MODELLING TIME-VARYING HAZARDS AND COVARIATE EFFECTS

Mike West
ISDS, Duke University

DP# 91-A09
MODELLING TIME-VARYING HAZARDS AND COVARIATE EFFECTS

MIKE WEST
Institute of Statistics & Decision Sciences
Duke University
Durham, NC 27706

ABSTRACT. Models for survival analysis in the presence of possibly time-varying covariate effects are constructed as piecewise proportional hazards models in which the regression parameters 'evolve' over the survival time axis. Suitable models for exploratory data analysis involve parameters changing according to random walk mechanisms, modelling the prior view that parameter values in consecutive intervals of the time axis are more likely to be close than those well separated in time. Analysis uses Bayesian updating to approximate posterior distributions for such time-varying parameters, providing estimates of their values over time - the parameter 'trajectories', and also approximations to posterior predictive survival distributions for further cases. Illustration is provided here in analyses of nasopharynx cancer survival data.

1. Model Description

The starting point for development is the familiar proportional hazards model, denoted PH for short. The survival time of an individual case under study has a distribution described via the hazard function \( \lambda(t) = \lambda_0(t) \exp(z^\top \beta) \), for survival time \( t > 0 \). Here \( z \) is the vector of covariates of the case at time \( t \) - often these are constant in time, as is implicit in the notation here, though this is not necessary; \( \beta \) is the vector of parameters determining the effects of covariates on the hazard, and \( \lambda_0(t) \) is the baseline hazard function - just the hazard for a case whose covariates are all zero, \( z = 0 \). In this standard model, the baseline hazard is arbitrary and the effects of the covariates assumed to be constant in time since \( \beta \) is not time dependent. In some applications, variation in covariate effects are encountered, consistent with \( \beta \) changing, or evolving over the survival time axis. Previous modelling of time-varying effects have included piecewise constant regression parameters, taking \( \beta \) constant within each interval of a prespecified grid over the time axis, essentially applying separate proportional hazard models within each such interval (Anderson and Senthilvelan (1982), Gore, Pocock and Kerr (1984), for example). This idea is natural, though allowing the regression parameters to change without restriction between intervals is undesirable; typically, effects will be expected to vary smoothly over time unless there is additional information about interventions or protocol changes that suggest otherwise. Thus the 'piecewise' proportional hazards model may be supplemented by a prior distribution, or class of priors, over the collection of regression parameters to introduce 'smoothing' of the time variation, or time 'trajectories', of the parameters. Building on dynamic Bayesian models for time
series (West and Harrison (1989)), Gamerman (1987, 1991) introduces classes of models involving a variety of such smoothness prior, and develops their Bayesian analyses; further development and applications appear in Gamerman and West (1987), Gamerman, Pole and West (1987), and West (1987). One particular model within this class, the simplest and most useful in exploratory data analysis, provides a discrete ‘random walk’ evolution of the regression parameters over time, as follows.

In a general time-varying model, indicate time dependence of effects by subscripting the parameter $\beta$ by $t$; then the individual’s hazard function may be written as $\lambda(t) = \exp(x'\theta_t)$ where $x' = (1, x')$ and $\theta'_t = (\log(\lambda_0(t)), \beta'_t)$. A piecewise proportional hazards model is defined over a prespecified grid of the survival time axis, based on intervals $I_j = (t_{j-1}, t_j]$ ($j = 1, 2, \ldots$) where the specified time points $t_j$ intervals are obviously ordered. Then when $t$ lies in the interval $I_j$, the regression parameter vector is $\theta_t = \theta_j$, the hazard function

$$\lambda(t) = e^{x'\theta_j}, \quad (t \in I_j),$$

(j = 1, \ldots).

Time variation in the effects of covariates is now allowed since the values $\theta_j$ may be different in different intervals $I_j$. Varying degrees of smoothness of the evolution of the parameter can be modelled by relating consecutive values of the regression parameter vector via

$$\theta_j = \theta_{j-1} + b_j,$$  \hspace{1cm} (1)

say, where $b_j$ is a random term, or evolution ‘error’ (cf. the time series genesis and terminology of West and Harrison (1989)). One simple but flexible possible structure has the $b_j$ independently distributed with zero means and some variance matrices controlling the amount of time variation; this random walk structure allows for changes but does not anticipate the directions of changes. If these variances are large, then the effects may change appreciably between intervals, otherwise the model is similar to the standard PH model; one of the most immediate prospects is that of specifying rather small variances so as to generalise to a small ‘neighbourhood’ of the PH model, providing opportunity to assess the adequacy of the latter. Let $B_j = V(b_j)$ be the variance of the change in interval $j$. Possible variance structures are considered in Gamerman, West and Pole (1987).

Note that changes between intervals in the individual’s regression vector may be accommodated, simply extending the vector $x$ to $x_j$ in $I_j$, as necessary in the application to unemployment studies in Gamerman and West (1987). We ignore this possibility here.

2. Analysis Summary

Suppose we have data $(y_i, c_i, x_i) (i = 1, \ldots, n)$ on $n$ individuals, where $y_i$ is the observed time; $c_i$ is binary with $c_i = 1$ implying $y_i$ survival/failure at time $y_i$ and $c_i = 0$ implying that the survival time is (non-informatively) censored at $y_i$; and $x_i$ the vector of covariates. Partition the survival time axis as above with some $s$ intervals $I_j = (t_{j-1}, t_j] (j = 1, \ldots, s)$, taking $t_0 = 0$ and $t_s = \infty$. Then $y_i$ has survival distribution defined by hazard function

$$\lambda_i(t) = e^{x'\theta_j}, \quad (t \in I_j),$$

(j = 1, \ldots).

Analysis developed in the preceding references is based on approximate updating of posterior distributions for the $\theta_j$, performed sequentially by processing data in each interval
separately. Define information sets \( D_j = \{ D_{j-1}, H_j \} \) \((j = 1, \ldots)\), where \( H_j \) is the data observed in \( I_j \) and \( D_0 \) represents initial prior information (which includes the values of the \( x_i \)). Thus, for each \( j \),

\[
H_j = \{ (y_i, c_i) \mid y_i \in I_j \}.
\]

Proceeding to analyse the data sequentially, we successively update posteriors \( p(\theta_j | D_j) \) only approximately, and partially, in terms of posterior moments

\[
m_j \overset{\text{def}}{=} E(\theta_j | D_j) \quad \text{and} \quad C_j \overset{\text{def}}{=} V(\theta_j | D_j) \quad (2)
\]

using the dynamic generalised linear modelling algorithm of Gamerman (1987, 1991), based on that in West, Harrison and Migon (1985). At the end of interval \( I_{j-1} \), suppose preceding analysis has provided approximate moments \( m_{j-1} \) and \( C_{j-1} \) for \( (\theta_{j-1} | D_{j-1}) \). Evolution to the next interval via equation (1) does not change the mean, but increases the variance by the additional factor \( B_j \), so that

\[
m_{j-1} = E(\theta_j | D_{j-1}) \quad \text{and} \quad C_j \overset{\text{def}}{=} V(\theta_j | D_{j-1}) = C_{j-1} + B_j; \quad (3)
\]

these are the moments of the parameter vector in \( I_j \) prior to observing and processing the data \( H_j \) in that interval. The dynamic generalised linear model updating calculations update these prior moments to the posterior moments \( m_j \) and \( C_j \); full details of the algorithm appear in section 4.2 of Gamerman (1991), and are not given here.

The posterior means \( m_j \) \((j = 1, \ldots, s)\) provide estimates of the time trajectories of the regression parameters, and uncertainties about these estimates may be derived from the sequence \( C_j \) \((j = 1, \ldots, s)\). Monitoring these trajectories provides indications of the extent of influence of the data in each interval on the posterior for the current parameter vector. Having processed all the data up to and including \( H_s \), the complete information set \( D_s \) contains information that is clearly relevant to inference about previous parameters \( \theta_j \) for \( j < s \), but that was not available for inference at time \( t_j \). Thus the on-line posterior summaries \( m_j \) and \( C_j \) may be revised to provide retrospective summaries of the ‘smoothed’ posterior distribution \( p(\theta_j | D_s) \) \((1 \leq j \leq s)\). This is done using smoothing, or filtering algorithms similar to the standard algorithms in dynamic linear models (West and Harrison, 1989); development based on linear Bayes’ estimation in dynamic generalised linear models is detailed again in Gamerman (1987). Performing these computations leads to revised summaries of the approximate posterior \( p(\theta_j | D_s) \) again in terms of moments,

\[
m_j^* \overset{\text{def}}{=} E(\theta_j | D_s) \quad \text{and} \quad C_j^* \overset{\text{def}}{=} V(\theta_j | D_s). \quad (4)
\]

Inferences about time variation in the parameter sequence should be based on the ‘smoothed’ estimates \( m_j^* \) and the associated uncertainties measured by \( C_j^* \).

Further inferences of key importance concern out-of-sample predictive distributions. For an hypothetical individual with survival time \( y \) and specified covariate vector \( x \), we require the predictive distribution \( P(y | D_s, x) \) \((y > 0)\). Useful approximations to these distributions are developed in section 5 of Gamerman (1991), where computational details may be found.

Finally, note that analysis requires initialisation using prior information \( D_0 \) summarised in terms of a prior distribution for \((\theta_1 | D_0)\). Consistent with the partial development of posteriors in terms of approximate means and variance matrices above, this initial prior
is specified in similar terms, requiring moments $m_0 = E(\theta_1|D_0)$ and $C_0^* = V(\theta_1|D_0)$ (in notation consistent with (3)).

3. Illustration

3.1. DATA AND MODEL

Data in West (1987) provide information on 181 nasopharynx cancer patients whose cancer careers, culminating in either death (127 cases) or censoring (54 cases), are recorded to the nearest month and range from 1 to 177 months. For each case a variety of covariates are available, none of which are viewed as subject to change during the career of the patient. Previous, alternative analyses of the data, and some further exploratory investigations, serve to indicate the importance of a subset of the covariates in describing the observed variation in survival times. Discussion here is restricted to analyses using only this reduced set of covariates.

The analyses are based on the following five covariates: (1) Sex, a classifying factor indicating the sex of the patient (0 for male, 1 for female); (2) Age, the age of the patient at time $t = 0$, the start of monitoring of the cancer career of that patient (standardised to have zero arithmetic mean and unit standard deviation across all patients in the study); (3) Dosel, an average measure of the extent of radiotherapy treatment to which the patient has been subjected (also standardised, as with Age); (4) Tumor1, a measurement of the extent of the cancer (in terms of an estimate of the number) of cancerous cells, taking values 1, 2, 3 or 4; (5) Tumor2, a measure similar to Tumor1 though from a different X-ray section, again taking values 1, 2, 3 or 4. These Tumor variables are measures of tumour growth at the start of monitoring, hence proxies for tumour lifetime. Higher levels of each are therefore expected to be consistent with increased hazards.

The data on the original variables, along with observed death times or times of censoring $y_i$, and the related death/censoring indicator variables $c_i$, are given in West (1987), and are available on request.

Complete model specification requires the discretisation of the time axis into intervals $I_j$, the evolution variance matrices $B_j$ and the initial prior moments (3) at $j = 1$. On time axis discretisation, various discussions can be found in the earlier references, and in related works including Breslow (1974), and Kalbfleisch and Prentice (1973). One feature of the cancer data, not uncommon in general, is the preponderance of deaths in the early stages, so such years will be informative about any change in parameters during those year; this suggests shorter intervals over the first few years in an attempt to identify any such changes. In later years, deaths are fewer so the data is less informative about covariate effects and longer intervals are appropriate. Although parameters are modelled as possibly varying according to (1), there is no information from such intervals relevant to estimating change and the posterior means will remain constant over such intervals, though the posterior variances will change slightly (Gamerman, 1991). Grids comprised of intervals whose endpoints are observed survival/failure times are commonly used in the literature, and discussed in the references above, on the basis that only at such times is potential change detectable. A version of this fixes intervals, typically of unequal lengths, to contain the same number of observed times; then early intervals will be short, later intervals much longer. This can be rationalised in advance of observing the data as a sensible prior model structure if the
total number of cases (eg. patients in the study, or items on test) is known in advance, and supported using the above argument about information content of data per interval versus ‘decay’ of information over the interval through the evolution equation. The analyses reported here use this discretisation, with interval endpoints defined at every eighth observed death; this gives 16 intervals, with endpoints 3,5,7,9,10,12,13,16,19,21,25,29,35,45,60 and ∞ (in months), the final interval including just seven observed death times, though many censored times. Other sensible discretisations lead to analyses differing in small detail quantitatively, though not in important ways qualitatively; in particular, the discretisation in West (1987) has \( t_j = 4 \) \((j = 1, \ldots, 10)\) and \( t_j = 12 \) for \( j > 10 \), and results in inferences very similar to those reported below in similar models.

3.2. INITIAL PRIORS

The parameter vector \( \theta_j \) has six elements: the baseline hazard, the effect of Sex (measuring the difference on the log-hazard scale between female and male patients), and the coefficients of the Age, Dosel, Tumor1 and Tumor2 variables. It is plausible to expect that Age, Tumor1 and Tumor2 will be positively associated with cancer hazards. Dosel, on the other hand, may be expected to be beneficial, having a negative coefficient. The effect of Sex is not anticipated initially in analyses here (though this might be altered if suitable expert opinion about a possible sex-link were available). These considerations are taken into account in the initial prior estimates assigned below in several analyses. We use the initial prior specified as follows — an appropriately diffuse prior but consistent with the above general views. Other priors might be used, of course; it is our purpose here to illustrate the use of the class of models in estimating parameter time trajectories and to compare models that differ primarily in the extent of time-variation modelled, so that a fixed prior across models is appropriate (so long as it is reasonably diffuse). Initial prior means and standard deviations are assigned to each of the parameters in the first interval \( I_1 \) as follows.

The baseline constant has a prior mean of \(-3\), with s.d. \(2.5\), implying a very wide likely range and representing an appropriately diffuse prior specification for the baseline hazard. The Sex effect has a prior mean of \(0\) with s.d. \(0.25\). Taking a 2 s.d. interval and converting to the hazard scale, this implies a likely interval of approximately 0.4 to 2.7 for the multiplicative effect of Sex, with most likely value 1. Thus, although no effect is expected, the prior is diffuse enough so that a halving or doubling of the hazards across Sexes is not implausible. Age effect has a prior mean of \(0.5\) with s.d. \(0.25\). As with Sex, the implied range of plausible multiples of the baseline hazard is large, but now the favoured values are positive, consistent with the prior expectation that hazards will tend to be higher for older patients. Dosel has a prior mean of \(-0.5\) with s.d. of \(0.25\). This prior for the coefficient of standardised Dosel measurements is similar to that for the Age coefficient, although favours negative values consistent with the expectation that the treatment is beneficial. The prior means of each of Tumor1 and Tumor2 are taken as \(0.5\) with common s.d \(0.25\). Alternative models might anticipate interaction between the two measures of tumour growth, or incorporate prior correlation between these parameters, though this is not pursued here. With this proviso, all prior covariances are taken at zero. This is rather innocuous in view of the diffuseness input via relatively large standard deviations for the parameters, and any other prior correlation structure would be rather difficult to quantify.
3.3. COMPARISON OF MODELS

The final component of model specification is the extent and nature of variation anticipated in the parameters over time, modelled through the evolution variance matrices $B_j$ of the terms $b_j$ in (1). Obviously these values are crucial to the analysis in so far as assessing time variation in effects is concerned. Sensitivity to the values is explored by examining the fit of classes of models that differ only in these variances. Model fit is assessed using standard Bayesian techniques, based on comparison of predictive performance between models. Any model analysis provides a single numerical measure of goodness of fit, namely the observed value of the predictive density of the data under that model, $p(y_1, \ldots, y_n|D_0)$; Gamerman (1987) describes calculation of this quantity in the sequential analysis of the model, and illustrates its use in selection of covariates for inclusion in the model. This is essentially a measure of how well the model predicts from one interval of the time axis to the next (rather than a measure of retrospective fit of the model to the data), whose value provides a point on the likelihood function over model space. Evaluating such quantities across models that differ only in certain parameter values allows a subjective comparison of the models to be made. The software of Gamerman, West and Pole (1987) provides these model (log-)likelihoods as a by-product of the sequential analysis. Though used here exclusively, note that this is just one simple, and crude, summary measure of model ‘fit’, and other approaches to model assessment and criticism might be developed in addition.

Comparison is made across models differing only through the degree of time-variation modelled in parameters; the included covariates, time axis discretisation and prior moments are fixed across models.

Model 1: We begin with the ‘static’ model in which $B_j = 0$ for all $j$ so that the parameters are fixed and we have an exponential baseline PH model. Following analysis, the model log-likelihood is computed as $-617.9$, parameter estimate summaries (just the means from $m_j$ and standard deviations from $C_j$) appear in the first row of Table 1.

Model 2: Models which allow the baseline hazard to vary but constrain the remaining parameters to be fixed over time have $B_j = diag(B_{j,1}, 0, 0, 0, 0, 0)$, where $B_{j,1} > 0$ is the evolution variance of the baseline log-hazard parameter between intervals $I_{j-1}$ and $I_j$. If this is appropriately large, then the baseline hazard is essentially unrestricted, allowed to change markedly between intervals. Such a structure is essentially a standard PH model with an otherwise unrestricted piecewise exponential baseline survival distribution. One such model is summarised here (and why this particular model is chosen is discussed below), with $B_{j,1} = 0.01$ for each $j$. The second row of Table 1 gives summaries of the posterior for the five fixed covariate parameters. The log-likelihood for this model is $-611.7$, over six units larger than that for the static Model 1. In terms of this raw measure of fit, Model 2 clearly dominates. Across a range of values of $B_{j,1}$ (fixed over $j$) this Model has high likelihood; as $B_{j,1}$ increases from zero (Model 1) the model likelihood increases, initially rapidly, up to a maximum around $B_{j,1} = 0.01$, decaying slowly thereafter; at $B_{j,1} = 1.0$, for example, coming closer to the usual, unrestricted PH model, the value is $-612.7$, not so different from Model 2 as illustrated.
TABLE 1. Fixed parameter estimates in Models 1 and 2

<table>
<thead>
<tr>
<th>Parameter:</th>
<th>Baseline</th>
<th>Sex</th>
<th>Age</th>
<th>Dosel</th>
<th>Tumor1</th>
<th>Tumor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1:</td>
<td>−5.07(.34)</td>
<td>−.34(.16)</td>
<td>.18(.09)</td>
<td>−.24(.08)</td>
<td>.21(.09)</td>
<td>.30(.09)</td>
</tr>
<tr>
<td>Model 2:</td>
<td>−.32(.16)</td>
<td>.17(.09)</td>
<td>−.32(.16)</td>
<td>.18(.09)</td>
<td>.26(.09)</td>
<td></td>
</tr>
</tbody>
</table>

Model 3: The third model summarised allows all parameters to vary over time according to a fixed evolution variance matrix $B_j = B = bI$ for all $j$, where $I$ is the identity matrix. With $b$ small, this is close to Model 1. Varying $b$ from zero upwards, the model log-likelihood function increases to a maximum of around $−606.6$ for $0.005 \leq b \leq 0.01$, and slowly decreases thereafter; see Table 2 for some summary values. Models with $b$ at 0.001 and 0.005 differ negligibly in resulting inferences, and represent the included ‘optimal’ range. Analysis with $b = 0.005$ is further summarised; note the clear dominance, in likelihood terms, over each of Model 1 and Model 2.

TABLE 2. Log-likelihood function for $b$ in Model 3.

<table>
<thead>
<tr>
<th>$b$:</th>
<th>0</th>
<th>.001</th>
<th>.005</th>
<th>.01</th>
<th>.025</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-likelihood:</td>
<td>$−617.9$</td>
<td>$−610.0$</td>
<td>$−606.6$</td>
<td>$−606.6$</td>
<td>$−608.7$</td>
<td>$−612.2$</td>
</tr>
</tbody>
</table>

The following figures provide some resulting inferences, particularly concerning time variation in the model parameters. The initial figures summarise the time variation in model parameters: the baseline hazard in Figure 1, the Sex effect in Figure 2, and so on. The full line in each figure provides the estimated time trajectory of the corresponding parameter. . . the posterior means from $m_j^\theta$ plotted over the survival time axis. The dashed lines in each figure represent uncertainty about the values of the parameter over time; at each time point, the dashed lines are simply one posterior standard deviation above and below the posterior mean for that parameter. Working through the figures one at a time, we have, on the basis of this model, the following tentative inferences, to be compared with the fixed parameter model summaries in Table 1.

(i) The baseline hazard in Figure 1 is very stable, indicating that the log-baseline hazard is estimated at about $−4.75$ right across the time axis. Note that the trajectory is very stable even though much more marked variation has been allowed through the evolution variances; the inference is then that the baseline hazard is roughly constant, corresponding to an approximately exponential baseline survival distribution.

(ii) A similarly stable trajectory appears in Figure 2 for the Sex effect parameter, though with a suspicion of a slight decrease in early months, estimated at roughly $−0.4$ across the time axis. Having concluded above that the baseline hazard is roughly constant, it is not, perhaps, surprising that the Sex effect should be so too — all other things being equal, we might expect the forms of the male and female survival distributions to be similar, and so then a constant baseline hazard suggests a constant Sex effect.

Note that the Sex effect is negative, indicating consistently lower hazard for female patients relative to males. The extent of the difference on the hazard and survival scales may be explored by computing predictive survival distribution for future male and female cases; though this is not done here, similar computations are performed below to explore the effects of the tumor measurement variables.
Figure 1. Estimated baseline hazard function (s.d. limits dashed).

Figure 2. Estimated trajectory of Sex effect (s.d. limits dashed).
(iii) The coefficient of Age in Figure 3 appears lower in the first two years than later, though uncertainties, as indicated by the dashed limits, are high. At a level of around 0.25, the estimated coefficient indicates the increased hazard across time for older patients.

![Figure 3. Estimated trajectory of Age coefficient (s.d. limits dashed).](image)

(iv) The time trajectory of the Dosel coefficient in Figure 4 is apparently stable, favouring values around $-0.15$ across time. The negative sign supports the beneficial nature of the treatment (across the range of values of Dosel applied to the patients under study) in decreasing the hazard consistently over a cancer career. Early apparent variation in the trajectory is insignificant, the coefficient being stable across the time axis, estimated at around $-0.15$, but with considerable uncertainty. Note that the values for the coefficient supported here are somewhat lower than the estimates from Model 1 and 2; if this model is accepted as dominant on the likelihood basis, and assuming a causal effect of treatment on survival time, this indicates the earlier models overestimate the strength of treatment effect in reducing hazards. Again, however, uncertainty is high and such issues are better addressed by comparing predictive survival distributions under the various models.

![Figure 4. Estimated trajectory of Dosel coefficient (s.d. limits dashed).](image)
(v) Tumor1 is a measure of the initial extent of cancer development, higher levels potentially predicting increased hazards. The graph in Figure 5 supports this in the early years of a cancer career, with the estimated coefficient around 0.2 during the first two years, but then the indication is that the coefficient is indeed time-varying, decaying to essentially zero after four years, or so. Thus higher levels of Tumor1 at the start of a cancer career are, naturally, strongly indicative of increased hazards during the first few years. However, conditional on survival beyond several years, a patient whose initial Tumor1 level was high appears to be at no greater risk than one with a lower level. This conclusion could be anticipated in qualitative form by examining those patients with observed death times exceeding six years; there are very few such patients and the death rates are apparently unrelated to the Tumor1 variable. Note that, in these later years, there are indeed patients surviving with a full range of Tumor1 values; the paucity of deaths leads to little evidence for or against the value of Tumor1 in conditional survival predution.

![Graph showing trajectory of Tumor1 coefficient](image)

Figure 5. Estimated trajectory of Tumor1 coefficient (s.d. limits dashed).

(vi) The plot in Figure 6 for the coefficient of the second cancer measure, Tumor2, is qualitatively very similar to that of Figure 5, as might have been anticipated following the above discussion. Tumor1 and Tumor2 are both measured on a scale of one to four and so the fact that the estimated trajectory of Tumor2 is higher than that of Tumor1 indicates a more potent predictor variable. The decay of the coefficient is similar to that of Tumor1, although the ultimate level to which this coefficient decays after six years or so is still positive, around 0.15, indicating a sustaining of the increased hazard with higher levels of Tumor2 in the later stages of a long cancer career – again, in dataset patients dying in those later years have higher levels of Tumor2.
Figure 6. Estimated trajectory of Tumor2 coefficient (s.d. limits dashed).

3.4. PREDICTIVE SURVIVAL CHARACTERISTICS

Implications of the apparent time variation in effects, particularly that of the Tumor1 and Tumor2 variables, and the practical relevance of the uncertainties associated with parameter values (evident in the trajectory plots, notably Figure 5 and 6), are perhaps best explored by reference to predictive inferences. We therefore consider prediction of survival times for (hypothetical) new patients with pre-specified covariates. Figures 7 to 10 inclusive provide features of such an exercise, designed primarily to give insight into the nature of the effects of the Tumor variables. Consider an hypothetical male patient, of average age and treated with the average level of Dosel; thus Sex is at level one and, since both Age and Dosel are standardised, this corresponds to zero for each. Figures 7 and 8 provide survival characteristics for such an individual, the five displayed survival curves $S(t)$ relating to the Tumor1 levels of 0, 1, 2, 3 and 4. Figure 8 displays the corresponding predictive hazard curves $h(t)$. As time increases, the decay to zero of the estimated effect of Tumor1 means that the latter eventually coincide, conditional predictive survival probabilities beyond about sixty months being essentially unaffected by Tumor1 levels. A minor increase in the hazards is apparent, as previously noted, in the first year or so. Similar graphs, but now for Tumor2, are given in Figures 9 and 10, the main additional point of interest being the wider separation of the survival functions reflecting the greater predictive relevance of the second tumour measurement.
Figure 7. Predictive survival functions across levels of Tumor1.
(The survival curves are ordered, from the top down, by levels Tumor1 = 0, 1, ..., 4; higher survival probabilities \( S(t) \) correspond to lower levels of Tumor1.)

Figure 8. Predictive hazard functions across levels of Tumor1.
(Corresponding to the survival curves in Figure 7, hazard functions ordered, from the bottom up, by levels Tumor1 = 0, 1, ..., 4; lower hazards \( h(t) \) correspond to lower levels of Tumor1.)
Figure 9. Predictive survival functions across levels of Tumor2.

(The survival curves are ordered, from the top down, by levels Tumor2 = 0, 1, ..., 4; higher survival probabilities \( S(t) \) correspond to lower levels of Tumor2.)

Figure 10. Predictive hazard functions across levels of Tumor2.

(Corresponding to the survival curves in Figure 9, hazard functions ordered, from the bottom up, by levels Tumor2 = 0, 1, ..., 4; lower hazards \( h(t) \) correspond to lower levels of Tumor2.)

Acknowledgements: The author is grateful to Professor Carla Rossi of the Department of Mathematics, II University of Rome, for providing the nasopharynx cancer data, and the opportunity to study the data. Research reported here was partially supported by CNR Italy, and under NSF grant DMS-8903842.
References