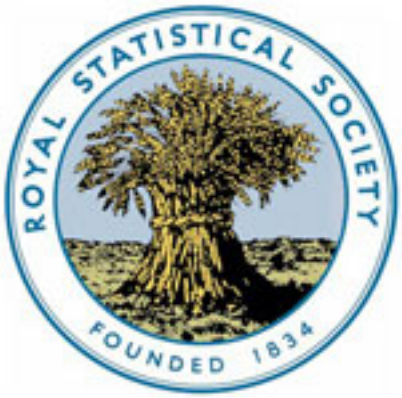

Appl. Statist. (1995)
44, *No.* 3, *pp.* 357–368

Assessing the Losses Caused by an Industrial Intervention: a Hierarchical Bayesian Approach

By ELJA ARJAS†

University of Oulu, Finland

and LIPING LIU

Peking University, Beijing, People's Republic of China

[Received March 1994. Final revision October 1994]

SUMMARY

A nonparametric multiplicative hazard model is proposed, and applied to the estimation of losses caused by a strike at a Quebec aluminium smelter in 1967. The Bayesian approach is adopted, and Gibbs sampling for the numerical computations. We obtain a posterior distribution of the hazard rate and the corresponding predictive distribution of the projected losses. The main merit of the approach proposed is that the uncertainties concerning the actual losses, both in terms of the total number of operating days lost and the excess number of damaged aluminium reduction cells, can be quantified with probabilities. The compensation of the losses is discussed, and the results are compared with some point estimates from earlier analyses of the same data.

Keywords: Bayes inference; Gibbs sampler; Hierarchical model; Numerical assessment of losses; Predictive distribution

1. Introduction

In 1967, during a labour dispute at a Quebec aluminium smelter, the electric power was cut off, resulting in an uncontrolled shut-down of the electrolytic aluminium reduction cells that were in use in the smelter pot rooms. The consequent cooling of cell contents and the difficulties of restarting were believed by the company to have damaged some or all of the cells in service at the time of the shut-down. Since the strike was deemed illegal, the company sued the worker's union and asked for corresponding compensation. The critical question was then how to evaluate the total loss of operating life of the cells, if any, which could be attributed to the intervention.

This question and the data set were presented by Gentleman and Whitmore (1982). Their paper contained two statistical analyses of the data, by John D. Kalbfleisch and Cyntha A. Struthers, and by Duncan C. Thomas. The data set

†*Address for correspondence*: Department of Mathematical Sciences, University of Oulu, Linnanmaa, SF-90570 Oulu, Finland.
E-mail: TIL-EA@FINOU.OULU.FI

 0035–9254/95/44357

contains the installation times, ages at the time of intervention and the failure times of 572 cells in total. The cells were divided into 34 types. All cells were installed before the intervention and 349 were in use at that time. Cells of the same type (subgroup) were installed at almost the same time.

In their brief description of the court case which followed the claim, Gentleman and Whitmore (1982) quoted several expert witnesses' estimates of the total loss of operating life. Two methodologies were employed to derive the estimates: one based on life-table methods and the other on the assumption that the data were censored data from a normal distribution. The highest of the point estimates quoted was 125228 lost days, and the lowest 30619 days. The corresponding estimates of lost cells ranged from 20.2 to 85.5. Kalbfleisch and Struthers (Gentleman and Whitmore, 1982) first reported that they had been unsuccessful in fitting parametric models to the data. They then went on to study the data by using the Cox proportional hazards model, reporting the point estimate 82653.5 days or 60.9 cells lost, saying that this was likely to be an underestimate of the actual losses. They also analysed the number of failures by using a simple point process model, arriving at the estimate of 156 cells lost, with a standard error of 44. Thomas (Gentleman and Whitmore, 1982) considered several different versions of the Cox model, and quoted 22601 days as 'the best point estimate' of the time lost. For more details on the analyses and data, including some graphical representations, we refer to the original paper by Gentleman and Whitmore (1982).

Today still, more than 10 years after the data and the analyses by Kalbfleisch and Struthers and Thomas were published, the Cox proportional hazards model would probably be the method of choice of most statisticians if they were facing a similar problem (see Volf (1993)). This would be despite the fact that it was originally designed for the estimation of relative risks and is much less handy in making actual survival predictions. The Cox model contains both a parametric and a nonparametric part, and the corresponding estimates (and their confidence intervals or bands) need to be combined to arrive at estimated survival probabilities and confidence limits.

Our approach is to use a hierarchical Bayesian model, and to apply Markov chain Monte Carlo techniques (see, for example, Smith and Roberts (1993) or Besag *et al.* (1995)) in the numerical estimation. An additional bonus of the Bayesian approach is that the predictive probabilities that are derived have a direct probabilistic interpretation in the context considered, as degrees of belief concerning uncertain future events. This is important particularly in situations, like the present problem, where the events considered are singular in that it is difficult to think of them as 'repeating a large number of times under similar circumstances', which would correspond to a frequentist interpretation of probabilities.

Our statistical model is introduced in Section 2. Section 3 discusses the estimation of losses, describes the algorithm and ends with numerical estimates. The paper concludes with some general remarks.

2. The Model

A quick analysis of the data (see, for example, Figs 1 and 2 in Gentleman and Whitmore (1982)) shows clearly that there were more frequent failures after the intervention than before. However, as evidenced by the earlier statistical analyses,

different model assumptions and different methods of statistical inference can give very different estimates of the damage attributable to the intervention. Here we follow the Bayesian approach, proposing a nonparametric multiplicative hazard model. We first give the general form of the model.

The aim is to estimate the hazard rate in the hypothetical situation where no intervention had taken place, by using the data before the intervention, and then use the estimate as a projection into the future. Our approach differs from the other analyses (see Gentleman and Whitmore (1982) and Volf (1993)) in that we do not model the hazard rate after the intervention. Indeed, since the actual failure times of those cells which were in use at the time of intervention are known, their values can be compared directly with the corresponding projected failure times arising where there was no intervention. Using the failure times after the intervention as a secondary source of data in an inferential problem concerning hazard rates would necessarily involve making some structural assumptions about how the intervention influences the hazard. In the absence of such knowledge, we simply estimate the pre-intervention hazard from data censored at the time of intervention. Such a choice leaves less room for speculation about model assumptions and their influence on the estimates of damage.

To accommodate a possible dependence of the hazards on calendar time (trends, seasonal variation), a 'realtime' approach is used. In what follows, t will always be the calendar time, except during the intervention when we deem t as stopped, for none of the cells were in use then.

There are many covariates that might affect the hazard rate. Here we consider two: usage time or age (when $t \geqslant I_i$) of cell i at time t, $Z_{1i}(t) = t - I_i$, where I_i is the installation time of cell i, and cell type $Z_{2i} = j$ indicating that cell i is of type j. Our basic (and, essentially, the only) structural assumption is that the covariates act on the hazard rate in a multiplicative way, i.e. the hazard rate of cell i at time t is of the form

$$\lambda_i(t) = f_0(t) f_1\{Z_{1i}(t)\} f_2(Z_{2i}). \tag{1}$$

Here f_0, f_1, and f_2 are all unknown non-negative functions and will be viewed as unknown (random) parameters of the model. The first factor f_0 is a base-line hazard describing possible changes over calendar time. The second factor $f_1\{Z_1(t)\}$ is defined to be 0 for negative $Z_1(t)$, meaning that the hazard rate is 0 before the cell was installed, and as time began to evolve after the cell's installation $f_1\{Z_1(t)\}$ reflects an age effect of that cell. Finally, f_2 is used to describe the relative risks of different cell types, assuming that they remain constant over time.

We now specify the prior distribution of the hazard rate (1), using a hierarchical model structure. To arrive at a finite parameterization we assume that f_0 and f_1 have a piecewise constant structure. The changepoints are not fixed, however, and will in general be different in different sample paths. The prior of f_2 is described in terms of a two-level gamma distribution structure. Finally, f_0, f_1 and f_2 are assumed to be independent according to the prior distribution.

Since the functions f_0, f_1 and f_2 are determined from equation (1) only up to proportionality, some parameters should be fixed. Here we let $f_0(0) = f_1(0) \equiv 1$. The piecewise constancy of f_0 and f_1 means that they can be expressed in the form

$$f_0(t) = \sum_{j \geqslant 0} \mathbf{1}_{\{S_j \leqslant t < S_{j+1}\}} a_j, \tag{2}$$

$$f_1(t) = \sum_{j \geqslant 0} \mathbf{1}_{\{T_j \leqslant t < T_{j+1}\}} b_j, \tag{3}$$

where $\mathbf{1}_A$ is the indicator of the event A, $0 = S_0 < S_1 < \ldots$ and $0 = T_0 < T_1 < \ldots$ are two increasing sequences of random changepoints, and where $\{a_0, a_1, \ldots\}$ and $\{b_0, b_1, \ldots\}$ are the corresponding non-negative random level sequences. Considering first $\{f_0(t);\ t \geqslant 0\}$, we suppose that the jump times $\{S_j;\ j \geqslant 1\}$ form a time homogeneous Poisson process with fixed hyperparameter μ, and that the level sequence $\{a_0, a_1, \ldots\}$ is independent of this. More specifically, we make the prior assumption that $a_0 = 1$ and that, given $a_0, \ldots, a_j$, the level a_{j+1} follows the gamma distribution $\gamma(\ ;\ \alpha_0, \alpha_0/a_j)$, where α_0 is a fixed hyperparameter specifying the shape, while α_0/a_j is the scale parameter. As a consequence, $E_{\text{prior}}(a_{j+1} | a_0, \ldots, a_j) = a_j$ for all j. We specify the prior of $\{f_1(t);\ t \geqslant 0\}$ in the same way, using fixed hyperparameters ν and α_1 for the jump times $\{T_1, T_2, \ldots\}$ and levels $\{b_1, b_2, \ldots\}$ $(b_0 = 1)$ respectively, and finally assume that $\{f_0(t);\ t \geqslant 0\}$ and $\{f_1(t);\ t \geqslant 0\}$ are independent. These definitions imply immediately that $\{f_0(t);\ t \geqslant 0\}$ and $\{f_1(t);\ t \geqslant 0\}$ are independent martingales with mean 1, with respect to the prior distribution and the internal history. This corresponds, loosely, to the prior assumption that f_0 and f_1 do not have trends. For a more detailed account behind these assumptions see Arjas and Gasbarra (1994).

We could of course argue that the cells deteriorate with age and that this should be taken into account by making the prior assumption that the sample paths of f_1 are increasing. Although this would be easy to do technically (see Arjas and Gasbarra (1994), section 4) it is also possible that heterogeneity in cell quality and a corresponding selection mechanism in failures result in hazard rates which are initially decreasing. Although the posterior analysis did not support such a hypothesis (see Fig. 3, later), we felt that, particularly considering that our task was to provide a neutral estimate of the losses, it was justified to make the conservative 'no trend' prior assumption.

As for f_2, a hierarchical model seems natural. According to Gentleman and Whitmore (1982), the 34 cell types can be divided into a standard design group of 20 types (A_1–A_{20}) and an experimental design group of 14 types. The latter can be divided further into two subgroups containing 10 (B–K) and four (L–O) types respectively. In this way we are led to consider three groups of types. It is reasonable, as a prior assumption, to think that the hazard rates of different cell types are more similar within each of the three groups than across groups. Corresponding to this, we specify the prior of f_2 in two steps. In the first step we draw the (random) parameters θ_1, θ_2 and θ_3 independently from $\gamma(\ ;\ \alpha_2, \beta_2)$, and in the second step, if cell type j belongs to the kth group, $f_2(j)$ is drawn from the gamma distribution $\gamma(\ ;\ \eta_k, \eta_k/\theta_k)$, $k = 1, 2, 3$. So the prior expectation of $f_2(j)$, given θ_k, is just θ_k, and the prior expectation of θ_k is α_2/β_2. Here α_2, β_2, η_1, η_2 and η_3 are fixed hyperparameters.

The choice of the hyperparameters reflects the analyst's opinion about the model parameters f_0, f_1 and f_2. Since f_0 is a function of calendar time and f_1 a function of cell age, we must control their variability over time. A useful guideline for such a prior assessment concerning f_0 follows readily from the expressions for the conditional mean and variance of the 'next level', given by $E_{\text{prior}}(a_{j+1} | a_0, \ldots, a_j) = a_j$ and $\text{var}_{\text{prior}}(a_{j+1} | a_0, \ldots, a_j) = a_j^2/\alpha_0$, $j = 1, 2, \ldots$. The corresponding

conditional coefficient of variation is therefore $1/\sqrt{\alpha_0}$. Similar formulae hold for f_1. Therefore, giving a large value to α_0 and a small value to α_1 corresponds to the perception that the hazards do not change much over calendar time but could be affected more strongly by the current cell age. Similarly, in setting up a prior for f_2 we gave smaller values to η_3 and η_2 than to η_1, as we thought that there could be more variability between the cell types belonging to the two experimental design groups than between those belonging to the standard design group.

We point out that, although the justification behind the assumption of a piecewise constant structure for f_0 and f_1 was essentially the need to arrive at a convenient finite parameterization of the hazard rates, the precise local behaviour of these functions seems rather unimportant. For example, continuous and piecewise linear functions, or splines, could have been employed easily with little change in the numerical results. This is essentially because the final assessment of predictive probabilities involves two integrations of the hazard rates: one over the individual functions f_0, f_1 and f_2, with respect to the joint posterior, and another over time.

As mentioned earlier, we use this model as a description of the hazard rate in a situation where no intervention has taken place. For this reason, when estimating the functions f_0, f_1 and f_2, we use failure data collected before the intervention, some of the observations being right censored at the time of intervention. Denoting by Y_i the failure time of the ith cell and by T the time of intervention, we therefore consider the data $\{Y_i^*, \delta_i\}$ where $Y_i^* = \min(Y_i, T)$ and $\delta_i = \mathbf{1}_{\{Y_i \leqslant T\}}$.

As a summary of the assumptions, we now form the prior density of the model parameters and the likelihood expression arising from the data. We first select a sufficiently large $T_{\max}$, so that all computations can be done in the time interval $(0, T_{\max}]$. Under the assumptions introduced above, the prior densities of f_0, f_1 and f_2 can be written respectively as

$$\mu^{m_0} \exp(-\mu T_{\max}) \prod_{0 \leqslant j < m_0} \gamma(a_{j+1};\ \alpha_0,\ \alpha_0/a_j), \tag{4}$$

$$\nu^{m_1} \exp(-\nu T_{\max}) \prod_{0 \leqslant j < n_0} \gamma(b_{j+1};\ \alpha_1,\ \alpha_1/b_j) \tag{5}$$

and

$$\prod_{k=1}^{3} \gamma(\theta_k;\ \alpha_2,\ \beta_2) \prod_{j=1}^{20} \gamma\{f_2(j);\ \eta_1,\ \eta_1/\theta_1\} \prod_{j=21}^{30} \gamma\{f_2(j);\ \eta_2,\ \eta_2/\theta_2\}$$

$$\times \prod_{j=31}^{34} \gamma\{f_2(j);\ \eta_3,\ \eta_3/\theta_3\}, \tag{6}$$

where $m_0 = \Sigma_{j \geqslant 1} \mathbf{1}_{\{S_j \leqslant T_{\max}\}}$ is the number of jumps that f_0 makes in the interval $(0, T_{\max}]$ and $m_1 = \Sigma_{j \geqslant 1} \mathbf{1}_{\{T_j \leqslant T_{\max}\}}$ has the same role for f_1. The first factor in expression (4) is an ordinary Poisson likelihood, corresponding to the assumption that the jump times in f_0 form a time homogeneous Poisson process with parameter μ. Note that it depends on $S_1, S_2, \ldots$ only through the counting variable m_0. The second (product) term in expression (4) is the contribution from the jump levels $a_1, a_2, \ldots, a_{m_0}$, obtained by a simple application of the chain rule. Similar remarks hold for expression (5) and f_1. By the assumed independence, the prior

density of the model parameter (f_0, f_1, f_2) is simply the product of expressions (4)–(6). The likelihood function can therefore be written as

$$\text{lik}(\text{data}; f_0, f_1, f_2) = \prod_{i=1}^{572} [f_0(Y_i^*) f_1\{Z_{1i}(Y_i^*)\} f_2(Z_{2i})]^{\delta_i} \times \exp\left[- \sum_{i=1}^{572} f_2(Z_{2i}) \int_0^{Y_i^*} f_0(s) f_1\{Z_{1i}(s)\} \, ds \right], \quad (7)$$

and the posterior density of the parameters is proportional to the product of expressions (4)–(7).

3. Estimation of Losses

3.1. *Numerical Algorithm*

Hierarchical Bayesian models typically involve parameters of a high dimension and, except in some rare cases, their estimation cannot be done analytically. Here we have the additional difficulty that, although both functions f_0 and f_1 can be parameterized in a finite manner in terms of m_0, m_1, $\{(S_j, a_j): 0 \leqslant j \leqslant m_0\}$ and $\{(T_j, b_j): 0 \leqslant j \leqslant m_1\}$, the numbers m_0 and m_1 themselves are random. By using a suitable form of Gibbs sampling, these difficulties can be solved numerically. Here we follow an approach introduced by Arjas and Gasbarra (1994).

To begin the algorithm, we first generate initial values of the functions f_0, f_1 and f_2 from the prior density, denoting them by f_0^0, f_1^0 and f_2^0. In a generic step of the iteration of the Gibbs sampler, let f_0^n, f_1^n and f_2^n be the 'current values', with f_0^n and f_1^n having the assumed piecewise constant structure. We first update f_0^n to f_0^{n+1}, conditioning on f_1^n, f_2^n and the data, then update f_1^n to f_1^{n+1}, conditioning on f_0^{n+1}, f_2^n and the data, and finally update f_2^n to f_2^{n+1}, conditioning on f_0^{n+1}, f_1^{n+1} and the data, completing this step of iteration.

When updating f_0^n and f_1^n, our algorithm is almost the same as that described by Arjas and Gasbarra (1994). The only difference is that here we have two piecewise constant functions f_0 and f_1, instead of only one, and an additional 34 parameters from f_2 to be updated. The rejection sampling method was used when updating a_j, b_j and θ_k (see Ripley (1987)).

Remark 1. Model (1) can be generalized easily to situations where more covariates are used. The general multiplicative form of the hazard rate of individual i, if alive at time t, would then be $f_0(t) \prod_{k=1}^{K} f_k\{Z_{ki}(t)\}$. Here the functions f_k are specified nonparametrically, in the same way as f_1 and f_2 above. In such a general nonparametric multiplicative hazard model, the corresponding Gibbs sampler is formulated easily and will in principle work in the same way as the algorithm described here. However, if the f_k are assumed to have the parametric form $f_k\{Z_k(t)\} = \exp\{\beta_k Z_k(t)\}$, the model collapses into the well-known Cox model with time-dependent covariates.

3.2. *Estimation of Losses*

Our original goal was to estimate the total loss of operating life, for all cells combined, which could be attributed to the intervention. This is done by considering

projected remaining survival times of the cells which were in use at the time of the intervention.

Let U be the index set of the cells which were at risk at the time of the strike, i.e. $U = \{i;\ Y_i > T\} = \{i;\ \delta_i = 0\}$. To make the comparison between what actually happened and the counterfactual event of 'no strike' more concrete, it is convenient to define a family of latent random variables $\{\tilde{Y}_i;\ i \in U\}$ for what would have been the residual lifetimes of those cells which were alive at T if the intervention had not taken place. So, at any calendar time t, $\tilde{Y}_i + T$ will have the hazard rate $\lambda_i(t)\,\mathbf{1}_{\{t \geqslant T\}}$, where $\lambda_i(t)$ is specified in model (1), with the same installation time I_i and type Z_{2i} as Y_i. $\tilde{Y}_i$ is not real in the sense that it would correspond to some physical quantity in the actual historical development of this case, and we can interpret it as the answer to the question 'If there were no intervention, how long after T could cell i still be used?'. The corresponding loss of operating life of cell i due to the intervention is then simply $\tilde{Y}_i - (y_i - T)$, so that the total projected loss of usage time is $\tilde{Y} = \Sigma_{i \in U}(\tilde{Y}_i + T - y_i)$. Here we have used the notation y_i $(i \in U)$ for the failure times Y_i which actually occurred in the data after the intervention, to emphasize the fact that in the assessment only the latent variables $\tilde{Y}_i$ are treated as random.

The predictive distribution of $\tilde{Y}_i$ at $y > 0$, corresponding to a single cell i, is equal to

$$P(\tilde{Y}_i > y \,|\, \text{data}) = \int_\Theta \exp\left\{ -\int_T^{T+y} \lambda_i^\theta(s)\,\mathrm{d}s \right\} \pi(\theta \,|\, \text{data})\,\mathrm{d}\theta, \tag{8}$$

where, for explicitness, we use the parameter notation $\theta = (f_0, f_1, f_2)$ and $\lambda_i = \lambda_i^\theta$, and denote by $\pi(\theta \,|\, \text{data})$ the corresponding posterior density. The $\tilde{Y}_i$ are generally not independent given the data, and therefore the predictive distribution of $\tilde{Y}$ cannot be computed as a convolution of expressions of this form. However, for any y, we can use the direct approximation

$$P(\tilde{Y} > y \,|\, \text{data}) = \int_\Theta P(\tilde{Y} > y \,|\, \theta)\,\pi(\theta \,|\, \text{data})\,\mathrm{d}\theta \approx \frac{1}{N} \sum_{n=1}^{N} P(\tilde{Y} > y \,|\, \theta^n), \tag{9}$$

where $\theta^n = (f_0^n, f_1^n, f_2^n)$ and N is the number of iterations of the Gibbs sampler. We use the usual justification for the Gibbs sampler based on the ergodic theorem: the sampled values form a Markov chain with the posterior $\pi(\theta \,|\, \text{data})$ as the limiting distribution, and therefore the sample path averages converge with probability 1 to the corresponding expected value. The terms $\tilde{Y}_i$ are conditionally independent given the parameter θ^n, and therefore the probabilities $P(\tilde{Y} > y \,|\, \theta^n)$ could in principle be determined analytically, as convolution integrals. But it is much simpler also to approximate these probabilities by simulated relative frequencies. Let $k_n \geqslant 1$ be an arbitrary integer. If we generate k_n random samples $\{\tilde{Y}_{ij}^n,\ i \in U,\ j = 1, \ldots, k_n\}$ from the distribution

$$P(\tilde{Y}_i \leqslant y \,|\, \theta^n) = 1 - \exp\left[-f_2^n(Z_{2i}) \int_T^y f_0^n(s)\, f_1^n\{Z_{1i}(s)\}\,\mathrm{d}s \right],$$

and denote

$$\tilde{Y}_{\cdot j}^n \triangleq \sum_{i \in U} (\tilde{Y}_{ij}^n + T - y_i),$$

then $\{\tilde{Y}^n_{\cdot j}, j = 1, \ldots, k_n\}$ is a random sample from the distribution $P(\tilde{Y} > y|\theta^n)$. So we can replace approximation (9) by the approximation

$$P(\tilde{Y} > y|\text{data}) \approx \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\mathbf{1}_{\{\tilde{Y}^n_{\cdot j} > y\}}\right). \tag{10}$$

The numerical values of this empirical distribution function can now be used directly in the assessment of losses, by viewing approximation (10) as the posterior probability of the event that the total number of operating days lost exceeds the value y. If we want to arrive at a single numerical estimate of the projected losses, we can use for example the corresponding median or mean.

3.3. *Alternative Way of Compensation*

In addition to reporting results based on the estimated losses of operating life (in days), Gentleman and Whitmore (1982) also quoted estimates expressed in terms of how many cells were damaged in excess of what would have been obtained if there had been no intervention. Here we try to find such a number, but giving it a slightly different interpretation. We ask 'How many new spare cells, used for replacement immediately after the failure times, would be needed to compensate for the losses?'. Such a number would perhaps also be a more direct answer to the dispute concerning financial losses than an estimate of operating days lost.

As the lifetimes of these new cells are random variables, we calculate the distribution of the total operating life of some number, say n^c, of such cells and compare it with the predictive distribution of the losses. In this way we can look for a 'reasonable' number n^c which would be acceptable to both parties. The answer depends of course on the cell type. Here we consider two such choices, using classes A_1 and A_3 as spares. We also assume that the spares would be used at the earliest failure times after the intervention. So, if we knew the number of spares, then their covariates would also be known, and we could calculate the distribution of the total life of these new cells as described above.

Let I^c_i and Y^c_i be the installation time and failure time of new cell i (used for compensation). Denote by C the index set of all new cells that would be used, so that n^c is the cardinality of C. As in Section 3.2, the approximate distribution of the sum of all lifetimes

$$Y^c \triangleq \sum_{i\in C}(Y^c_i - I^c_i)$$

could be determined the same way as in approximation (10).

Our next task is to compare the losses $\tilde{Y}$ with the compensation Y^c. Both are unknown, and the above analysis can only give us the posterior marginals $\tilde{F}(y) = P(\tilde{Y} \leqslant y|\text{data})$ and $F^c(y) = P(Y^c \leqslant y|\text{data})$. In this situation a natural idea is to look for a coupling $(\tilde{Y}^*, Y^{c*})$, say, where $\tilde{Y}^*$ and Y^{c*} are distributed according to $\tilde{F}$ and F^c, but where the comparison between $\tilde{Y}^*$ and Y^{c*} could be made pointwise. The standard method is to let $\tilde{Y}^* = \tilde{F}^{-1}(U^*)$ and $Y^c = (F^c)^{-1}(U^*)$, where F^{-1} denotes the inverse of F and U^* is a (0, 1) uniformly distributed random variable, representing a 'common source of randomness'. Both $\tilde{Y}^*$ and Y^{c*} are then increasing functions of U^*. Furthermore, the probability that $Y^{c*} > \tilde{Y}^*$, say,

is directly the proportion of the us in the unit interval such that $(F^c)^{-1}(u) < \tilde{F}^{-1}(u)$. This can be calculated for any given number n^c of spares. Perhaps the most reasonable value of n^c is the value that makes this probability closest to 0.5. In the present case the distribution functions $\tilde{F}$ and F^c cross each other exactly once, so that this amounts to finding the n^c for which F^c and $\tilde{F}$ have (approximately) the same median. Denote this number by n_0^c. Then, if such a hypothetical coupling were used, we could make the assessment that, if more than n_0^c spare cells were given to the company free of charge, it would have at least a 50% chance to be compensated for all its losses, and less than a 50% chance if fewer were used. A symmetric reasoning applies to the worker's union.

An alternative would be to look for a value of n^c such that $\tilde{F}$ and F^c have the same mean, corresponding to expected total residual lifetimes in the two cases. The result would appear to be more sensitive to the tail behaviour of the two distributions, however, and because long survival times are sparse in the data it could be rather strongly influenced by the specification of the prior.

3.4. *Numerical Results*

As always, the posterior distribution of parameters and therefore the predictive distribution of the total losses will depend on the prior distribution. Different values of hyperparameters were tried in this case, representing 'vague prior knowledge', and leaving relatively more room for the effect of the likelihood. In this way the final result is not very sensitive to the choice of hyperparameters, especially when expressed in terms of n_0^c. This is understandable because the random samples of $\tilde{Y}$ and Y^c were generated by using the same sample of hazards. Usually 2000 iterations of the Gibbs sampler are enough for convergence. In the simulations we used systematically the value $k_n = 1$ in approximation (10).

As a first trial we gave the hyperparameters the values $\mu = \nu = 0.002$, $\alpha_0 = 10$, $\alpha_1 = 5$, $\alpha_2 = 5$, $\beta_2 = 5000$, $\eta_1 = 10$, $\eta_2 = 8$ and $\eta_3 = 7$. Thus the mean $E_{\text{prior}}[f_0(t) \times f_1\{Z_1(t)\}f_2(Z_2)]$, which could be viewed as the prior hazard rate of a cell at calendar time t, becomes $\alpha_2/\beta_2 = 0.001$, if the cell is in use then. The corresponding approximate predictive distribution of $\tilde{Y}$, based on 3000 iterations, is displayed in Fig. 1. Its mean and median are 238915 and 230574 (days) respectively.

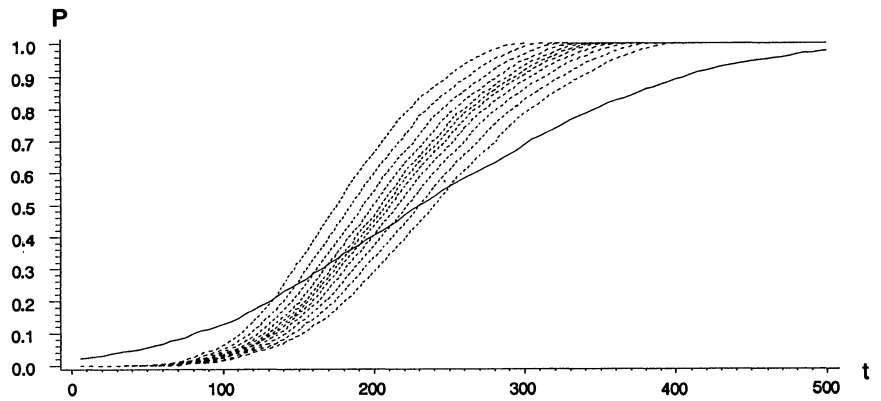

Fig. 1. Distribution of losses (———) and compensation (--------) (t is in units of 1000 days)

In the same figure we have drawn the predictive distribution curves of Y^c, as described above in Section 3.3, with the number of spares taking values $n^c = 110$, 116, 121, 125, 128, 130, 132, 135, 139, 144, 150. All the spares were of type A_1. From our numerical calculation we obtain the value $n_0^c = 146$. We did the same calculation by using the hyperparameter values $\alpha_0 = 5$ and $\alpha_1 = 10$, and keeping the other values as above. The mean and median of $\tilde{Y}$ were then 232045 and 208586. We do not display the predictive distributions here in a graphical form because they coincide almost exactly with those in Fig. 1. If type A_3 were used in the compensation, we would obtain $n_0^c = 138$. Therefore it seems that type A_3 cells are of better quality than type A_1 cells. In both simulations, the posterior probability that $\tilde{Y}$ is non-positive was small (1.8% and 2.8%). So apparently the intervention was detrimental.

Compared with the collection of point estimates presented by Gentleman and Whitmore (1982), our assessment of days lost due to the intervention seems rather large. The smallest figure cited, 22601 cell-days presented in the analysis by Thomas, would correspond approximately to the 3.4% quantile of the posterior distributions displayed in Fig. 1. Similarly, the figure 82653.5 cell-days, which, however, Kalbfleisch and Struthers thought likely to be an underestimate of actual losses, corresponds approximately to the 10.1% quantile. Corresponding differences can be seen when the estimates are based on the excess number of failures (or on the number of spares used for compensation). Our estimates of n_0^c are, however, remarkably close to the estimated excess of 156 (with a standard error of 44) cells, which was derived by Kalbfleisch and Struthers by applying a simple Poisson process model.

Perhaps the main observation concerning this predictive distribution, apart from its location, is that it is spread over a very wide interval. This is an expression of the fact that knowledge here is very imprecise. It also explains, at least in part, why there is so little consistency between the point estimates reported by Gentleman and Whitmore (1982), including several which were presented by expert witnesses in the court case, and later by Volf (1993).

4. Concluding Remarks

Some additional results can be obtained as by-products of our analysis. For example, if we want to compare the qualities of types A_1 and A_3, we can use the simulated sequences $\{f_2^n(1), f_2^n(3)\}$ to obtain the posterior (marginal) distributions of these two parameters. Fig. 2 shows these curves, with distribution functions of types A_1, A_3, C and L. Such curves can be drawn for any type, so that a comparison of any pair is easy. Fig. 2 also tells us that type A_3 is better than type A_1 in the sense of stochastic ordering, which agrees with our numerical results in Section 3.4.

As f_0 and f_1 are here viewed as random functions, their posterior distributions cannot be described easily. Here we just give the averaged sample paths in Fig. 3, corresponding to drawing the posterior mean of $f_0(t)$ and $f_1(t)$ for all t. The age of a cell clearly has a stronger influence on the hazard rate than does the calendar time.

The main purpose of this paper has been to demonstrate how the full Bayesian approach, involving hierarchical modelling, relatively few distributional assump-

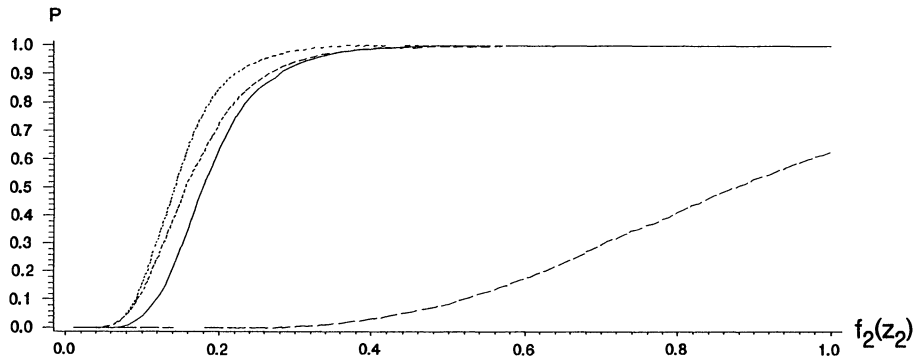

Fig. 2. Comparison of four different cell types: posterior marginal cumulative density functions of the relative risk $f_2(z_2)$, where z_2 refers to type A_1 (———), type A_3 (· · · · · ·), type C (--------) and type L (------)

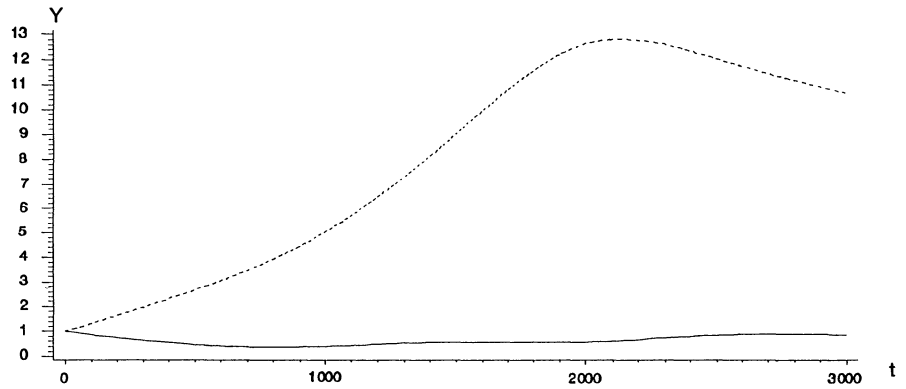

Fig. 3. Averaged sample paths of f_0 (———) and f_1 (--------)

tions and an iterative algorithm to do the numerical calculations, can be used as a direct means of providing a reasonable answer to the, in some sense impossible, question 'How big were the losses resulting from the strike?'. We have limited the amount of data analysis here on purpose, to emphasize that the aluminium smelter case-study serves as an illustration of a more general approach, rather than being the goal itself. No conclusive numerical answer is therefore aimed at. If the robustness of the numerical results became a critical issue (which would be likely to happen if the original legal dispute were still unsettled and we were hired to present these views as expert witnesses), we would no doubt have to examine in more detail how our answer depends on the choice of the hyperparameters and on the assumed multiplicative form (1) of the hazard rates. However, we believe that this general approach, properly adjusted, can be applied in a wide variety of cases where one must provide, on the basis of non-experimental data, a quantitative answer to a question of the type 'How much difference would it have made if . . .?'.

Acknowledgements

We are grateful to two referees for their many helpful suggestions to improve the presentation. This work was supported by grants from the University of Helsinki and the Academy of Finland.

References

Arjas, E. and Gasbarra, D. (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statist. Sin.*, **4**, 505–524.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Spatial statistics, image analysis and Bayesian inference. *Statist. Sci.*, to be published.

Gentleman, J. F. and Whitmore, G. A. (1982) Measuring the impact of an intervention on equipment lives (case studies in data analysis no. 1, with contributions by Kalbfleisch, J. D. & Struthers, C. A., and Thomas, D. C.). *Can. J. Statist.*, **10**, 237–259.

Ripley, B. P. (1987) *Stochastic Simulation*. New York: Wiley.

Smith, A. F. M. and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 3–23.

Volf, P. (1993) Nonparametric estimation of hazard regression with application to a case study. *Research Report 1776*. Academy of Sciences of the Czech Republic, Prague.