Adaptive Bayesian regresion splines in semiparametric generalized linear models', *Journal of Computational Statistics and Data Analysis*, **12**, 122–140 (2000).

12. Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. *Markov chain Monte Carlo in Practice*, Chapman and Hall/CRC, 1996.

13. Roberts, G. O. and Tweedie, R. L. 'Exponential convergence of Langevin distributions and their discrete approximations', *Bernoulli*, **24**, 341–363 (1996).

14. Roberts, G. O. and Rosenthal, J. S. 'Optimal scaling of discrete approximations to Langevin diffusions', *Journal of the Royal Statistical Society, Series B*, **60**, 225–268 (1998).

15. Verweij, P. J. M. and Van Houwelingen, H. C. 'Cross-validation in survival analysis', *Statistics in Medicine*, **12**, 2305–2314 (1993).

16. Gamerman, D. 'Efficient sampling from the posterior distribution in generalized linear models', *Statist. Comput.*, **7**, 57–68 (1997).

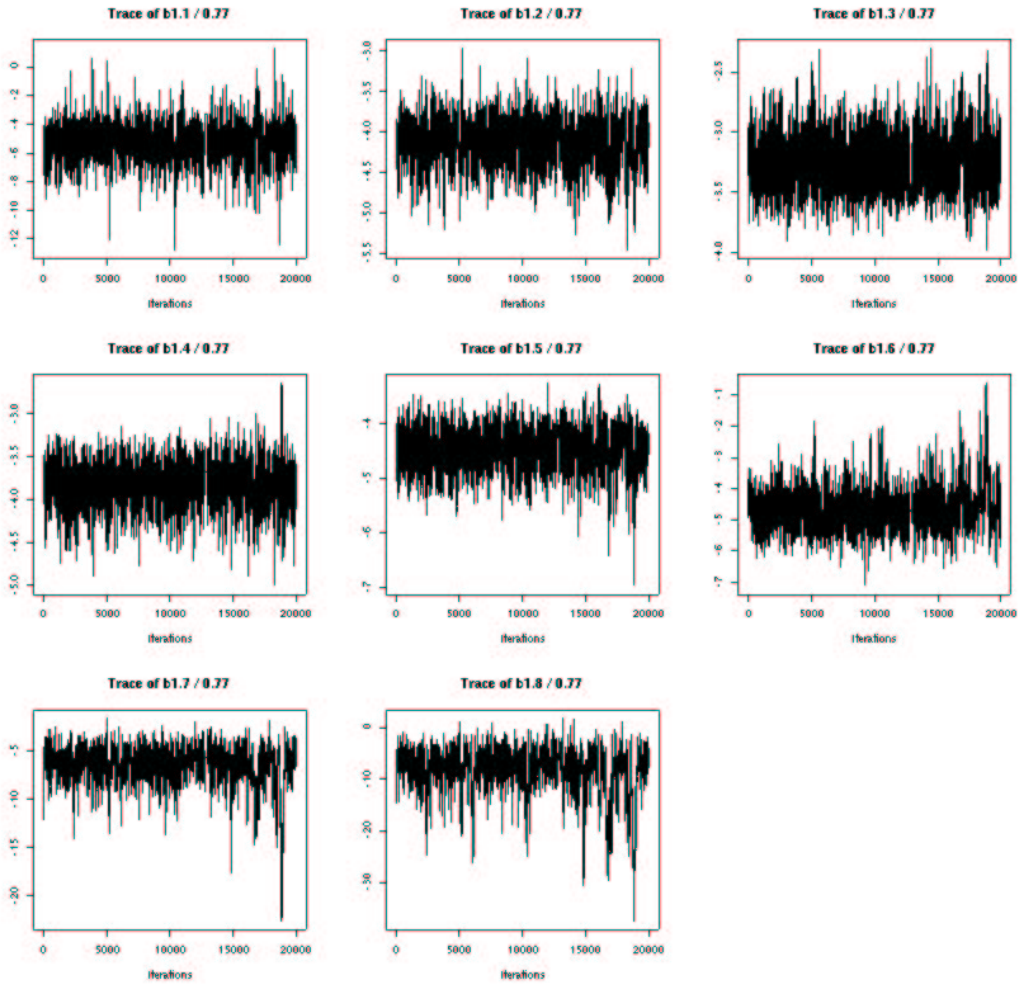


Figure 1: Langevin-Hastings chain for the baseline hazard B-splines parameters.

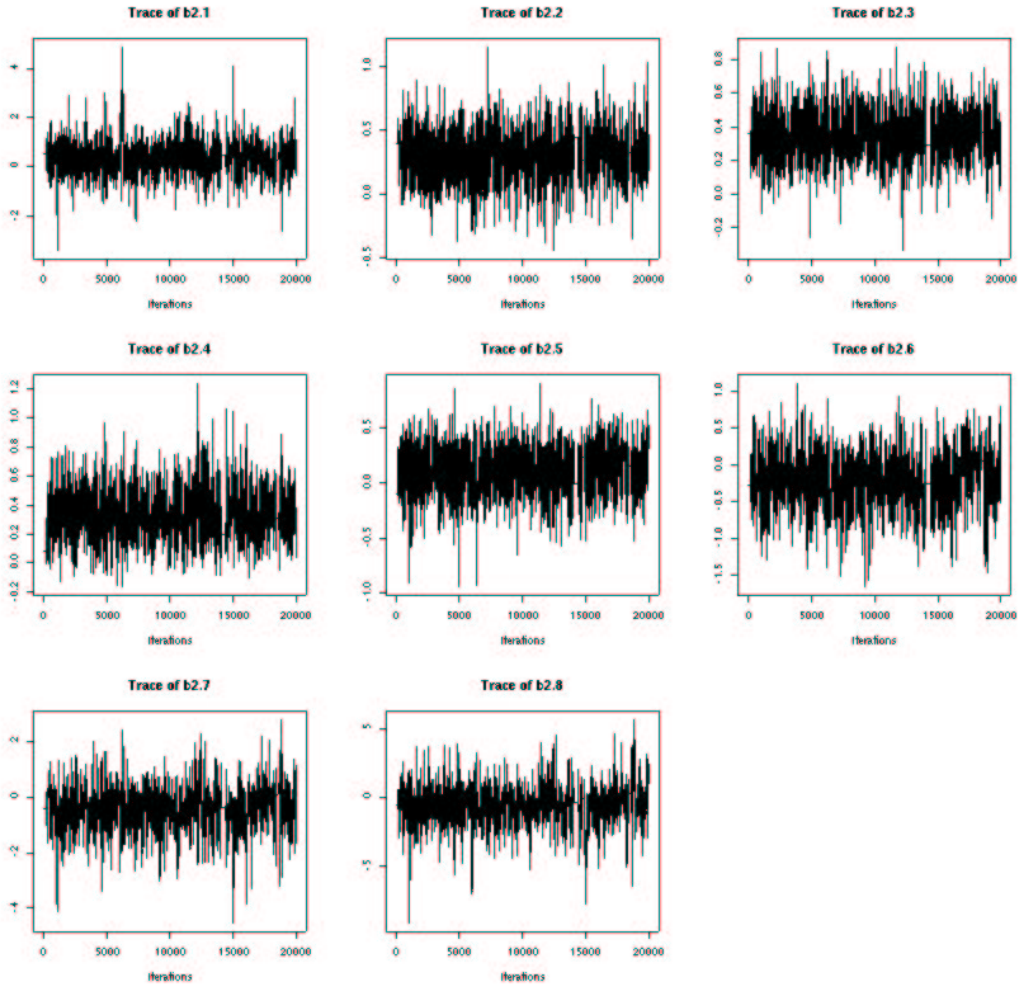


Figure 2: Langevin-Hastings chain for the B-splines parameters associated to covariate `diam`.

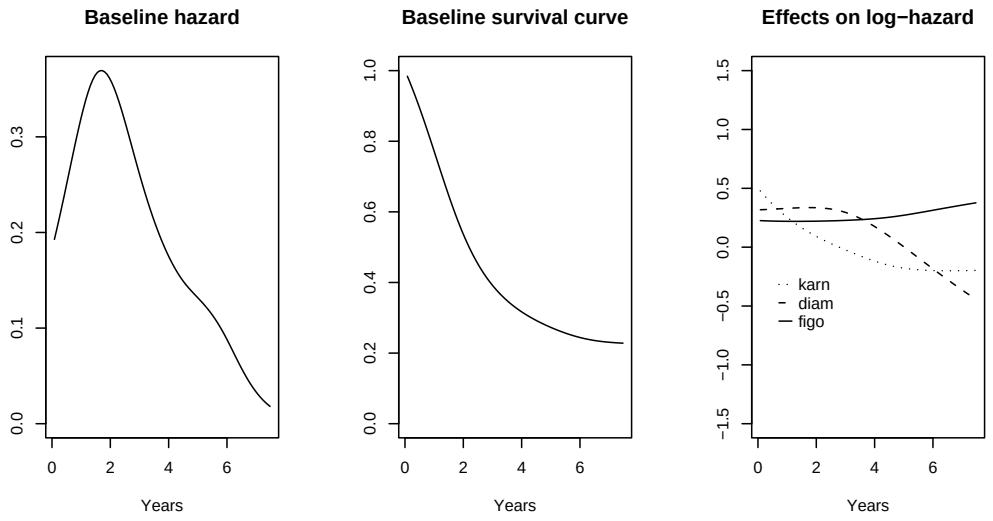


Figure 3: Fitted baseline hazard, baseline survival curve and time-varying regression coefficients.

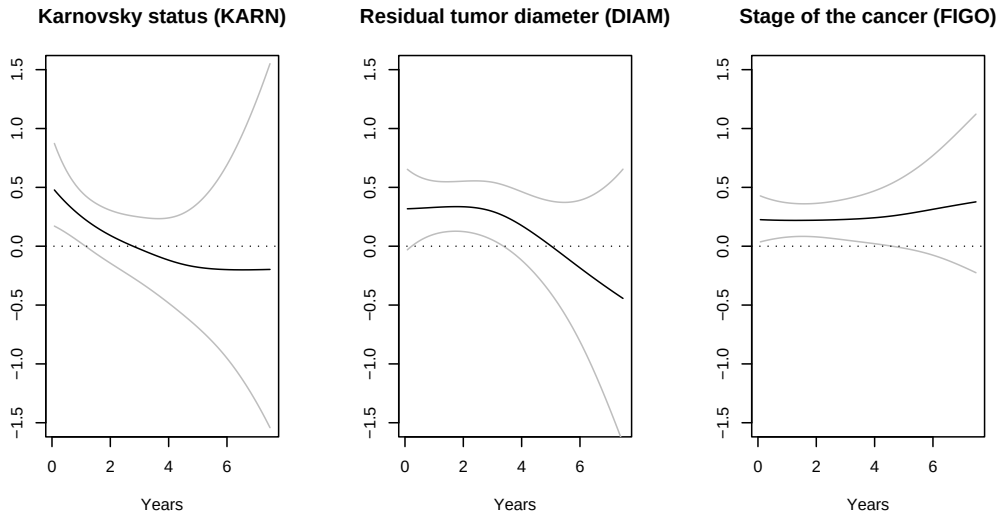


Figure 4: 90% credibility envelope for the time-varying regression coefficients: posterior mean (solid black line) ; posterior 5.0% and 95.0% quantiles (solid grey lines).