Bayesian Analysis of Hazard Regression Models under Order Restrictions on Covariate Effects and Ageing

Bhattacharjee, Arnab and Bhattacharjee, Madhuchhanda
University of St Andrews

2007

Bayesian Analysis of Hazard Regression Models under Order Restrictions on Covariate Effects and Ageing*

Arnab Bhattacharjee
School of Economics and Finance
University of St. Andrews, UK.

Madhuchhanda Bhattacharjee
Department of Mathematics and Statistics
Lancaster University, UK.

April 2007
(Preliminary, Second Draft)

Abstract

We propose Bayesian inference in hazard regression models where the baseline hazard is unknown, covariate effects are possibly age-varying (non-proportional), and there is multiplicative frailty with arbitrary distribution. Our framework incorporates a wide variety of order restrictions on covariate dependence and duration dependence (ageing). We propose estimation and evaluation of age-varying covariate effects when covariate dependence is monotone rather than proportional. In particular, we consider situations where the lifetime conditional on a higher value of the covariate ages faster or slower than that conditional on a lower value; this kind of situation is common in

*Corresponding Author: M. Bhattacharjee, Department of Mathematics and Statistics, Fylde College Building, Floor B, Lancaster University, Lancaster LA1 4YF, UK. Tel: +44 1524 593066. e-mail: m.bhattacharjee@lancaster.ac.uk
The authors thanks Ananda Sen, Debasis Sengupta, and participants at the IISA Joint Statistical Meeting and International Conference (Cochin, India, Jan. 2007) for their valuable comments and suggestions.

applications. In addition, there may be restrictions on the nature of ageing. For example, relevant theory may suggest that the baseline hazard function decreases with age. The proposed framework enables evaluation of order restrictions in the nature of both covariate and duration dependence as well as estimation of hazard regression models under such restrictions. The usefulness of the proposed Bayesian model and inference methods are illustrated with an application to corporate bankruptcies in the UK.

Keywords: Bayesian nonparametrics; Nonproportional hazards; Frailty; Age-varying covariate effects; Ageing.

1 Introduction

Understanding the nature of covariate dependence and ageing are the main objectives of regression analysis of lifetime or duration data. In many applications, relevant underlying theory or preliminary analysis may suggest that there are important order restrictions either covariate dependence, or the shape of the baseline hazard, or both. Parametric inference in such situations can be conducted by making functional form or distributional assumptions that impose the above order restrictions. However, such assumptions can be very restrictive and lead to weak inference. Instead, one may aim to conduct order restricted nonparametric analysis under the constraints implied by theory or past experience. In fact, such inference can also be used to judge the validity of the order restrictions themselves.

In this paper, we propose Bayesian models to conduct order restricted nonparametric inference in applications with single spell lifetime data. Specifically, our framework for inference in hazard regression models incorporates three important features. First, we do not assume proportional hazards with respect to all covariates included in the analysis. It is well-known that the proportionality assumption underlying the Cox proportional hazards model does not hold in many applications. On the other hand, credible inference under the model depends crucially on the validity of the proportionality assumption. Further, the effect of a covariate is often monotone, in the sense that the lifetime (or duration) conditional on a higher value of the covariate ages faster or slower than that conditional on a lower value (Bhattacharjee, 2006). In particular, we consider relative ageing in the nature of convex or concave ordering (Kalashnikov and Rachev, 1986) of lifetime distributions conditional on different values of the covariate in question. Ordered departures of this kind are common in applications, and the models provide useful and intuitively appealing descriptions of covariate dependence in nonproportional situations. Further, ordered departures of the above kind can be convenienty studied in a Cox type regression model with age-varying covariate effects (Bhattacharjee, 2004), where positive ageing for higher covariate values implies that the age-varying effect of the covariate is a nondecreasing function of lifetime. Thus, in this paper, order restriction in covariate dependence will be taken as monotone age-varying covariate effects for some selected covariates.

Second, in addition to order restricted covariate dependence, we will allow for constraints on the shape of the baseline hazard function. These order restrictions will typically be in the nature of monotone (increasing/ decresing) hazard rates. They could also be characterised by weaker notions of ageing, such as "new better than used". As discussed above, these kind of ordering are important in many applications, and reflect the inherent structural nature of the ageing process not related to differences in observed or unobserved covariates.

The third characteristic feature of our work is in the treatment of unobserved heterogeneity. In our approach, unobserved covariates induce hazard rates to vary across individuals in two different ways. Unobserved covariates that act at the group level (and are therefore identified by group membership) are incorporated in our model as fixed effects heterogeneity. In addition, we allow a scalar unobserved covariate independent of the included regressors which has a completely unspecified distribution. Our approach is in contrast of much of the literature that specifies a parametric frailty distribution. The nonparametric approach to modeling frailty (Heckman and Singer, 1984) operates through a sequence of discrete multinomial distributions. Each of these distributions comprises a set of mass points along with the probabilities of a subject being located at each mass point. By progressively increasing the number of mass points, we are able to approximate any arbitrary frailty distribution.

The remainder of the paper is organised as follows. Section 2 presents a selective review of the literature. We describe our model in Section 3 and our application is presented and discussed in Section 4. Finally, Section 5 concludes.

2 Literature

This paper is in the area of order restricted Bayesian semiparametric inference in the context of hazard regression models. The work is rather unique in that there is very little prior literature in the area. However, there is literature in several related areas, both in a Bayesian paradigm as well as frequentist inference. We survey the literature in these areas briefly with a view towards placing our work within the context of the literature and highlighting the distinctive nature of our approach.

2.1 Bayesian semiparametric inference in hazard regression models

Semiparametric approaches to Bayesian inference in hazard regression models usually assume the Cox proportional hazards model

$$\lambda\left(t|\underline{z}_i(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}^T.\underline{z}_i(t)\right], \qquad i = 1, \ldots, n \tag{1}$$

where $\underline{z}_i(t)$ is the p-dimensional vector of (time varying) covariates for the i-th subject at time $t > 0$, $\underline{\beta}$ is the (fixed) vector of unknown regression coefficients, and $\lambda_0(t)$ is the unknown baseline hazard function. Various Bayesian formulations of the model differ mainly in the nonparametric specification of $\lambda_0(t)$.

2.1.1 Prior specification for hazard regression models

A model based on an independent increments gamma process was proposed by Kalbfleisch (1978) who studied its properties and estimation. Extensions of this model to neutral to the right processes was discussed in Wild and Kalbfleisch (1981). In the context of multiple event time data, Sinha (1993) considered an extension of Kalbfleisch's (1978) model for $\lambda_0(t)$. The proposal assumes the events are generated by a counting process with intensity given by a multiplicative expression similar to (Equation 1), but including an indicator of the censoring process, and individual frailties to accommodate the multiple events occurring per subject.

Several other modelling approaches based on the Cox PH model have been studied in the literature. Laud et al. (1998) consider a Beta process prior for $\Lambda_0(t)$ and propose an MCMC implementation for full posterior inference. Nieto-Barajas and Walker (2002a) propose their flexible Lévy driven Markov process to model $\lambda_0(t)$, and allowing for time dependent covariates. Full posterior inference is achieved via substitution sampling.

2.1.2 Bayesian survival data models

While Bayesian formulation of the Cox proportional hazards model has been rather narrow in the specification of the baseline hazard function, sevral other models have been used more generally in Bayesian survival analysis. These models can be used in the context of hazard regression models to specify the baseline hazard or baseline cumulative hazard functions.

Many stochastic process priors that have been proposed as nonparametric prior distributions for survival data analysis belong to the class of neutral to the right (NTTR) processes. A random probability measure $F(t)$ is an NTTR process on the real line, if it can be expressed as $F(t) = 1 - exp(-Y(t))$, where $Y(t)$ is a stochastic process with independent increments, almost surely right-continuous and non-decreasing with $P\{Y(0) = 0\} = 1$ and $P\{\lim_{t \longrightarrow \infty} Y(t) = 1\} = 1$ (Doksum 1974). The posterior for a NTTR prior and i.i.d. sampling is again a NTTR process. Ferguson and Phadia (1979) showed that for right censored data the class of NTTR process priors remains closed, i.e., the posterior is still a NTTR process.

NTTR processes are used in many approaches that construct probability models for the hazard function $\lambda(t)$ or the cumulative hazard function $\Lambda(t)$. Dykstra and Laud (1981) define the extended gamma process as a model for $\lambda(t)$, generalizing the independent gamma increments process studied in Ferguson (1973). Dykstra and Laud (1981) show that the resulting function $\lambda(t)$ is monotone, making it useful for modeling ageing in the nature of monotone hazard rates.

An alternative Beta process prior on $\Lambda(t)$ was proposed by Hjort (1990), where the baseline hazard comprises piecewise constant independent beta distributed increments. This model is closed under prior to posterior updating as the posterior process is again of the same type. Full Bayesian inference for a model with a Beta process prior for the cumulative hazard function using Gibbs sampling can be found in Damien et al. (1996). Walker and Mallick (1997) specify a similar structure for the prior, but use independently distributed gamma hazards.

While the above models for $\Lambda(t)$ are based on independent hazard increments $\{\lambda_j\}$, considering dependence provides a different modeling perspective. A convenient way to introduce dependence is a Markovian process prior on $\{\lambda_j\}$. Gamerman (1991) proposes the following model: $\ln(\lambda_j) = \ln(\lambda_{j-1}) + \varepsilon_j$ for $j \geq 2$, where $\{\varepsilon_j\}$ are independent with $E(\varepsilon_j) = 0$ and $Var(\varepsilon_j) = \sigma^2 < \infty$. Later, Gray (1994) used a similar prior process but directly on the hazards $\{\lambda_j\}$, without the log transformation. A further generalization involving a martingale process was proposed in Arjas and Gasbarra (1994). More recently, Nieto-Barajas and Walker (2002b) proposed a model based on a latent process $\{u_j\}$ such that $\{\lambda_j\}$ is included as

$$\lambda_1 \longrightarrow u_1 \longrightarrow \lambda_2 \longrightarrow u_2 \longrightarrow \dots$$

and the pairs (u, λ) are generated from conditional densities $f(u|\lambda)$ and $f(\lambda|u)$ implied by a specified joint density $f(u, \lambda)$. The main idea is to ensure linearity in the conditional expectation: $E(\lambda_{j+1}|\lambda_j) = a_j + b_j\lambda_j$. Nieto-Barajas and Walker (2002b) show that both the gamma process of Walker and Mallick (1997) and the discrete Beta process of Hjort (1990) are obtained as special cases of their construction, under appropriate choices of $f(u, \lambda)$.

2.1.3 Unobserved heterogeneity

Accounting for unobserved heterogeneity is important in the analysis of hazard regression models. With single survival data and individual-level frailty, estimation of individual frailties is not possible but their distribution can be inferred on. Clayton (1991) and Walker and Mallick (1997) both consider Bayesian inference in the Cox proportional hazards model with gamma frailty distribution, but with different priors on the baseline hazard function. Sinha (1993) also assumes gamma distributed frailties, but in multiple event survival data. Extensions of this model to the case of positive stable frailty distributions and a correlated prior process with piecewise exponential hazards can be found in Qiou et al. (1999).

In its ability to deal with potentially large number of latent variables, the Bayesian framework offers the possibility of a more nonparametric approach to modeling individual level frailty. Based on repeated failures data, Bhattacharjee et al. (2003) and Arjas and Bhattacharjee (2003) have proposed a hierarchical Bayesian model based on a latent variable structure for modeling unobserved heterogeneity; the model is very powerful and shown to be useful in applications.

Since our application here is based on single failure per subject data, we use a latent variable structure but with the objective of inferring on the frailty distribution rather than the latent variables themselves. We model frailty in two different ways. First, we divide the subjects into groups and incorporate fixed effects unobserved heterogeneity across these different groups. Second, we model individual level frailty in a more nonparametric tradition (see Heckman and Singer, 1984) by introducing a sequence of multinomial frailty distributions with increasing number of support points; for a related Bayesian implementation, see Campolieti (2001).

2.1.4 Order restricted inference

The literature on order restricted Bayesian inference, with restrictions either on the shape of the baseline hazard function or on the nature of covariate depence, is indeed very sparse. Notable contributions to the literature in this area are Arjas and Gasbarra (1996), Sinha et al. (1999), Gelfand and Kottas (2001) and Dunson and Herring (2003); all these papers are related to the current work.

Arjas and Gasbarra (1996) develop models of the hazard rate processes in two samples under the restriction of stochastic ordering. They define their prior on the space of pairs of hazard rate functions; the unconstrained prior in this space consists of piecewise constant gamma distributed hazards which incorporate path dependence. The constrained prior is then constructed by restricting to a subspace on which the maintained order restriction holds. In their work, Arjas and Gasbarra (1996) propose a coupled and constrained Metropolis-Hastings algorithm for posterior elicitation based on the order restriction and also for Bayesian evaluation of the stochastic ordering assumed in the analysis. For the same problem, Gelfand and Kottas (2001) developed an alternative prior specification and computational algorithm. The Bayesian model in Arjas and Gasbarra (1996), in combination with the general treatment of Bayesian order restricted inference (for example, in Gelfand et al., 1992), is related to the current paper.

Sinha et al. (1999) develop Bayesian analysis and model selection tools using interval censored data where covariate dependence is possibly nonproportional. They model the baseline hazard function using an independent Gamma prior and age varying covariate effects are endowed with a Markov type property $\beta_{k+1}|\beta_1, \ldots, \beta_k \sim N(\beta_k, 1)$. While Sinha et al. (1999) do not explicitly consider order restrictions either on covariate dependence or on ageing, they provide Bayesian inference procedures to infer on the validity of the proportional hazards assumption.

In other work related to this paper, Dunson and Herring (2003) consider an order restriction on covariate dependence in hazard regression models. They develop Bayesian methods for inferring on the restriction that the effect of an ordinal covariate is higher for higher levels of the covariate; in other words, they conduct inference on trend in conditional hazard functions. We work with restrictions on covariate dependence which are different in two respects. First, the covariate is continuous in our case and not categorical. Second, our order restriction is related to convex/ concave partial ordering of conditional hazard functions rather than trend. Consequently, we express our constraints in terms of monotonic age-varying covariate effects, and propose a different methodology for Bayesian inference.

2.2 Order restricted frequentist inference

Order restrictions relating both to the shape of the baseline hazard function (ageing) as well as the effect of covariates (covariate dependence) are important in the study of hazard regression models. However, the literature on frequentist order restricted inference in hazard regression models deal mainly with covariate dependence.

In the two sample (binary covariate) setup, testing for proportionality of hazards against some notion of relative ageing (such as, monotone hazard ratio, or monotone ratio of cumulative hazards) has been an active area of research (Gill and Schumacher, 1987; Deshpande and Sengupta, 1995; Sengupta et al., 1998). Order restricted estimation in two samples under the corresponding partial orderings (convex ordering and star ordering) has not been considered in the literature. However, estimation in two samples with right-censored survival data under the stronger constraint of stochastic ordering has been considered in Dykstra (1982), and extended to uniform conditional stochastic ordering in the k-sample setup by Dykstra et al. (1991). These inference procedures are, however, not very useful in the hazard regression context, where covariates are typically continuous in nature.

In a recent contribution, Bhattacharjee (2006) extended the notion of monotone hazard ratio in two samples to the situation when the covariate is continuous, and proposed tests for proportional hazards against ordered alternatives. Specifically, the alternative hypothesis here states that, lifetime conditional on a higher value of the covariate is convex (or concave) ordered with respect to that conditional on a lower covariate value:

$$\begin{aligned} IHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_1)/\lambda(t|x_2) \uparrow t (\equiv (T|X=x_1) \prec_c (T|X=x_2) \\ DHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_2)/\lambda(t|x_1) \uparrow t (\equiv (T|X=x_2) \prec_c (T|X=x_1) \end{aligned} \quad (2)$$

where x_1 and x_2 are two distinct values of the covariate under study, $\prec_c$ denotes convex ordering, and IHRCC (DHRCC) are acronyms for "Increasing (Decreasing) Hazard Ratio for Continuous Covariates". Bhattacharjee (2004) shows that, in the absence of unobserved heterogeneity, monotone covariate dependence of this type can be nicely represented by monotonic age varying covariate effects, so that

$$\begin{aligned} IHRCC &: \lambda(t|x_i) = \lambda_0(t). \exp\left[\beta(t).x_i\right], \beta(t) \uparrow t \quad (3) \\ DHRCC &: \lambda(t|x_i) = \lambda_0(t). \exp\left[\beta(t).x_i\right], \beta(t) \downarrow t. \end{aligned}$$

Thus, the above partial orders can be conveniently studied using age-varying covariate effects; using this representation, Bhattacharjee (2004) proposed biased bootstrap methods (like data tilting and local adaptive bandwidths) to estimate hazard regression models under these order restrictions. Bhattacharjee (2007) extended the test for proportionality to a regression model with individual level unobserved heterogeneity with completely unrestricted frailty distribution.

In this paper, we will consider order restrictions on the shape of the baseline hazard function in addition to constraints on covariate dependence. This kind of ordering is relevant in many applications. For example, relevant theory may suggest that the the effect of a covariate is positive but decreases to zero with age. In addition, the baseline hazard function decreases with age.

3 Our Bayesian model

The Bayesian framework offers several advantages in conducting order restricted inference in the current problem. First, inference on order restrictions jointly on covariate dependence and ageing is a challenging problem, and the Bayesian setup is better equipped to deal with such difficult problems. Second, prior beliefs can be explicitly incorporated in the model, including beliefs on order restrictions. Third, the framework provides very good flexibility where frailty of different kinds can be included and inferred on.

The major challanges, on the other hand, are (a) appropriate representation of prior beliefs in the model, and (b) ensuring numerical tractability of posterior simulations.

As mentioned earlier, the inference procedures in this paper are developed with reference to an application to firm exits due to bankruptcy in the UK. The major objective of our empirical analysis is to understand the effect of macroeconomic conditions on business failure. Age of the firms is measured in years post-listing. The lifetime data are right censored, left truncated and contain delayed entries. Most of the covariates included in the regression model (firm-specific and macroeconomic) are time-varying. In addition, our data includes industry dummies which are fixed over age.

Initially, we consider the Cox proportional hazards model with time varying covariates, fixed regression coefficients and completely unrestricted baseline hazard function (Equation 1). We will incorporate into the model additional features of our analysis: (a) *order restricted covariate dependence* –

time varying (and possibly monotonic) covariate effects, (b) *unobserved heterogeneity* – fixed effects heterogeneity and frailty, and (c) *order restrictions on ageing.*

To facilitate analysis and presentation, we partition the time axis $[0, \infty)$ into a finite number of disjoint intervals (in our case, in years), say $I_1, I_2, \ldots, I_{g+1}$, where $I_j = [a_{j-1}, a_j)$ for $j = 1, 2, \ldots, g+1$ with $a_0 = 0$ and $a_{g+1} = \infty$. We assume the baseline hazard function to be constant within each of these intervals (taking values $\lambda_1, \lambda_2, \ldots, \lambda_{g+1}$), and the age-varying covariate effects to be similarly piecewise constant.

3.1 Order restricted covariate dependence

Like many other applied disciplines, economic theory does not usually imply functional forms or exact distributions, but rather order restrictions such as monotonicity, convexity, homotheticity etc. In the context of survival models, there are many applications where there is evidence of order restrictions of the kind described in (Equation 2) or (Equation 3) on the nature of covariate dependence.

For example, Metcalf *et. al.* (1992) and Card and Olson (1992) observed that the impact of real wage changes varied with duration of strikes, and the variation was in the nature of ordered departures. In particular, Card and Olson (1992) found that, while longer duration strikes (lasting more than 4 weeks) were most common for strikes with wage changes of less than 15 per cent, shorter duration strikes (1 to 3 days) were most frequent for wage changes above 15 per cent. Similarly, Narendranathan & Stewart (1993) observe that the effect of unemployment benefits on unemployment durations decreases the closer one is to the termination of benefits.

In a previous study using the current dataset, the impact of macroeconomic instability on business exit is found to decrease with age of the firm post-listing (Bhattacharjee et al., 2002). Such evidence of monotonic covariate effects are not confined to economic applications. For survival with malignant melanoma, for example, Andersen *et. al.* (1993) observe that, while “hazard seems to increase with tumor thickness” (pp. 389), the plot of estimated cumulative baseline hazards for patients with ‘2mm $\leq$ tumor thickness $<$ 5mm’ and ‘tumor thickness $\geq$ 5mm’ against that of patients with ‘tumor thickness $<$ 2mm’ reveal “concave looking curves indicating that the hazard ratios decrease with time” (pp. 544–545).

Based on the above discussion, covariates with both fixed and age-varying covariate effects are included in our analysis. For some covariates with nonproportional hazards, the age-varing effects monotonically increase with age while for some others, the effect decreases as time goes on.

3.2 Unobserved heterogeneity

We account for unobserved covariate effects in two distinct ways. First, there are unobserved covariates at the industry level which create variation in exit rates across industries (other factors remaining constant). Since industry membership is observed for all firms, these factors can be incorporated by including fixed effects heterogeneity. In essence, we include a dummy variable for each industry in our regression model. The estimates for these fixed effects will then be interpreted as the effect of all unobserved regressors at the industry level.

Second, we include a multiplicative frailty variable that is independent of all other included or industry level covariates. Unlike previous Bayesian studies, the frailty distribution is fully nonparametric in our case. We implement this feature using a method suggested by Heckman and Singer (1984), where the unknown distribution is approximated by a sequence of multinomial distributions based on progressively increasing number of mass points. For example, with two mass points, log-frailty is assumed to have a two point distribution (say, with mass at $m_1 = 0$ and m_2, and corresponding probabilities π_1 and $\pi_2 = 1 - \pi_1$); one of the mass points is set at zero because of scaling. The number of mass points is increased sequentially until no substantial improvement in the model is observed. At that point, the multinomial distribution approximates the unknown frailty distribution reasonably well.

Modeling frailty distribution in this way offers excellent opportunities for inference and interpretation. For example, a two support point distribution with $\pi_1 = 0.25$ would indicate that, with respect to the unobserved covariate, there are two types of subjects. 25% of these subjects draw a lower value from the population and consequently have a lower hazard rate. Contrast this with a gamma distributed frailty; similar inferences on the estimates of the frailty distribution are not so readily derived.

3.3 Ageing

In addition to covariate dependence, it is often reasonable to expect order restrictions on the shape of the baseline hazard function. For example, in a similar application based on the current data, Bhattacharjee et al. (2002) find that the baseline hazard function exhibits some negative ageing. However, this evidence is not in the nature of a decreasing hazard rate, but perhaps a weaker form of partial order. This indicates a weak form of learning not related to other observed covariates. This would suggest an additional order restriction, perhaps in the nature of a "new worse than used" lifetime distribution.

We incorporate such order restrictions in our application to evaluate any evidence on ageing.

Incorporating the above three features in the Cox PH model (Equation 1), we have the following hazard regression model:

$$\lambda\left(t|\underline{J^{(d)}}_i, \underline{z^{(f)}}_i(t), \underline{z^{(v)}}_i(t), \nu_i\right) = \lambda_0(t).\exp\left[\underline{\beta^{(d)}}^T.\underline{J^{(d)}}_i + \underline{\beta^{(f)}}^T.\underline{z^{(f)}}_i(t) + \underline{\beta^{(v)}}(t)^T.\underline{z^{(v)}}_i(t)\right].\nu_i, \tag{4}$$

where $\lambda_0(t)$ is the unknown baseline hazard function which could potentially have order restrictions on ageing, $\underline{J^{(d)}}_i$ is a vector of dummy variables indicating membership in the various industry groups, $\underline{z^{(f)}}_i(t)$ are covariates with proportional effects on the hazard function, $\underline{z^{(v)}}_i(t)$ are covariates with non-proportional effects potentially represented by order restrictions on covariate dependence, and ν_i is an individual-level multiplicative frailty variable with arbitrary distribution.

3.4 Prior specification

We explore several models with different specifications for the prior distributions. These prior distributions are related to models considered in the literature, for example in Sinha et al. (1999). However, our models are unique in that they explicitly consider order restrictions in covariate dependence and ageing, in the presence of individual level multiplicative frailty. Below we describe specification of priors for the three main categories of parameters for our model: covariate effects, baseline hazard and frailty.

3.4.1 Covariate effects

We use three alternative prior distributions for modeling the covariate effects:

1. Truncated normal, with truncation reflecting whether the covariate effect is expected to be positive or negative. For the industry fixed effects, there is no truncation, and the distribution is centered at zero.

2. Truncated normal, with variance proportional to the number at risk (for age-varying covariate effects)

3. Exponential prior. Like above, for age-varying effects, parameter is made proportional to number at risk.

For the covariates with potentially age varying effects, we model order restrictions in three different ways:

1. Initially, no order restriction is imposed, leaving the effects free to assume any value (positive or negative). However, a first order smoothing condition is assumed: $E\left[\beta\left(t_k\right)|\beta\left(t_{k-1}\right)\right] = \beta\left(t_{k-1}\right).$ Further, variance was set at 10 for β's up to age 35, and variance was set at 1 thereafter – this was to control for the cumulative uncertainty effect due to the smoothing assumption.

2. Order restrictions in the posterior mean

3. Stochastic ordering: For example, for decreasing covariate effects, mean set at a reasonable level initially, decreasing by a step each year. Steps have exponential distributions, with parameter proportional to number at risk.

We make use of the well known consistency property of Bayesian updating procedures that if the prior is supported completely by a subset of the parameter space, then so is the posterior.

3.4.2 Baseline hazard

Four different specifications for the baseline hazard prior are explored.

1. Gamma independent increments

2. Truncated normal independent increments

3. Neutral to the right gamma process

4. Gamma independent increments till age 10, stochastically decreasing thereafter (this reflects a weak form of negative ageing)

3.4.3 Frailty

Our empirical work in the following Section is based on a two-point support frailty distribution. Since we do not find substantial evidence of individual level frailty, we have not extended the analysis to frailty distributions with higher number of support points.

3.5 Model Implementation

We have formulated the model in the Bugs language and performed parameter estimation using WinBUGS 1.4 (Spiegelhalter et al., 1999).

4 Results and discussion

Bhattacharjee et al. (2002) have analysed firm exits in the UK over the period 1965 to 1998. The data pertain to around 4300 listed manufacturing companies covering approximately 48,000 company years, and include 206 exits due to bankruptcy. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1965, and contain delayed entries. A major focus of the analysis is on the effect of macroeconomic conditions and instability on business failure. Age is measured in years post-listig, and all time varying covariatesare measured at an annual frequency. Industry dummies are included in the analysis – these are fixed covariates.

Since the data includes delayed entries, our inference will be based solely on the partial likelihood based on an appropriate definition of risk sets. Partial likelihood inference is valid in a wide range of situations with delayed entries (Andersen et al., 1993), even though some standard properties of counting processes do not hold here.

Four measures of macroeconomic conditions and instability are considered: (a) US business cycle (Hodrick-Prescott filter of US output per capita), (b) instability in foreign currency markets (maximum monthly change, year on year for each month, in exchange rates over a year), (c) instability in prices (similar to exchange rates, but measured in terms of RPI inflation), and (d) a measure of business cycle turnaround (measured by the curvature, or second order difference, of the annual Hodrick-Prescott filtered series of UK output per capita). Theory suggests that the effect of the first and the fourth measure on bankruptcy may be negative, and the second and third ones positive. Because of learning effects, the adverse impact of instability is expected to decline in the age of the firm, post-listing. Similarly, the effect of the US business cycle, negative initially, may also rise with age.

A firm level variable – size, measured as logarithm of gross fixed assets in real terms – is also included as a covariate.

Industry dummies are used as fixed effects control for unobserved factors at the industry level.

We now report the results of two models under different specifications of the prior distribution and different order restrictions, and corresponding model estimates.

4.1 Model A

For the i-th subject (in this case company), let the corresponding counting process be denoted by $N_i(t)$. We model the process as having increments $dN_i(t)$ in the time interval $[t, t + dt)$ distributed as independent Poisson random variables with means $\Lambda_i(t)dt$.

For computational simplicity we use the conjugacy property of Poisson-Gamma distributions in this context and model the baseline hazard function as a Gamma distributed random variable for each distinct age (measured in years). In our implementation, we model the baseline hazard $\lambda_0(t)$ using a Gamma process prior with unit mean.

Two time varying macroeconomic indicators are included as covariates, namely instability in exchange rates and business cycle turnaround. Note that these indicators are calender time specific, while their effect on a company could potentially depend on the age of the company. Therefore, these two covariates are assumed to have age varying effects; we denote the covariates by $Z_e^v(t)$ and $Z_t^v(t)$ respectively.

Further information on company size, industry code, etc. are available but not used in the current preliminary model. Also, no order restriction on ageing is included in the model.

Annual unbalanced panel data on 4320 listed companies over the period 1965 to 2000 are used for the analysis, accumulating to a total of 45546 company years. The maximum age observed in this data was 50 years. As mentioned above, calender year specific data on exchange rates and US business cycle were included in the analysis.

A total of 166 exits due to bankruptcy (involuntary liquidation) were observed for these 4320 companies. Age at exit ranges form 1 year to 48 years. However, very few exits were observed after the age of 35 years. The lack of failure data on the age range between 35-48 years requires a slightly stronger modelling assumption in order to obtain usable inference.

The distributional assumptions for the likelihood and priors for this model are described in the following

$$\begin{aligned} dN_i(t) &\sim Poisson\left[\Lambda_i(t)dt\right], \\ \Lambda_i(t)dt &= d\Lambda_0(t) \times \exp\left[\beta_e^v(t) \times Z_e^v(t) + \beta_t^v(t) \times Z_t^v(t)\right], \\ d\Lambda_0(t) &\sim Gamma(1,1), \quad \text{for } t = 1, \ldots, 50. \end{aligned} \tag{5}$$

where $d\Lambda_0(t) = \Lambda_0(t)dt$ is the increment in the integrated baseline hazard function during the time interval $[t, t + dt)$, with Z's and β's being the corresponding (age varying) covariates and (possibly age varying) regression coefficients.

Economic intuition, and prior empirical evidence, indicates that the effect of the business cycle on bankruptcy hazard is negative while te covariate effect of exchange rate instability is positive. Further, these effects are strong for a newly listed firm but gradually wane off with age (Bhattacharjee et al., 2002). As mentioned above we will not assume any order restrictions on the covariate effects explicitly, however we would like to infer on the direction of effect and variation of covariate effects with age. This structure is incorporated in the prior distributions as follows:

a) $\beta_e^v(1) \sim Normal(25, 0.1)$ and $\beta_t^v(1) \sim Normal(-25, 0.1)$. Note that the second parameter of normal indicates precision (i.e. inverse variance) and not variance.

b) $\beta_k^v(t) \sim Normal(\beta_k^v(t-1), 0.1)$ where $k = e, t$ and $t = 1, \ldots, 35$.

c) $\beta_k^v(t) \sim Normal(\beta_k^v(t-1), 1)$ where $k = e, t$ and $t = 36, \ldots, 50$. Note that, data for later ages do not contain as much information as earlier ones. The precision is accordingly set at a higher value to adjust for the lack of data and to control the compounding propagation of uncertainty through the first order model.

The posterior distributions, based on Model A, for the age varying covariate effects and the baseline hazard function offer useful and intuitively appealing interpretation. The baseline hazard estimates do not show any apparent trend. In other words, no substantial ageing is evident in the data, after accounting for covariate effects of exchange rate instability and business cycle turnaround.

However noticeable trend over time is evidenced in the regression coefficients. The posterior estimates strongly reflect the age-varying nature of the effect of exchange rate instability (Figure 1). There is a strong positive effect on exits when the firm is newly listed, but the effect decreases with age and dies out at about the age of 13 years post-listing.

Similarly, the age varying effect of business cycle turnaround is negative initially and rises to zero with age (Figure 2).

It is worth noting that these observed trends in the posterior is actually a contribution from the data and not from the prior. In fact, other than setting positive or negative direction for only the initial starting values for regression coefficients of the two covariates no further structural assumptions were made.

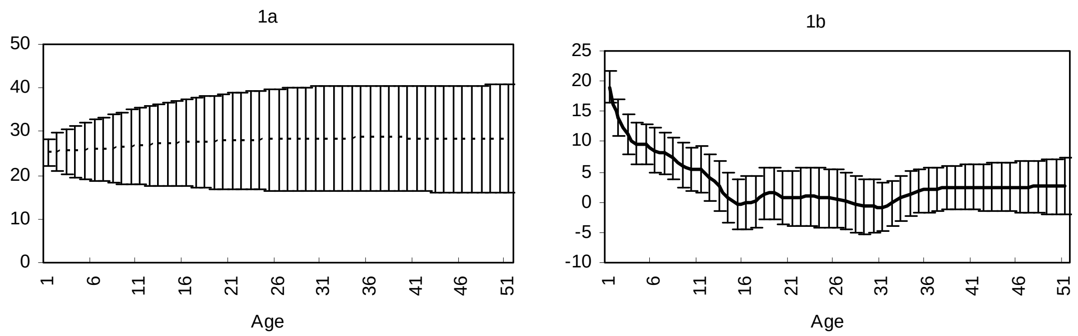

Figure 1: Age varying covariate effects for exchange rate volatility:
(a) Prior (b) Posterior

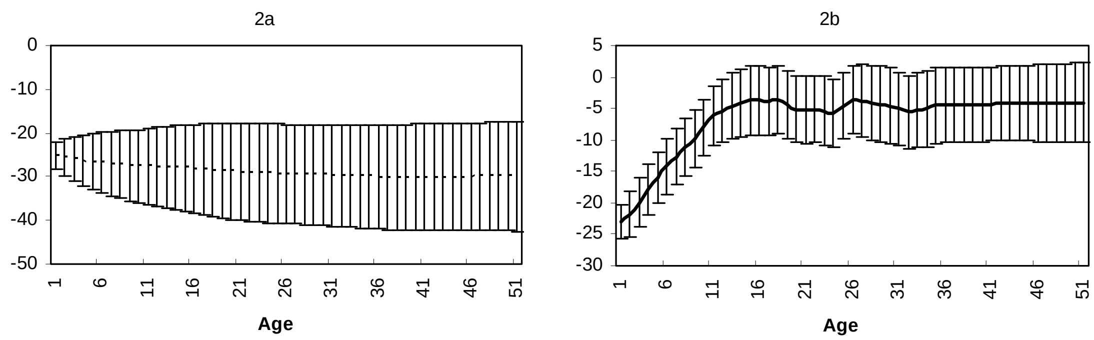

Figure 2: Age varying covariate effects for business cycle turnaround:
(a) Prior (b) Posterior

Therefore the results confirm the economic intuition and prior evidence on order restrictions in covariate dependence. In summary, the model which is rather simplistic nevertheless seems to yield meaningful and useful results.

4.2 Model B

Having experimented with a rather simplistic hazard regression model in the preceding subsection, we now enhance the model in several important ways. First, in addition to macroeconomic factors, we include covariate effect in an important firm level covariate – size (measured by the log of gross fixed assets). Second, we drop business cycle turnaround and include instability in price and the US business cycle as covariates. Third, we include several industry dummies to account for unobserved fixed effects heterogeneity at the industry level. Fourth, and in addition to the above, we include a multiplicative frailty term representing unobserved heterogeneity orthogonal to observed covariates. The frailty distribution is modeled as a two support point multinomial distribution. Fifth, we now measure age in years since inception, rather than years post-listing. This change is motivated partly by the lack of evidence on negative ageing in the baseline hazard function, with age measured in years post listing. The current definition of age is more in line with prior research in empirical industrial organisation, where negative ageing is interpreted as evidence of learning.

Because our model now includes individual level frailty, our dataset needs to be modified to ensure that all included firms contain data for at least two years. We also include two additional years of data on UK listed firms; our data now covers the period 1965 to 2002. Further, as discussed above, we now measure age in years since inception. The data includes 4117 companies with 48176 company years. The maximum age of any company covered in these data is 186 years and maximum exit age is 113 years. The data includes 208 exits due to bankruptcy, of which 203 exits occur by the age of 50 years post listing.

As before we continue to exploit the conjugacy property of Poisson-Gamma distributions and the baseline hazard function is modelled as a Gamma distributed random variable in each year. However the prior distribution for the baseline hazard is adjusted to reflect the availability of information at different ages. This is achieved by allowing the variance to depend on the number at risk at the specified age.

We model the base line hazard $\lambda_0(t)$ using a Gamma process prior, with the parameter depending on the number at risk at each age. The prior distribution is defined as follows:

a) $d\Lambda_0(1) \sim Gamma(1,1)$,

b) $d\Lambda_0(t) \sim Gamma\,[\alpha_1(t), \alpha_2(t)]$, for $t = 2, \ldots, 50$ where $\alpha_1(t)$ and $\alpha_2(t)$ such that the mean is $d\Lambda_0(t-1)$ and variance $Y(t)/100$ ($Y(t)$ being the number at risk at age t), and

c) $d\Lambda_0(t) = d\Lambda_0(t-1)$ for $t > 50$.

We implement the hazard regression model with fixed and age-varying covariate effects, with fixed effects heterogeneity, and with individual level frailty (Equation 4) as follows:

$$\Lambda_i(t)dt = d\Lambda_0(t) \times exp\left[\begin{array}{c} \sum_{j=1}^{J} \beta_j^{(d)}.J_{ji}^{(d)} + \beta_s^{(f)}.z_{si}^{(f)}(t) + \beta_y^{(f)}.z_{yi}^{(f)}(t) \\ +\beta_e^{(v)}(t).z_{ei}^{(v)}(t) + \beta_\pi^{(v)}(t).z_{\pi i}^{(v)}(t) + \theta_i \end{array}\right] \quad (6)$$

The following covariates are included in the model:

1. *Industry dummies*, $J_{ji}^{(d)}$ (J distinct industries, $j = 1, \ldots, J$), are included in the analysis as fixed covariates.with corresponding age constant fixed effects coefficients $\beta_j^{(d)}$,

2. *Covariates with proportional hazards (with age constant covariate effects)*: $z_{si}^{(f)}(t)$ is size of the firm and $z_{yi}^{(f)}(t)$ is a measure of the US business cycle (Hodrick-Prescott filter of output per capita), with corresponding coefficients $\beta_s^{(f)}$ and $\beta_y^{(f)}$,

3. *Covariates with age varying coefficients*: $z_{ei}^{(v)}(t)$ and $z_{\pi i}^{(v)}(t)$ denote exchange rate and price instability, with corresponding nonproportional covariate effects $\beta_e^{(v)}(t)$ and $\beta_\pi^{(v)}(t)$ respectively (the covariate effects are expected to be positive initially and decreasing with age), and

4. $\nu_i = exp(\theta_i)$ is an individual level multiplicative frailty term with a two point support distribution.

The prior distribution for log-frailty (θ_i) is modeled as having two support points $m_1 = 0$ and m_2, with corresponding probabilities p_1 and $p_2 = 1 - p_1$; m_1 is fixed at zero because of scaling. We assume a standard normal distribution for the prior of m_2. The population assignment of a company is then given by a latent variable, here assumed to have a multinomial distribution with a Dirichlet prior for the probability p_1. Our implementation, which is similar to Campolieti (2001), has two major advantages. First, it exploits the Multinomial-Dirichlet conjugacy property which helps in computations. Second, the model is easily extendible to a larger number of support points for the frailty distribution.

Standard normal priors were considered for the industry fixed effects.

For the time constant coefficients nearly half normal distributions were considered as priors, with a slight shift from zero:

$$\beta_s^{(f)}, \beta_y^{(f)} \sim Normal(-0.01, 10) \text{ truncated on } (-\infty, 0).$$

For the age varying coefficients decreasing with age, Gamma distributed increments were taken away from the coefficient at the previous age to maintain monotonicity in the prior distributions:

a) $\beta_k^{(v)}(1) \sim Normal(0.25, 1),\ k = e, \pi;$

b) For $t \in (2, 50)$, $\beta_k^{(v)}(t) = \beta_k^{(v)}(t-1) - \left[b_k^0(t-1) \times \frac{Y(t)}{c}\right]$, where $b_k^0(t-1) \sim Gamma(0.01, 1)$, $Y(t)$ is the number at risk at age t, and c is the maximum number at risk at any age in the data.

c) For $t > 50$ $\beta_k^{(v)}(t) = \beta_k^{(v)}(t-1)$

The posterior estimates for the baseline hazard function (Figure 3a) do not show any obvious evidence of ageing. This is a bit surprising since earlier work has found evidence of negative ageing. This observation, however, does not seem to be feature of the current data. In fact, estimates of the baseline hazard function based on the partial likelihood estimates also show a very similar age-varying pattern to the posterior mean (Figure 3a).

The age varying covariate effects for exchange rate and price instability (Figures 4 and 5 respectively) indicate strong evidence of non-proportionality. The age-specific coefficients are positive when the firm is newly listed, but decline to zero as the firm gets older.

The usefulness of our model of unobserved heterogeneity, in terms of fixed effects heterogeneity at the industry level combined with individual level frailty with distribution on a finite number of support points, is emphasized by the empirical results. The posterior distributions of the industry level fixed effects demonstrate evidence of substantial unobserved heterogeneity (Figure 3b). Other factors being equal, high technology industries such as

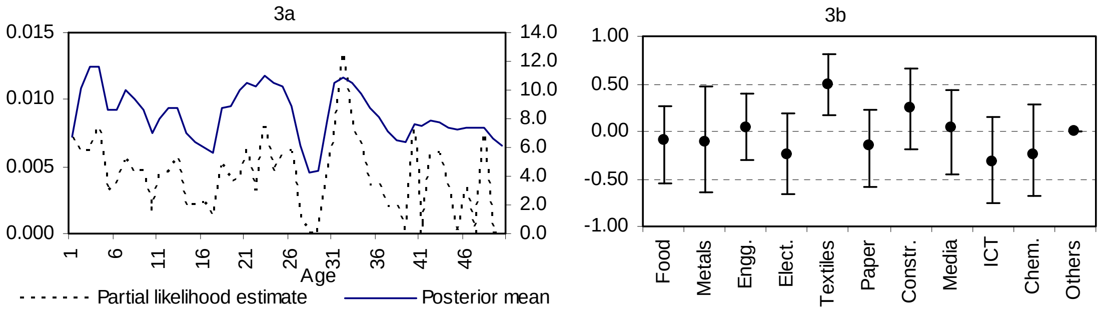

Figure 3: Posterior Estimates: (3a) Baseline hazard, (3b) Industry fixed effects (with 95% posterior intervals)

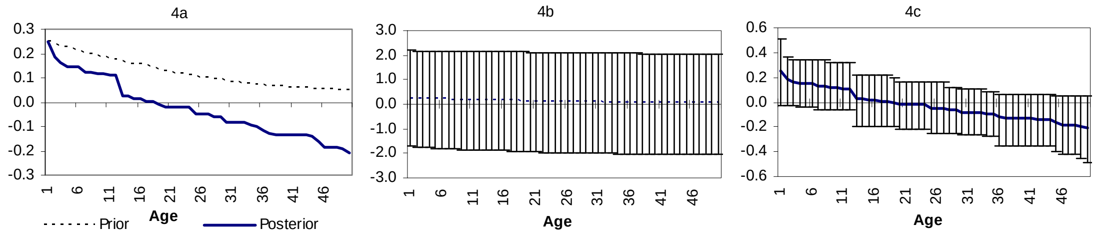

Figure 4: Age varying covarite effects for exchange rate volatility:
(a) Prior and posterior mean (b) prior mean and 95% interval (c) posterior mean and 95% interval

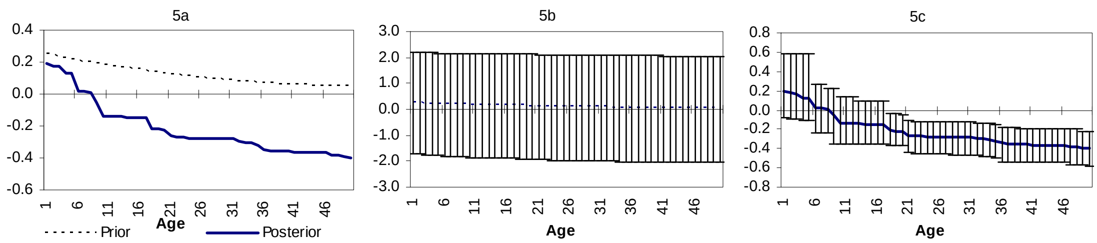

Figure 5: Age varying covarite effects for price instability:
(a) Prior and posterior mean (b) prior mean and 95% interval (c) posterior mean and 95% interval

"ICT" and "Electronics and Electricals" have a lower hazard rate of exit due to bankruptcy, while the "Textiles" industry attracts a substantially higher hazard. This is in reasonable agreement with economic intuition and prior empirical evidence.

At the same time, we do not evidence of multiplicative frailty at the level of the individual firm. In fact, the posterior distribution of frailty converges to a single mass point. From an economic point of view, this evidence is not surprising, because unobserved human capital may be rather homogeneous in a sample of successful listed firms.

In summary, we find strong support for the order restrictions on covariate dependence, but not much evidence of expected shape in the baseline hazard function. We also find that the models and priors developed here are useful for inference on order restricted covariate dependence and ageing, as well as on the effect of unobserved heterogeneity.

5 Conclusion

There has not been much research on order restricted Bayesian inference in survival models. In this paper, we make contributions to this literature by proposing a Bayesian framework for order restricted inference in hazard regression models in the presence of unobserved heterogeneity. We consider constraints on covariate dependence; these constraints are in the nature of convex (concave) ordering of lifetime distributions conditional on distinct covariate values. Our proposed methods are very useful in understanding covariate dependence in situations where the proportional hazards assumption does not hold.

In addition to covariate dependence, we also discuss order restrictions on the shape of the baseline hazard function. These order restrictions inform about ageing properties of the lifetime distributions, holding observed covariates and frailty constant.

Our methodology pays special attention to the modeling of frailty. In addition to fixed effects unobserved heterogeneity, we model individual level frailty nonparametrically using an expanding sequence of multinomial distributions. This is in sharp contrast to the existing literature where frailties are assumed to have parametric distributions that do not offer additional insights.

The analysis of corporate failure data using our methdology offers interesting new evidence on the nature of covariate dependence. In particular, we find that the macroeconomic environment has a strong effect on the hazard rate of firm exits die to bankruptcy. Further, the effect of adverse economic conditions which is quite drastic on young firms decreases to zero as the firm gains in experience. However, in our application, we do not find much evidence on ageing characteristics in the baseline hazard function.

While we observe substantial fixed effects unobserved heterogeneity at the industry level, evidence points to absence of significant multiplicative frailty at the level of the individual firm.

References

[1] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models based on Counting Processes*. Springer-Verlag, New York.

[2] Arjas, E. and Bhattacharjee, M. (2003). Modelling heterogeneity: Hierarchical Bayesian approach. In Mazzuchi, T.A., Singpurwalla, N.D. and Soyer, R. (Eds.), Mathematical Reliability: An Expository Perspective, Kluwer Academic Publishers.

[3] Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, **4**, 505–524.

[4] Arjas, E. and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association*, **91**, 1101–1109.

[5] Bhattacharjee, A. (2004). Estimation in hazard regression models under ordered departures from proportionality. *Computational Statistics and Data Analysis*, **47**(3), 517–536.

[6] Bhattacharjee, A. (2006). Testing proportionality in duration models with respect to continuous covariates. *Mimeo*.

[7] Bhattacharjee, A. (2007). A simple test for the absence of covariate dependence in hazard regression models. *Mimeo*.

[8] Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2002). Macro economic instability and business exit: Determinants of failures and acquisitions of large UK firms. DAE Working Paper No. **0206**, Department of Applied Economics, University of Cambridge.

[9] Bhattacharjee, M., Arjas, E. and Pulkkinen, U. (2003). Modelling heterogeneity in nuclear power plant valve failure data. In Lindqvist, B. and Doksum, K. (Eds.), Mathematical and Statistical Methods in Reliability, World Scientific Publishing.

[10] Campolieti, M. (2001). Bayesian semiparametric estimation of discrete duration models: An application of the Dirichlet process prior. *Journal of Applied Econometrics*, **16**, 1–22.

[11] Card, D. and Olson, C.A. (1992). Bargaining power, strike duration, and wage outcomes: An analysis of strikes in the 1880s. Working Paper No. **4075**, National Bureau of Economic Research, Cambridge, MA.

[12] Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **64**, 141–151.

[13] Damien, P., Laud, P. and Smith, A. (1996). Implementation of Bayesian nonparametric inference using Beta processes. *Scandinavian Journal of Statistics*, **23**, 27–36.

[14] Deshpande, J.V. and Sengupta, D. (1995). Testing for the hypothesis of proportional hazards in two populations. *Biometrika* **82**, 251–261.

[15] Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Annals of Probability*, **2**, 183–201.

[16] Dykstra, R.L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association*, **77**, 621–628.

[17] Dykstra, R.L., Kochar, S. and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics*, **19**, 870–888.

[18] Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Annals of Statistics*, **9**, 356–367.

[19] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

[20] Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Annals of Statistics*, **7**, 163–186.

[21] Gamerman, D. (1991). Dynamic bayesian models for survival data. *Applied Statistics*, **40**, 63–79.

[22] Gelfand, A.E. and Kottas, A. (2001). Nonparametric Bayesian modeling for stochastic ordering. *Annals of the Institute of Statistical Mathematics*, **53**, 865–876.

[23] Gelfand, A.E., Smith, A.F.M. and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using the Gibbs sampler. *Journal of the American Statistical Association*, **87**, 523–532.

[24] Gill, R.D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika*, **74**, 289–300.

[25] Gray, R. J. (1994). A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, **50**, 244–253.

[26] Heckman, J.J. and Singer, B. (1984). A method for minimising the impact of distributional assumptions in econometric models for duration data. *Econometrica*, **52**, 271–320.

[27] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, **18**, 1259–1294.

[28] Kalashnikov, V.V. and Rachev, S.T. (1986). Characterisation of queuing models and their stability, in Probability Theory and Mathematical Statistics (eds. Yu.K. Prohorov et. al.), VNU Science Press, 2, 37–53.

[29] Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society* Series B, **40**, 214–221.

[30] Laud, P., Damien, P. and Smith, A. F. M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In Dey, D., Müller, P. and Sinha, D. (Eds.) *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer-Verlag, 213–225.

[31] Metcalf, D., Wadsworth, J. and Ingram, P. (1992). Do strikes pay?, Centre for Economic Performance, Discussion Paper No. **92**, ESRC Research Centre, London School of Economics.

[32] Narendranathan, W. and Stewart, M.W. (1993). How does the benefit effect vary as unemployment spells lengthen?. *Journal of Applied Econometrics*, **8**, 361–381.

[33] Nieto-Barajas, L. and Walker, S. G. (2002a). Bayesian nonparametric survival analysis via Lévy driven Markov processes. Technical report, Department of Mathematical Sciences, University of Bath.

[34] Nieto-Barajas, L. and Walker, S. G. (2002b). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, **29**, 413–424.

[35] Qiou, Z., Ravishanker, N. and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics*, **55**, 637–644.

[36] Sengupta, D., Bhattacharjee, A., and Rajeev, B. (1998). Testing for the proportionality of hazards in two samples against the increasing cumulative hazard ratio alternative. *Scandinavian Journal of Statistics* **25**, 637–647.

[37] Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, **88**, 979–983.

[38] Sinha, D., Chen, M.-H. and Ghosh, S.K. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, **55**(2), 585–590.

[39] Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1999). WinBUGS Version 1.2 User Manual. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

[40] Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society* Series B, **59**, 845–860.

[41] Wild, C. J. and Kalbfleisch, J. D. (1981). A note on a paper by Ferguson and Phadia. *Annals of Statistics*, **9**, 1061–1065.

———x———