Fábio Nogueira Demarqui

Uma classe mais flexível de modelos semiparamétricos para dados de sobrevivência

Belo Horizonte - MG

Junho / 2010

Fábio Nogueira Demarqui

Uma classe mais flexível de modelos semiparamétricos para dados de sobrevivência

Tese de doutorado apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estatística.

Orientadora: Rosangela H. Loschi

Co-orientadores: Enrico A. Colosimo & Dipak K. Dey

Doutorado em Estatística
Departamento Estatística
Centro de Ciências Exatas
Universidade Federal Minas Gerais

Belo Horizonte - MG

Junho / 2010

Ao meu irmão,
Edgar.

Agradecimentos

A minha família, pelo exemplo e apoio incondicional. Vocês são a razão da minha força e inspiração para seguir sempre em frente.

À Graici, por simplesmente fazer parte da minha vida.

A Rosangela e o Enrico, pela orientação, confiança e, sobre tudo, pela amizade que se formou no decorrer desses quase 6 anos de trabalho em conjunto. Fico com a certeza de que ainda temos muito trabalho pela frente.

Ao professor Dipak Dey, pela excelente acolhida, e, principalmente, pela grande oportunidade que me proporcionou ao aceitar-me em seu departamento. Tenho aprendido muito com o Sr., Dr. Dey. Gostaria de agradecer também aos funcionários e colegas do departamento de Estatística da UCONN por toda hospitalidade e ajuda.

Aos professores do Departamento de Estatística da UFMG, em especial a Denise, Frederico, Gregório, Luiz, Marcelo, Mercedes e Renato. Gostaria também de agradecer a todos os funcionários do Departamento de Estatística, em particular a Cristina, Marcinha, Rogéria e Rose.

À Maristela, amiga de todas as horas, meu agradecimento especial, pelo companheirismo, por ter me ajudado em tudo um pouco, e pelas nossas infindáveis discussões. Se estas, na maioria das vezes não nos levou a lugar algum, pelo menos nos divertimos a valer.

Ao Marcos, Rafael (Waldo) e Nathalie, por toda a ajuda durante o período em que visitei os EUA, e, principalmente, pela amizade que se formou entre nós. Guardo vocês no coração.

Ao Max, Rodrigo e Juan, pelas inumeras discussões sobre o trabalho da tese, ou simplesmente por me ouvirem nos meus momentos de devaneios.

Aos demais amigos que, independentemente da distância, acompanharam-me nesta jornada. Não citarei nomes, pois a lista seria muito extensa para ser colocada aqui. Além disso, assim não correrei o risco de acabar esquecendo de alguém.

À CAPES, pelo suporte financeiro concedido para o desenvolvimento desta pesquisa, sobretudo, pela bolsa de doutorado sanduíche que possibilitou a minha ida aos Estados Unidos. Gostaria de agradecer também a FAPEMIG e ao CNPq pelo auxílio financeiro provido em diversas ocasiões.

"A razão cardeal de toda a superioridade humana é sem dúvida a vontade. O poder nasce do querer. Sempre que o homem aplicar a veemência e a perseverante energia de sua alma a um fim, ele vencerá os obstáculos e, se não atingir o alvo, pelo menos fará coisas admiráveis."

José de Alencar

Resumo

O modelo exponencial por partes (MEP) é uma alternativa aos modelos paramétricos bastante flexível e popular em análise de sobrevivência. Embora paramétrico em um sentido estrito, o MEP pode ser pensado como um modelo não-paramétrico, uma vez que a sua função risco não apresenta restrições quanto a sua forma. Por esta razão, o MEP tem sido bastante utilizado na literatura relativa à análise de sobrevivência e confiabilidade. Apesar de sua popularidade, o maior desafio de se trabalhar com o MEP reside na especificação da grade de tempos necessária para o seu ajuste. Nesta tese de doutorado são apresentadas extensões da abordagem introduzida por Demarqui *et al.* (2008) para a modelagem da grade do MEP. As contribuições mais importantes deste trabalho dizem respeito à generalização e unificação da abordagem inicialmente proposta por Demarqui *et al.* (2008) com outras abordagens já consagradas na literatura. A metodologia proposta se encaixa em diferentes contextos de análise de sobrevivência, e pode ser aplicada a dados do tipo tempo até a ocorrência de um evento de interesse oriundos de qualquer área do conhecimento. Em particular, a metodologia apresentada é utilizada para a estimação da função risco, bem como o ajuste de modelos de regressão e modelos com fração de cura. A abordagem proposta pode, ainda, ser estendida para a modelagem de dados de sobrevivência multivariados e/ou espacialmente correlacionados. O mecanismo usado para modelar a aleatoriedade da grade do MEP possui algumas características interessantes que não são compartilhadas por outros métodos que têm sido propostos para resolver o problema. Em especial, as restrições impostas ao conjunto de possíveis grades torna possível a sua estimativa após especificações a *priori* adequadas. Além disso, a existência de pelo menos um tempo de falha em cada intervalo aleatório induzido pelas grades aleatórias fica garantida. Outra vantagem da abordagem proposta é que a disposição dos tempos de falha sobre o eixo do tempo é levada em conta na modelagem da grade. Os modelos resultantes incluem outros modelos consagrados na literatura como casos especiais, e fornece uma estrutura flexível para a modelagem de dados de sobrevivência. Propriedades dos modelos propostos são discutidas e a utilização da nova metodologia é exemplificada através da análise de conjuntos de dados reais. Para fins de comparação, os resultados obtidos são comparados com aqueles fornecidos por outros métodos existentes na literatura. Os modelos propostos neste trabalho são bastante gerais, e podem ser utilizados para a modelagem de dados de sobrevivência resultantes de qualquer campo área do conhecimento, desde que os tempos de sobrevivência sejam censurados à direita e ao mecanismo de censura seja não informativo.

Abstract

The piecewise exponential model (PEM) is a quite attractive and popular alternative to parametric models in survival analysis. Although parametric in a strict sense, the PEM can be thought of as a nonparametric model as far as its hazard function does not have a closed shape. For this reason the PEM has been widely used in the literature to model time-to-event data. Despite its popularity, the greatest challenge of working with the PEM is the specification of the time grid needed to fit this model. In this thesis we present some extensions of the approach introduced by Demarqui *et al.* (2008) to fit the PEM with random time grid. The contributions of this work are twofold. First, we provide a more flexible framework for modeling the randomness of the time grid of the PEM. Then, we show how this new framework can be extended to accommodate accordingly other well-known approaches available in the literature. The new methodology is suitable for time-to-event data arising from any field of knowledge. It can be used in problems involving hazard function estimation, as well as to fit regression and cure rate models. The proposed approach can be further extended to model multivariate and/or spatially correlated survival times. The mechanism used to model the randomness of the time grid of the PEM has some nice features not shared by other approaches that have been proposed in the literature. In particular, the constrained imposed on the set of possible time grids for the PEM turns possible its estimation after directly prior elicitation. It further guarantees the existence of at least one failure time in each random interval induced by the random time grids. In addition, the arrangement of the failure times on the time axes is taking into account in the modeling of the time grid. The resultant models include other models established in the literature as special cases and provides a flexible framework for survival data modeling. Properties of the proposed models are discussed and the use of the new methodology is exemplified through the the analysis of a real data sets. For comparison purposes, the results obtained are compared with those provided by other methods existing in the literature. The new approaches introduced in this work are quite general and can be applied to model time to event data arising from any field of science, provided that survival times are right censored and the censoring mechanism is not informative.

Introdução

O avanço computacional observado principalmente nas duas últimas décadas, aliado à disponibilidade de algoritmos computacionais mais eficientes, tem possibilitado sobremaneira o desenvolvimento de métodos estatísticos para a análise de dados com estruturas cada vez mais complexas. Em análise de sobrevivência, este recente desenvolvimento tecnológico tem impulsionado o desenvolvimento de modelos semiparamétricos e não-paramétricos como alternativas mais flexíveis e robustas aos modelos paramétricos. Em particular, abordagens Bayesianas para a modelagem de dados de sobrevivência tem recebido lugar de destaque na literatura recente. Seguindo essas tendências, e adotando o enfoque Bayesiano, nesta tese de doutorado são propostas algumas extensões, com enfoque Bayesiano, do Modelo Exponencial por Partes (MEP), incluindo a modelagem direta da função risco, modelos de regressão semi-parametricos e modelos com fração de cura.

O MEP é um dos modelos mais populares para a modelagem de dados de sobrevivência. Embora paramétrico num senso estrito, o MEP pode ser pensado como um modelo não-paramétrico, uma vez que a sua função risco não apresenta restrições quanto a sua forma, como ocorrem nos casos das distribuições Weibull e exponencial, por exemplo. Essa interessante característica do MEP permite-nos aproximar de maneira satisfatória funções riscos de variadas formas, sem que haja a necessidade de impormos muitas restrições para obtermos um ajuste adequado do modelo aos dados. Devido a essa flexibilidade, o MEP tem sido largamente utilizado para se modelar dados do tipo "tempo até a ocorrência de um evento de interesse" em uma gama de contextos, tais como situações clínicas incluindo infecção de rins (Sahu *et al.* (1997)), transplante de coração (Aitkin *et al.* (1983)), mortalidade hospitalar (Clark e Ryan (2002)), e estudos envolvendo diferentes tipos de câncer como leucemia (Breslow (1974)), câncer gástrico (Gamerman (1991)), câncer de mama (Ibrahim *et al.* (2001b), veja também para uma aplicação do MEP em dados com censura intervalar), melanoma (Kim *et al.* (2006)) e câncer da nasofaringe (McKeague e Tighiouart (2000)), entre outros. O MEP também tem sido utilizado em engenharia, em estudos de confiabilidade (Kim e Proschan (1991) e Gamerman (1994)), e problemas econômicos (Gamerman (1991) e Bastos e Gamerman (2006)). Para uma aplicação em dados com censura intervalar, veja Sinha *et al.* (1999).

O MEP é caracterizado pela aproximação da função risco por segmentos de retas cujos comprimentos são determinados por uma grade de pontos que divide o eixo do tempo em um número finito, digamos b, de intervalos. Para construirmos este modelo, precisamos especificar uma grade de pontos, digamos $\tau = \{s_0, s_1, ..., s_b\}$, tal que $0 = s_0 < s_1 < \cdots < s_b < \infty$. Esta grade τ induz uma partição do eixo do tempo em intervalos contíguos da forma $I_j = (s_{j-1}, s_j]$, dentro dos quais assumimos uma taxa de falha constante, λ_j, $j = 1, \cdots, b$. Logo, a função risco, avaliada num tempo $t > 0$, é aproximada por:

$$h(t) = \lambda_j, \ t \in I_j = (s_{j-1}, s_j], \quad j = 1, \cdots, b.$$

Existem vários trabalhos que discutem propriedades e extensões do MEP. Friedman (1982) faz uso do MEP para modelar a função de risco de base do modelo de Cox, e apresenta condições para a existência e distribuição assintótica dos estimadores de máxima verossimilhanca (EMV) para as taxas de falha e coeficientes da regressão. Kim e Proschan (1991) discutem algumas vantagens do MEP sobre o estimador de Kaplan-Meyer para a função de sobrevivência para dados sem covariáveis. Barbosa *et al.* (1996) aplicam métodos usados em modelos lineares generalizados para ajustar dados de tempos de vida acelerados utilizando o MEP. O método da máxima verossimilhanca também é empregado por Chen e Ibrahim (2001) para ajustar modelos com fração de cura utilizando o MEP na presença de dados faltantes. A literatura Bayesiana relacionada ao MEP também é bastante extensiva. Gamerman (1991) estende o modelo de Cox propondo uma abordagem dinâmica para modelar dados de sobrevivência em que os efeitos de variáveis explicativas variam ao longo do tempo. Sahu *et al.* (1997) usa o MEP para modelar dados de sobrevivência multivariados. Em tal trabalho, fragilidades (i.e., efeitos aleatórios) são consideradas para acomodar a correlação inerente aos tempos de sobrevivência de elementos pertencentes a um mesmo grupo, e um processo correlacionado é assumido para modelar a incerteza sobre as taxas de falha do MEP. Uma breve revisão sobre algumas abordagens com enfoque Bayesiano para o MEP pode ser encontrada em Ibrahim *et al.* (2001b).

A escolha da grade $\tau = \{s_0, s_1, ..., s_J\}$ desempenha um papel central na qualidade do ajuste do MEP. Uma grade contendo um número muito grande de intervalos produz estimativas instáveis para as taxas de falha, enquanto um número muito pequeno de intervalos, em geral, leva a uma aproximação grosseira da função de sobrevivência. Na prática, desejamos uma grade capaz de produzir boas aproximações tanto para a função risco quanto para função de sobrevivência. Esta questão tem sido um dos grandes desafios para se trabalhar com o MEP. Embora exista uma vasta literatura relacionada a este

modelo, a grade $\tau = \{s_0, s_1, ..., s_b\}$ tem sido escolhida de maneira arbitrária na maioria dos trabalhos disponíveis, entre os quais os citados anteriormente. Kalbfleisch e Prentice (1973) sugerem que a escolha da grade deveria ser feita independentemente dos dados, mas não apresentam um procedimento para tanto. Breslow (1974) propõe que se tome os pontos finais s_j dos intervalos $I_j = (s_{j-1}, s_j]$ como sendo iguais aos tempos de falha observados. Outras discussões heurísticas com respeito a possíveis escolhas para a grade $\tau = \{s_0, s_1, ..., s_b\}$ podem ser encontradas em Gamerman (1991), Sahu *et al.* (1997) e Qiou *et al.* (1999), entre outros.

Na prática, o problema de especificarmos uma grade apropriada para o ajuste do MEP pode ser resolvido assumindo-se que $\tau = \{s_0, s_1, ..., s_b\}$ é uma quantidade aleatória. O primeiro esforço efetivo nesta direção é devido a Arjas e Gasbarra (1994). Estes autores assumen que os pontos finais dos intervalos $I_j = (s_{j-1}, s_j]$ são definidos de acordo com um processo de saltos que seguem uma estrutura de martingal, a qual é incluída no modelo através das distribuições *a priori*. Extensões desta abordagem podem ser encontradas em McKeague e Tighiouart (2000) and Kim *et al.* (2006).

Apesar do uso de um processo de Poisson homogêneo (PPH) fornecer uma solução para a questão da modelagem do MEP com grade aleatória, esta estratégia apresenta alguns inconvenientes. Primeiro, tal abordagem não possibilita a modelagem direta da grade que define o MEP, não sendo possível, desta forma, que a mesma seja estimada. Além disso, na ausência de qualquer informação *a priori* sobre λ_j, para que λ_j seja estimável é necessário que cada intervalo $I_j = (s_{j-1}, s_j]$ contenha pelo menos um tempo de falha. Embora esta condição possa ser relaxada assumindo-se um processo correlacionado para a função risco (veja, por exemplo, Gamerman (1991, 1994) e Ibrahim *et al.* (2001b)), intervalos contendo pelo menos um tempo de falha são preferíveis no processo de estimação. Ademais, a disposição dos tempos de falha sob o eixo do tempo não é diretamente considerada nesta abordagem. Finalmente, assumindo de um PPH para gerarmos o número de intervalos e seus respectivos pontos finais, o número total de parâmetros do modelo passa a ser aleatório, tornando necessário o uso de um algoritmo MCMC com saltos reversíveis.

Uma alternativa inovadora para a modelagem do MEP com grade aleatória é introduzida por Demarqui *et al.* (2008). Nesse trabalho, a estrutura de agrupamento do Modelo Partição Produto (MPP) proposto por Barry e Hartigan (1992) é utilizada para modelar diretamente a aleatoriedade de $\tau = \{s_0, s_1, ..., s_b\}$. Desta forma, a abordagem proposta por Demarqui *et al.* (2008) é completamente Bayesiana, e permite a estimação da grade que define o MEP, tornando a modelagem muito mais atrativa e elegante. Sob

tal abordagem, a modelagem da grade é feita levando-se em conta a disposição dos tempos de falha sob o eixo do tempo, e a existência de pelo menos um tempo de falha em cada intervalo induzido pela grade aleatória do MEP é garantida. Além disto, apesar de o número de parâmetros a serem estimados poder variar, esta forma de modelagem mantem fixado o número máximo de parâmetros do modelo. Além disso, a abordagem proposta por Demarqui *et al.* (2008) pode ser estendida em diversas direções, como por exemplo no ajuste de modelos com covariáveis dependentes do tempo e/ou efeitos das covariáveis mudando ao longo do tempo, modelos com fração de cura, modelos de fragilidade para dados de sobrevivência multivariados e/ou espacialmente correlacionados, entre outros, abrindo, desta forma, um novo leque de pesquisa envolvendo o MEP.

Esta tese é composta por quatro artigos. Em todos esses trabalhos, extensões da abordagem proposta por Demarqui *et al.* (2008) para a modelagem da grade do MEP são consideradas. A metodologia proposta neste trabalho é bastante geral, podendo ser aplicada para dados do tipo tempo até a ocorrência de um evento de interesse oriundos de qualquer área do conhecimento. Entretando, é dado enfoque nesta tese à aplicações oriundas da área médica, especificamente, de dados de sobrevivência referentes a indívíduos diagnosticados com câncer. A suposição básica para a aplicabilidade dos modelos aqui propostos é que os tempos de sobrevivência sejam censurados à direita, ou seja, os tempos até a ocorrência do evento de interesse são maiores ou iguais aos tempos observados, e que o mecanismo gerador das censuras seja não informativo. A abordagem proposta também é apropriada para dados de sobrevivência com a presença de empates. A contribuição de cada artigo é resumida a seguir.

Os dois primeiros artigos apresentados nesta tese versam sobre a situação mais simples encontrada na prática em análise de sobrevivência, que consiste na modelagem direta da função risco na ausência de variáveis explicativas, efeitos aleatórios, etc, assumindo o MEP com grade aleatória. No primeiro artigo uma análise Bayesiana objetiva é proposta, assumindo-se como especificações *a priori* a distribuição de Bayes-Laplace para a grade, e a distribuição de Jeffreys para as taxas de falha do MEP. O tempo de vida de pacientes diagnosticados com câncer no cérebro no condado de Windham-CT, EUA, é analisado utilizando-se a metodologia proposta. Os resultados são comparados com aqueles fornecidos pela abordagem frequentista proposta em Kim e Proschan (1991), e uma discussão sobre as vantagens e desvantagens de cada abordagem é apresentada. No segundo artigo, uma classe ampla e flexível de distribuições *a priori* para a modelagem do MEP é introduzida, e quatro diferentes especificações *a priori* para a taxa de falha são discutidas em detalhe. A classe proposta inclui como casos particulares o modelo proposto por Demar-

qui *et al.* (2008), o modelo apresentado no primeiro artigo, e as abordagens introduzidas por Gamerman (1994) e Ibrahim *et al.* (2001a). A análise dos tempos de sobrevivência de indivíduos diagnosticados com câncer no cérebro no condado americano de Windham-CT, EUA, é novamente apresentada como ilustração dos modelos propostos. As comparações entre os modelos ajustados considerando-se grades fixas e aleatórias, bem como diferentes distribuições *a priori* para a taxa de falha, são realizadas utilizando-se como medida de comparação a média do logaritmo da pseudo-verossimilhança marginal (veja Ibrahim *et al.* (2001b) e referências).

O terceiro artigo trata-se de modelos de regressão semiparamétricos com variáveis explicativas com efeito agindo multiplicativamente na função risco. O modelo proposto neste artigo pode ser visto como uma extensão do modelo de riscos proporcionais proposto por Cox (1971), em que é permitido que covariáveis e ou seus correspondentes efeitos variem ao longo do tempo, e incluí como caso especial o modelo dinâmico introduzido por Gamerman (1991). A análise dos tempos de sobrevivência de indivíduos diagnosticados com câncer no cérebro no condado americano de Windham-CT é novamente apresentada como ilustração, considerando-se o sexo dos pacientes como variável explicativa. Para avaliar-se o desempenho do modelo segundo diferentes especificações a *priori*, considerando-se grades fixas e aleatórias para o MEP, são utilizados o fator de Bayes e as probabilidades *a posteriori* de cada modelo (veja Kass e Raftery (1995) e Ibrahim *et al.* (2001b), entre outros).

Finalmente, no último artigo que compõe esta tese, as abordagens propostas por Ibrahim *et al.* (2001a) e Demarqui *et al.* (2008) são apresentadas de maneira unificada, dando origem a uma nova abordagem semi-parametrica para dados de sobrevivência incluindo uma fração de cura. O modelo resultante difere fundamentalmente do modelo proposto por Kim *et al.* (2006), que também assume o MEP com grade aleatória no contexto de modelos de fração de cura, em dois aspectos. Primeiramente, nossa abordagem modela diretamente a grade de tempos, permitindo, desta forma, a sua estimação, enquanto que, no modelo proposto por Kim *et al.* (2006), o uso de um PPH para a modelagem do MEP com grade aleatória não permite a estimação da mesma. Segundo, embora em ambas as abordagens uma função risco acumulada associada a um modelo paramétrico seja utilizada na especificação a *priori* dos modelos, nossa abordagem modela diretamente as taxas de falha do MEP, ao contrário do modelo proposto por Kim *et al.* (2006), em que distribuições a *priori* são especificadas para o logaritmo das taxas de falha, através de uma abordagem dinâmica. O estudo clínico E1673 conduzido pela ECOG (Eastern Cooperative Oncology Group), EUA, envolvendo pacientes diagnostica-

dos com melanoma, é usado para ilustrar a aplicação deste modelo. O banco de dados E1673 tem sido investigado na literatura por outros autores (veja Ibrahim *et al.* (2001b) e Kim *et al.* (2006)), possibilitando, desta forma, a comparação do modelo proposto com modelos concorrentes. Por este motivo, a soma do logaritmo da pseudo-verossimilhança marginal é utilizada neste caso como medida de comparação para avaliar o desempenho do modelo proposto, de acordo com diferentes especificações *a priori*, bem como estabelecer comparações com o modelo concorrente introduzido por Kim *et al.* (2006)).

Modeling survival data using the piecewise exponential model with random time grid

Fabio N. Demarqui[1] [*], Dipak K. Dey[2],
Rosangela H. Loschi[1] and Enrico A. Colosimo[1]

[1] Universidade Federal de Minas Gerais, Brazil,
email:{fndemarqui, loschi, enricoc}@est.ufmg.br
[2]University of Connecticut, USA,
emai: dey@stat.uconn.edu

Abstract

In this paper we present a fully Bayesian approach to model survival data using the piecewise exponential model with random time grid. We assume a joint noninformative improper prior distribution for the time grid and the failure rates of the PEM, and show how the clustering structure of the product partition model can be adapted to accommodate improper prior distributions in the framework of the PEM. Properties of the model are discussed and the use of the proposed methodology is exemplified through the analysis of a real data set. For comparison purposes, the results obtained are compared with those provided by other methods existing in the literature.

Keywords: Bayesian inference, Gibbs sampler, piecewise exponential model, product partition model, survival analysis.

1 Introduction

In many practical situations, specially those including medical data, it is often not possible to control a significant part of the sources of variation of an experiment. These uncontrolled sources of variation, when present, may compromise considerably one or more assumptions of a given parametric model assumed to fit the data, which may lead to misleading conclusions.

The Piecewise Exponential Model (PEM) arises as a quite attractive alternative to parametric models for the analysis of time to event data. Although parametric in a strict sense,

[*]To whom all correspondences should be addressed. Departament of Statistics, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6.627, Pampulha, 31270-901 - Belo Horizonte - MG, Brazil, Phone: (+55 31) 3409 5938. Fax: (+55 31) 3409 5924.

the PEM can be thought as a nonparametric model as far as it does not have a closed form for the hazard function. This nice characteristic of the PEM allows us to use this model to approximate satisfactorily hazard functions of several shapes. For this reason, the PEM has been widely used to model time to event data in different contexts, such as clinical situations including kidney infection Sahu et al. (1997), heart transplant data Aitkin et al. (1983), hospital mortality data Clark and Ryan (2002), and cancer studies including leukemia Breslow (1974), gastric cancer Gamerman (1991), breast cancer Ibrahim et al. (2001) (see also Sinha et al. (1999) for an application to interval-censored data), melanoma Kim et al. (2006) and nasopharynx cancer McKeague and Tighiouart (2000), among others. The PEM has also been used in reliability engineering (Kim and Proschan, Kim and Proschan (1991), Gamerman, Gamerman (1994)), and economics problems Gamerman (1991), Bastos and Gamerman (2006).

In order to construct the PEM, we need to specify a time grid which divides the time axis into a finite number of intervals. Then, for each interval induced by that time grid, we assume a constant failure rate. Thus, we have a discrete version, in the form of a step function, of the true and unknown hazard function.

The time grid $\tau = \{s_0, s_1, ..., s_J\}$ plays a central role in the goodness of fit of the PEM. It is well know that a time grid having a too large number of intervals might provide unstable estimates for the failure rates, whereas a time grid based on just few intervals might produce a poor approximation for the true survival function. In practice, we desire a time grid which provides a balance between good approximations for both the hazard and survival functions. This issue has been one of the greatest challenges of working with the PEM. Although there exist a vast literature related to the PEM, the time grid $\tau = \{s_0, s_1, ..., s_J\}$ has been arbitrarily chosen in most of those works. Kalbfleisch and Prentice (1973) suggest that the selection of the time grid $\tau = \{s_0, s_1, ..., s_J\}$ should be made independent of the data, but they do not provide any procedure to do such. Breslow (1974) proposes defining the endpoints s_j of the intervals $I_j = (s_{j-1}, s_j]$ as the observed failure times. We shall refer the PEM constructed based on such time grid to as nonparametric PEM. Other heuristic discussions regarding adequate choices for the time grid of the PEM can be found in Gamerman (1991), Sahu et al. (1997) and Qiou et al. (1999), to cite few.

In practice, the problem of specifying a suitable time grid to fit the PEM can be overcome by assuming that $\tau = \{s_0, s_1, ..., s_J\}$ is itself a random quantity to be estimated using the data information. The first effective effort in this direction is due to Arjas and Gasbarra (1994). In that work it is assumed that the endpoints of the intervals $I_j = (s_{j-1}, s_j]$ are defined according to a jump process following a martingale structure which is included into the model through the prior distributions. Similar approaches to modeling the time grid of the PEM are considered by McKeague and Tighiouart (2000) and Kim et al. (2006). Independently from those works, Demarqui et al. (2008) also propose an approach which considers a random grid for the PEM. Based on the usual assumptions for the time grid and

assuming independent gamma prior distributions for the failure rates, they prove that the prior distribution for the time grid has a product form, and use the structure of the Product Partition Model (PPM) proposed by Barry and Hartigan (1992) to handle the problem. By considering such approach, the use of the reversible jump algorithm to sample from the posteriors is avoided althought the dimension of the parametric space is not fixed.

In this paper we extend the approach proposed by Demarqui et al. (2008) by deriving a noninformative joint prior distribution for ($\boldsymbol{\lambda}, \tau$). Specifically, we assume a discrete uniform prior distribution for the random time grids of the PEM and then, conditionally on those random time grids, we build the joint Jeffreys's prior for the failure rates. Conditions regarding the properties of the joint posterior distribution of (λ, τ) are discussed. Finally, we illustrate the usefulness of the proposed methodology by analyzing the survival time of patients diagnosed with brain cancer in Windham-CT, USA, obtained from the SEER (Surveillance, Epidemiology and End Results) database. For comparison purposes, the results are compared with those provided by other methods existing in the literature.

This paper is organized as follows: the proposed model is introduced in Section 2. The new methodology is illustrated with the analysis of a real data set in Section 3. Finally, in Section 4 some conclusion about the proposed model are draw.

2 Model construction

In this section we introduce a piecewise exponential model(PEM) which time grid is a random variable. We start our model presentation reviewing the piecewise exponential distribution.

2.1 Piecewise exponential distribution and the likelihood

Let T be a non-negative random variable. Assume, for instance, that T denotes the time to the event of interest. In order to obtain the probability density function of the PEM we need first to specify a time grid $\tau = \{s_0, s_1, ..., s_J\}$, such that $0 = s_0 < s_1 < s_2 < ... < s_J = \infty$, which induces a partition of the time axis into J intervals I_1 , ..., I_J , where $I_j = (s_{j-1}, s_j]$, for $j = 1, ..., J$. Then, we assume a constant rate for each interval induced by τ, that is:

$$h(t) = \begin{cases} \lambda_1, \text{if } t \in I_1; \\ \lambda_2, \text{if } t \in I_2; \\ \vdots \\ \lambda_J, \text{if } t \in I_J. \end{cases} \tag{1}$$

To conveniently define the cumulative hazard function and the survival and density func-

tions as well, we define:

$$t_j = \begin{cases} s_{j-1}, & \text{if } t < s_{j-1}, \\ t, & \text{if } t \in I_j, \\ s_j, & \text{if } t > s_j, \end{cases} \tag{2}$$

where $I_j = (s_{j-1}, s_j]$, $j = 1, ..., J$.

The cumulative hazard function of the PEM is computed from (1) and (2), yielding:

$$H(t|\ \boldsymbol{\lambda}, \tau) = \sum_{j=1}^{J} \lambda_j (t_j - s_{j-1}). \tag{3}$$

Consequently, it follows from the identity $S(t) = \exp\{-H(t)\}$ that the survival function of the PEM is given by:

$$S(t|\ \boldsymbol{\lambda}, \tau) = \exp\left\{ -\sum_{j=1}^{J} \lambda_j (t_j - s_{j-1}) \right\}. \tag{4}$$

The density function of T is obtained by taking minus the derivative of (4). Thus, we say that the random variable T follows a piecewise exponential model with time grid τ and vector parameter $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_J)'$, denoted by $T \sim PED(\tau,\ \boldsymbol{\lambda})$, if its probability density function is given by:

$$f(t|\ \boldsymbol{\lambda}, \tau) = \lambda_j \exp\left\{ -\sum_{j=1}^{J} \lambda_j (t_j - s_{j-1}) \right\}, \tag{5}$$

for $t \in I_j = (s_{j-1}, s_j]$ and $\lambda_j > 0$, $j = 1, ..., J$.

Let us assume that n individuals were observed independently. Let X_i be the survival time under study for the i-th element, $i = 1., ..., n$. Also, assume that there is a right-censoring scheme working independently of the failure process. Denote by C_i the censored time for the the i-th element, and assume that $C_i \sim G$, for some continuous distribution G defined on the semi-positive real line. Then, the complete information associated to the process is (T_i, δ_i), where $T_i = min\{X_i, C_i\}$ and $\delta_i = I(X_i \leq C_i)$ are, respectively, the observable survival time, and the failure indicator function for the i-th element. Suppose that $(T_i|\tau,\ \boldsymbol{\lambda}) \sim PEM(\tau,\ \boldsymbol{\lambda})$, with τ and $\boldsymbol{\lambda}$ as defined before.

In order to properly construct the likelihood function over the J intervals induced by $\tau = \{s_0, s_1, ..., s_J\}$, assume t_{ij} as defined in (2). Further, define $\delta_{ij} = \delta_i \nu_j^{(i)}$, where $\nu_j^{(i)}$ is the indicator function assuming value 1 if the survival time of the i-th element falls in the j-th interval, and 0 otherwise. It follows that the contribution of the survival time $t_i \in I_j = (s_{j-1}, s_j]$ for the likelihood function of the PEM is $\lambda_j^{\delta_{ij}} \exp\{-\sum_{j=1}^{J} \lambda_j (t_{ij} - s_{j-1})\}$. Then, given a time grid $\tau = \{s_0, s_1, ..., s_J\}$, we have that the complete likelihood function is

given by:

$$
\begin{aligned}
L(D|\lambda,\tau) &= \prod_{i=1}^{n}\left(\prod_{j=1}^{J}\lambda_j^{\delta_{ij}}\exp\left\{-\lambda_j(t_{ij}-s_{j-1})\right\}\right) \\
&= \prod_{j=1}^{J}\lambda_j^{\nu_j}\exp\left\{-\lambda_j\xi_j\right\}, \qquad (6)
\end{aligned}
$$

where the number of failures, $\nu_j = \sum_{l=1}^{n}\delta_{ij}$, and the total time under test observed at each interval I_j, $\xi_j = \sum_{i=1}^{n}(t_{ij}-s_{j-1})$, are sufficient statistics for λ_j, $j = 1, ..., J$.

It is noticeable that, given τ, the likelihood function given in (6) naturally factors into a product of kernels of gamma distributions. As we shall see in the following, along with mild conditions on the joint distribution of the time grid and failure rates, it allow us to use the structure of the PPM proposed by Barry and Hartigan (1992) to model the randomness of the time grid of the PEM.

2.2 Priors and the clustering structure

Following Demarqui et al. (2008), we start our model formulation by imposing some constrains on the set of possible time grids for the PEM. Specifically, we assume the time grid associated to the nonparametric approach as the finest possible time grid for the PEM. We further assume that only time grids whose endpoints are equal to distinct observed failure times are possible. These assumptions guarantee that at least one failure time falls at each interval induced by the random time grid of the PEM.

The randomness of the time grid of the PEM is modeled through the clustering structure of the PPM as follows. Let $\mathscr{F} = \{0, y_1, ..., y_m\}$ be the set formed by the origin and the m distinct ordered observed failure times from a sample of size n. Then, $\mathscr{F}$ defines a partition of time into disjoint intervals I_j, $j = 1, \ldots, m$, as defined previously. Further, denote by $\mathcal{I} = \{1, ..., m\}$ the set of indexes related to such intervals. Let $\rho = \{i_0, i_1, ..., i_b\}$, $0 = i_0 < i_1 < ... < i_b = m$, be a random partition of $\mathcal{I}$, which divides the m initial intervals into $B = b$ new disjoint intervals. The random variable B denotes the number of clustered intervals related to the random partition ρ. Finally, let $\tau = \tau(\rho) = \{s_0, s_1, ..., s_b\}$ be the time grid induced by the random partition ρ, where

$$
s_j = \begin{cases} 0, & \text{if } j = 0, \\ y_{i_j}, & \text{if } j = 1, \ldots, b, \end{cases} \qquad (7)
$$

for $b = 1, ..., m$. Then, it follows that the clustered intervals induced by $\rho = \{i_0, i_1, ..., i_b\}$ are given by:

$$
I_\rho^{(j)} = \cup_{r=i_{j-1}+1}^{i_j} I_r = (s_{j-1}, s_j], \quad j = 1, ..., b. \qquad (8)
$$

Conditionally on $\rho = \{i_0, i_1, ..., i_b\}$, we assume that:

$$h(t) = \lambda_r \equiv \lambda_\rho^{(j)}, \tag{9}$$

where $\lambda_\rho^{(j)}$ denotes the common failure rate related to the clustered interval $I_\rho^{(j)}$, for $i_{j-1} < r \leq i_j$, $r = 1, ..., m$ and $j = 1, ..., b$.

In order to complete the model specification, we need to specify the joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$. This is done hierarchically by first specifying a prior distribution for the random partition ρ, and then eliciting prior distributions for $\boldsymbol{\lambda}_\rho$, conditioning on ρ.

Under the assumption that there is no prior information available regarding the time grid, we elicit the Bayes-Laplace prior for $\rho = \{i_0, i_1, ..., i_b\}$, that is,

$$\pi(\rho = \{i_0, i_1, ..., i_b\}) = \frac{1}{2^{m-1}}. \tag{10}$$

This prior distribution puts equal mass onto the 2^{m-1} possible partitions associated with the time grids formed by time-points belonging to $\mathscr{F}$, reflecting our lack of information about the time grid. Observe that, if we set $P(\rho = \{i_0, i_1, ..., i_b\}) = 1$ for a particular partition, we return to the usual model that assumes a fixed time grid for the PEM.

Remember that we are defining a random time grid for the PEM in terms of a random partition of the intervals I_j. Furthermore, we are considering that only contiguous intervals are possible, and that the endpoint i_j of each clustered interval $I_\rho^{(j)}$ depends only upon the previous endpoint i_{j-1}. Thus, it follows that the prior distribution (10) can be written as the product prior distribution proposed by Barry and Hartigan (1992):

$$\pi(\rho = \{i_0, i_1, ..., i_b\}) = \frac{1}{K} \prod_{i=1}^{b} c_{I_\rho^{(j)}}, \tag{11}$$

with prior cohesions $c_{I_\rho^{(j)}} = 1$, $\forall\ (i_{j-1}, i_j) \in \mathcal{I}$ and $K = \sum_{\mathscr{C}} \prod_{j=1}^{b} c_{I_\rho^{(j)}} = 2^{m-1}$, where $\mathscr{C}$ denotes the set of all possible partitions of the set $\mathcal{I}$ into b disjoint clustered intervals with endpoints $i_1, ..., i_b$, satisfying the condition $0 = i_0 < i_1 < ... < i_b = m$, for all $b \in \mathcal{I}$.

Conditionally on ρ, we assume the Jeffreys's prior distribution as a joint noninformative prior distribution for $\boldsymbol{\lambda}_\rho$. Let $I(\boldsymbol{\lambda}_\rho)$ denote the Fisher information matrix for $\boldsymbol{\lambda}_\rho$. Then, the joint prior distribution for $\boldsymbol{\lambda}_\rho$ is defined as:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho|\rho) &\propto |I(\boldsymbol{\lambda}_\rho)|^{\frac{1}{2}} \\ &\propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{-1}. \end{aligned} \tag{12}$$

One attractive characteristic of the Jeffreys's prior is that, regardless of the nature of vector of parameters of model under consideration, this noninformative prior distribution

is invariant under one-to-one transformations of those parameter, i.e., the Jeffreys's prior is invariant to parameterizations. In particular, the product form of (12) also induces independence among the failure rates into different intervals.

It follows from (11) and (12) that the (improper) joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ is given by:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho, \rho) &\propto \pi(\boldsymbol{\lambda}_\rho|\rho)\pi(\rho) \\ &\propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{-1}. \end{aligned} \tag{13}$$

Hence, conditionally on $\rho = \{i_0, i_1, ..., i_b\}$, from the product form of (6) and (13) we have that the joint distribution of the observations has also a product form, given by:

$$\begin{aligned} f(\mathbf{D}|\rho) &= \prod_{j=1}^{b} \int \left(\lambda_\rho^{(j)}\right)^{\eta_j - 1} \exp\left\{-\xi_j \lambda_\rho^{(j)}\right\} d\lambda_\rho^{(j)} \\ &= \prod_{j=1}^{b} \frac{\Gamma(\eta_j)}{\xi_j^{\eta_j}}, \end{aligned} \tag{14}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Thus, the joint distribution of the observations given in (14) satisfies the product condition required for applying the clustering structure of the PPM to model the randomness of the time grid of the PEM. Bayes inference under noninformative priors for the baseline hazard distribution was also conisered by Sinha et al. (2003), but from different modeling perspective.

2.3 Posterior distributions and related inference

The joint posterior distribution of $(\boldsymbol{\lambda}_\rho, \rho)$ is given by:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D}) &= L(\mathbf{D}|\, \boldsymbol{\lambda}_\rho, \rho)\pi(\boldsymbol{\lambda}_\rho|\rho)\pi(\rho) \\ &\propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{\nu_j - 1} \exp\left\{-\lambda_\rho^{(j)} \xi_j\right\}. \end{aligned} \tag{15}$$

It is easy to see that (15) is always a proper joint posterior distribution. This is an immediate result of the model formulation we are proposing. Notice that (15) corresponds to a product of kernels of gamma distributions, since that $\nu_j > 0$ and $\xi_j > 0$, $\forall\, j$, can always be verified, regardless of the random time grid of the PEM.

The posterior distribution of $\rho = \{i_0, i_1, ..., i_b\}$ is obtained after integrating out (15) with

respect to $\boldsymbol{\lambda}_\rho$, that is:

$$\begin{aligned}\pi(\rho|\mathbf{D}) &= \int_{\boldsymbol{\lambda}_\rho} L(\mathbf{D}|\ \boldsymbol{\lambda}_\rho, \rho)\pi(\ \boldsymbol{\lambda}_\rho|\rho)\pi(\rho)d\ \boldsymbol{\lambda}_\rho \\ &= \frac{1}{K^*}\prod_{i=1}^{b} c^*_{I_\rho^{(j)}}, \end{aligned} \tag{16}$$

where $c^*_{I_\rho^{(j)}} = (\Gamma(\eta_j)/\xi_j^{\eta_j})$ denotes the posterior cohesion associated with the j-th clustered interval $I_\rho^{(j)}$, and $K^* = \sum_{\mathscr{C}} \prod_{j=1}^{b} c^*_{I_\rho^{(j)}}$.

Following the structure of the PPM, we have that the posterior distribution for λ_k, $k = 1, ..., m$, is given by the mixture of distributions:

$$\pi(\lambda_k|\mathbf{D}) = \sum_{i_{j-1}<k\leq i_j} \pi(\lambda_\rho^{(j)}|\mathbf{D})R(I_\rho^{(j)}|\mathbf{D}), \tag{17}$$

where $R(I_\rho^{(j)}|\mathbf{D})$ is named as posterior relevance, and which denotes the probability of each clustered interval $I_\rho^{(j)}$ to belong to the random partition ρ, and $\pi(\lambda_\rho^{(j)}|\mathbf{D})$ denotes the posterior distribution of the common parameter $\lambda_\rho^{(j)}$, $j = 1, ..., b$.

Assuming the squared-error loss function, we have that the product estimate for the failure rate (PEFR) λ_k is given by:

$$\hat{\lambda}_k = \sum_{i_{j-1}<k\leq i_j} E(\lambda_\rho^{(j)}|\mathbf{D})R(I_\rho^{(j)}|\mathbf{D}), \tag{18}$$

for $j = 1, ..., b$ and $k = 1, ..., m$.

Finally, the posterior survival function for a new element, assumed to be independent of the observed data set, is obtained by averaging the conditional survival function $S(y|\mathbf{D}, \rho)$ over all random partitions $\rho = \{i_0, i_1, ..., i_b\}$, that is:

$$S(y|\mathbf{D}) = \sum_{\rho} S(y|\mathbf{D}, \rho)\pi(\rho|\mathbf{D}), \tag{19}$$

where

$$\begin{aligned} S(y|\mathbf{D}, \rho) &= \int S(y|\lambda_\rho^{(j)}, \rho)\pi(\lambda_\rho^{(j)}|\mathbf{D})d\lambda_\rho^{(j)} \\ &= \left(1 + \frac{y - s_{j-1}}{\gamma_j}\right)^{-\alpha_j} \prod_{r=1}^{j-1}\left(1 + \frac{s_r - s_{r-1}}{\gamma_r}\right)^{-\alpha_r}, \end{aligned} \tag{20}$$

for $\lambda_\rho^{(j)} = (\lambda_\rho^{(1)}, ..., \lambda_\rho^{(b)})'$ and $y \in I_\rho^{(j)}$, $j = 1, ..., b$. Throughout this paper we shall refer to (20) as product estimate for the survival function (PESF).

3 Numerical Illustration

In this section we use the proposed model to analyze the survival times (in months) of 231 individuals diagnosed with brain cancer in Windham-CT, USA, during 1995 to 2004. This data set was obtained from SEER database. Our interest relies on investigate the performance of our model in estimating both the hazard and survival function. The computational procedures needed to fit the proposed model can be found in Demarqui et al. (2008).

From the 231 patients diagnosed with brain cancer, we observed 134 failures and 97 censored times, resulting in a percentage of failures of the order of 58%. It is also noteworthy that, as the survival times were measured in months, only 32 of the 134 observed failure times correspond to distinct failure times. Thus, under the setup we are proposing, these 32 distinct failure times compose the finest possible time grid for the PEM.

We first examined the performance of the proposed model in estimating the hazard function by comparing the PEFR with the estimates provided by the competing piecewise exponential estimators (PEXE) for the failure rates (Kim and Proschan (1991)), namely, the estimates associated with the nonparametric PEM, obtained via maximum likelihood approach.

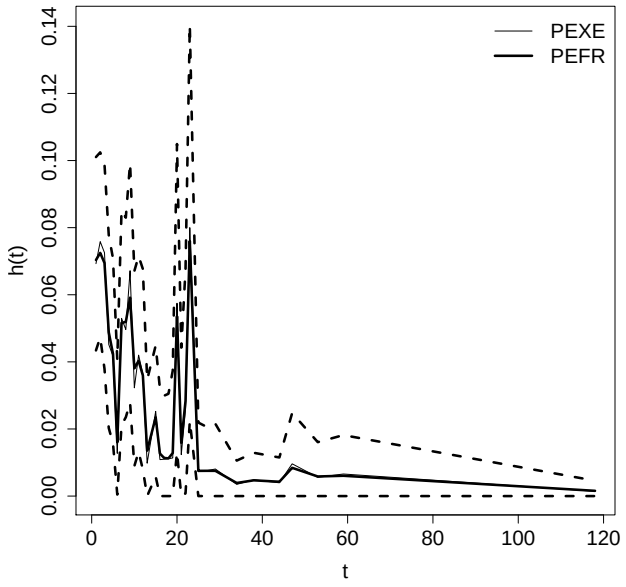

Figure 1: Estimated hazard function (solid lines) for patients with brain cancer and the 95% HPD interval (dashed lines) provided by the PEM with random time grid.

When there are no ties among the observed survival times, the PEXE for the failure rates are not consistent, once we have only one failure time at each interval, regardless of the sample size n. On the other hand, in the presence of ties, asymptotic results could hold only under the assumption that the number of ties in each interval increases without bond, which is not a realistic assumption in practice. Therefore, asymptotic confidence

intervals for the failure rates should not be computed in these cases. Furthermore, in the finite sample size scenario, little is known theorically about the PEXE (Kim and Proschan (1991)). These drawbacks of the maximum likelihood approach are easily overcome in the setting we are proposing, since HPD intervals do not rely on the sample size n, and can be obtained straightforwardly.

In Figure 1 we present the PEFR and the PEXE for the failure rates along with the 95% HPD intervals provided by the proposed model. Notice that PEFR and the PEXE for the failure rates are quite similar, and yield essentially the same estimated hazard function for the current data set. From the clinical point of view, the estimated failure rates displayed in Figure 1 indicate a decreasing hazard function, suggesting that the risk of death by brain cancer decreases through time.

The well known Kaplan-Meyer product limiting estimator (KME) arises as a standard estimator for the survival function. In practice, the PEXE yields a smoothing version of the KME for the survival function. Moreover, as shown in Kitchin et al. (1980), the KME and PEXE for the survival function are asymptotically equivalent. Thus, for the sake of simplicity, we compare the PESF with the KME.

Figure 2 displays the estimated survival functions provided by the PESF and the KME, along with their corresponding 95% confidence and HPD intervals. The similar performance of the two competing estimators in both point and interval estimation for the survival function is also evident.

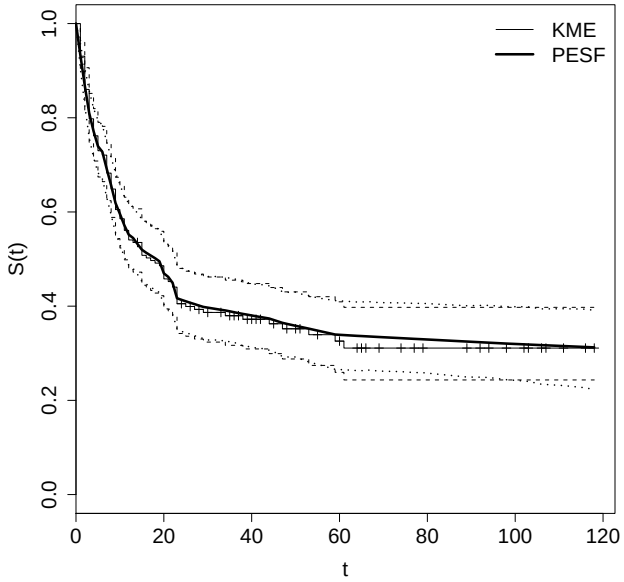

Figure 2: Estimated survival function (solid lines) for patients with brain cancer and their corresponding 95% confidence and HPD intervals (dashed lines) provided by the PEM with random time grid.

Another advantage of the proposed model is that it allows us to enrich the analysis by

making inferences about the time grid and the number of intervals used to fit the PEM. For instance, the posterior most probable number of intervals is $B = 25$, with probability 0.174. The estimated 95% HPD intervals for B is $[21, 29]$. Other characteristics of the posterior sample of the number of intervals of the PEM are given in Table 1 and Figure 3.

Table 1: Summary of sample of the posterior distribution of the number of intervals.

	mean	sd	min	max
	24.779	2.292	15	32
$P_{2.50}$	$P_{25.0}$	$P_{50.0}$	$P_{75.0}$	$P_{97.50}$
20	23	25	26	29

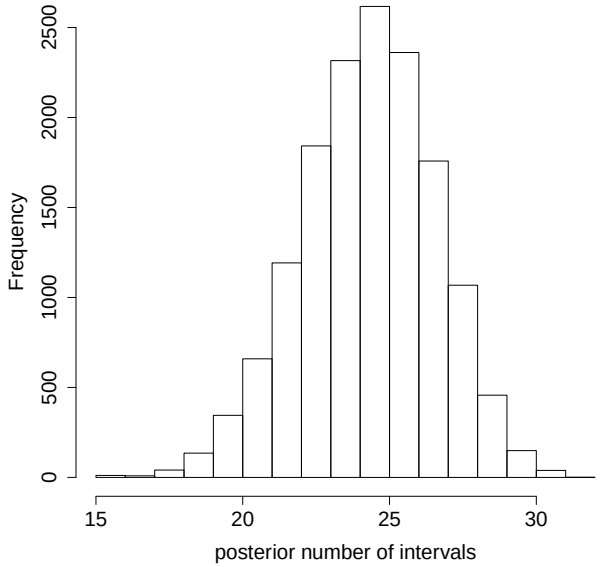

Figure 3: Posterior distribution of the number of intervals.

4 Conclusions

In this paper we presented a fully Bayesian approach to model time to event data using the piecewise exponential model with random time grid. Extending the previous work due to Demarqui et al. (2008), we elicited noninformative priors for both the time grid and the failure rates. Fixed the time grid, the Jeffreys' prior for the failure rates is a product distribution favoring the use of the structure of the PPM. It also induces independence among the failure rates in different intervals. Finally, we conducted the analysis of a real data set to illustrate the performance of the proposed model.

The results obtained from the analysis of the brain cancer data set suggest that the estimates provided by the proposed model are comparable with the those yielded by other estimators established in the literature such as the PEXE and the KME. However, interval estimation is straightforward under the framework we are proposing, and it does not rely on asymptotic approximations such as the PEXE and the KME do. Other advantage of the proposed model is that it enriches the analysis by enabling inferences for the time grid of the PEM. Furthermore, the assumption of a joint noninformative prior distribution for the failure rates and the time grids is quite attractive in situations where there is no prior information available.

Acknowledgments

The authors would like to express their gratitude to the editor and the referees for the careful refereeing of the paper. Fabio N. Demarqui's research was sponsored by CAPES(*Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior*) of the Ministry for Science and Technology of Brazil. Dipak K. Dey ... Rosangela H. Loschi's research has been partially funded by CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 306085/2009-7, 304505/2006-4. Enrico A. Colosimo's research has been partially funded by the CNPq, grants 150472/2008-0 and 306652/2008-0.

References

Aitkin, M., Laird, N., and Francis, B. (1983). A reanalysis of the stanford heart transplant data (with discussion). *J Am Stat Assoc* **78,** 264–292.

Arjas, E. and Gasbarra, D. (1994). Nonparametric bayesian inference from right censored survival data. *Stat Sinica* **4,** 505–524.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems. *Ann Statist* **20,** 260–279.

Bastos, L. S. and Gamerman, D. (2006). Dynamic survival models with spatial frailty. *Lifetime Data Anal* **12,** 441–460.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30,** 89–99.

Clark, D. E. and Ryan, L. M. (2002). Concurrent prediction of hospital mortality and length of stay from risk factors on admission. *Health Services Res* **37,** 631–645.

Demarqui, F. N., Loschi, R. H., and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Anal* **14,** 333–356.

Gamerman, D. (1991). Dynamic bayesian models for survival data. *Appl Stat* **40,** 63–79.

Gamerman, D. (1994). Bayes estimation of the piece-wise exponential distribution. *IEEE Trans Reliab* **43,** 128–131.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian survival analysis.* Springer-Verlag, New York.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on cox's regression and life models. *Biometrika* **60,** 267–278.

Kim, J. S. and Proschan, F. (1991). Piecewise exponential estimator of the survival function. *IEEE Trans Reliab* **40,** 134–139.

Kim, S., Chen, M. H., Dey, D. K., and Gamerman, D. (2006). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Anal* **13,** 17–35.

Kitchin, J., Langberg, N. A., and Proschan, F. (1980). A new method for estimating life distributions from incomplete data. Technical report, Department of Statistics, Florida State University.

McKeague, I. W. and Tighiouart, M. (2000). Bayesian estimators for conditional hazard functions. *Biometrics* **56,** 1007–1015.

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

Sahu, S. K., Dey, D. K., Aslanidu, H., and Sinha, D. (1997). A weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Anal* **3,** 123–137.

Sinha,D., Chen, M.H. and Ghosh,S.K.(1999). Bayesian Analysis and Model Selection for Interval-Censored Survival Data. *Biometrics* **55,** 585–590.

Sinha,D., Ibrahim, J.G., and Chen, M.H. (2003). A Bayesian Justification of Cox's Partial Likelihood. *Biometrika* **90,** 629–641.

Surveillance and Epidemiology and End Results (SEER). *Surveillance Research Program, National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 6.5.1. Database: Incidence - SEER 17 Regs Limited-Use Nov 2006 Sub (1973-2004 varying).*

Flexible Piecewise Exponential Model in Survival Analysis and Beyond

Fabio N. Demarqui[1,2], Rosangela H. Loschi[1],
Enrico A. Colosimo[1], Dipak K. Dey[2]
[1] Universidade Federal de Minas Gerais, Brazil
[2]University of Connecticut, USA

Abstract

In this paper we present semiparametric Bayesian approaches for modeling survival data using the piecewise exponential model (PEM). We assume that the time grid needed to fit the PEM is a random quantity and propose a flexible class of prior distributions for modeling jointly the time grid and its corresponding failure rates. The mechanism used to model the randomness of the time grid of the PEM has several advantages over other approaches that have been proposed in the literature. The resultant approach includes other models established in the literature as special cases and provides a flexible framework for survival data modeling. Properties of the model are discussed in the paper. The use of the proposed methodology is exemplified through the analysis of the survival times of patients diagnosed with brain cancer in Windham-CT, USA, obtained from the SEER (Surveillance, Epidemiology and End Results) database.
Keywords: Bayesian inference, Gibbs sampler, product partition model, survival analysis.

1 Introduction

The Piecewise Exponential Model (PEM) arises as a quite attractive alternative to parametric models for the analysis of time to event data. Although parametric in a strict sense, the PEM can be thought as a nonparametric model as far as it does not have a closed form for the hazard function. This nice characteristic of the PEM allows us to use this model to approximate satisfactorily hazard functions of several shapes. For this reason, the PEM has been widely used to model time to event data in different contexts, such as clinical situations including kidney infection, Sahu et al. (1997), heart transplant data, Aitkin et al. (1983), hospital mortality data, Clark and Ryan (2002), and cancer studies including leukemia, Breslow (1974), gastric cancer, Gamerman (1991), breast cancer, Ibrahim et al. (2001b) (see

also Sinha et al. (1999) for an application to interval-censored data), melanoma, Kim et al. (2006), and nasopharynx cancer, McKeague and Tighiouart (2000), among others. The PEM has also been used in reliability engineering (Kim and Proschan (1991), Gamerman (1994), and Barbosa et al. (1996)), and economics problems, Gamerman (1991), and Bastos and Gamerman (2006).

Let T be a non-negative random variable representing a survival time of interest. In order to construct the PEM, we first need to specify a time grid $\tau = \{s_0, s_1, ..., s_b\}$, such that $0 = s_0 < s_1 < s_2 < ... < s_b < \infty$, which induces a partition of the time axis into b intervals $I_j = (s_{j-1}, s_j]$, for $j = 1, ..., b$. Then, we assume a constant failure rate for each interval induced by τ, that is:

$$h(t) = \lambda_j, \ t \in I_j, \quad j = 1, \cdots, b.$$

In order to express the cumulative hazard, the survival and the density functions, we define:

$$t_j = \begin{cases} s_{j-1}, & \text{if } t < s_{j-1}, \\ t, & \text{if } t \in I_j, \\ s_j, & \text{if } t > s_j, \end{cases} \tag{1}$$

for $j = 1, ..., b$.

The cumulative hazard function of the PEM is computed from (1), yielding:

$$H(t|\boldsymbol{\lambda}, \tau) = \sum_{j=1}^{b} \lambda_j (t_j - s_{j-1}). \tag{2}$$

It follows from the well-knwon identity $S(t) = \exp\{-H(t)\}$ that the survival function of the PEM is given by:

$$S(t|\boldsymbol{\lambda}, \tau) = \exp\left\{ -\sum_{j=1}^{b} \lambda_j (t_j - s_{j-1}) \right\}. \tag{3}$$

The density function associated with T is obtained by taking minus the derivative of (3). Thus, we say that the random variable T follows a PEM with time grid τ and vector of parameter $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_b)'$, denoted by $T \sim PED(\tau, \boldsymbol{\lambda})$, if its probability density function is given by:

$$f(t|\boldsymbol{\lambda}, \tau) = \lambda_j \exp\left\{ -\sum_{j=1}^{b} \lambda_j (t_j - s_{j-1}) \right\}, \tag{4}$$

for $t \in I_j$ and $\lambda_j > 0$, $j = 1, ..., b$.

The time grid $\tau = \{s_0, s_1, ..., s_J\}$ plays a central role in the goodness of fit of the PEM. A time grid with a large number of intervals might provide unstable estimates for the failure

rates, whereas a time grid based on just a few intervals might produce a poor approximation for the true survival function. In practice, we desire a time grid which provides a balance between good approximations for both the hazard and survival functions. This issue has been one of the greatest challenges of working with the PEM. Although the PEM has been widely used in the literature, the time grid $\tau = \{s_0, s_1, ..., s_b\}$ has been arbitrarily chosen in most of those works. Kalbfleisch and Prentice (1973) suggested that the selection of the time grid $\tau = \{s_0, s_1, ..., s_b\}$ should be made independent of the data, but they did not provide any procedure to do such. Breslow (1974) proposed defining the endpoints s_j of the intervals I_j as all the observed failure times. We shall refer the PEM constructed based on such time grid to as nonparametric PEM. Other heuristic discussions regarding adequate choices for the time grid of the PEM can be found in Gamerman (1991), Sahu et al. (1997) and Qiou et al. (1999), to cite few.

In practice, the problem of specifying a suitable time grid to fit the PEM can be overcome by assuming that $\tau = \{s_0, s_1, ..., s_b\}$ is itself a random quantity to be estimated using the data information. The first effective effort in this direction is due to Arjas and Gasbarra (1994). In that work it is assumed that the endpoints of the intervals I_j are defined according to a jump process following a martingale structure which is included into the model through the prior distributions. Similar approaches for modeling the time grid of the PEM were considered by McKeague and Tighiouart (2000) in the context of regression models and Kim et al. (2006) to fit cure rate models. Independently from those works, Demarqui et al. (2008) also proposed an approach which considers a random grid for the PEM. Based on the usual assumptions for the time grid and assuming independent Gamma prior distributions for the failure rates, they proved that the prior distribution for the time grid has a product form, and have used the structure of the Product Partition Model (PPM) proposed by Barry and Hartigan (1992) to handle the problem. By considering such approach, the the reversible jump algorithm can be used to sample from the posterior distributions of interest.

In this paper we introduce a general framework for modeling survival data using the PEM via PPM. The proposed methodology offers more flexibility in fitting the PEM by allowing a broad class of prior distributions for $(\boldsymbol{\lambda}, \tau)$. The resultant model includes as special cases the models proposed by Demarqui et al. (2008), Gamerman (1994) and Ibrahim et al. (2001a), and any other model whose prior process satisfies the (conditional) independence condition for the failure rates and yields a closed form expression for the marginal distribution of the data. Finally, we illustrate the usefulness of the proposed methodology by analyzing the survival times of patients diagnosed with brain cancer in Windham-CT, USA, obtained from the SEER (Surveillance, Epidemiology and End Results) database. The results obtained by using our approach are compared with those provided by other methods existing in the literature.

This paper is organized as follows: the proposed model is introduced in Section 2. The new methodology is illustrated with the analysis of a real data set in Section 3. Finally, in

Section 4 some conclusions are drawn about the proposed model.

2 A general class for the PEM via PPM

In the following sequence, we formally introduce the general framework for fitting the PEM via PPM. The notation, as well as the conditions under which our new methodology relies on, are addressed in Section 2.1. Particular cases arising from objective and subjective prior specifications for the failure rates of the PEM are properly discussed in Sections 2.2 and 2.3, respectively.

2.1 Model formulation

We start our model formulation by assuming that only time grids whose endpoints are equal to distinct observed failure times are possible. This mild constrain on the set of possible time grids can be justified by the plausible argument that the observed failure times are themselves good candidates for the endpoints of the intervals needed to fit the PEM. In addition, such constrain also guarantees the existence of at least one failure time at each interval induced by the random time grid of the PEM.

Formally speaking, let $\mathscr{F} = \{0, y_1, ..., y_m\}$ be the set formed by the origin and the m distinct ordered observed failure times from a sample of size n. Consider the time grid $\tau' = \{0, y'_1, ..., y'_{m'}\}$ satisfying $\tau' \subseteq \mathscr{F}$, where m', $1 \leq m' \leq m$, denotes the maximum number of intervals admitted a *priori*. Then, the time grid $\tau' = \{0, y'_1, ..., y'_{m'}\}$ induces the following set of disjoint intervals:

$$I_j = \begin{cases} (0, y'_1], & \text{if } j = 1, \\ (y'_{j-1}, y'_j], & 1 < j \leq m'. \end{cases} \tag{5}$$

Further, denote by $\mathcal{I} = \{1, ..., m'\}$ the set of indexes related to the intervals defined in (5). Let $\rho = \{i_0, i_1, ..., i_b\}$, $0 = i_0 < i_1 < ... < i_b = m'$, be a random partition of $\mathcal{I}$ dividing the m' initial intervals given in (5) into $B = b$ new disjoint intervals, where the random variable B denotes the number of clustered intervals related to the random partition ρ. Finally, let $\tau = \tau(\rho) = \{s_0, s_1, ..., s_b\}$ be the time grid induced by the random partition ρ, where

$$s_j = \begin{cases} 0, & \text{if } j = 0, \\ y'_{i_j}, & \text{if } j = 1, \ldots, b, \end{cases} \tag{6}$$

for $b = 1, ..., m'$. Hence, it follows from (5) and (6) that the clustered intervals induced by $\rho = \{i_0, i_1, ..., i_b\}$ are given by:

$$I_\rho^{(j)} = \cup_{r=i_{j-1}+1}^{i_j} I_r = (s_{j-1}, s_j], \quad j = 1, ..., b. \tag{7}$$

Then, given $\rho = \{i_0, i_1, ..., i_b\}$, we assume that:

$$h(t) = \lambda_r \equiv \lambda_\rho^{(j)}, \tag{8}$$

where $\lambda_\rho^{(j)}$ denotes the common failure rate associated with the clustered interval $I_\rho^{(j)}$, $i_{j-1} < r \leq i_j$, $r = 1, ..., m'$ and $j = 1, ..., b$.

In order to properly construct the likelihood function over the b intervals induced by $\tau(\rho) = \{s_0, s_1, ..., s_b\}$, we redefine (1) in order to accommodate accordingly the information of the n elements belonging to the observed sample as:

$$t_{ij} = \begin{cases} s_{j-1}, & \text{if } t_i < s_{j-1}, \\ t_i, & \text{if } t_i \in I_\rho^{(j)}, \\ s_j, & \text{if } t_i > s_j, \end{cases} \tag{9}$$

for $j = 1, ..., b$ and $i = 1, ..., n$. Further, we also define $\delta_{ij} = \delta_i \nu_j^{(i)}$, where $\nu_j^{(i)}$ is the indicator function assuming value 1 if the survival time of the i-th element falls in the j-th interval, and 0 otherwise.

Due to the one-to-one relationship between τ and ρ, it is possible to express the likelihood function in terms of the random partition ρ. Then, under the assumption that there is a right-censoring scheme working independently of the failure process it follows that the contribution of a survival time (either failure or censored) $t_i \in I_\rho^{(j)}$ for the likelihood function of the PEM reduces to $\prod_{j=1}^{b} (\lambda_\rho^{(j)})^{\delta_{ij}} \exp\{-\lambda_\rho^{(j)}(t_{ij} - s_{j-1})\}$. Thus, the entire likelihood function can be written as:

$$\begin{aligned} L(\boldsymbol{\lambda}_\rho, \rho | D) &= \prod_{i=1}^{n} \left(\prod_{j=1}^{b} (\lambda_\rho^{(j)})^{\delta_{ij}} \exp\left\{-\lambda_\rho^{(j)}(t_{ij} - s_{j-1})\right\} \right) \\ &= \prod_{j=1}^{b} (\lambda_\rho^{(j)})^{\eta_j} \exp\left\{-\lambda_\rho^{(j)} \xi_j\right\}, \end{aligned} \tag{10}$$

where the number of failures, $\eta_j = \sum_{l=1}^{n} \delta_{ij}$, and the total time under test observed at each interval $I_\rho^{(j)}$, $\xi_j = \sum_{i=1}^{n} (t_{ij} - s_{j-1})$, are sufficient statistics for $\lambda_\rho^{(j)}$, $j = 1, ..., b$.

The joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ is specified hierarchically by first eliciting a prior distribution for the random partition ρ, and then specifying a prior distribution for $\boldsymbol{\lambda}_\rho$, conditionally on ρ, that is,

$$\pi(\boldsymbol{\lambda}_\rho, \rho) = \pi(\boldsymbol{\lambda}_\rho | \rho) \pi(\rho). \tag{11}$$

Demarqui et al. (2008) presented arguments justifying that the prior distribution of ρ

can be written as the product form:

$$\pi(\rho = \{i_0, i_1, ..., i_b\}) \propto \prod_{i=1}^{b} c_{I_\rho^{(j)}}, \tag{12}$$

for $0 = i_0 < i_1 < ... < i_b = m'$ and $b \in \mathcal{I}$. Here, $c_{I_\rho^{(j)}}$, called prior cohesion, is a positive quantity representing the degree of similarity among the intervals being clustered. Moreover, accordingly normalized, the prior cohesions $c_{I_\rho^{(j)}}$ can be interpreted as the one-step transition probability of the Markov Chain defined by the endpoints of the clustered intervals $I_\rho^{(j)}$, $j = 1, ..., b$.

We further assume that, conditionally on ρ, the joint prior distribution for $\boldsymbol{\lambda}_\rho$ can be expressed by the following product form:

$$\pi(\boldsymbol{\lambda}_\rho|\rho) = \prod_{j=1}^{b} \pi(\lambda_\rho^j). \tag{13}$$

This assumption is equivalent to assume that, for a given ρ, the common failure rates are independent *a priori*. Hence, conditionally on ρ, it follows from (10) and (13) that the joint (marginal) distribution of the observations reduces to the product form:

$$f(\mathbf{D}|\rho) \;=\; \prod_{j=1}^{b} \int \left(\lambda_\rho^{(j)}\right)^{\eta_j} \exp\left\{-\xi_j \lambda_\rho^{(j)}\right\} \pi(\lambda_\rho^{(j)}) d\lambda_\rho^{(j)}. \tag{14}$$

According to Barry and Hartigan (1992), any joint distribution of the observations and partitions that satisfies the product condition for the partitions and the independence condition for observations, given the partition, follows a PPM. Consequently, the product condition follows for the joint distribution of the observations given in (14), provided the integrals are well defined (where we mean by well defined finite intervals), and the clustering structure of the PPM can be extended for modeling the randomness of the time grid of the PEM.

The joint posterior distribution of $(\boldsymbol{\lambda}_\rho, \rho)$ is:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D}) &\propto L(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D})\pi(\boldsymbol{\lambda}_\rho|\rho)\pi(\rho) \\ &\equiv \pi(\boldsymbol{\lambda}_\rho|\rho, \mathbf{D})\pi(\rho|\mathbf{D}). \end{aligned} \tag{15}$$

Then, it follows from (12) and (13) that the posterior distribution of the random partition ρ is given by:

$$\begin{aligned} \pi(\rho|\mathbf{D}) &\propto f(\mathbf{D}|\rho)\pi(\rho) \\ &\propto \prod_{i=1}^{b} c^*_{I_\rho^{(j)}}, \end{aligned} \tag{16}$$

where $c^*_{I_\rho^{(j)}} = \left(\int_{\lambda_\rho^{(j)}} (\lambda_\rho^{(j)})^{\nu_j} \exp\left\{-\lambda_\rho^{(j)}\xi_j\right\} \pi(\lambda_\rho^{(j)}) d\lambda_\rho^{(j)}\right) c_{I_\rho^{(j)}}$ denotes the posterior cohesion associated with j-th clustering interval $I_\rho^{(j)}$.

Following the structure of the PPM, we have that the posterior distribution of λ_k is given by the following mixture of distributions:

$$\pi(\lambda_k|\mathbf{D}) = \sum_{i_{j-1}<k\leq i_j} \pi(\lambda_\rho^{(j)}|\mathbf{D},\rho) R(I_\rho^{(j)}|\mathbf{D}), \tag{17}$$

for $k = 1, ..., m'$, where $R(I_\rho^{(j)}|\mathbf{D})$ is named as posterior relevance, and denotes the probability of each clustered interval $I_\rho^{(j)}$ to belong to the random partition ρ, and $\pi(\lambda_\rho^{(j)}|\mathbf{D},\rho)$ denotes the full conditional posterior distribution of the common parameter $\lambda_\rho^{(j)}$, $j = 1, ..., b$.

Then, assuming the squared-error loss function, it follows that the product estimate for the failure rate (PEFR) λ_k is given by:

$$\hat{\lambda}_k = \sum_{i_{j-1}<k\leq i_j} E(\lambda_\rho^{(j)}|\mathbf{D},\rho) R(I_\rho^{(j)}|\mathbf{D}), \tag{18}$$

for $j = 1, ..., b$ and $k = 1, ..., m'$.

Under this structure, the posterior survival function for a new element, assumed to be independent of the observed data set, is obtained by averaging the conditional survival function $S(y|\mathbf{D},\rho)$ over all random partitions $\rho = \{i_0, i_1, ..., i_b\}$, that is:

$$S(y|\mathbf{D}) = \sum_{\rho} S(y|\mathbf{D},\rho)\pi(\rho|\mathbf{D}), \tag{19}$$

where

$$S(y|\mathbf{D},\rho) = \int S(y|\boldsymbol{\lambda}_\rho,\rho)\pi(\boldsymbol{\lambda}_\rho|\mathbf{D},\rho)d\boldsymbol{\lambda}_\rho, \tag{20}$$

for $\boldsymbol{\lambda}_\rho = (\lambda_\rho^{(1)}, ..., \lambda_\rho^{(b)})'$. Throughout this paper we shall refer to (20) as product estimate for the survival function (PESF).

2.2 Objective prior elicitation

A full objective analysis can be performed by assuming the Bayes-Laplace prior distribution for ρ, and, the Jeffreys prior distribution for $\boldsymbol{\lambda}_\rho$, given ρ.

The Bayes-Laplace prior distribution for ρ corresponds to the special case of the prior distribution (12) in which $c_{I_\rho^{(j)}} = 1$, $\forall\ j$ such that $(i_{j-1}, i_j) \in \mathcal{I}$. Now, given ρ, let $I(\boldsymbol{\lambda}_\rho)$ denote the Fisher information matrix associated with $\boldsymbol{\lambda}_\rho$. The Jeffreys prior distribution for

$\boldsymbol{\lambda}_\rho$ is defined as:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho|\rho) &\propto |I(\boldsymbol{\lambda}_\rho)|^{\frac{1}{2}} \\ &\propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{-1}. \end{aligned} \tag{21}$$

An attractive characteristic of the Jeffreys prior is that it is invariant under parameterizations, a desired feature of objective prior distributions. Moreover, it also induces independence among the $\lambda_\rho^{(j)}$'s. Thus, the product condition required to use the structure of the PPM to model the PEM with random time grid is satisfyied.

It follows from (12) and (21) that the joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ is given by:

$$\pi(\boldsymbol{\lambda}_\rho, \rho) \propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{-1}, \tag{22}$$

which is an improper prior distribution.

Replacing (22) in (15), the joint posterior distribution of $(\boldsymbol{\lambda}_\rho, \rho)$ becomes:

$$\pi(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D}) \propto \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{\nu_j - 1} \exp\left\{-\xi_j \lambda_\rho^{(j)}\right\}, \tag{23}$$

which is a proper posterior distribution. This is a direct consequence of our model formulation, which ensures that both $\nu_j > 0$ and $\xi_j > 0$, $\forall\ j = 1, ..., b$, regardless of ρ. Moreover, after a simple algebra, the posterior distribution of ρ in (16) becomes:

$$\pi(\rho|\mathbf{D}) \propto \prod_{j=}^{b} \frac{\Gamma(\nu_j)}{\xi_j^{\eta_j}},$$

where $\Gamma(\cdot)$ denotes the gamma function.

2.3 Subjective prior elicitation

The gamma family of distributions has been widely used as prior distributions for the failure rates of the PEM. This family of distributions corresponds to the conjugate family of prior distributions for the exponential sample family, and facilitates the inference processs regarding the PEM. In addition, structures of dependence between successive intervals can be easily introduced through its hyper-parameters (Arjas and Gasbarra (1994), Gamerman (1994) and Ibrahim et al. (2001a), among others).

In this section we consider three different ways to build prior distributions for the failure rates of the PEM using Gamma distributions.

2.3.1 Independent priors

Assume that the components of $\boldsymbol{\lambda}_\rho$ are conditionally independent, with each $\lambda_\rho^{(j)}$ following a gamma distribution with shape and scale parameters α_j and γ_j, respectively, for $j = 1, ..., b$. Then, considering (12), the joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ becomes:

$$\pi(\boldsymbol{\lambda}_\rho, \rho) \quad \propto \quad \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{\alpha_j - 1} \exp\left\{-\gamma_j \lambda_\rho^{(j)}\right\} c_{I_\rho^{(j)}}. \tag{24}$$

In the case that information regarding the mean of the failure rates and their variances is available, informative prior distributions for the $\lambda_\rho^{(j)}$'s can be obtained by solving the system of equations formed by $E(\lambda_\rho^{(j)}) = \alpha_j/\gamma_j$ and $V(\lambda_\rho^{(j)}) = \alpha_j/\gamma_j^2$. Furthermore, by replacing (24) in (15), the posterior distribution of $(\boldsymbol{\lambda}_\rho, \rho)$ becomes:

$$\pi(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D}) \propto \prod_{j=1}^{b} \frac{\gamma_j^{\alpha_j}}{\Gamma(\alpha_j)} \left(\lambda_\rho^{(j)}\right)^{\alpha_j + \eta_j - 1} \exp\left\{-[\gamma_j + \xi_j]\lambda_\rho^{(j)}\right\} c_{I_\rho^{(j)}}. \tag{25}$$

It is straightforward to show that the posterior distribution for the random partition ρ given in (16) reduces to:

$$\pi(\rho|\mathbf{D}) \propto \prod_{j=}^{b} \frac{\gamma_j^{\alpha_j}}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \eta_j)}{(\gamma_j + \xi_j)^{\alpha_j + \eta_j}} c_{I_\rho^{(j)}},$$

and, not surprisingly, $(\lambda_\rho^{(j)}|\mathbf{D}, \rho) \sim Ga(\alpha_j + \eta_j, \gamma_j + \xi_j)$, for all $j = 1, ..., b$.

Although the gamma family of distributions allows for subjective elicitation of the prior distribution of each common failure rate $\lambda_\rho^{(j)}$, due to the randomness of ρ, it is not possible, in practice, to set a "realistic" informative prior distribution for each component of $\boldsymbol{\lambda}_\rho$ by a direct specification of the hyper-parameters α_j and γ_j. In this fashion, in the original formulation of the PEM via PPM, Demarqui et al. (2008) assumed that the common failure rates were i.i.d. with $\lambda_\rho^{(j)} \sim Ga(\alpha_0, \gamma_0)$, $\forall\ j = 1, ..., b$. Nevertheless, considering this specification, the prior means indicate a constant hazard function through time (equals to α_0/γ_0), which may not be in accordance with a gamma of hazard functions associated with many applications. Therefore, this drawback turns necessary the search for more comprehensive mechanism for the prior elicitation of λ_ρ. In the next two subsections we discuss two approaches which overcome this drawback.

2.3.2 Dynamic priors

In the following we show how the time grid of the PEM can be modeled assuming a product distribution for the random partition ρ as well as a dynamic approach for the components of $\boldsymbol{\lambda}_\rho$. In this setup, the dynamic model originally proposed by Gamerman (1994) arises as

a special case when we set $P(\rho = \{i_0, i_1, ..., i_b\}) = 1$ for a particular partition.

Let $D_\rho^{(0)}$ be the subjective information that we have before the data has been observed. Denote by $D_\rho^{(j)}$ the set of all information gathered up to the j-th clustered interval $I_\rho^{(j)}$, $j = 1, ..., b$. We perform the following sequential analysis:

i) Set $j = 1$, and specify $(\lambda_\rho^{(j)}|D_\rho^{(j-1)}, \rho) \sim Ga(\alpha_j; \gamma_j)$;

ii) Update the prior available information contained in $(\lambda_\rho^{(j)}|D_\rho^{(j-1)}, \rho)$ with the data information provided by $I_\rho^{(j)}$, yielding $(\lambda_\rho^{(j)}|D_\rho^{(j)}, \rho) \sim Ga(\alpha_j + \eta_j; \gamma_j + \xi_j)$;

iii) Perform the parametric evolution (through time) by setting $(\lambda_\rho^{(j+1)}|D_\rho^{(j)}, \rho) \sim Ga(\alpha_{j+1}; \gamma_{j+1})$, where $\alpha_{j+1} = (\alpha_j + \eta_j)\phi$ and $\gamma_{j+1} = (\gamma_j + \xi_j)\phi$, with $0 < \phi \leq 1$;

iv) Increment j, return to step (ii), and repeat the cicle until all data information has been processed.

In the literature of dynamic models, the complete conditional posterior distribution $(\lambda_\rho^{(j)}|D_\rho^{(j)}, \rho)$ is often referred to as online distribution. The quantity ϕ, called discount factor, controls the passage of information through successive intervals. As closer to one is ϕ, as much information is allowed to pass from one interval to another. On the other hand, as $\phi \to 0$, no information passes to the next interval, providing b independent estimations. Moreover, by setting $(\lambda_\rho^{(j)}|D_\rho^{(j-1)}, \rho) \sim Ga([\alpha_{j-1} + \eta_{j-1}]\phi; [\gamma_{j-1} + \xi_{j-1}]\phi)$, we have a prior distribution for $\lambda_\rho^{(j)}$ which has the same mean of $(\lambda_\rho^{(j-1)}|D_\rho^{(j-1)}, \rho)$, but with a variance larger than $V(\lambda_\rho^{(j-1),\rho}|D_\rho^{(j-1)}, \rho)$.

The dependence between $\lambda_\rho^{(j)}$'s associated with adjacent clustered intervals is specified through the hyper-parameters of their corresponding prior distributions. In this fashion, the components of $\boldsymbol{\lambda}_\rho$ are conditionally independent, and, consequently, the joint prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ can be written in the product form given in (24). Therefore, the conditions stated earlier in this text to model the randomness of the time grid of the PEM still hold in the present case.

The sequential analysis described above is equivalent to assume that the common failure rates $\lambda_\rho^{(j)}$'s are related via the following evolution equation:

$$\log(\lambda_\rho^{(j)}) = \log(\lambda_\rho^{(j-1)}) + \varepsilon_j, \tag{26}$$

where ε_j is a stochastic error with mean zero and variance given by:

$$\sigma_j^2 = \left(\frac{1}{\phi} - 1\right) V[\log(\lambda_\rho^{(j-1)})|D_\rho^{(j-1)}, \rho].$$

From equation (26) it is possible to obtain the complete conditional posterior distributions of each common failure rate $\lambda_\rho^{(j)}$ given all the available information contained in $D_\rho^{(b)}$, denoted by $(\lambda_\rho^{(j)}|D_\rho^{(b)}, \rho)$, $j = 1, ..., b$. These distributions, referred to as smoothed distributions, are

more informative than the online distributions $(\lambda_\rho^{(j)}|D_\rho^{(j)})$'s, which are based only upon the information available up to the current interval $I_\rho^{(j)}$, and therefore, should be preferred when making inferences for the components of $\boldsymbol{\lambda}_\rho$. The recursive (smoothing) algorithm used to obtain the smoothed distributions $(\lambda_\rho^{(j)}|D_\rho^{(b)}, \rho)$'s can be found in Gamerman (1991).

2.3.3 Structural priors

In this section we consider the structural (hierarchical) framework, first introduced by Ibrahim et al. (2001a) in the context of cure rate models, to set a prior distribution for λ_ρ. We refer such approach to as structural because the prior distribution for the $\lambda_\rho^{(j)}$'s are built under the assumption that both their means and variances are functions of a cumulative hazard function associated with a parametric distribution. Such structural modeling approach is particularly suitable for situations in which the estimation of the right tail survival curve is affected by the lack of data information.

Assume that the components of $\boldsymbol{\lambda}_\rho$ are independent a *priori* so that each $\lambda_\rho^{(j)}$ has a Gamma prior distribution with mean and variance, respectively, given by:

$$\mu_j = E(\lambda_\rho^{(j)}|\lambda_0, \kappa) = \frac{H_0(s_j) - H_0(s_{j-1})}{s_j - s_{j-1}} \tag{27}$$

and

$$\sigma_j^2 = V(\lambda_\rho^{(j)}|\lambda_0, \kappa) = \mu_j \kappa^j, \tag{28}$$

where $H_0(\cdot|\lambda_0)$ is a cumulative hazard function of a parametric survival distribution $F_0(\cdot|\lambda_0)$.

The choice of $F_0(\cdot|\lambda_0)$ is arbitrary and must reflect someone's opinion about the right tail of the survival curve. Note that, by taking $\kappa \to 0$, the prior process approaches $h_0(t|\lambda_0) = \frac{d}{dy}H_0(t|\lambda_0)$, once $\sigma_j^2 \to 0$ and we have that $h_0(t|\lambda_0) \approx \frac{H_0(s_j)-H_0(s_{j-1})}{s_j - s_{j-1}}$, for $t \in I_j = (s_{j-1}, s_j]$, $j = 1, ..., b$. Therefore, κ controls the degree of parametricity along the time axis. In particular, when $j \to \infty$, the right tail of the survival curve converges to the parametric model $F_0(\cdot|\lambda_0)$, regardless of the value assumed for κ (Ibrahim et al. (2001a)).

Considering the third parameter λ_0 (possible vector valued), the posterior distribution given in (15) becomes:

$$\begin{aligned} \pi(\boldsymbol{\lambda}_\rho, \lambda_0, \rho|\mathbf{D}) &\propto L(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D})\pi(\boldsymbol{\lambda}_\rho|\lambda_0, \rho)\pi(\rho)\pi(\lambda_0) \\ &\equiv \pi(\boldsymbol{\lambda}_\rho|\lambda_0, \rho, \mathbf{D})\pi(\lambda_0, \rho|\mathbf{D}), \end{aligned}$$

where $\pi(\lambda_0)$ corresponds to the prior distribution of λ_0, and does not depend on λ_ρ and ρ.

Sampling from the posterior distribution $\pi(\boldsymbol{\lambda}_\rho|\lambda_0, \rho, \mathbf{D})$ is straightforward since we have $(\lambda_\rho^{(j)}|\lambda_0, \rho, \mathbf{D}) \sim Ga(\mu_j\kappa^{-j} + \eta_j; \kappa^{-j} + \xi_j)$, $1 \leq j \leq b$. In order to sample from $\pi(\lambda_0, \rho|\mathbf{D})$, we

use the following complete conditional posteriors distributions:

$$\pi(\rho|\lambda_0, \mathbf{D}) \propto \prod_{j=1}^{b} \left(\frac{(1/\kappa^j)^{\mu_j \kappa^{-j}}}{\Gamma(\mu_j \kappa^{-j})} \frac{\Gamma(\mu_j \kappa^{-j} + \eta_j)}{(\kappa^{-j} + \xi_j)^{\mu_j \kappa^{-j} + \eta_j}} \right) c_{\rho}^{(j)} \tag{29}$$

and

$$\pi(\lambda_0|\rho, \mathbf{D}) \propto \left(\prod_{j=1}^{b} \frac{(1/\kappa^j)^{\mu_j \kappa^{-j}}}{\Gamma(\mu_j \kappa^{-j})} \frac{\Gamma(\mu_j \kappa^{-j} + \eta_j)}{(\kappa^{-j} + \xi_j)^{\mu_j \kappa^{-j} + \eta_j}} \right) \pi(\lambda_0). \tag{30}$$

The similarity between the complete conditional posterior distributions given in (26) and (29) is evident, since the second one depends on λ_0 only through the μ_j's. Therefore, the Gibbs sampling algorithm proposed by Barry and Hartigan (1993) can still be used to sample from (29).

We emphasize here that, although that structural modeling approach has been considered before by Ibrahim et al. (2001a), an approach considering simultaneously that hierarchical structure along with a random time grid for the PEM have not been considered in the literature yet.

3 Analysis of real data set

We use the proposed methodology in this section to analyze the survival times (in months) of 231 individuals diagnosed with brain cancer in Windham-CT, USA, during 1995 to 2004. This data set was obtained from SEER database. Our interest here relies on investigate the performance of the models discussed in the previous sections in estimating both the hazard and survival function. We further carry out a sensitivity analysis taking into account different specifications for m', ϕ and κ as well.

From the 231 patients diagnosed with brain cancer in this period, it was observed 134 deaths and 97 censored times, resulting in a percentage of failures of the order of 58%. It is also noteworthy that, since the survival times were measured in months, only 32 of the 134 observed failure times were distinct failure times. In order to set the finest time grid for the PEM we proceed as follows. Given m', we obtain k and r such that $m = km' + r$. The elements of τ' are then chosen so that the first $m' - r$ intervals have k failure times and the remaining r intervals have $k+1$ failure times. We note that this procedure for specifying the finest time grid assumed for the PEM is appealing in the sense that it allows more failure times to be placed in the last intervals, where less data information is available.

In the following analysis, we will restrict our attention to discrete uniform prior distributions for the random partition ρ. Such prior distributions can be obtained by setting $c_{I_\rho^{(j)}} = 1$ for all $j = 1, ..., b$, $b \in \mathcal{I}$. Other prior specifications for the random partition ρ, can be found in Demarqui et al. (2008). We refer to as Model 1 the model based on the objective

Jeffreys prior distribution for λ_ρ introduced in Section 2.2. The model originally proposed by Demarqui et al. (2008), briefly discussed in Section 2.3, is referred to as Model 2. For that specific model, it is assumed that $\lambda_\rho^{(j)} \sim Ga(0.001; 0.001)$ for all $j = 1, ..., b$, and $b \in \mathcal{I}$, so that the $\lambda_\rho^{(j)}$'s have a flat prior distribution with mean 1 and variance equals to 1000. The dynamic model described in Section 2.3.2 is referred to as Model 3. For that case, we assume that $(\lambda_\rho^{(1)}|D_\rho^{(0)}) \sim Ga(0.001; 0.001)$. Finally, Model 4 is referred to as the structural model discussed in Section 2.3.3. For that model the Weibull distribution with cumulative hazard function $H(t|\lambda_0) = \exp(\beta)t^\zeta$, where $\lambda_0 = (\zeta, \beta)'$ was considered. Consequently, it follows that $\mu_j = \exp(\beta)(s_j^\zeta - s_{j-1}^\zeta)/(s_j - s_{j-1})$ and $\sigma_j^2 = \kappa^j \exp(\beta)(s_j^\zeta - s_{j-1}^\zeta)/(s_j - s_{j-1})$. In this case, we assumed that ζ and β are independent a *priori*, with $\zeta \sim Ga(0.001; 0.001)$ and $\beta \sim N(0; 1000)$, respectively.

Samples from the quantities of interest were obtained by using MCMC techniques. Specifically, the Gibbs sampler algorithm proposed by Barry and Hartigan (1993) was considered to sample from the posterior distribution of the random partition ρ, and the adaptive rejection Metropolis sampling (ARMS) algorithm introduced by Gilks et al. (1995) was used to sample from the posterior distribution of $\lambda_0 = (\zeta, \beta)'$, whereas samples from the posterior distribution of λ_ρ were obtained straightforwardly. For all scenarios considered in our analysis, it was considered single chains of size 100,000. Samples of size 1,000 were obtained after considering a burn-in period of 50,000 iterations and a lag of 50 to eliminate correlations. All computational procedures were implemented in the Object-oriented Matrix Programming Language Ox version 5.1 (Doornik; 2007), and are available from the first author upon request.

In order to evaluate the performance of the models under investigation, we considered the average of the logarithm of the pseudo-marginal likelihood as a measure to assess the goodness of fit of each model. Such measure, which we shall refer to as B-statistic, is based on the conditional predictive ordinate (CPO) statistic (see Ibrahim et al. (2001b) and references therein), and is defined as

$$B = \frac{1}{n}\sum_{i=1}^{n} \log(CPO_i). \tag{31}$$

Here, the quantity CPO_i corresponds to the posterior predictive density of t_i if t_i is a failure time, and the posterior predictive probability of the event $(T > t_i)$ if t_i is a censored time. In both cases, it can be shown that the CPO_i can be well approximated by

$$\widehat{CPO}_i = M\left\{\sum_{l=1}^{M}[L(t_i|\boldsymbol{\lambda}_{\rho^l}^l, \rho^l)]^{-1}\right\}^{-1}, \tag{32}$$

where $(\boldsymbol{\lambda}_{\rho^l}^l, \rho^l)$ corresponds to the l-th draw of the the posterior distribution $\pi(\boldsymbol{\lambda}_\rho, \rho|\mathbf{D})$, $l = 1, ..., M$, and M is the size of the posterior sampled distribution of the parameters.

Table 1: B-statistic for all fitted models.

	random			fixed		
	$m' = 10$	$m' = 20$	$m' = 32$	$m' = 10$	$m' = 20$	$m' = 32$
Model 1	**-2.503350**	-2.516439	-2.606023	**-2.506534**	-2.532250	-2.63161
Model 2	-2.513367	**-2.510599**	-2.511427	**-2.506528**	-2.532218	-2.669160
Model 3	random			fixed		
ϕ	$m' = 10$	$m' = 20$	$m' = 32$	$m' = 10$	$m' = 20$	$m' = 32$
0.05	-2.496495	-2.49401	-2.502240	-2.498875	-2.518719	-2.599979
0.25	-2.496463	**-2.491982**	-2.49662	-2.493988	-2.493242	-2.518089
0.50	-2.506806	-2.496618	-2.494191	-2.502427	**-2.492373**	-2.495125
0.75	-2.548372	-2.518222	-2.503464	-2.540568	-2.509052	-2.495988
0.95	-2.674377	-2.65032	-2.621556	-2.65979	-2.626306	-2.587857
Model 4	random			fixed		
$\kappa^{m'}$	$m' = 10$	$m' = 20$	$m' = 32$	$m' = 10$	$m' = 20$	$m' = 32$
1×10^{-15}	-2.512682	-2.508459	-2.508458	-2.514249	-2.518155	-2.517366
1×10^{-10}	-2.506334	-2.502427	-2.508058	-2.504214	-2.519977	-2.513245
1×10^{-05}	-2.501480	**-2.495374**	-2.512812	-2.501460	-2.515992	-2.522181
1×10^{-01}	-2.502772	-2.501773	-2.516225	**-2.500776**	-2.519597	-2.570711

It can be noticed from Table 1 and Figure 1 that, when random time grids were assumed to fit Model 1, it provided better fits than the models with fixed time grids, independently of the choice for m'. Looking at Figure 1, we further noticed that, either when fixed or random time grids were assumed to fit this model, poorer fits were obtained as the number of initial intervals m' was increased.

Figure 2 displays the estimated hazard functions assuming fixed and random time grids for the PEM with $m' = 10$. We further show in Figure 2 the estimated survival functions provided by the PEM with random time grid considering $m' = 10$, 20 and 32, and the Kaplan & Meier estimate (KME). By analyzing separately the estimates for both the hazard and survival functions regarding each scenario considered for Model 1, we found that, given m', Model 1 yielded essentially the same estimates for both the hazard and survival functions of the PEM when either fixed or random time grids were assumed.The corresponding estimated survival functions were quite similar, showing slight differences only in the last months of follow up.

According to Table 1, Model 2 showed its best performance when we set a fixed time grid for the PEM along with $m' = 10$. Nevertheless, despite the best performance associated with this specific scenario, we can observe from Figure 1 that, if m' was set equal to either 20 or 32, better fits were obtained when random time grids were assumed for Model 2. It is also noticeable that Model 2 presented similar performances for all choices of m' when random time grids were considered, whereas the goodness of fit associate with the models fitted assuming fixed time grids were strongly affected by larger values of m'.

In opposite to what was found for Model 1, regardless of m', the estimated hazard functions, displayed in Figure 3, provided by Model 2 differed substantially when random and fixed time grids were assumed to fit the PEM. Specifically, for all choices of m' we observed

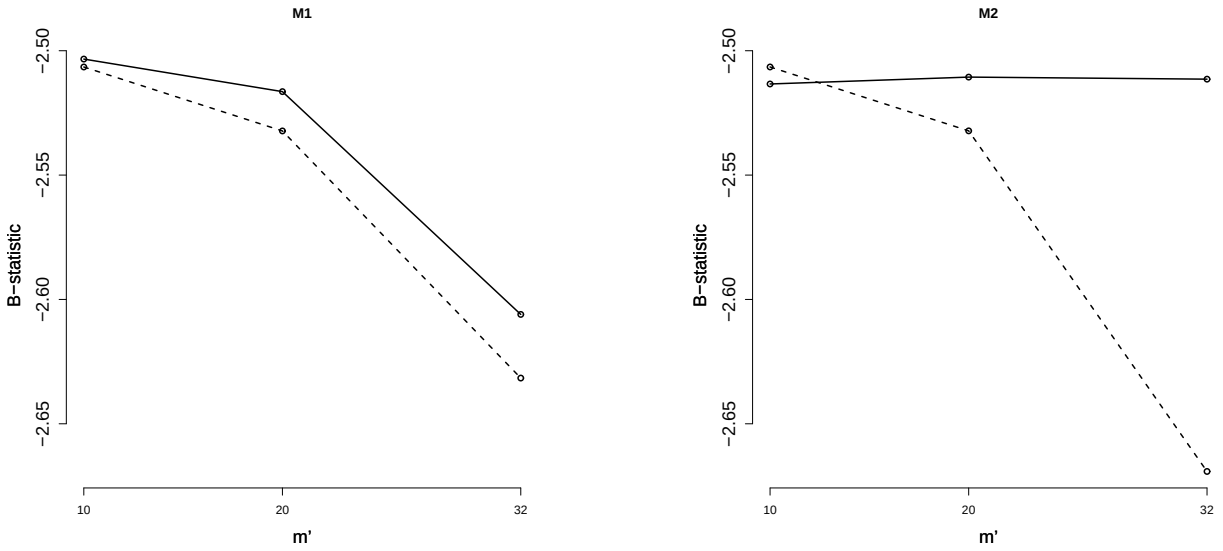

Figure 1: B-statistic for Models 1 and 2: random time grid (solid line) and fixed time grid (dashed line).

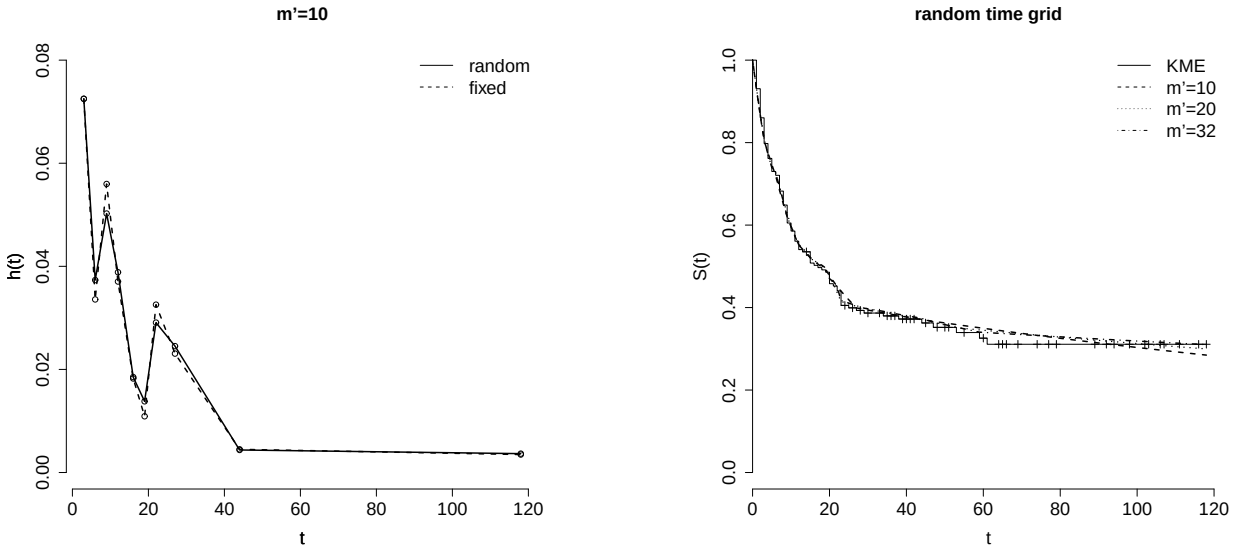

Figure 2: Estimated hazard and survival functions provided by Model 1.

that the estimated hazard functions associated with random time grid were considerably smoother than those obtained when fixed time grids were considered.

In order to evaluate the effect of different choices of ϕ on the performance of Model 3, we considered ϕ equals to 0.05, 0.25, 0.50, 0.75 and 0.95. As shown in Table 1, Model 3 presented the best performances among all fitted models. However, the performance of this model was substantially influenced by different choices of ϕ and m'. Specifically, by examining Figure 4 when random time grids were assumed, it is noteworthy a well defined association between the performance of Model 3 and ϕ (with better performances for small values of ϕ), regardless of the choice of m'. Figure 4 also indicates that large values of m'

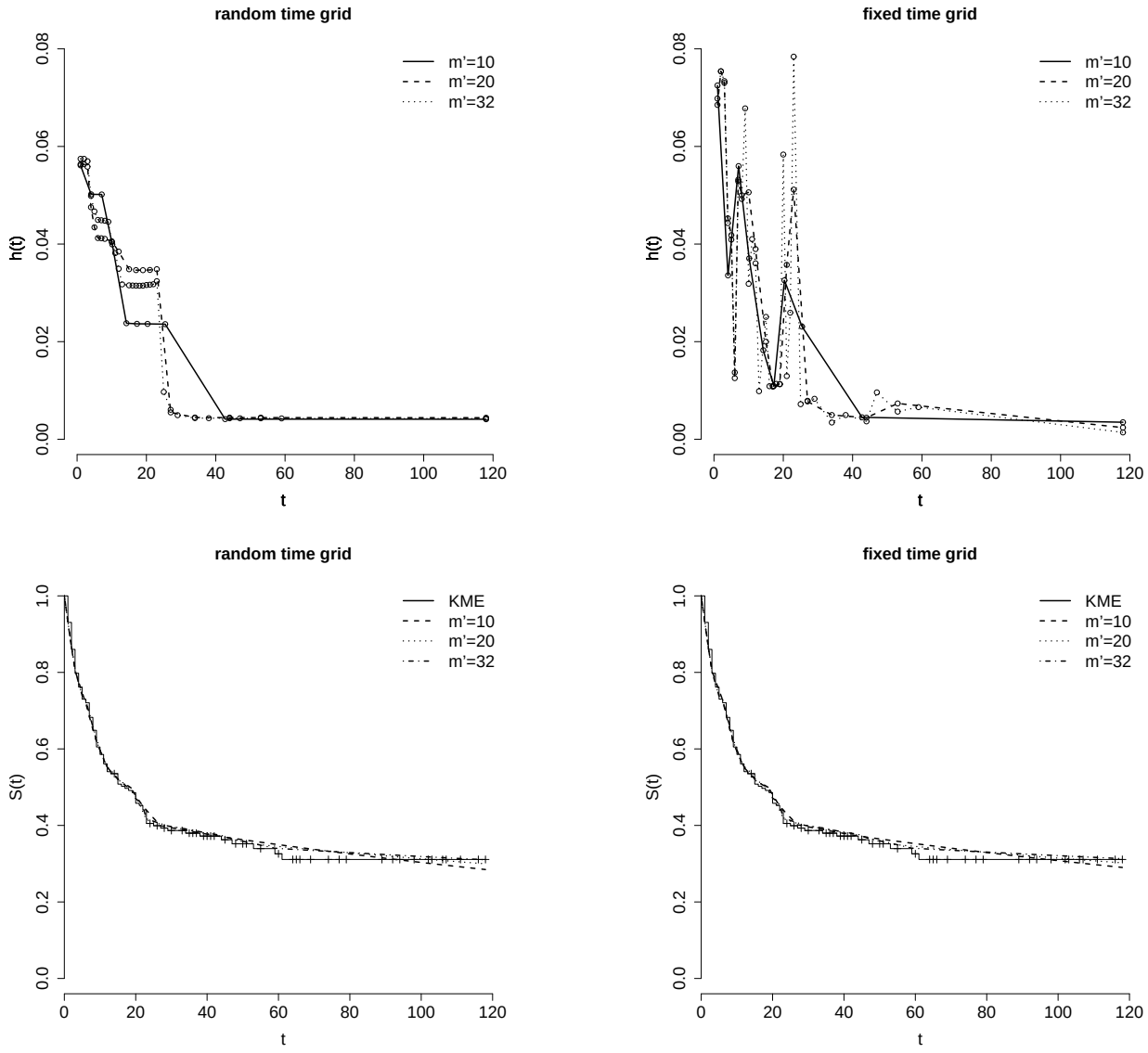

Figure 3: Estimated hazard and survival functions provided by Model 2.

tend to produce poorer fits. Moreover, when fixed time grids were assumed, the effect of ϕ on the goodness of fit of the model seems to depend strongly on the number of initial intervals m' than as it was observed for the random time grid scenario, with better performances for intermediate values of ϕ.

In Figures 5 and 6 we display the estimated hazard and survival functions, respectively, provided by Model 3 considering the extreme choices for ϕ, say ϕ equals to 0.05 and 0.95, for $m' = 10$, 20 and 32. It is possible to observe in Figure 5 that there is relationship between the choices of m' and ϕ and the smoothness of the hazard functions provided by Model 3. Specifically, by increasing the value of ϕ, we got the smoother estimates for the hazard function for each choice of m', whereas small values of ϕ provided less smooth estimates. Regardless of the choice of m', for both choices of ϕ, when a random time grid was assumed for Model 3, it provided smoother estimates for the hazard function than the corresponding

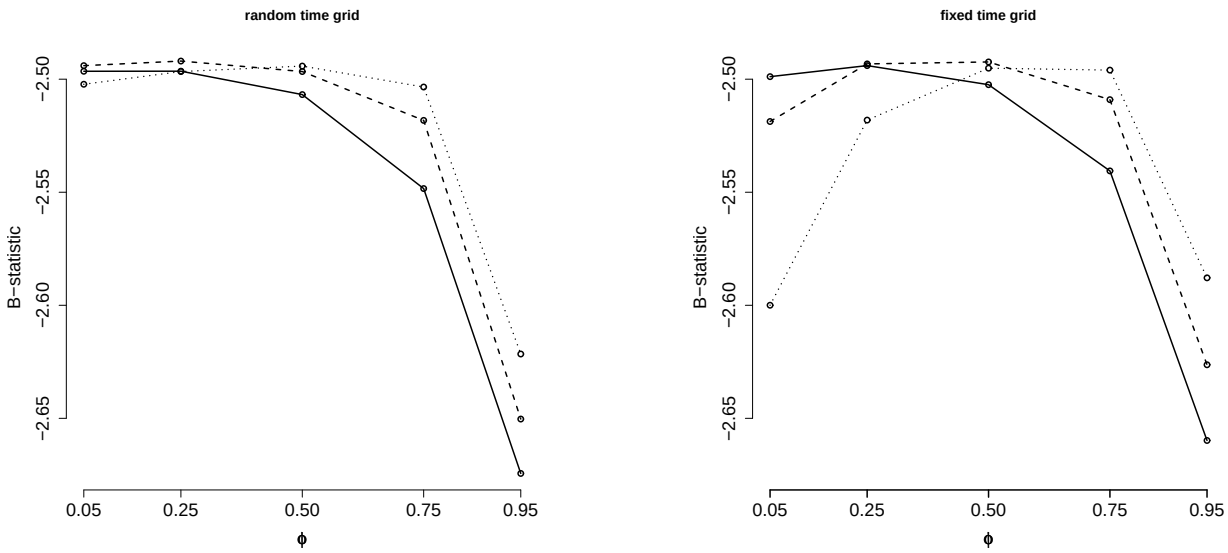

Figure 4: B-statistic for Model 3 - $m' = 10$ (solid line), $m' = 20$ (dashed line), and $m' = 32$ (dotted line).

model with fixed time grid. Indeed, this result was verified for all choices of ϕ. This fact was already expected since that by assuming a random time grid for the PEM, we are allowing some intervals associated with the finest time grid to be clustered, which reduces the number of intervals and, consequently, leads to smoother estimates for the hazard function.

The estimated survival functions provided by Model 3 assuming $\phi = 0.05$ were quite similar, with slight differences observed only in the last months of follow up, independently of the choice of m'. Such estimates are also in agreement with the KME of the survival function, indicating that the models seems to be well fitted. By either assuming fixed and random time grids for the PEM we also obtained similar estimated survival functions for $\phi = 0.95$. Nevertheless, we observed that the corresponding estimated survival functions for $m' = 32$ were substantially different from those estimates obtained assuming the finest time grids with 10 and 20 intervals. In addition, such estimates were considerably set apart from the KME of the survival function, which corroborates with the lack of goodness of fit indicated by the analysis of the B-statistic displayed in Table 1. This can be understood as a result of the effect of the over-smoothed estimated hazard functions shown in Figure 5.

Turning our attention to Model 4, we investigated the role played by the smoothing parameter κ on the goodness of fit of this model. We performed a sensitivity analysis considering different values of κ. With this aim we set κ so that $\kappa^{m'}$ was equal to 10^{-01}, 1×10^{-05}, 1×10^{-10} and 1×10^{15}, respectively. In the random time grid scenario we observed from Figure 7 that, regardless of the choice of $\kappa^{m'}$, the best fitted models were obtained when $m' = 20$. Close results were obtained by assuming $m' = 10$, and, in both cases, the values of κ so that $\kappa^{m'}$ was equal to 10^{-05} provided the best fits. For $m' = 32$ we observed a decreasing in the goodness of fit as $\kappa^{m'}$ got larger. Differently from what was observed in

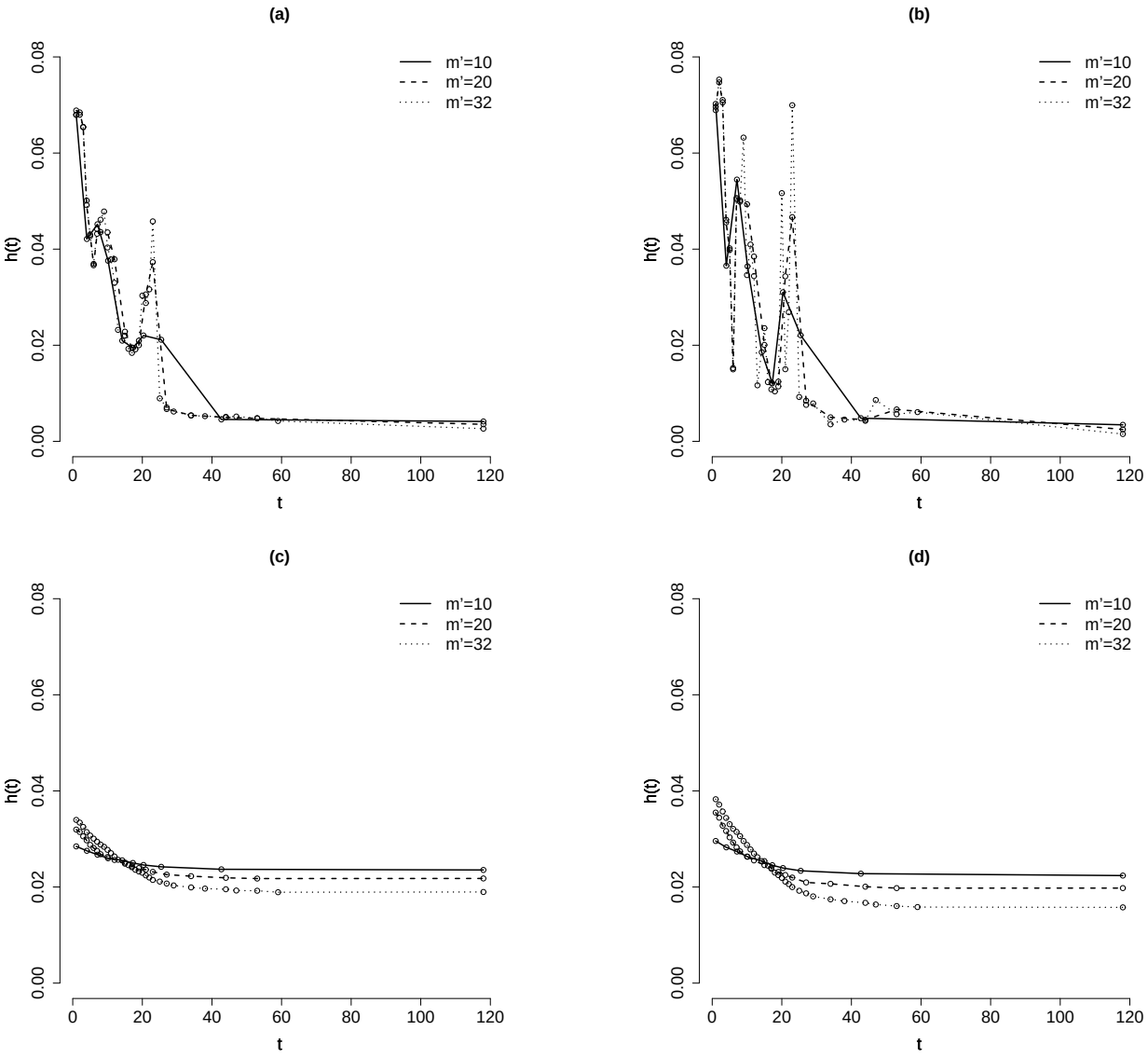

Figure 5: Estimated hazard functions provided by Model 3: random time grid and $\phi = 0.05$ (a), fixed time grid and $\phi = 0.05$ (b), random time grid and $\phi = 0.95$ (c), and fixed time grid and $\phi = 0.95$ (d),

the random time grid scenario, in the fixed one it can be noticed in Figure 7 that the best fitted models were obtained by setting $m' = 10$. In addition, when we set $m' = 32$, we can observe that the goodness of fit of the model was considerably affected as the value of $\kappa^{m'}$ is changed. The worst performance was obtained for $\kappa^{m'} = 1 \times 10^{-01}$.

We show in Figures 8 and 9 the plots regarding the estimated hazard and survival functions provided by Model 4 considering $m' = 10$, 20 and 32, for $\kappa^{m'} = 1 \times 10^{-01}$ and $\kappa^{m'} = 1 \times 10^{-15}$. Taking into account scenarios (a) and (b) in Figure 8, it can be observed that the estimated hazard functions provided by Model 4 assuming a random time grid for the PEM were smoother than those yielded by the model with fixed time grid, regardless of the choice of m'. However, when we look at scenarios (c) and (d) in Figure 8, we can see that,

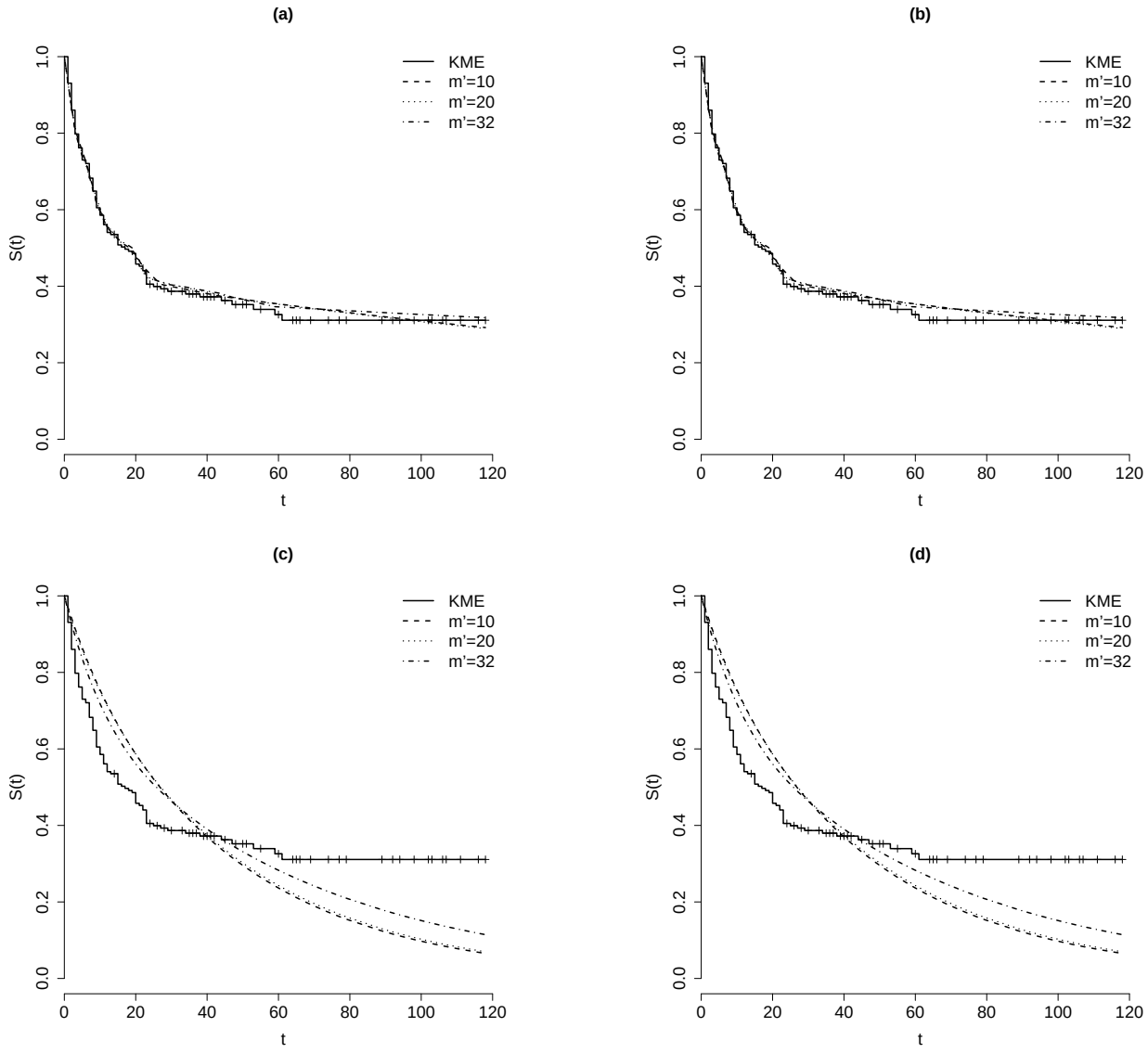

Figure 6: Estimated survival functions provided by Model 3: random time grid and $\phi = 0.05$ (a), fixed time grid and $\phi = 0.05$ (b), random time grid and $\phi = 0.95$ (c), and fixed time grid and $\phi = 0.95$ (d),

in opposite to what was observed for scenarios (a) and (b), for all choices of m' smoother estimates for the hazard function were obtained when the model with fixed grid was considered. In order to understand such result, we need to have in mind that, as the value of κ was fixed in advance, whenever a random time grid for the PEM was assumed, $b \leq m'$ and, consequently, $\kappa^{m'} \leq \kappa^b$, which yields "less" informative prior distributions for the $\lambda_\rho^{(j)}$'s. Hence, when $\kappa^{m'}$ was too small (it turns out that this was the case for $\kappa^{m'} \leq 1 \times 10^{-05}$ for the current data set), the model with fixed time grid was forced to be more parametric than the model with random time grid, which explains the results displayed in Figure 8.

As we have already pointed out, one of the many attractive features of the approach we are proposing is that it enable us to make inferences regarding both the time grid of the

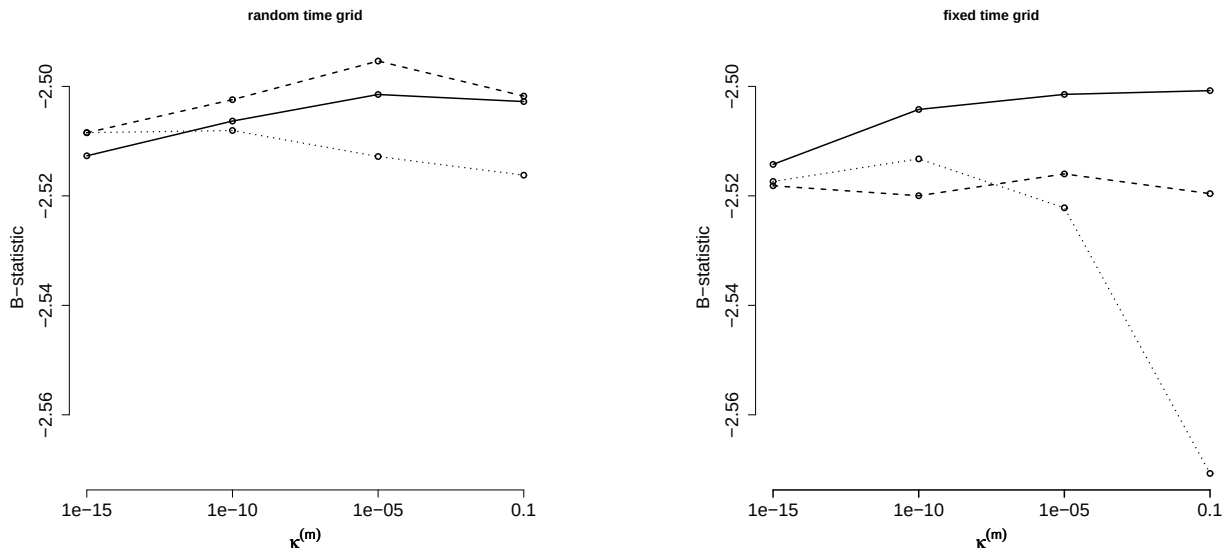

Figure 7: B-statistic for Model 4 - $m' = 10$ (solid line), $m' = 20$ (dashed line), and $m' = 32$ (dotted line).

PEM as well as its corresponding number of (clustered) intervals. In this fashion, we show in Table 2 the most probable time grids associated with the best fitted models according to Table 1. The corresponding posterior distributions of the number of clustered intervals are displayed in Figure 10, for which some summary statistics are presented in Table 3. Finally, Figure 11 provides an insight of the reduction in the number of initial intervals assumed for the PEM due to the clustering structure used to model the randomness of the time grid.

Table 2: Most probable time grids associated with the best fitted models.

Model	time grid	probability
1	$\tau = \{0, 3, 6, 12, 16, 19, 22, 27, 44, 118\}$	0.088
2	$\tau = \{0, 23, 118\}$	0.290
3	$\tau = \{0, 2, 3, 4, 10, 12, 17, 21, 23, 27, 53, 118\}$	0.002
4	$\tau = \{0, 5, 6, 7, 12, 19, 21, 23, 34, 44, 118\}$	0.002

Except for Model 1, whose best fit was obtained by considering $m' = 10$, the other models shown in Table 2 were obtained when $m' = 20$. Among the models presented in Table 2, Model 2 yielded the estimated time grid with the greatest reduction regarding the initial number of intervals. In addition, estimated time grid provided by Model 2 had by far the highest posterior probability.

In terms of the reduction in the number of initial intervals, we observed from Figure 11 that Model 1 presented an overall low reduction (around 20%) whereas Models 2 showed the greatest reduction among all fitted models ($> 70\%$ for all choices of m'). The reduction associated with Model 3 ranged from 10% to 70%, depending on specific combinations of ϕ and m', and a reduction falling between 40% and 72% was observed for model 4, according

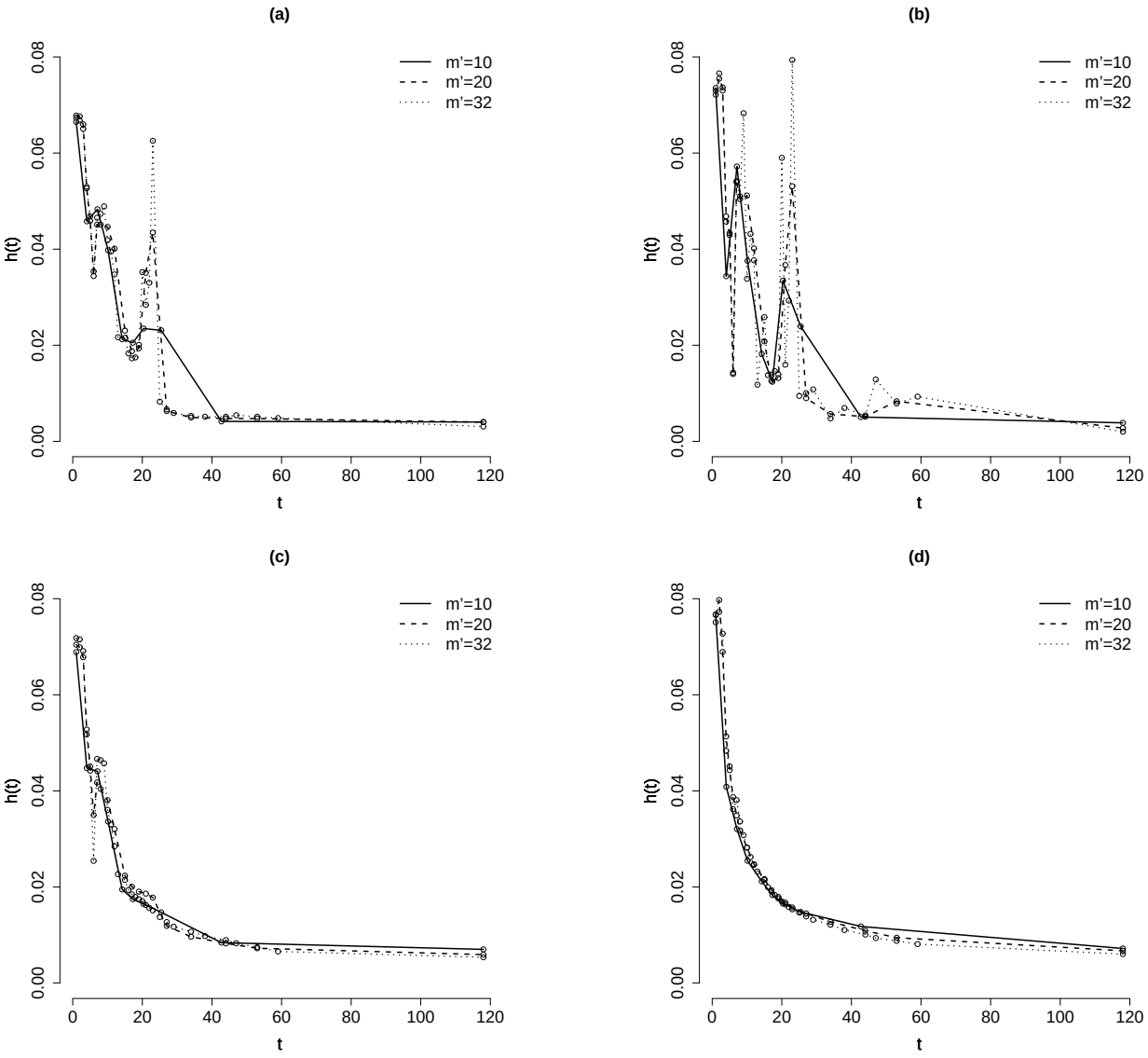

Figure 8: Estimated hazard functions provided by Model 4: random time grid and $\kappa^{m'} = 1 \times 10^{-01}$ (a), fixed time grid and $\kappa^{m'} = 1 \times 10^{-01}$ (b), random time grid and $\kappa^{m'} = 1 \times 10^{-15}$ (c), and fixed time grid and $\kappa^{m'} = 1 \times 10^{-15}$ (d).

Table 3: Summary statistics for the number of clustered intervals associated with the best fitted models.

Model	m'	median	mean	sd	95% HPD-interval	mode (prob.)
1	10	8	7.92	1.197	[6; 10]	8 (0.334)
2	20	3	2.705	0.496	[2; 3]	3 (0.670)
3	20	11	11.094	1.738	[8; 14]	11 (0.247)
4	20	10	9.845	1.951	[6; 13]	9 (0.222)

to different choices of $\kappa^{m'}$ and m'.

We complete our analysis discussing, in general lines, the improvement in the goodness of fit as a result of the assumption of random time grids for the PEM. Based on Table 4,

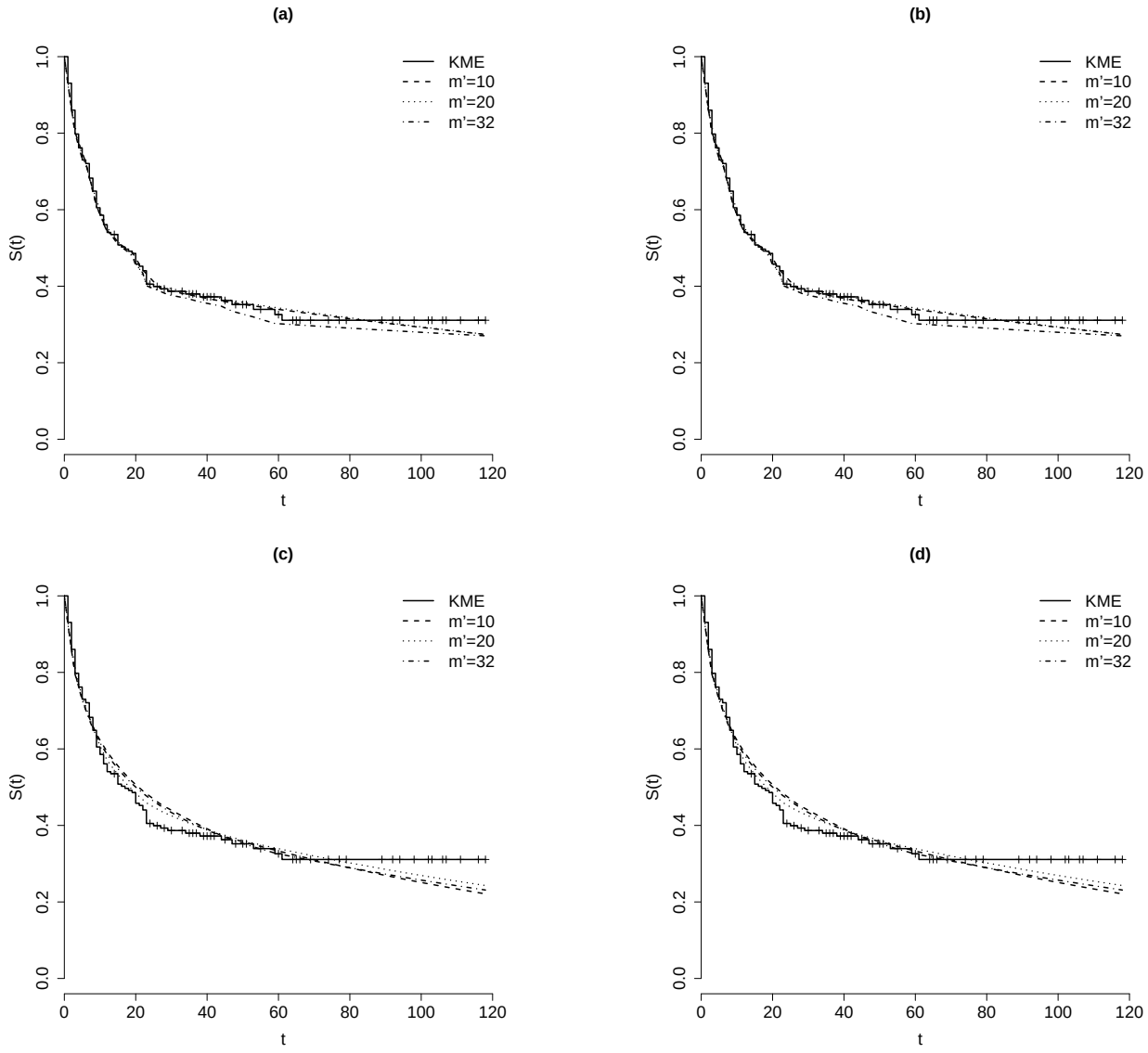

Figure 9: Estimated survival functions provided by Model 4: random time grid and $\kappa^{m'} = 1 \times 10^{-01}$ (a), fixed time grid and $\kappa^{m'} = 1 \times 10^{-01}$ (b), random time grid and $\kappa^{m'} = 1 \times 10^{-15}$ (c), and fixed time grid and $\kappa^{m'} = 1 \times 10^{-15}$ (d).

which summarizes the performance of the PEM with random time grid, we observe that the assumption of random time grid provided better fits in 100% of the scenarios considered for Model 1, and in 66.66% of those considered for Model 2. Despite the overall performance of Model 3 with random time grids was lower than we expected (better fitted models in 40% of the cases), we remark that Model 3 with a random time grid provided the best fitted model among all models considered. Therefore, we strongly suggest the reader to performing a sensitivity analysis taking into account different specifications for both m' and ϕ when fitting Model 3. With regard to Model 4, it was observed an improvement of the order of 75% in the goodness of fit when a random time grid was assumed for the PEM. A sensitivity analysis considering different choices of κ is also recommended in this case.

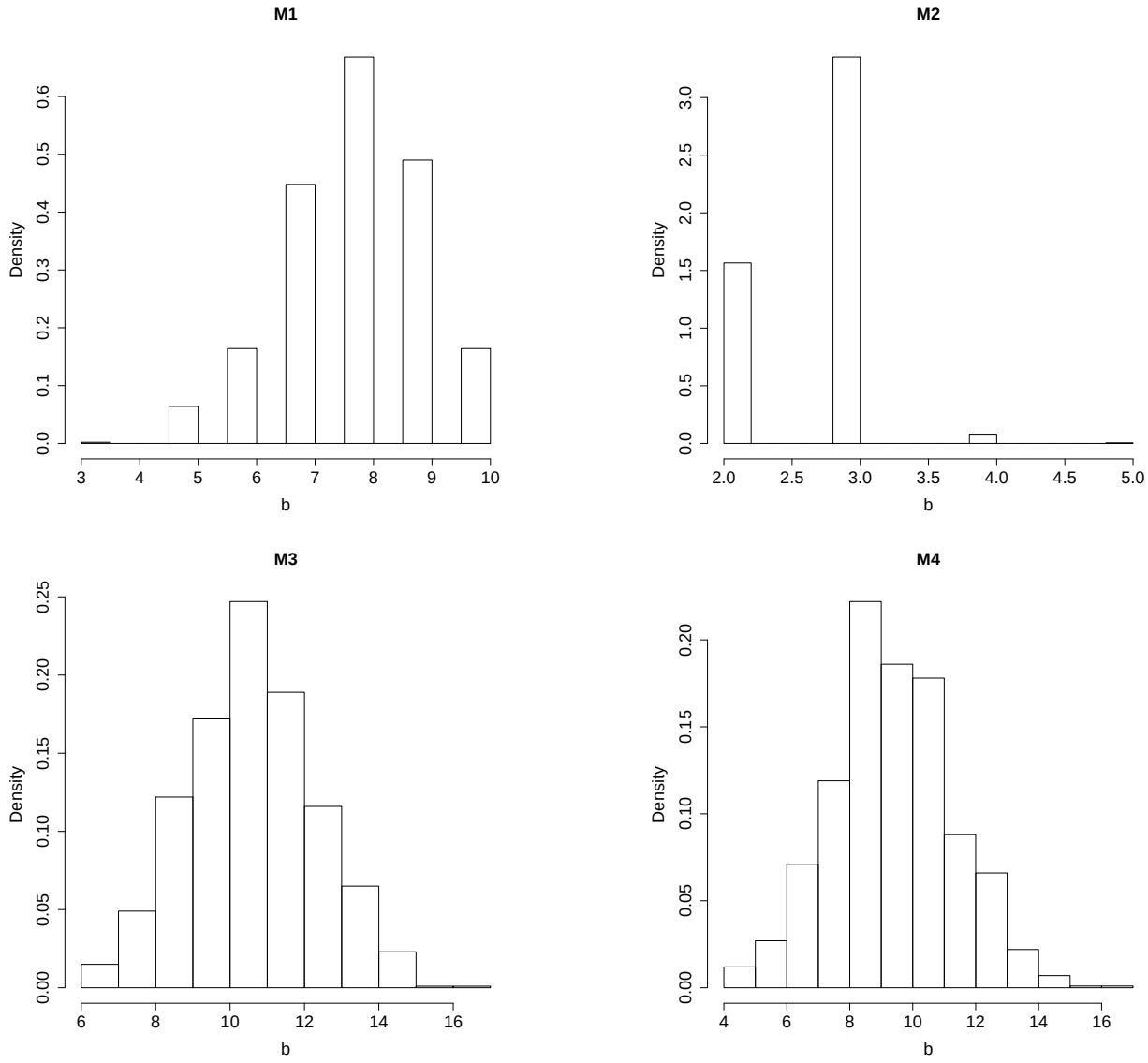

Figure 10: Posterior distributions of the number of intervals for the best fitted models.

Finally, considering all fitted models, the PEM with random time grid yielded better fits in 60% of the scenarios analyzed, indicating that the usefulness of the proposed approach.

Table 4: Performance of the PEM with random time grid.

Model	better fit	total	%
1	3	3	100.00
2	2	3	66.66
3	6	15	40.00
4	9	12	75.00
All models	20	33	60.60

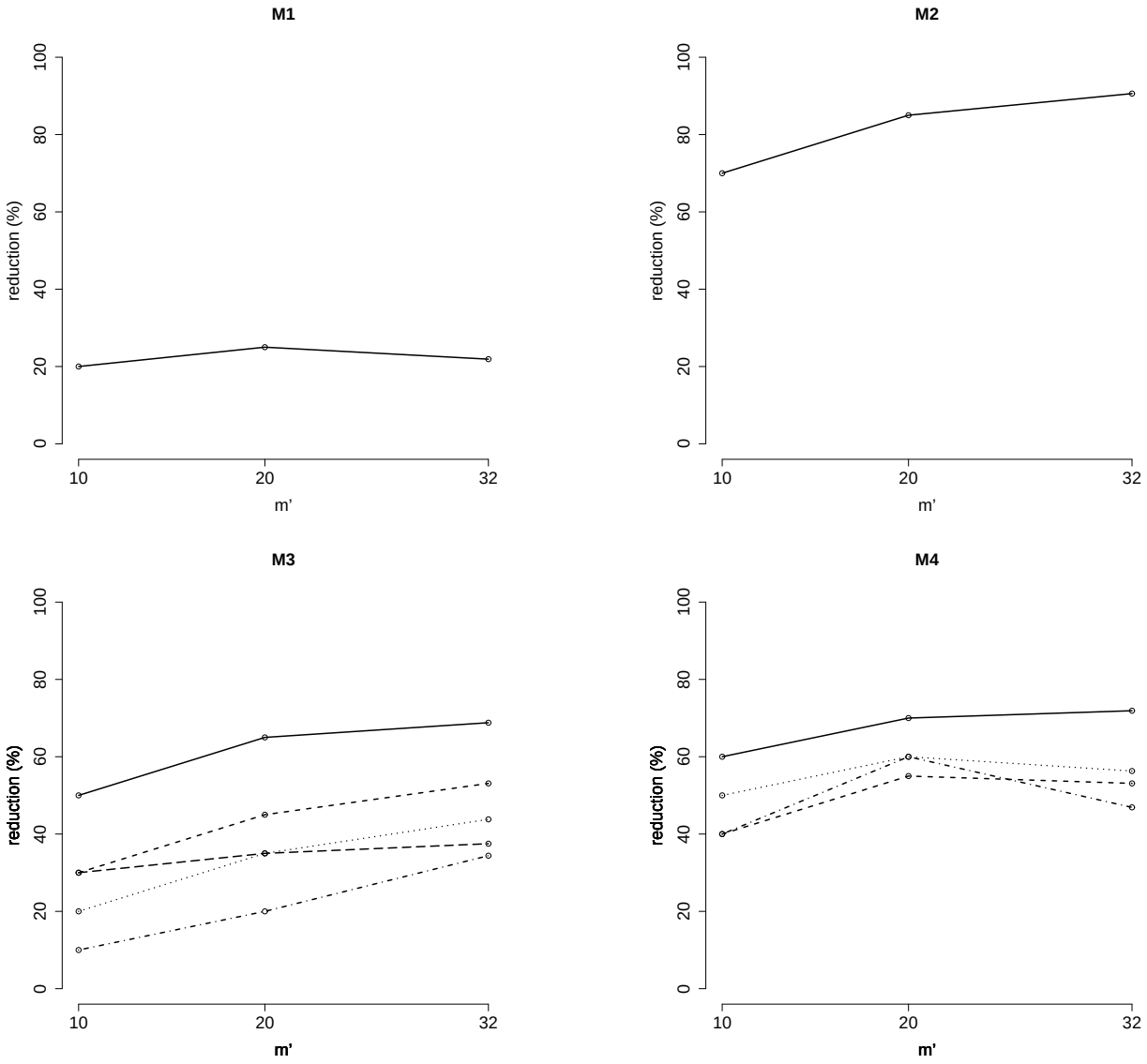

Figure 11: Reduction due to clustering structure: Model 3 - $\phi = 0.05$ (solid line), $\phi = 0.25$ (dashed line), $\phi = 0.50$ (dotted line), $\phi = 0.75$ (dashed-dotted line), and $\phi = 0.95$ (double-dashed line); Model 4 - $\kappa^{m'} = 1 \times 10^{-01}$ (solid line), $\kappa^{m'} = 1 \times 10^{-05}$ (dashed line), $\kappa^{m'} = 1 \times 10^{-10}$ (dotted line) and $\kappa^{m'} = 1 \times 10^{-15}$.

4 Conclusions

We have introduced a class of semiparametric Bayesian approaches for modeling survival data assuming the PEM with random time grid. The proposed model extends other models discussed in the literature by allowing more flexibility in the modeling of both the time grid and the failure rates of the PEM, and is suitable for modeling right censored time-to-event data arising from any area of knowledge.

The proposed approach has several advantages over other ones that are available in the literature. It allows the direct modeling of the time grid of the PEM after suitable prior elicitation. In particular, the mechanism assumed to generate the random time grids for

the PEM depends only upon a set of observed failure times. This fact enables us to model the time grid taking into account the arrangement of the observed failure times over the time axis, as well as to control the maximum number of failure times associated with all the random intervals induced by the generated time grids. Furthermore, the proposed approach accommodates a wide class of prior distributions for both the time grid and the failure rates of the PEM (not all considered in this paper), and includes other models established in the literature as special cases, which provides a flexible framework for survival data modeling.

Based on our experience, we recommend Models 3 and 4 as the best models under the framework we have proposed. Such models have shown to provide better performances than Models 1 and 2, as was illustrated in the analysis of the brain cancer data. Moreover, we understand that the dynamic and structural prior distributions add more flexibility in survival data modeling. Nevertheless, since Models 3 and 4 have shown to be sensible to different prior specifications, a sensitivity analysis taking into account different prior specifications is strongly recommended.

Future works involve carrying out a study with artificial data sets in order to evaluate to confirm the good performance the proposed methodology. In addition, the proposed approach can be readily extended to the proportional hazards framework as well as to model survival data with cure fraction. Further possible extensions include introducing frailties into the model to accounting for multivariate and/or spacial correlated survival data.

Acknowledgements

Fabio N. Demarqui's research has been sponsored by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) of the Ministry for Science and Technology of Brazil, and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais). Rosangela H. Loschi's research has been partially funded by CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 306085/2009-7, 304505/2006-4. Enrico A. Colosimo's research has been partially funded by the CNPq, grants 150472/2008-0 and 306652/2008-0. Part of this work was made during the period the first author was visiting the Department of Statistics of the University of Connecticut, Storrs-CT, USA.

References

Aitkin, M., Laird, N. and Francis, B. (1983). A reanalysis of the Stanford heart transplant data (with discussion), *Journal of the American Statistical Association* **78**: 264–292.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, *Statistica Sinica* **4**: 505–524.

Barbosa, E. P., Colosimo, E. A. and Louzada-Neto, F. (1996). Accelerated life tests analyzed by a piecewise exponential distribution via generalized linear models, *IEEE Transactions on Reliability* **45**: 619–623.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems, *Ann Statist* **20**: 260–279.

Barry, D. and Hartigan, J. A. (1993). Bayesian analysis for change point problems, *Journal of the American Statistical Association* **88**: 309–319.

Bastos, L. S. and Gamerman, D. (2006). Dynamic survival models with spatial frailty, *Lifetime Data Analysis* **12**: 441–460.

Breslow, N. E. (1974). Covariance analysis of censored survival data, *Biometrics* **30**: 89–99.

Clark, D. E. and Ryan, L. M. (2002). Concurrent prediction of hospital mortality and length of stay from risk factors on admission, *Health Services Research* **37**: 631–645.

Demarqui, F. N., Loschi, R. H. and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model, *Lifetime Data Analysis* **14**: 333–356.

Doornik, J. A. (2007). *Ox 5 - An Object-oriented Matrix Programming Language*, Timberlake Consultants Ltd.

Gamerman, D. (1991). Dynamic Bayesian models for survival data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **40**: 63–79.

Gamerman, D. (1994). Bayes estimation of the piece-wise exponential distribution, *IEEE Transactions on Reliability* **43**: 128–131.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within gibbs sampling, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **44**: 455–472.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001a). Bayesian semiparametric models for survival data with a cure fraction, *Biometrics* **57**: 383–388.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001b). *Bayesian survival analysis*, Springer-Verlag, New York.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life models, *Biometrika* **60**: 267–278.

Kim, J. S. and Proschan, F. (1991). Piecewise exponential estimator of the survival function, *IEEE Transactions on Reliability* **40**: 134–139.

Kim, S., Chen, M. H., Dey, D. K. and Gamerman, D. (2006). Bayesian dynamic models for survival data with a cure fraction, *Lifetime Data Analysis* **13**: 17–35.

McKeague, I. W. and Tighiouart, M. (2000). Bayesian estimators for conditional hazard functions, *Biometrics* **56**: 1007–1015.

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

Sahu, S. K., Dey, D. K., Aslanidu, H. and Sinha, D. (1997). A weibull regression model with gamma frailties for multivariate survival data, *Lifetime Data Analysis* **3**: 123–137.

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

A class of dynamic Piecewise Exponential Models with Random Time Grid

Fabio N. Demarqui[1] *, Rosangela H. Loschi[1],
Dipak K. Dey[2] and Enrico A. Colosimo[1]

[1] Universidade Federal de Minas Gerais, Brazil,
email:{fndemarqui, loschi, enricoc}@est.ufmg.br
[2]University of Connecticut, USA,
emai: dey@stat.uconn.edu

Abstract

A novel fully Bayesian approach for modeling survival data with explanatory variables using the Piecewise Exponential Model (PEM) with random time grid is proposed. We consider a class of correlated Gamma prior distributions for the failure rates. Such prior specification is obtained via the dynamic generalized modeling approach jointly with a random time grid for the PEM. A product distribution is considered for modeling the prior uncertainty about the random time grid, turning possible the use of the structure of the Product Partition Model (PPM) to handle the problem. A unifying notation for the construction of the likelihood function of the PEM, suitable for both static and dynamic modeling approaches, is considered. Procedures to evaluate the performance of the proposed model are presented. The use of the new methodology is exemplified by the analysis of a real data set of survival times of patients with brain cancer obtained from SEER (Surveillance Epidemiology and End Results) database. For comparison purposes, the data set is also fitted using the dynamic model with fixed time grid established in the literature. The results obtained show the superiority of the proposed model..

Keywords: Bayesian inference; MCMC methods; Model comparison; Product Partition Model; Survival Analysis.

*To whom all correspondences should be addressed. Departament of Statistics, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6.627, Pampulha, 31270-901 - Belo Horizonte - MG, Brazil, Phone: (+55 31) 3409 5938. Fax: (+55 31) 3409 5924.

1 Introduction

The Piecewise Exponential Model (PEM) is one of the most popular and useful models in Reliability and Survival Analysis. This model provides good approximations to many continuous distributions and, in practical situations, can describe accordingly hazard functions of several shapes. An interesting practical feature of this model is that it is not needed to specify a distribution for the data which has a closed form expression for the hazard function. This characteristic brings great flexibility in modeling. For this reason, the PEM has been widely used to model time to event data in different contexts, such as in reliability engineering (Kim and Proschan, 1991, Gamerman, 1994), clinical situations such as kidney infection (Sahu et al.; 1997), heart transplant data (Aitkin et al.; 1983), hospital mortality data (Clark and Ryan; 2002), economics (Bastos and Gamerman; 2006), and cancer studies including leukemia (Breslow; 1974), gastric cancer (Gamerman; 1991), breast cancer (Ibrahim et al.; 2001b), melanoma (Kim et al.; 2006) and nasopharynx cancer (McKeague and Tighiouart; 2000), among others

In order to construct the PEM we need to specify a finite time grid $\tau = \{s_0, s_1, ..., s_J\}$ which divides the time axis into J disjoint intervals. Given this time grid, we assume a constant failure rate λ_j for each interval $I_j = (s_{j-1}, s_j]$, $j = 1, ..., J$. Consequently, we have a discrete version, in the form of a step function, of the true and unknown hazard function. Therefore, it is evident that a good approximation for the true hazard function depends on a suitable choice of the time grid.

There are many previous works related to the PEM. Friedman (1982) considers the PEM to model the baseline hazard function of the Cox model, and derives conditions regarding the existence and asymptotic properties of the maximum likelihood estimators (MLEs) for both the failure rates and the regression coefficients. Later, Kim and Proschan (1991) discuss some advantages of the PEM over the well-known Kaplan-Meyer estimator (KME) for the survival function in the absence of covariates. Barbosa et al. (1996) apply generalized linear model methods to fit accelerated life tests data using the PEM. The maximum likelihood method for fitting cure rate models using the PEM in the presence of missing covariates is considered by Chen and Ibrahim (2001). The Bayesian literature related to the PEM is also extensive. Gamerman (1991) extends the Cox model by proposing a dynamic approach to fit survival data in the presence of explanatory variables with effects changing through time. Sahu et al. (1997) use the PEM to model multivariate survival data. In that work, frailties are considered to account for correlation of survival times of elements within a same specific group, and a correlated process for the failure rates of the PEM is assumed. An overview of some Bayesian approaches considering independent gamma prior distributions, Jeffreys's prior and some correlated processes for the failure rates of the PEM, is given in Ibrahim et al. (2001b). The PEM with a correlated process for the failure rates governed by a smoothing parameter is considered by Ibrahim et al. (2001a) to fit a cure rate model

based upon biological motivations. Dynamic models accommodating spatial correlation and time varying covariate effects are proposed by Bastos and Gamerman (2006).

Despite the vast existing literature, the time grid $\tau = \{s_0, s_1, ..., s_J\}$ has been arbitrarily chosen in most of the works related to the PEM, including all those cited previously. Kalbfleisch and Prentice (1973) suggest that the selection of the time grid $\tau = \{s_0, s_1, ..., s_J\}$ should be made independent of the data, but they do not provide any procedure to do such. Breslow (1974) proposes defining the endpoints s_j of the intervals $I_j = (s_{j-1}, s_j]$ as the observed failure times. This choice of time grid for the PEM, which we shall refer in this paper to as the nonparametric approach, has been commonly used to fit the PEM. Other heuristic discussions regarding adequate choices for the time grid of the PEM can be found in Gamerman (1991), Sahu et al. (1997) and Qiou et al. (1999), and a lot of others. The first effective effort to model the time grid of the PEM is due to Arjas and Gasbarra (1994). In their work it is assumed that the endpoints of the intervals $I_j = (s_{j-1}, s_j]$ are defined according to a jump process following a martingale structure which is specified through the prior distributions. Latter, McKeague and Tighiouart (2000) introduce a general class of models in which the jump times s_j's are generated from a time-homogeneous Poisson process and, conditionally on the jump times s_j's, the log-baseline hazard is modeled as a Gaussian Markov random field. Such approach includes as special cases the models proposed by Gamerman (1991) and Arjas and Gasbarra (1994). Kim et al. (2006) also consider the PEM with a random grid to model the baseline hazard function of cure rate models. Similar to McKeague and Tighiouart (2000), Kim et al. (2006) assume a time-homogeneous Poisson process to describe the uncertainty about the endpoints of the time grid.

The assumption of a time-homogeneous Poisson process to model the time grid $\tau = \{s_0, s_1, ..., s_J\}$ has some drawbacks. First, as pointed out by Chen et al. (1999), in the absence of any prior information about λ_j, at least one failure is supposed to occur at each interval $I_j = (s_{j-1}, s_j]$ so that λ_j may be estimable. Although this condition can be relaxed by assuming a correlated or smoothed hazard function for the PEM (see Gamerman; 1991, 1994; Ibrahim et al.; 2001b, for instance), it is not possible to establish conditions on the time grid under that assumption. Besides, by assuming a time-homogeneous Poisson process to generate the endpoints of the time grid $\tau = \{s_0, s_1, ..., s_J\}$, the number of parameters in the model becomes random, and a reversible jump Markov chain Monte Carlo algorithm (Green; 1995) is needed to sample from the posterior distributions, which leads to an increase in algorithmic complexity. Moreover, it seems to be more reasonable to set a time grid to fit the PEM which is somehow related to the failure times.

Based on the arguments presented above, Demarqui et al. (2008) introduce an approach to model the time grid of the PEM that overcome those drawbacks. By imposing a constraining on the set of possible time grids, they prove that the prior distribution of $\tau = \{s_0, s_1, ..., s_J\}$ can be written as a product form, and show how the structure of the Product Partition Model (PPM) proposed by Barry and Hartigan (1992) can be used to handle the problem.

In their approach, however, a regression structure on the hazard function is not considered, and independent gamma prior distributions are assumed for the failure rates induced by the random time grids.

In this paper we introduce a general class of PEM in which the time grid is random and both the failure rates and the regression coefficients in different intervals are correlated. This new approach extends static models such as in Ibrahim et al. (2001b) and Demarqui et al. (2008), as well as the dynamic model based on an arbitrary time grid proposed by Gamerman (1991). By considering a dynamic generalized modeling approach along with a product distribution for the random time grid, we assume a class of correlated Gamma prior distributions for the failure rates and use the clustering structure of the PPM to model the randomness of the time grid of the PEM. We further develop procedures to evaluate the performance of the proposed model and carry out sensitivity analysis for different prior specifications. Finally, to illustrate the use of the proposed model, the analysis of survival times of patients with brain cancer is performed. Such data set is obtained from SEER (Surveillance, Epidemiology and End Results) database. The results are compared with those obtained by fitting the dynamic model proposed by Gamerman (1991).

This paper is organized as follows: in Section 2 we develop the formulation of the model and the necessary notation needed. The new methodology is applied to analyze a real data set in Section 3, and in Section 4 some conclusions and perspectives of future works are presented.

2 Model formulation

In survival analysis, covariate effects are usually specified through the hazard function. In this spirit, the proportional hazard (PH) model introduced by Cox (1972) is one of the mostly used survival models to incorporate covariate effects into the analysis. The PH model is given by:

$$h(t) = h_0(t)g(\mathbf{x}\,\boldsymbol{\beta}), \tag{1}$$

where $h_0(t)$ is called the baseline hazard function, $\mathbf{x}$ is a $1 \times p$ vector of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, and $g(.)$ is a non-negative function.

Under the framework of PH models, the PEM appears as an attractive alternative to model the baseline hazard function $h_0(t)$. Assuming the PEM, a more general structure of model (1) can be obtained by allowing the regression coefficients to vary through the intervals induced by the time grid. In this setup, Gamerman (1991) proposes a dynamic Bayesian approach to model time to event data, but a method for specifying accordingly the time grid need to fit the PEM is not established in his work. In this section a twofold extension of the approach introduced by Demarqui et al. (2008) is described. We provide a suitable way to fit the PEM assuming a random time grid when covariates are available, and

also consider the dynamic structure proposed by Gamerman (1991) to add in the model, the correlation among successive intervals.

By assuming a dynamic structure and a random time grid, estimates and predictions are obtained by first conditioning on a particular realization of the random time grid, and then averaging over all possible time grids. Thus, we start the model presentation by reviewing the generalization of model (1) proposed by Gamerman (1991), considering the time grid $\tau = \{s_0, s_1, ..., s_J\}$ fixed. The mechanism used to deal with the randomness of time grid of the PEM is then introduced in Subsection 2.2. In Subsection 2.3 the dynamic structure for the time varying regression coefficients is specified. Finally, in Subsection 2.4 the issue regarding model comparison is addressed.

Following Gamerman (1991), giving a time grid $\tau = \{s_0, s_1, ..., s_J\}$, and setting $\mathbf{x} = (1, x_1, x_2, ..., x_p)$ and $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})'$, with $\lambda_j = \exp\{\beta_{0j}\}$, model (1) becomes:

$$h(t) = \exp\{\mathbf{x}\,\boldsymbol{\beta}_j\}, \tag{2}$$

for $t \in I_j = (s_{j-1}, s_j]$, $j = 1, ..., J$.

It is noticeable that, conditionally on $\tau = \{s_0, s_1, ..., s_J\}$, model (2) has no longer a constant hazard ratio over time. In fact, the constant hazard ratio assumption of the PH model is satisfied only within each interval $I_j = (s_{j-1}, s_j]$, unless the $\boldsymbol{\beta}_j$'s are all equal. However, model (2) presents several advantages over model (1). First, correlated or smoothed hazard functions are, in general, more suitable than hazard functions specified via processes with independent increments. Second, in many practical situations, the assumption of proportional hazards cannot be verified, and accommodation of temporal influences of covariates is not possible for those cases. Third, a dynamic structure to model covariate effects in the hazard function also enables to model time-dependent covariates straightforwardly. Fourth, the class of dynamic models includes the class of static models as a particular case.

2.1 Likelihood function

Suppose we observe an independent sample of size n. Let Y_i be the survival time under study for the i-th element, $i = 1., ..., n$. Assume that there is a right-censoring scheme working independently of the failure process. Denote by C_i the censored time for the the i-th element. The complete information associated to the i-th element is (T_i, δ_i), where $T_i = \min\{Y_i, C_i\}$ and $\delta_i = \mathbf{1}(Y_i \leq C_i)$ are, respectively, the observable survival time, and the failure indicator function.

Denote by $\mathbf{D} = (\mathbf{t},\ \boldsymbol{\delta})$ the observed data set, where $\mathbf{t}' = (t_1, ..., t_n)$ and $\boldsymbol{\delta}' = (\delta_1, ..., \delta_n)$.

In order to properly construct the likelihood function, given $\tau = \{s_0, s_1, ..., s_J\}$ we define:

$$t_{ij} = \begin{cases} s_{j-1}, & \text{if } t_i < s_{j-1}, \\ t_i, & \text{if } t_i \in I_j, \\ s_j, & \text{if } t_i > s_j, \end{cases} \tag{3}$$

where $I_j = (s_{j-1}, s_j]$, $i = 1, ..., n$ and $j = 1, ..., J$. Consider $\delta_{ij} = \delta_i \nu_j^{(i)}$, where $\nu_j^{(i)}$ is the indicator function assuming value 1 if the i-th element experiences an event (either failure or censored) at the j-th interval, and 0 otherwise. Also consider the set $\boldsymbol{\lambda} = \{\lambda_{ij} : \lambda_{ij} = \exp\{\mathbf{x}_i \boldsymbol{\beta}_j\}, i = 1, ..., n, j = 1, ..., J\}$, where λ_{ij} is the hazard function for the i-th element at the j-th interval. Hence, the pair (t_{ij}, δ_{ij}) corresponds to the contribution of the i-th element at the j-th interval, for which the hazard function is given by λ_{ij}.

In the dynamic framework for the PEM proposed by Gamerman (1991), the data information, as well as the parametric evolution, is processed by each interval induced by $\tau = \{s_0, s_1, ..., s_J\}$. In this context, the information contained in the pair (t_{ij}, δ_{ij}) is incorporated into the likelihood function through the conditional piecewise exponential distribution via $\lambda_{ij} \exp\{-\lambda_{ij}(t_{ij} - s_{j-1})\}$, if the i-th element has experienced an event (either failure or censored) at the j-th interval, or via $\exp\{-\lambda_{ij}(t_{ij} - s_{j-1})\}$, otherwise. Then, we have that the complete likelihood function is given by:

$$L_\tau(\boldsymbol{\lambda}; \mathbf{D}) = \prod_{j=1}^{J} \prod_{i=1}^{n} (\lambda_{ij})^{\delta_{ij}} \exp\{-\lambda_{ij}(t_{ij} - s_{j-1})\}. \tag{4}$$

The definition of the survival times given in (3) provides a quite attractive way for constructing the likelihood function of the PEM. Note that the likelihood function for the static PEM (Ibrahim et al.; 2001b) can be easily obtained by exchanging the product order given in (4), and redefining the hazard function given in (2) accordingly, say, assuming $\boldsymbol{\beta}_j = \boldsymbol{\beta}$, $j = 1, ..., J$. Therefore, definition (3) provides a simple and unified notation for both static and dynamic approaches for the PEM.

Finally, we observe that the likelihood function in (4) depends on $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J)'$ only through the link function $\lambda_{ij} = \exp\{\mathbf{x}_i \boldsymbol{\beta}_j\}$, $i = 1, ..., n$ and $j = 1, ..., J$. Furthermore, it naturally factors into a product form, which is a required condition for applying the structure of the PPM to fit the PEM with random time grid. See details in Subsection 2.2.

2.2 Clustering structure

One of the greatest challenges of working with the PEM is to find an appropriate time grid to obtain a well fitted model. It is well known that a time grid with a large number of intervals might provide unstable estimates for the failure rates, whereas a small number of intervals might produce a poor approximation for the true survival function (Sahu et al.;

1997; Qiou et al.; 1999). Therefore, there is a need for seeking to a time grid for the PEM which provides a balance between good approximations for both the hazard and survival functions. The approach proposed by Demarqui et al. (2008) to fit the PEM with random time grid is extended in the following to fit the dynamic model in (2).

Although unnecessary, to guarantee that at least one failure time falls at each interval induced by the random time grid, Demarqui et al. (2008) started the model construction by imposing the time grid associated to the nonparametric approach as the finest possible time grid for the PEM. Then, the structure of the PPM is used to cluster the intervals induced by that initial time grid. We generalize such approach by clustering the intervals induced by a coarser initial time grid. Such grid assumes that intervals containing more than one failure time are possible *a priori*. This allow us to control the maximum number of intervals, and consequently, the maximum number of parameters in the model.

Let $\mathscr{F} = \{0, y_1, ..., y_m\}$ be the set formed by the origin and the m distinct observed failure times from a sample of size n. Consider the time grid $\tau' = \{0, y'_1, ..., y'_{m'}\}$ satisfying $\tau' \subseteq \mathscr{F}$, where m', $1 \leq m' \leq m$, denotes the maximum number of intervals admitted *a priori*. Then, $\tau' = \{0, y'_1, ..., y'_{m'}\}$ defines the following set of disjoint intervals:

$$I_j = \begin{cases} (0, y'_1], & \text{if } j = 1, \\ (y'_{j-1}, y'_j], & 1 < j \leq m'. \end{cases} \tag{5}$$

The degree of parametricity of the model is controlled by m'. As m' gets closer to m, the model becomes more nonparametric. On the other hand, when m' approaches 1, the model becomes more parametric. A completely parametric model is obtained whenever $m' = 1$. In this case, the PEM reduces to the exponential model and no clustering structure is needed.

Denote by $\mathcal{I} = \{1, ..., m'\}$ the set of indexes related to the intervals defined in (5). Let $\rho = \{i_0, i_1, ..., i_b\}$, $0 = i_0 < i_1 < ... < i_b = m'$, be a random partition of $\mathcal{I}$, which divides the m' initial intervals given in (5) into $B = b$ new disjoint intervals. The random variable B denotes the number of clustered intervals related to the random partition ρ. Now, let $\tau_\rho = \tau(\rho) = \{s_0, s_1, ..., s_b\}$ be the time grid induced by the random partition ρ, where

$$s_j = \begin{cases} 0, & \text{if } j = 0, \\ y'_{i_j}, & \text{if } j = 1, \ldots, b, \end{cases} \tag{6}$$

for $b = 1, ..., m'$. Then, it follows from (5) and (6) that the clustered intervals induced by $\rho = \{i_0, i_1, ..., i_b\}$ are given by:

$$I_\rho^{(j)} = \cup_{r=i_{j-1}+1}^{i_j} I_r = (s_{j-1}, s_j], j = 1, ..., b.$$

Observe that we are defining a random time grid for the PEM in terms of a random partition of the intervals given in (5). Besides, since we are considering that only contiguous

intervals are possible, it is reasonable to assume that the endpoint i_j of each clustered interval $I_\rho^{(j)}$ depends only upon the previous endpoint i_{j-1}. Thus, under this assumption, it is unavoidable to assume a product distribution to describe the prior uncertainty about $\rho = \{i_0, i_1, ..., i_b\}$. Then, we assume the following prior distribution for $\rho = \{i_0, i_1, ..., i_b\}$:

$$P(\rho = \{i_0, i_1, ..., i_b\}) = \frac{1}{K} \prod_{j=1}^{b} c_{I_\rho^{(j)}},$$

where $c_{I_\rho^{(j)}} > 0$ denotes the prior cohesion related to the clustered interval $I_\rho^{(j)}$, and represents the degree of similarity among the observations into the intervals I_r, $i_{j-1} < r \leq i_j$, for $r = 1, ..., m'$, and $K = \sum_{\mathscr{C}} \prod_{j=1}^{b} c_{I_\rho^{(j)}}$, with $\mathscr{C}$ denoting the set of all possible partitions of the set $\mathcal{I}$ into b disjoint clustered intervals with endpoints $i_1, ..., i_b$, satisfying the condition $0 = i_0 < i_1 < ... < i_b = n$, for all $b \in \mathcal{I}$. Notice that if $P(\rho = \{i_0, i_1, ..., i_b\}) = 1$ for a particular partition, we return to the usual model that assumes fixed time grid.

Let $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, ..., \beta_{kp})'$ be the vector of regression coefficients associated with the k-th interval defined in (5). Then, conditionally on $\rho = \{i_0, i_1, ..., i_b\}$, we assume that $\boldsymbol{\beta}_k \equiv \boldsymbol{\beta}_\rho^{(j)}$, where $\boldsymbol{\beta}_\rho^{(j)} = (\beta_\rho^{(j0)}, \beta_\rho^{(j1)}, ..., \beta_\rho^{(jp)})'$ denotes the common vector of regression coefficients related to the clustered interval $I_\rho^{(j)}$.

In order to complete the model specification, we need to specify the prior distributions for both $\boldsymbol{\beta}_\rho = (\boldsymbol{\beta}_\rho^{(1)}, .., \boldsymbol{\beta}_\rho^{(b)})'$ and $\boldsymbol{\lambda}_\rho = \{\lambda_\rho^{(ij)} : \lambda_\rho^{(ij)} = \exp\{\mathbf{x}_i \boldsymbol{\beta}_\rho^{(j)}\}, i = 1, ..., n, j = 1, ..., b\}$. For the failure rates $\lambda_\rho^{(ij)}$, we assume conjugate gamma prior distributions $Gamma(\alpha_{ij}, \gamma_{ij})$. Further, we also assume that the failure rates $\lambda_\rho^{(ij)}$ are conditionally independent, given α_{ij} and γ_{ij}. The prior specification for $\boldsymbol{\beta}_\rho$ is made according to the dynamic modeling approach we shall see in the next subsection. In Subsection 2.3, we also show how the hyperparameters α_{ij} and γ_{ij} can be obtained from the prior distributions of the common regression coefficients $\boldsymbol{\beta}_\rho$. This approach introduces temporal correlation between the common parameters of successive intervals.

Hence, under the conditions presented above, conditionally on $\rho = \{i_0, i_1, ..., i_b\}$ we have that the joint distribution of the observations has the following product form:

$$f(\mathbf{D}|\rho) = \prod_{j=1}^{b} f_{I_\rho^{(j)}}(\mathbf{D}), \tag{7}$$

where

$$\begin{aligned} f_{I_\rho^{(j)}}(\mathbf{D}) &= \prod_{i=1}^{n} \frac{\gamma_{ij}^{\alpha_{ij}}}{\Gamma(\alpha_{ij})} \int \left(\lambda_\rho^{(ij)}\right)^{\alpha_{ij}+\delta_{ij}-1} \exp\left\{-\lambda_\rho^{(ij)} \left[\gamma_{ij} + (t_{ij} - a_{j-1})\right]\right\} d\lambda_\rho^{(ij)} \\ &= \prod_{i=1}^{n} \frac{\gamma_{ij}^{\alpha_{ij}}}{\Gamma(\alpha_{ij})} \frac{\Gamma(\alpha_{ij} + \delta_{ij})}{\left[\gamma_{ij} + (t_{ij} - a_{j-1})\right]^{(\alpha_{ij}+\delta_{ij})}} \end{aligned}$$

is named data factor, and corresponds to the marginal distribution of the data associated with the clustered interval $I_\rho^{(j)}$.

Thus, the joint distribution of the observations given in (7) satisfies the product condition required for applying the clustering structure of the PPM. Then, following the structure of the PPM, the posterior distribution of $\rho = \{i_0, i_1, ..., i_b\}$ is:

$$P(\rho = \{i_0, i_1, ..., i_b\}|\mathbf{D}) = \frac{1}{K^*} \prod_{i=1}^{b} c^*_{I_\rho^{(j)}},$$

where $c^*_{I_\rho^{(j)}} = f_{I_\rho^{(j)}} c_{I_\rho^{(j)}}$ denotes the posterior cohesion associated with the j-th clustered interval $I_\rho^{(j)}$, and $K^* = \sum_{\mathscr{C}} \prod_{j=1}^{b} c^*_{I_\rho^{(j)}}$.

The posterior distribution for $\boldsymbol{\beta}_k$, $k = 1, ..., m'$, is given by the following mixture of distributions:

$$\pi(\boldsymbol{\beta}_k|\mathbf{D}) = \sum_{i_{j-1}<k\leq i_j} \pi(\boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}) R(I_\rho^{(j)}|\mathbf{D}), \tag{8}$$

where $R(I_\rho^{(j)}|\mathbf{D})$ is named posterior relevance, and which denotes the probability of each clustered interval $I_\rho^{(j)}$ belongs to the random partition ρ, and $\pi(\boldsymbol{\beta}_\rho^{(j)}|\mathbf{D})$ denotes the posterior (smoothed) distribution of the common parameter $\boldsymbol{\beta}_\rho^{(j)}$, $j = 1, ..., b$, which are obtained in Subsection 2.3. Then, assuming the squared-error loss function, we have that the product estimate for $\boldsymbol{\beta}_k$ is given by:

$$\hat{\boldsymbol{\beta}}_k = \sum_{i_{j-1}<k\leq i_j} E(\boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}) R(I_\rho^{(j)}|\mathbf{D}), \tag{9}$$

for $j = 1, ..., b$ and $k = 1, ..., n$.

Finally, the posterior survival function for a new element, assumed to be independent of the observed data set, and having a vector of covariates x^*, is obtained by averaging over all random partitions $\rho = \{i_0, i_1, ..., i_b\}$ as follows:

$$S(y|\mathbf{D}) = \sum_{\rho} S(y|\mathbf{D}, \rho)\pi(\rho|\mathbf{D}),$$

where

$$\begin{aligned} S(y|\mathbf{D}, \rho) &= \int S(y|\, \boldsymbol{\lambda}_\rho^{(x^*)}, \rho)\pi(\, \boldsymbol{\lambda}_\rho^{(x^*)}|\mathbf{D}) d\, \boldsymbol{\lambda}_\rho^{(x^*)} \\ &= \left(1 + \frac{y - s_{j-1}}{\gamma_j^{(x^*)}}\right)^{-\alpha_j^{(x^*)}} \prod_{r=1}^{j-1} \left(1 + \frac{s_r - s_{r-1}}{\gamma_r^{(x^*)}}\right)^{-\alpha_r^{(x^*)}}, \end{aligned} \tag{10}$$

for $\boldsymbol{\lambda}_\rho^{(x^*)} = (\lambda_\rho^{(x^*;1)}, ..., \lambda_\rho^{(x^*;b)})'$ and $y \in I_\rho^{(j)}$, $j = 1, ..., b$.

2.3 Dynamic modeling approach

In the previous section we showed how the structure of the PPM can be used to fit the PEM with a random time grid. In this section we present an overview of the dynamic modeling approach for the PEM proposed by Gamerman (1991). This approach can be viewed as an extension of the class of dynamic general linear models (DGLMs) proposed by West et al. (1985). All the results derived are obtained after conditioning on $\rho = \{i_0, i_1, ..., i_b\}$.

The analysis of dynamic models is performed sequentially. In order to move from one step to another (i.e., from one clustered interval to another), we first need to specify the evolution system, which defines the way the parametric evolution is performed. Second, we construct the likelihood function at each step, and then we perform the parametric estimation or updating.

As in Gamerman (1991), we assume that the common parameters of different clustered intervals are related through the following evolution equation:

$$\boldsymbol{\beta}_{\rho}^{(j)} = \mathbf{G}_j(\Delta_j)\ \boldsymbol{\beta}_{\rho}^{(j-1)} + \ \boldsymbol{\epsilon}_j, \tag{11}$$

where $\mathbf{G}_j$ is a matrix of known coefficients that defines the systematic evolution of $\boldsymbol{\beta}_{\rho}^{(j)}$ over time, Δ_j is the length of the current interval j, and $\boldsymbol{\epsilon}_j$ is a stochastic error whose distribution has mean zero and covariance matrix $\mathbf{W}_j$. The stochastic error $\boldsymbol{\epsilon}_j$ is further assumed to be independent of time as well as of the observed data set.

In the approach proposed by West et al. (1985), for each step in the dynamic analysis, prior and posterior distributions are partially specified in terms of their moments. Let $\mathbf{D}_{\rho}^{(j-1)}$ be the set of all the available data information up to interval $I_{\rho}^{(j-1)}$. Assume that the posterior distribution for the common vector of regression coefficients $\beta_{\rho}^{(j-1)}$ has mean vector $\mathbf{m}_{j-1}$ and covariance matrix $\mathbf{C}_{j-1}$. Such posterior is denoted by $(\ \boldsymbol{\beta}_{\rho}^{(j-1)}|\mathbf{D}_{\rho}^{(j-1)}) \sim (\mathbf{m}_{j-1}, \mathbf{C}_{j-1})$. Then, using (11), we can show that the prior distribution of $\boldsymbol{\beta}_{\rho}^{(j)}$, given $\mathbf{D}_{\rho}^{(j-1)}$, is

$$(\ \boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D}_{\rho}^{(j-1)}) \sim (\mathbf{a}_j, \mathbf{P}_j),$$

where $\mathbf{a}_j = \mathbf{G}_j(\Delta_j)\mathbf{m}_{j-1}$ and $\mathbf{P}_j = \mathbf{G}_j(\Delta_j)\mathbf{C}_{j-1}\mathbf{G}_j^{'}(\Delta_j) + \mathbf{W}_j$.

Let r_j denote the number of elements that are still at risk at the beginning of the interval $I_{\rho}^{(j)}$. Denote by $\mathbf{D}_{\rho}^{(i-1;j-1)}$ the set of all data information available up to interval $I_{\rho}^{(j-1)}$, along with the processed information associated with the first $i-1$ elements that still at risk at the beginning of interval $I_{\rho}^{(j)}$. Then, assuming $(\ \boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D}_{\rho}^{(i-1;j-1)}) \sim (\mathbf{a}_{ij}, \mathbf{P}_{ij})$, we have that the joint prior distribution for $(\ \boldsymbol{\beta}_{\rho}^{(j)}, \log\lambda_{\rho}^{(ij)})$ is given by:

$$\left[\begin{pmatrix} \boldsymbol{\beta}_{\rho}^{(j)} \\ \log\lambda_{\rho}^{(ij)} \end{pmatrix} |D_{\rho}^{(i-1;j-1)}\right] \sim \left[\begin{pmatrix} \mathbf{a_{ij}} \\ f_{ij} \end{pmatrix}; \begin{pmatrix} \mathbf{P}_{ij} & \mathbf{s}_{ij} \\ \mathbf{s}'_{ij} & q_{ij} \end{pmatrix}\right], \tag{12}$$

where $\mathbf{s}_{ij} = \mathbf{P}_{ij}\mathbf{x}'_i$, $q_{ij} = \mathbf{x}_i\mathbf{s}_{ij}$ and $f_{ij} = \mathbf{x}_i\mathbf{a}_{ij}$.

The prior distribution $(\lambda_\rho^{(ij)}|\mathbf{D}_\rho^{(i-1;j-1)})$ is the gamma distribution $Gamma(\alpha_{ij}, \gamma_{ij})$ assumed in Subsection 2.2. The hyperparameters α_{ij} and γ_{ij} of this distribution are obtained from the moments of $\lambda_{ij} = \exp(\mathbf{x}_i\ \boldsymbol{\beta}_\rho^{(j)})$. Such moments are approximations derived from (12) by using a first order Taylor series expansion, which yields $\alpha_{ij} = q_{ij}^{-1}$ and $\gamma_{ij} = q_{ij}^{-1}\exp\{-f_{ij}\}$.

Once the prior distribution $(\lambda_\rho^{(ij)}|\mathbf{D}_\rho^{(i-1;j-1)}) \sim Gamma(\alpha_{ij}, \gamma_{ij})$ becomes available, the posterior distribution $(\lambda_\rho^{(ij)}|\mathbf{D}_\rho^{(i;j-1)}) \sim Gamma(\alpha_{ij} + \delta_{ij}, \gamma_{ij} + (t_{ij} - s_{j-1}))$ can be obtained straightforwardly by using the information provided by the i-th element at risk in $I_\rho^{(j)}$. Then, applying linear Bayesian methods (West et al.; 1985), we update $(\ \boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}^{(i-1;j-1)}) \sim (\mathbf{a}_{ij}, \mathbf{P}_{ij})$ into the posterior $(\ \boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}^{(i;j-1)}) \sim (\mathbf{m}_{ij}, \mathbf{C}_{ij})$, where

$$\mathbf{m}_{ij} = \mathbf{a}_{ij} + \frac{\mathbf{s}_{ij}}{q_{ij}} \log\left\{\frac{1 + q_{ij}\delta_{ij}}{1 + q_{ij}(t_{ij} - s_{j-1})\exp(f_{ij})}\right\}$$

and

$$\mathbf{C}_{ij} = \mathbf{P}_{ij} - \frac{\delta_{ij}}{1 + q_{ij}\mathbf{s}_{ij}\mathbf{s}'_{ij}}.$$

At the beginning of each clustered interval $I_\rho^{(j)}$ the updating starts by setting $\mathbf{a}_{1j} = \mathbf{a}_j$, $\mathbf{P}_{1j} = \mathbf{P}_j$ and $\mathbf{D}_\rho^{(0,j-1)} = \mathbf{D}_\rho^{(j-1)}$. As there is no evolution within intervals, after processing the information of the i-th element at risk in $I_\rho^{(j)}$, we set $\mathbf{a}_{i+1;j} = \mathbf{m}_{ij}$ and $\mathbf{P}_{i+1;j} = \mathbf{C}_{ij}$ and continue this cycle until the data information associated with the r_j elements is processed. Then, we set $\mathbf{m}_{r_j,j} = \mathbf{m}_j$, $\mathbf{C}_{r_j,j} = \mathbf{C}_j$ and $\mathbf{D}_\rho^{(r_j,j)} = \mathbf{D}_\rho^{(j)}$, and perform the parametric evolution through equation (11), and start a new updating cycle, until all the data information $\mathbf{D}_\rho^{(b)} = \mathbf{D}$ has been processed.

In practice, due to the randomness of the time grid of the PEM, the elicitation of the covariance matrices $\mathbf{W}_j$ in (11) requires an automated mechanism. This can be done by using the concept of discount factor discussed in West and Harrison (1997). Specifically, we assume that $P_j = \frac{1}{\phi}C_{j-1}$, for the discount factor $0 < \phi \leq 1$. Notice that proceeding in such a way, the increase of uncertainty inherent to the parametric evolution is preserved, and the variance matrix $\mathbf{W}_j$ can be easily recovered by setting $\mathbf{W}_j = (\frac{1}{\phi} - 1)C_{j-1}$.

The online distribution $(\ \boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}_\rho^{(j)})$ retains all the required information to perform the parametric evolution over the successive intervals induced by the partition $\rho = \{i_0, i_1, ..., i_b\}$. However, this distribution is obtained using only the data information available up to interval $I_\rho^{(j)}$. Therefore, a posterior distribution of $\boldsymbol{\beta}_\rho^{(j)}$ given all data information $\mathbf{D}$, known as smoothed distribution, is more informative than the online distribution $(\ \boldsymbol{\beta}_\rho^{(j)}|\mathbf{D}_\rho^{(j)})$, and should be preferable for making inferences about the common regression coefficients $\boldsymbol{\beta}_\rho^{(j)}$'s, as well as to predict survival times of future observations. For this reason, these distributions are used to obtain (8), (9) and (10) in Subsection 2.2.

The smooth distributions $(\ \boldsymbol{\beta}_\rho^{(j)}|\mathbf{D})$'s are obtained via a recursive algorithm which redis-

tribute all the data information smoothly for the parameters. This algorithm is an extension of the smoothing algorithm used in standard dynamic linear models (DLMs) proposed by West (1982), and its proof is presented in Gamerman (1991).

Denote by $(\boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D}) \sim (\mathbf{m}_j^{\mathbf{D}}; \mathbf{C}_j^{\mathbf{D}})$ the smoothed distribution of the parameters associated to the interval $I_{\rho}^{(j)}$. Initialize with $\mathbf{m}_b^{\mathbf{D}} = \mathbf{m}_b$ and $\mathbf{C}_b^{\mathbf{D}} = \mathbf{C}_b$. Then, the moments of the smooth distribution $(\boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D})$ are given by the following equations:

$$\mathbf{m}_j^{\mathbf{D}} = E[\boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D}] = \mathbf{m}_j + \mathbf{C}_j\mathbf{G}'_{j+1}(\Delta_{j+1})\mathbf{P}_{j+1}^{-1}(\mathbf{m}_{j+1}^{\mathbf{D}} - \mathbf{a}_{j+1})$$

and

$$\mathbf{C}_j^{\mathbf{D}} = V[\boldsymbol{\beta}_{\rho}^{(j)}|\mathbf{D}] = \mathbf{C}_j - \mathbf{C}_j\mathbf{G}'_{j+1}(\Delta_{j+1})\mathbf{P}_{j+1}^{-1}(\mathbf{P}_{j+1}\mathbf{C}_{j+1}^{\mathbf{D}})\mathbf{G}_{j+1}(\Delta_{j+1})\mathbf{C}_j,$$

for $j = 1, ..., b-1$.

2.4 Model comparison

In Bayesian theory for model comparison (see Kass and Raftery (1995) and Kadane and Lazar (2004), among others), Bayes factor (BF) and posterior model probability (PMP) are standard measures used to compare competing models. In this subsection we describe how these measures can be computed in the framework we are proposing.

Let $\mathcal{M}_{\tau'}$ and $\mathcal{M}_{\rho}$ be the PEM with arbitrary and random time grids, respectively. The marginal distribution $f(\mathbf{D}|\mathcal{M}_{\rho})$ is obtained by firstly computing the marginal distribution of $\mathbf{D}$ conditioning on the partitions, and then averaging over all the partitions, that is:

$$f(\mathbf{D}|\mathcal{M}_{\rho}) = \sum_{\rho} \int L(\boldsymbol{\lambda}_{\rho}, \rho; \mathbf{D})\pi(\boldsymbol{\lambda}_{\rho}|\mathbf{D}, \rho)\pi(\rho|\mathbf{D})d \boldsymbol{\lambda}_{\rho}.$$

The predictive distribution $f(\mathbf{D}|\mathcal{M}_{\tau'})$ is obtained straightforward. Consequenltly, the Bayes factor to compare models $\mathcal{M}_{\tau'}$ and $\mathcal{M}_{\rho}$ becomes:

$$BF(\mathcal{M}_{\rho}, \mathcal{M}_{\tau'}) = \frac{\sum_{\rho} \int L(\boldsymbol{\lambda}_{\rho}, \rho; \mathbf{D})\pi(\boldsymbol{\lambda}_{\rho}|\mathbf{D}, \rho)\pi(\rho|\mathbf{D})d \boldsymbol{\lambda}_{\rho}}{\int L_{\tau'}(\boldsymbol{\lambda}; \mathbf{D})\pi(\boldsymbol{\lambda}|\mathbf{D})d \boldsymbol{\lambda}}.$$

When we are interested in comparing more than two models, say $\mathcal{M}_i$, $i =, ..., M$ and $2 < M < \infty$, the usual procedure consists of computing the posterior model probability of each model $\mathcal{M}_i$:

$$\pi(\mathcal{M}_i|\mathbf{D}) = \frac{f(\mathbf{D}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{i=1}^{M} f(\mathbf{D}|\mathcal{M}_i)p(\mathcal{M}_i)},$$

where $p(\mathcal{M}_i)$ is the prior probability associated to model $\mathcal{M}_i$, satisfying $\sum_{i=1}^{M} p(\mathcal{M}_i) = 1$,

and $f(\mathbf{D}|\mathcal{M}_i)$ is the corresponding marginal distribution of $\mathbf{D}$. Then, one may choose that model with the highest posterior probability among all models under consideration.

3 Analysis Real data sets

In this section we use the model discussed in the previous sections to analyze the survival times (in months) of 231 individuals diagnosed with brain cancer in Windhan-CT, USA, during 1995 to 2004. This data set was obtained from SEER database. Our main interest relies on comparing the inferences that were made using both the PEM with fixed and random time grids, and to perform a sensitivity analysis for different choices of the initial number of time itervals m' and the discount factor ϕ. We also investigate the effect of the risk factor age (in years) on the survival times (in months) of those individuals diagnosed with brain cancer. Another study (not included in this paper) using the proposed model, considering other risk factor for brain cancer such as gender and race, showed that those risk factors are not statistically significant for the particular population under investigation.

For the brain cancer data set we observed that the median survival time is 16.25 months. The percentiles of order 2.5% and 97.5% are 14.42 and 18.93 months, respectively. The mean and the standard deviation for the covariate age is 49.52 and 23.33 years, respectively. For this data set, it was observed a total of 134 failure times, resulting in percentage of failures of the order of 58%. From those total of failure times, we observed 32 distinct failure times. In order to specify the finest time grids τ' admitted for the PEM, we considered m' equals to 20, 25 and 32. Given these values of m', we proceeded as follows. First, we obtained values k and r from the identity $m = km' + r$. Then, we chose the elements of τ' so that the first $m' - r$ intervals had k failure times and the remaining r intervals had $k + 1$ failure times. This procedure can be justified by arguing that it allows more failure times to be in the last intervals, where less data information is available.

Prior elicitation for the regression coefficients $\boldsymbol{\beta}_\rho$ was made as follows. For the regression coefficients of the first clustered interval $I_\rho^{(1)}$ we assumed that, given $\rho = \{i_0, i_1, ..., i_b\}$ $(\beta_\rho^{(1)}|\mathbf{D}_\rho^{(0)}) \sim [0, 100\mathbf{I}_{p+1}]$. It was also assumed a random walk for the parametric evolution through time by setting in (11) $G_\rho^{(j)} = \mathbf{I}_{p+1}$, $j = 1, ..., b$. The covariance matrices $\mathbf{W}_j$ were specified according to the discount factor strategy briefly discussed in Subsection 2.3. Finally, we elicited a discrete uniform prior distribution for $\rho = \{i_0, i_1, ..., i_b\}$ by assuming the following prior cohesions: $c_{I_\rho^{(j)}} = 1$, for $j = 1, ..., b$ and $b \in \mathcal{I}$. A detailed study about other prior cohesions choices can be found in Demarqui et al. (2008).

The Gibbs Sampler algorithm given in Demarqui et al. (2009) was used to sample from the posterior quantities of intereste. Some diagnostic procedures discussed in Cowles and Carlin (1996) were used to check the convergence. Single chains of size 160,000 were considered. Posterior samples of size 15,000 were obtained after considering a burn-in period of 10,000 iterations and a lag of 10 to eliminate correlations. All computational procedures were

implemented in Ox version 5.1 (Doornik; 2007).

Table 1 provides the Bayes factor for the PEM with random time grid over the one with fixed time grid. Three choices of m' and 16 values ϕ were considered, thus, for each m', 32 models were compared. The corresponding posterior probabilities of those models are displayed in Figure 1. The superiority of the PEM with random time grid over the PEM with fixed time grid can be observed by analyzing simultaneously Table 1 and Figure 1. For discount factor $\phi \leq 0.5$, except if $m' = 20$ and $\phi = 0.5$ or $\phi = 0.45$, BF indicates strong ($BF \geq 10$) or decisive ($BF \geq 100$) evidence for the PEM with random time grid. On the other hand, for $\phi \geq 0.7$ the PEM with fixed time grid must be prefered. In this cases, however, Figure 1 shows that the posterior probabilities for PEM with random time grid do not differ substantially of those obtained for models with fixed time grid. It is also noteworthy that, in those scenarios in which the PEM with fixed time grid performs better than the PEM with random time grid, the posterior model probabilities associated with those models are considerably lower than those in which the PEM with random time grid has a superior performance.

Table 1: Bayes Factor (random/fixed time grids) for different choices of m' and ϕ.

ϕ	0.1	0.15	0.2	0.25
$m' = 32$	3.5E+24	7.1E+17	2.9E+13	1.9E+10
$m' = 25$	1.4E+17	1.3E+12	7.8E+08	4.1E+06
$m' = 20$	6.6E+11	1.0E+08	4.4E+05	10663
ϕ	0.3	0.35	0.4	0.45
$m' = 32$	7.7E+07	1.1E+06	40405	2810.8
$m' = 25$	90247	5133.1	559.18	94.71
$m' = 20$	781.29	113.42	26	8.12
ϕ	0.5	0.55	0.6	0.65
$m' = 32$	318.64	51.57	10.78	2.7
$m' = 25$	22.74	6.83	2.4	0.93
$m' = 20$	3.13	1.39	0.67	0.35
ϕ	0.7	0.75	0.8	0.85
$m' = 32$	0.76	0.23	0.08	0.03
$m' = 25$	0.38	0.17	0.08	0.06
$m' = 20$	0.19	0.12	0.08	0.06

Figure 1 also provides some insight of the effect of ϕ in the model fitting. Considering the PEM with fixed time grid in scenario (a) we can observe a smooth increase in the performance of the fitted models until $\phi = 0.75$, followed by a smooth decrease in their performance for the remaining values of ϕ. Similar behavior is shared by the PEM with fixed time grid in scenarios (b) and (c), with the best models obtained when $\phi = 0.65$ and $\phi = 0.55$, respectively. For the PEM with random time grid, the posterior model probabilities also have a unique modal value. The best models are obtained by setting $\phi = 0.50$ in scenario

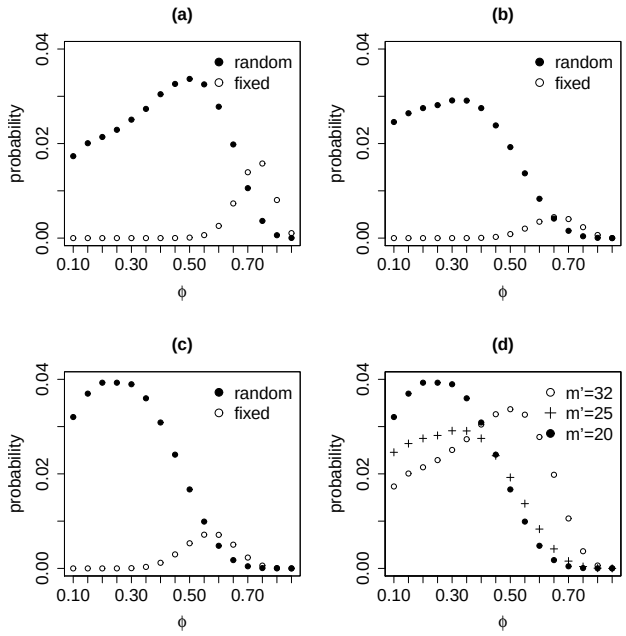

Figure 1: Posterior model probabilities for $m' = 32$ (a), $m' = 25$ (b), $m' = 20$ (c), and random time grid with $m' = 20, 25, 32$, all together (d)

(a), and $\phi = 0.20$ and $\phi = 0.25$ in scenarios (b) and (c), respectively.

Notice that there exist a well defined pattern between the choices of m' and ϕ, and the performance of the fitted models. By reducing m', and consequently, allowing more data information, say, more failure times into each interval, one can observe that the models with both fixed and random time grids perform better if smaller values for ϕ are chosen. As a result of the assumptions made for the random time grids of the PEM, this pattern is even more clear when we look at the PEM with random time grid. A reasonable explanation for this is that, if more data information is allowed to each interval, considering values for ϕ closer to 1 may over smooth the parameters of the model.

For the remaining of this section, we concentrate our attention on making inferences for the quantities of interest based on the PEM with random time grid with $m' = 20$ and $\phi = 0.25$, which corresponds to the most probable model *a posteriori* among those considered in our sensitivity analysis.

One advantage of the PEM with random time grid over the PEM with fixed time grid is that it enables us to make inferences about the time grid and the number of intervals used to fit the PEM. For the current model under investigation, the posterior most probable number of intervals is $B = 9$, with probability 0.241. The estimated 95% HPD intervals for B is $[6, 12]$. There is a significant reduction on the number of intervals when compared with the maximum number of intervals assumed a prior ($m' = 20$, for this case). Other summary measures related to the posterior sample of the number of intervals of the PEM are given in Table 2 and Figure 2.

The estimated mean trajectory of the risk factor age on the survival times of those

Table 2: Summary of sample of the posterior distribution of the number of intervals.

	mean	sd	min	max
	9.335	1.648	4	16
$P_{2.50}$	$P_{25.0}$	$P_{50.0}$	$P_{75.0}$	$P_{97.50}$
6	8	9	10	13

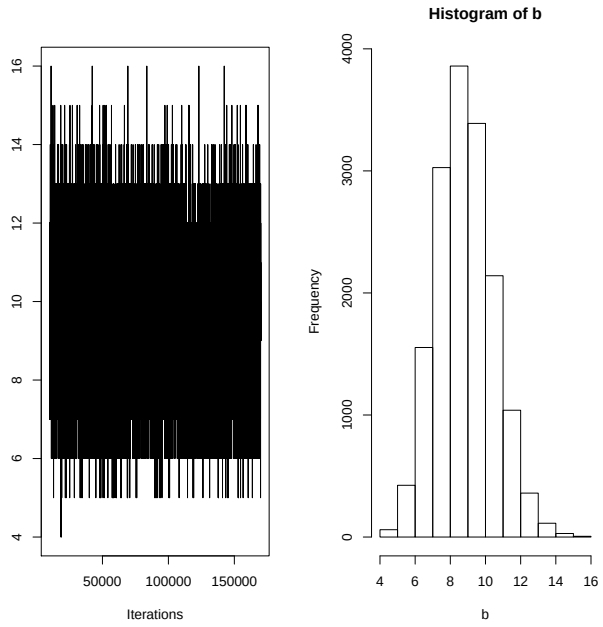

Figure 2: Posterior distribution of the number of intervals associated with the PEM with random time grid with $m' = 20$ and $\phi = 0.25$.

patients diagnosed with brain cancer in Windhan, CT, is shown in Figure 3, along with its corresponding 95% HPD interval. Notice that the effect due to age changes through time, showing a decreasing trend as the time goes by. It is also noteworthy that after a period of approximately 25 months, this effect becomes statistically nonsignificant (as the HPD interval turns to include zero).

In Figure 4 we present the posterior survival curves for two individuals with 10 and 53 years along with the 95% HPD intervals for $S(y|\ \boldsymbol{\lambda}_\rho, \rho, \mathbf{D})$. It is noticeable that those patients are more likely to die in the period of approximate 25 months after the disease is diagnosed. We also can observe from Figure 4 that older patients are considerable likely to die than younger ones.

4 Final Remarks

This paper presented a novel framework to model time to event data using a dynamic piecewise exponential model with random time grid. The issue of modeling the time grid of

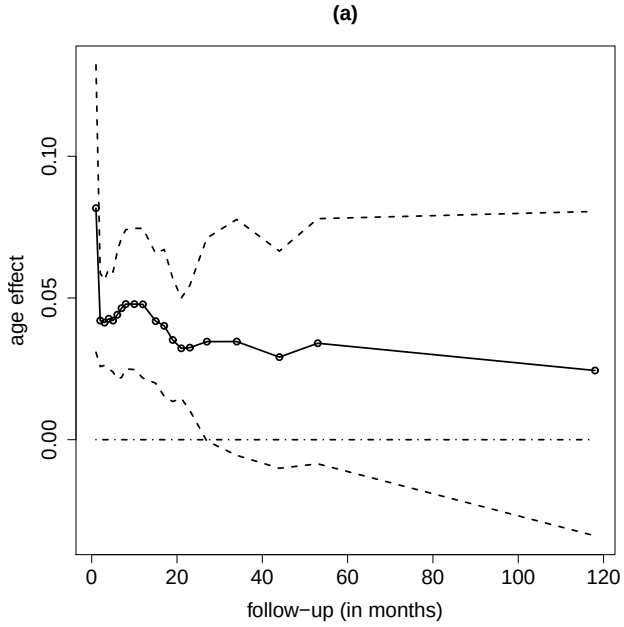

Figure 3: Estimated mean trajectory (solid line) of the effect of age on patients with brain cancer and its corresponding 95% HPD interval (dashed line) provided by the PEM with random time grid assuming $m' = 20$ and $\phi = 0.25$.

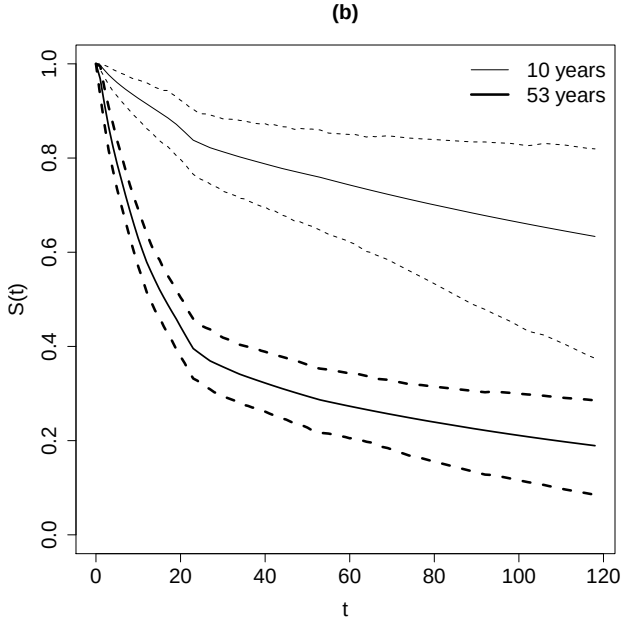

Figure 4: Posterior survival functions for 10 and 53 years old individuals provided by the PEM with random time grid with $m' = 20$ and $\phi = 0.25$.

the PEM was addressed by firstly assuming a finest time grid for the PEM such that only intervals with endpoints given by distinct observed failure times were possible, and then using the clustering structure of the PPM to generate the random time grids. Conditionally on those generated time grids, a correlated process for the common set of parameters (both

the failure rates and the regression coefficients) was also assumed for the successive intervals through time. We further introduced a unified notation suitable for both static and dynamic approaches for the PEM. The proposed model also includes the approaches proposed by Gamerman (1991, 1994) and Demarqui et al. (2008) as special cases. Procedures (Bayes factor and posterior model probabilities) to evaluate the performance of the model were derived, and a sensitivity analysis, using a real data set obtained from the SEER database, was carried out.

The mechanism we considered in this paper to model the time grid of the PEM has clearly advantages over other approaches existing in the literature. By constraining the finest possible time grid for the PEM to have only changing points belonging to a subset of the set of observed failure times, we ensure that all intervals induced by the random time grids contain at least one failure time. This increases considerably the consistency of the estimates of the parameters of interest. Notice that it is not possible to impose such restriction on the intervals generated by the time-homogeneous Poisson process assumed in Arjas and Gasbarra (1994), McKeague and Tighiouart (2000), and Kim et al. (2006). Another nice characteristic of the proposed model is that, although the random time grid of the PEM is assumed to be a random quantity, the final number of parameters of the model remains fixed. In addition, the posterior distributions of those parameters are given by a mixture of posterior distributions of the set of common parameters associated with each clustered interval, weighted according to the generator mechanism of the random time grids. This provides smoother estimates for the parameters of interest and, consequently, leads to more accurate inferences.

The maximum number of intervals admitted *a priori* for the PEM, m', plays a central role in the goodness of fit of the proposed model by determining the degree of parametricity of the model. It also controls the smoothness of the estimates of the parameters along the successive intervals induced by the finest time grid assumed for the PEM by determining the number of failures in those intervals. Moreover, it is particularly useful for handling large data sets. In these situations, many failure times are often available, and a clustering structure over a large number of intervals is usually not needed, and also computationally expensive. In practice, we suggest the user to perform a sensitivity analysis taking into account different values for m'. It is also noticeable that there exist a close relationship between the maximum number of intervals admitted a prior and the choice of the discount factor, indicating that smaller values for the discount should be preferable when the maximum number of intervals assumed for the PEM is small.

The new methodology we introduced is quite general and can be applied in a variety of practical problems. The proposed model is specially suitable for modeling data sets in which the proportional hazard assumption of the well known Cox's model (Cox; 1972) cannot be verified, such as data sets including time-dependent covariates and/or covariate effects changing through time, as the brain cancer data set analyzed in this paper. As expected,

the results obtained from the analysis of this data set suggests that the PEM with random time grid performs better than the PEM with fixed time grid. From the medical point of view, it was verified that age is a significant risk factor for individuals diagnosed with brain cancer. We further observed that the effect of age on the survival time of those patients with brain cancer changes through time according to a decreasing pattern. The study we conducted also showed that there is a high probability of death for patients within the first 20 to 30 months after diagnose, specially for patients above the age of 50 years old.

Future works include the extension of the proposed model to accommodate random effects (or frailties). Survival models with random effects have become quite popular in the literature to model multivariate survival time as well as to introduce geographical information to account for spatial correlation. Other model diagnostic tools such as CPO can also be derived.

References

Aitkin, M., Laird, N. and Francis, B. (1983). A reanalysis of the Stanford heart transplant data (with discussion), *Journal of the American Statistical Association* **78**: 264–292.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, *Statistica Sinica* **4**: 505–524.

Barbosa, E. P., Colosimo, E. A. and Louzada-Neto, F. (1996). Accelerated life tests analyzed by a piecewise exponential distribution via generalized linear models, *IEEE Transactions on Reliability* **45**: 619–623.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems, *Ann Statist* **20**: 260–279.

Bastos, L. S. and Gamerman, D. (2006). Dynamic survival models with spatial frailty, *Lifetime Data Analysis* **12**: 441–460.

Breslow, N. E. (1974). Covariance analysis of censored survival data, *Biometrics* **30**: 89–99.

Chen, M. H. and Ibrahim, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates, *Biometrics* **57**: 43–52.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a survival fraction, *Journal of the American Statistical Association* **94**: 909–919.

Clark, D. E. and Ryan, L. M. (2002). Concurrent prediction of hospital mortality and length of stay from risk factors on admission, *Health Services Research* **37**: 631–645.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostic: a comparative review, *Journal of the American Statistical Association* **91**(No. 434): 883–904.

Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)* **34**: 187–220. with discussion.

Demarqui, F. N., Dey, D. K., Loschi, R. H. and Colosimo, E. A. (2009). A dynamic approach for the piecewise exponential model with random time grid, *Technical Report ID 20*, Department of Statistics, University of Connecticut, Storrs, Connecticut, http://www.stat.uconn.edu/www/research/techreport/2009.php.

Demarqui, F. N., Loschi, R. H. and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model, *Lifetime Data Analysis* **14**: 333–356.

Doornik, J. A. (2007). *Ox 5 - An Object-oriented Matrix Programming Language*, Timberlake Consultants Ltd.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates, **10**: 101–103.

Gamerman, D. (1991). Dynamic Bayesian models for survival data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **40**: 63–79.

Gamerman, D. (1994). Bayes estimation of the piece-wise exponential distribution, *IEEE Transactions on Reliability* **43**: 128–131.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* **82**: 711–732.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001a). Bayesian semiparametric models for survival data with a cure fraction, *Biometrics* **57**: 383–388.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001b). *Bayesian survival analysis*, Springer-Verlag, New York.

Kadane, J. B. and Lazar, N. A. (2004). Methods and criterias for model selection, *Journal of the American Statistical Association* **99**(465): 279–290.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life models, *Biometrika* **60**: 267–278.

Kass, R. E. and Raftery, A. E. (1995). Bayes factor, *Journal of the American Statistical Association* **90**: 773–795.

Kim, J. S. and Proschan, F. (1991). Piecewise exponential estimator of the survival function, *IEEE Transactions on Reliability* **40**: 134–139.

Kim, S., Chen, M. H., Dey, D. K. and Gamerman, D. (2006). Bayesian dynamic models for survival data with a cure fraction, *Lifetime Data Analysis* **13**: 17–35.

McKeague, I. W. and Tighiouart, M. (2000). Bayesian estimators for conditional hazard functions, *Biometrics* **56**: 1007–1015.

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

Sahu, S. K., Dey, D. K., Aslanidu, H. and Sinha, D. (1997). A weibull regression model with gamma frailties for multivariate survival data, *Lifetime Data Analysis* **3**: 123–137.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, Springer-Verlag, New York.

West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting (with discussion), *Journal of the American Statistical Association* **80**: 73–97.

A full semiparametric Bayesian approach for modeling survival data with cure fraction

Fabio N. Demarqui[1], Dipak K. Dey[2], Rosangela H. Loschi[1],
Enrico A. Colosimo[1]
[1] Universidade Federal de Minas Gerais, Brazil
[2]University of Connecticut, USA

Abstract

In this paper we consider a piecewise exponential model (PEM) with random time grid to develop a full semiparametric Bayesian cure rate model. An elegant mechanism enjoying several attractive features for modeling the randomness of the time grid of the PEM is assumed. To model the prior behavior of the failure rates of the PEM we assume a structural modeling approach which allows us to control the degree of parametricity in the right tail of the survival curve. Some properties of the proposed model are discussed. In particular, we consider improper non-informative prior distributions for the regression coefficients and state conditions under which the resulting posterior distribution is proper. We further develop an efficient collapsed Gibbs sampler algorithm for carrying out posterior computation. A Bayesian diagnostic method for accessing goodness of fit and perform model comparisons is briefly discussed. Finally, we illustrate the usefulness of the new methodology with the analysis of a melanoma clinical trial which has been discussed in the literature.

Keywords: Bayesian inference, Gibbs sampler, piecewise exponential model, survival analysis.

1 Introduction

Due to the recent medical advances and the increasing number of early diagnosis experienced specially in the last two decades, studies involving several types of cancer including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma and head and neck cancers, have shown that a significant fraction of the population has been cured. Therefore, survival models incorporating a fraction of cured individuals in the population, known as cure rate models or long-term survival models, have received considerable attention in the literature.

Despite of the recent popularity of the cure rate models, the concept of cure fraction was introduced almost 60 years ago by Berkson and Gage (1952). In the original formulation of the model, Berkson and Gage (1952) assumed that a fraction of the population is cured. Further, the survival times of the remaining non-cured individuals of the population are modeled according to some distribution function F^*. In this fashion, the survival function for the entire population, denoted by $S^*_{pop}(t)$, is given by:

$$S^*_{pop}(t) = \pi + (1 - \pi)S^*(t), \tag{1}$$

where $S^*(t) = 1 - F^*(t)$.

The binary mixture model proposed by Berkson and Gage (1952) has played a major role in the literature on cure rate models, and it has been discussed by many authors, including Farewell (1982, 1986), Kuk and Chen (1992), Maller and Zhou (1996), Peng and Dear (2000) and Sy and Taylor (2000), among others. An alternative parametric cure rate model, known as promotion time cure rate model, is first discussed in Yakovlev et al. (1993) and Chen et al. (1999). In this approach, which has a strong biological appeal, a latent random variable N representing the number of competing causes (or risks) which may lead to the occurrence of the event of interest is introduced into the model. Given N, it is further considered that Z_k, $k = 1, ..., N$, as non-observable random variables denoting the time-to-event (the promotion time) associated with the k-th competing cause. The Z_k's are assumed to be independent of N, and also independent and identically distributed random variables with a distribution F. The number of competing causes as well as their corresponding promotion times are not observed, and the observable survival time is defined as the minimum promotion time.

A Bayesian formulation of the promotion time cure rate model, along with a detailed discussion about its relationship with the mixed cure rate model proposed by Berkson and Gage (1952), is given in Chen et al. (1999). Latter, Tsodikov et al. (1998) introduce the concept of bounded cumulative hazard (BCH) models, which contains the promotion time cure rate model proposed by Yakovlev et al. (1993) as a particular case. An excellent discussion of all those models can be found in Tsodikov et al. (2003). Unified approaches considering both the mixture and promotion time cure rate models are given in Yin and Ibrahim (2005), Rodrigues et al. (009a) and Rodrigues et al. (009b).

The way the hazard functions associated with either F^* or F is modeled is of major importance in the context of the cure rate models, since those are closely related to the survival function of the entire population. The exponential and Weibull distributions are common choices for modeling those hazard functions. However, those distributions are restrictive with respect to the shape of their hazard functions. A quite attractive alternative model to address this issue is to use the Piecewise Exponential Model (PEM) to deal with the problem. Although parametric in a strict sense, the PEM can be thought as a nonparametric model as far as it does not have a closed form for the hazard function. This nice

characteristic of the PEM allows us to use this model to approximate satisfactorily hazard functions of several shapes, providing great flexibility in survival data modeling.

In order to construct the PEM we need to specify a time grid which divides the time axis into a finite number b of intervals, say $\tau = \{a_0, a_1, ..., a_b\}$. Given $\tau = \{a_0, a_1, ..., a_b\}$, we assume a constant failure rate λ_j for each interval $I_j = (a_{j-1}, a_j]$, $j = 1, ..., b$. Consequently, we obtain a discrete version, in the form of a step function, of the true and unknown hazard function. As a result, a good approximation for the true hazard function depends on a suitable choice of the time grid. For this reason, appropriate specifications for the time grid τ plays one of the greatest challenges in working with the PEM.

In the context of the alternative cure rate model, Ibrahim et al. (2001a) consider a hierachical Bayesian approach for modeling the baseline hazard function of the PEM by introducing a smoothing parameter which controls the degree of parametricity in the right tail of the survival distribution. However, in that approach an arbitrary time grid is assumed for the PEM. Extending some previous works, Kim et al. (2006) consider a time-homogeneous Poisson process (THPP) to generate random time grids for the PEM. Similarly to Ibrahim et al. (2001a), conditionally on the generated time grids, Kim et al. (2006) further specify a prior process which controls the degree of parametricity in the right tail of the survival curve by assuming a correlated process for the baseline hazard of the PEM. Although the randomness of both b and the time jumps a_j's yields a nonparametric hazard function, the mechanism used by Kim et al. (2006) to generate the random time grid $\tau = \{a_0, a_1, ..., a_b\}$, permits no control of how the failure times are distributed over the random intervals $I_j = (a_{j-1}, a_j]$.

In this paper we propose an alternative semiparametric full Bayesian approach for modeling the promotion time cure rate model introduced by Chen et al. (1999) assuming the PEM with random time grid. Extending the approaches proposed by Ibrahim et al. (2001a) and Demarqui et al. (2008), we consider in the context of cure rate models a unified approach for modeling the time grids for the PEM which depends only upon a set of observed failure times. This allows us to model the time grid taking into account the arrangement of the observed failure times over the time axis, as well as to control the maximum number of failure times associated with all the random intervals induced by the random time grid. Besides, although time grid to fit the PEM is considered a random quantity, under the proposed model the number of parameters remains fixed, and the posteriors distributions of the failure rates are given by a mixture of distributions, providing smoother estimates for the failure rates.

This paper is organized as follows. In Section 2 we present an overview of the promotion time cure rate model as discussed in Chen et al. (1999). The proposed model is presented in Section 3. A Bayesian diagnostic method for accessing goodness of fit and perform model selection is considered in Section 4. The analysis of the E1673 data from a melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG) using the proposed

methodology is presented in Section 5. Some conclusions about the proposed model in Section 6 ends the paper.

2 The Promotion Time Cure Rate Model

Suppose that for each individual belonging to a sample of size n we observe the triple $(t_i, \delta_i, \mathbf{x}_i)$, where t_i denotes the observed survival time, which may be right censored, δ_i is the censoring indicator, assuming value 1 if t_i is a failure time, and 0 otherwise, and $\mathbf{x}_i' = (x_{i1}, ..., x_{ip})$ is the vector of covariates associated with the i-th individual, $i = 1, ..., n$. According to Chen et al. (1999), the promotion time cure rate model can be built based on the following biologic motivation. Let N_i be the non-observable number of metastasis-competent tumor cells which remains active in the i-th individual after the initial treatment. Assume that N_i follows a Poisson distribution with mean $\theta_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ is the corresponding vector of regression coefficients. Further, denote by Z_{ik} the non-observable random time for the production of detectable metastatic disease by the k-th metastasis-component tumor cell associated with the i-th individual, that is, the promotion time for the k-th tumor cell of the i-th individual. Given N_i, the promotion times Z_{ik}, $k = 1, ..., N_i$, are assumed to be independent and identically distributed with a common c.d.f. $F(t|\psi)$, which depends only on the parameter (possibly vector valued) ψ. The time to relapse to cancer of the i-th individual, which is in fact the observable quantity, is defined as the random variable $T_i = \min\{Z_{ik}, 0 \leq k \leq N_i\}$, with $P(Z_{i0} = \infty) = 1$.

It can be shown that, given $N_i > 0$, the conditional survival and hazard functions of T_i are, respectively, given by

$$h(t|\psi, N_i > 0) = N_i h(t|\psi), \tag{2}$$

and

$$S(t|\psi, N_i > 0) = S(t|\psi)^{N_i}. \tag{3}$$

The conditional hazard function given in (2) has a biological appeal, since the number of metastasis-competent tumor cells acts multiplicatively on (2), suggesting that the higher the N_i in the body, the lower will be the chance of cure.

Using (3), it follows that the populational survival function for the i-th individual can be expressed as

$$\begin{aligned} S_{pop}(t) &= P(\text{no metastatic cancer by time } t) \\ &= \sum_{k=0}^{\infty} P(N_i = k) S(t)^k = \exp\{-\theta_i F(t|\psi)\}. \end{aligned} \tag{4}$$

Some comments about (4) are in order. By taking $\theta_i \to \infty$, the cure fraction tends to 0, whereas if θ_i approaches 0, the cure fraction gets closer to 1. Furthermore, when $t \to \infty$, $S_{pop}(t)$ converges to $P(N_i = 0) = \exp(-\theta_i)$, which can be interpreted in this formulation as the cure fraction of the mixture cure rate model proposed by Berkson and Gage (1952). Therefore, the survival function (4) is not properly a survival function.

After some algebraic calculations, it can be shown that the populational hazard function is given by:

$$h_{pop}(t) = \theta_i f(t|\psi), \tag{5}$$

where $f(t|\psi) = \frac{d}{dt}F(t|\psi)$. Therefore, the effect of covariates is multiplicative in (5), and the proportional hazard structure holds for the populational hazard function. Such structure is quite attractive in survival analysis, since it provides an appealing interpretation of the covariate effects on the hazard function, and is also very convenient from the computational point of view. Indeed, as shown in Rodrigues et al. (009a), the proportional hazard structure in (5) holds if, and only if, the latent random variables N_i's follow a Poisson distribution.

Let $D_{obs} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{X})$ and $D = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{X}, \mathbf{N})$ be the observed and complete data sets, respectively, where $\mathbf{t} = (t_1, ..., t_n)'$, $\boldsymbol{\delta} = (\delta_1, ..., \delta_n)'$, $\mathbf{X}$ is the $n \times p$ matrix of covariates with i-th row $\mathbf{x}_i' = (x_{i1}, ..., x_{ip})$, and $\mathbf{N} = (N_1, ..., N_n)'$ is the corresponding vector of latent variables. Then, the complete data likelihood function for the cure rate model can be written as:

$$\begin{aligned} L(\beta, \psi|D) &= \left(\prod_{i=1}^{n} [N_i h(t_i|\psi)]^{\delta_i} S(t_i|\psi)^{N_i} \right) \\ &\times \exp\left\{ \sum_{i=1}^{n} [N_i(\mathbf{x}_i'\beta) - log(N_i!) - \exp(\mathbf{x}_i'\beta)] \right\}, \end{aligned} \tag{6}$$

and the likelihood function based only on the observed data is obtained by summing out $\mathbf{N}$ in expression (6), that is,

$$L(\beta, \psi|D_{obs}) = \sum_{\mathbf{N}} L(\beta, \psi|D) = \prod_{i=1}^{n} [\theta_i f(t_i|\psi)]^{\delta_i} \exp\{-\theta_i[1 - S(t_i|\psi)]\}. \tag{7}$$

As a final remark on the alternative cure rate described above, we emphasize that the latent variables N_i and $\mathbf{Z}_i = (Z_{i1}, ..., Z_{iN_i})'$ do not necessarily need to have any physical or biological interpretation as long as they are introduced into the model only to facilitate the model formulation and computational implementation as well. Therefore, despite its biological motivation, the alternative cure rate model is suitable for modeling any type of time-to-event data whenever the cure fraction assumption is reasonable and the survival times T_i can be thought of as a result of observing an unknown number N_i of competing

risks $Z_{i1}, ..., Z_{iN_i}$, $i = 1, ..., n$.

3 Full Semiparametric Bayesian Cure Rate Model

In the sequel, we present our formulation for the cure rate model addressed in the previous section. The PEM with random time grid will be assumed for modeling the distribution of the promotion times Z_{ik}'s. We emphasize here that, although the use of the PEM with random time grid to fitting cure rate models has been considered previously by Kim et al. (2006), our approach differs fundamentally from that by the way such random time grids are generated, as can be seen in the following.

3.1 Random Time Grid Modeling

We start our model formulation by specifying the finest time grid for the PEM, say τ', with possibly a moderate to large number of intervals. Then we introduce a clustering structure on the set of intervals induced by τ' in order to get a more parsimonious model. Our approach is full Bayesian since it avoids the use of ad-hoc procedures for choosing the time grid τ. In addition, by allowing the direct modeling of uncertainty about τ through suitable prior, we get a more flexible cure rate model.

The finest time grid is chosen in such a way that only distinct observed failure times are allowed to belong to τ'. This constrain enables us to have total control over the sample space of the random time grids, turning possible its estimation. It further guarantees that all random intervals induced by the random time grids contain at least one failure time, which, as pointed out before, consists of a desired property of τ.

Denote by $\mathscr{F} = \{y_1, ..., y_m\}$ the set of the m distinct observed failure times from a sample of size n. Let $\tau' = \{y'_1, ..., y'_{m'}\}$ be the initial time grid satisfying $1 \leq m' \leq m$ and $\tau' \subseteq \mathscr{F}$, where m' is the maximum number of initial intervals admitted *a priori*. Then, the time grid $\tau' = \{y'_1, ..., y'_{m'}\}$ induces the following set of disjoint intervals

$$I_j = \begin{cases} (0, y'_1], & \text{if } j = 1, \\ (y'_{j-1}, y'_j], & 1 < j \leq m'. \end{cases} \tag{8}$$

The maximum number of initial intervals m' plays an important role in the goodness of fit of the PEM, since it controls the degree of parametricity of the model. As m' gets closer to m, the model becomes more nonparametric. On the other hand, when m' approaches to 1, the model becomes more parametric. Particularly, when $m' = 1$ we have a completely parametric model as the PEM reduces to the Exponential model. Consequently, no clustering structure is needed in this case. In practice, we suggest performing a sensitive analysis considering different values for m'.

In this paper we propose the intervals associated with the finest time grid can be specified as follows. For a given m' we obtain values k and r such that $m = km' + r$. Then, we chose the elements of τ' such that the first $m' - r$ intervals have k failure times and the remaining r intervals have $k + 1$ failure times. This procedure allows more failure times to be in the last intervals, where few data information is usually available.

After $\tau' = \{y'_1, ..., y'_{m'}\}$ is properly specified, the clustering structure is performed as follows. Denote by $\mathcal{I} = \{1, ..., m'\}$ the set of indexes related to the intervals defined in (8). Let $\rho = \{i_0, i_1, ..., i_b\}$, $0 = i_0 < i_1 < ... < i_b = m'$, be a random partition of the set of indexes $\mathcal{I}$, which divides the m' initial intervals given in (8) into $B = b$ new disjoint intervals, where B is a random variable denoting the number of clustered intervals associated with the random partition ρ. Further, let $\tau = \tau(\rho) = \{a_0, a_1, ..., a_b\}$ be the time grid induced by the random partition ρ, where

$$a_j = \begin{cases} 0, \text{ if } j = 0, \\ y'_{i_j}, \text{ if } j = 1, ..., b. \end{cases} \tag{9}$$

Then, the clustered intervals induced by $\rho = \{i_0, i_1, ..., i_b\}$ are given by

$$I_\rho^{(j)} = \cup_{r=i_{j-1}+1}^{i_j} I_r = (y'_{r-1}, y'_r], j = 1, ..., b. \tag{10}$$

To properly construct the likelihood function, conditionally on $\rho = \{i_0, i_1, ..., i_b\}$, we assume that

$$h(t) = \lambda_r \equiv \lambda_\rho^{(j)}, \tag{11}$$

where $\lambda_\rho^{(j)}$ denotes the common failure rate related to the clustered interval $I_\rho^{(j)}$, for $i_{j-1} < r \leq i_j$, $r = 1, ..., m'$ and $j = 1, ..., b$. We further define

$$t_{ij} = \begin{cases} a_{j-1}, & \text{if } t_i < a_{j-1}, \\ t_i, & \text{if } t_i \in I_\rho^{(j)}, \\ a_j, & \text{if } t_i > a_j, \end{cases} \tag{12}$$

for $i = 1, ..., n$ and $j = 1, ..., b$.

Thus, given a particular random partition $\rho = \{i_0, i_1, ..., i_b\}$, survival function associated with the PEM can be written as

$$S(t|\boldsymbol{\lambda}_\rho, \rho) = \exp\left\{ -\sum_{j=1}^{b} \lambda_\rho^{(j)} (t_{ij} - a_{j-1}) \right\}, \tag{13}$$

where $\boldsymbol{\lambda}_\rho = (\lambda_\rho^{(1)}, ..., \lambda_\rho^{(b)})'$ is the vector of clustered failure rates. Thus, replacing (11) and

(13) in the likelihood function given in (6), the complete data likelihood function reduces to:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D) &= \prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{\eta_j} \exp\left\{-\xi_j \lambda_\rho^{(j)}\right\} \\ &\times \exp\left\{\sum_{i=1}^{n} [\nu_i \log(N_i) + N_i(\mathbf{x}_i'\boldsymbol{\beta}) \right. \\ &\left. -log(N_i!) - \exp(\mathbf{x}_i'\boldsymbol{\beta})]\right\}, \end{aligned} \tag{14}$$

where $\delta_{ij} = \delta_i \nu_j^{(i)}$, $\nu_j^{(i)}$ is the indicator function assuming value 1 if the i-th observation falls at the j-th interval, and 0 otherwise, $\eta_j = \sum_{i=1}^{n} \delta_{ij}$ and $\xi_j = \sum_{i=1}^{n} N_i(t_{ij} - a_{j-1})$, for $j = 1, ..., b$. Then, the likelihood based on the observed data D_{obs} is obtained after summing out N in (14), yielding

$$\begin{aligned} L(\beta, \lambda_\rho, \rho | D_{obs}) &= \exp\left\{\sum_{i=1}^{n} \nu_i(\mathbf{x}_i'\beta) + \sum_{i=1}^{n}\sum_{j=1}^{b} \left[\delta_{ij} \log \lambda_\rho^{(j)} - \nu_i \lambda_\rho^{(j)}(t_{ij} - a_j)\right]\right. \\ &+ \left.\sum_{i=1}^{n} \exp(\mathbf{x}_i'\beta) \left[1 - \exp\left(-\sum_{j=1}^{b} \lambda_\rho^{(j)}(t_{ij} - a_j)\right)\right]\right\}. \end{aligned} \tag{15}$$

3.2 Prior specifications

In order to complete the model specification, we need to specify the prior distributions for the parameters of interest, $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho)$. The prior distribution for $(\boldsymbol{\lambda}_\rho, \rho)$ will be assigned hierarchically by first specifying a prior distribution for the random partition ρ, and then eliciting a joint prior distribution for $\boldsymbol{\lambda}_\rho$, conditioning on ρ. The joint prior distribution for $\boldsymbol{\beta}$ will be set independently of $(\boldsymbol{\lambda}_\rho, \rho)$.

The random time grid for the PEM is defined in terms of random partitions of the intervals given in (8). Moreover, we assumed that only contiguous intervals are possible, and that the endpoint i_j of each clustered interval $I_\rho^{(j)}$ depends only upon the previous endpoint i_{j-1}. Therefore, the prior distribution for ρ can be written as the product prior distribution proposed by Barry and Hartigan (1992):

$$\pi(\rho = \{i_0, i_1, ..., i_b\}) = \frac{1}{K} \prod_{i=1}^{b} c_{I_\rho^{(j)}}, \tag{16}$$

where $c_{I_\rho^{(j)}} > 0$ denotes the prior cohesion related to the clustered interval $I_\rho^{(j)}$, and represents the degree of similarity among the initial intervals which compose $I_\rho^{(j)}$, $j = 1, ..., b.$, where $\mathscr{C}$ denotes the set of all possible partitions of the set $\mathcal{I}$ into b disjoint clustered intervals with endpoints $i_1, ..., i_b$, satisfying the condition $0 = i_0 < i_1 < ... < i_b = m'$, for all $b \in \mathcal{I}$.

The gamma family of distributions has been widely used as prior distributions for the failure rates of the PEM. It is a flexible class which also includes non-informative prior

distributions and from which correlated process can be easily constructed (see Arjas and Gasbarra (1994), Gamerman (1994) and McKeague and Tighiouart (2000), among others). We consider a structural prior distribution for $\boldsymbol{\lambda}_\rho$ by assuming independent Gamma prior distributions for the failure rates $\lambda_\rho^{(j)}$'s, with mean and variance given, respectively, by

$$\mu_j = E(\lambda_\rho^{(j)}|\lambda_0, \rho) = \frac{H_0(a_j) - H_0(a_{j-1})}{a_j - a_{j-1}} \tag{17}$$

and

$$\sigma_j^2 = V(\lambda_\rho^{(j)}|\lambda_0, \rho) = \mu_j \kappa^j, \tag{18}$$

where κ, $0 < \kappa < 1$, is a smoothing parameter which controls the degree of parametricity in the right tail of the survival curve, and $H_0(\cdot|\lambda_0)$ is a cumulative hazard function of a parametric survival distribution $F_0(\cdot|\lambda_0)$. The function $F_0(\cdot|\lambda_0)$ is arbitrarily chosen, and must reflect someone's opinion about the right tail of the survival curve.

The prior specification described above was introduced by Ibrahim et al. (2001a) and has some nice properties. Note that, if $\kappa \to 0$, then $\sigma^2 \to 0$, and the prior process approximates $h_0(t|\lambda_0) = \frac{d}{dy} H_0(t|\lambda_0)$ as $h_0(t|\lambda_0) \approx \frac{H_0(a_j) - H_0(a_{j-1})}{a_j - a_{j-1}}$, for $t \in I_j = (a_{j-1}, a_j]$, $j = 1, ..., b$. It is also noticeable that when $j \to \infty$, $\sigma^2 \to 0$ at the rate κ, implying more parametricity in the right tail of the survival curve, regardless of any choice of κ. Moreover, as shown in Ibrahim et al. (2001a), by assuming $(a_{j-1} + a_j)/2 \to t$ as $a_j - a_{j-1} \to 0$, it follows that $E(\lambda_\rho^{(j)}|\lambda_0, \rho) \to h_0(t|\lambda_0)$.

The model is completely specified by eliciting the prior distributions for $\boldsymbol{\beta}$ and λ_0. In this spirit, we will assume that

(i) $\boldsymbol{\beta}$ has the improper prior distribution $\pi(\boldsymbol{\beta}) \propto 1$, and

(ii) λ_0 is independent of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_\rho$, and has a prior distribution $\pi(\lambda_0)$ specified according to the choice of $F_0(\cdot|\lambda_0)$.

Under these assumptions, the joint prior distribution for $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho)$ is

$$\pi(\beta, \lambda_\rho, \lambda_0, \rho) \propto \left[\prod_{j=1}^{b} \left(\left(\lambda_\rho^{(j)}\right)^{\mu_j \kappa^{-j} - 1} \exp\left\{-\kappa^{-j} \lambda_\rho^{(j)}\right\} \right) c_\rho^{(j)} \right] \pi(\lambda_0). \tag{19}$$

Although such an approach has been previously considered before by Ibrahim et al. (2001a), as far as we know, an approach considering simultaneously this structural prior for the components of $\boldsymbol{\lambda}_\rho$ along with a random time grid for the PEM has not been proposed yet. In particular, our model includes as a special case the model proposed in Ibrahim et al. (2001a) by setting $P(\rho = \{i_0, i_1, ..., i_b\}) = 1$ for a particular partition ρ. Moreover, the proposed model differs fundamentally from Kim et al. (2006), which also consider a random time grid for the PEM to fit the cure rate model reviewed in Section 2, in two directions. First, our approach provides an elegant way for modeling the randomness of the time grid

of the PEM through the direct specification of suitable prior distributions for ρ, whereas, by assuming a THPP for modeling the PEM with random time as in Kim et al. (2006), such time grid cannot be estimated. Second, although both approaches consider a cumulative risk function associated to a parametric model to build the prior elicitation, in our approach prior distributions are directly assigned for the failure rates of the PEM, and in the model considered in Kim et al. (2006) prior distributions are dynamically specified for the logarithm of the failure rates of the PEM.

3.3 Posterior Inference and Computations

Considering the likelihood in (15) and the prior specifications provided in Section 3.2, the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho)$ becomes

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho | D_{obs}) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D_{obs}) \pi(\boldsymbol{\lambda}_\rho | \lambda_0, \rho) \pi(\lambda_0) \pi(\rho) \\
&\propto \exp\left\{ \sum_{i=1}^{n} \delta_i (\mathbf{x}_i' \beta) + \sum_{i=1}^{n} \sum_{j=1}^{b} \left[\delta_{ij} \log \lambda_\rho^{(j)} - \delta_i \lambda_\rho^{(j)} (t_{ij} - a_j) \right] \right. \\
&+ \left. \sum_{i=1}^{n} \exp(\mathbf{x}_i' \beta) \left[1 - \exp\left(- \sum_{j=1}^{b} \lambda_\rho^{(j)} (t_{ij} - a_j) \right) \right] \right\} \\
&\times \left(\prod_{j=1}^{b} \left(\lambda_\rho^{(j)} \right)^{\mu_j \kappa^{-j} - 1} \exp\left\{ -\kappa^{-j} \lambda_\rho^{(j)} \right\} c_{I_\rho^{(j)}} \right) \pi(\lambda_0). \qquad (20)
\end{aligned}$$

Since we consider a improper prior distribution for $\boldsymbol{\beta}$, it is necessary to prove that the joint posterior distribution for $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho)$ is proper. In the following theorem we state conditions under which such property of the posterior distribution in (20) is satisfied.

Theorem: *Let X^* be an $m \times p$ matrix with the i-th row given by $\delta_i \mathbf{x}_i'$ and consider the prior specifications for ρ and $\boldsymbol{\lambda}_\rho$ discussed in the previous section. If i) when $\delta_i = 1$, then $t_i > 0$; ii) $m \geq p$ and X^* is of full rank; and iii) the prior distribution $\pi(\lambda_0)$ in (19) is proper, regardless of the choice of $F_0(\cdot|\lambda_0)$, then, the joint posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho | D_{obs})$ is proper.*

The proof of this theorem is given in the appendix. The conditions established in the previous theorem are, in practice, satisfied for most data sets. This enables us to perform noninformative Bayesian inference for the regression coefficients as well as to perform comparisons with the maximum likelihood approach.

In order to sample from (20), we introduce the latent variables $\mathbf{N} = (N_1, ..., N_n)$. Thus,

from (14) we obtain the corresponding joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, \mathbf{N})$:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, \mathbf{N}|D_{obs}) &\propto \left(\prod_{j=1}^{b} \left(\lambda_\rho^{(j)}\right)^{\eta_j+\mu_j\kappa^{-j}} \exp\left\{-[\xi_j+\kappa^{-j}]\lambda_\rho^{(j)}\right\} c_{I_\rho^{(j)}}\right) \\ &\times \exp\left\{\sum_{i=1}^{n}[N_i(\mathbf{x}_i'\boldsymbol{\beta}) - log(N_i!) - \exp(\mathbf{x}_i'\boldsymbol{\beta})]\right\}\pi(\lambda_0). \quad (21)\end{aligned}$$

It is noteworthy that the posterior distribution in (21) does not have a closed form. Hence, MCMC methods are required to carry out our Bayesian analysis. We will consider a collapsed Gibbs sampler to generate from the posterior full conditional (p.f.c.) distributions $(\boldsymbol{\beta}, \mathbf{N}|\boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs})$ and $(\lambda_\rho, \lambda_0, \rho|\boldsymbol{\beta}, \mathbf{N}, D_{obs})$. This strategy helps to improve the performance of the MCMC algorithm by reducing the computational time as well as the correlation among the parameters *a posteriori*. Further details regarding this procedure can be found in Chen et al. (1999), Ibrahim et al. (2001b), and references therein.

To sample from $(\boldsymbol{\beta}, \mathbf{N}|\boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs})$, we consider the p.f.c. distributions $(\boldsymbol{\beta}|\boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs})$ and $(\mathbf{N}|\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs})$, which are given, respectively, by

$$\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs}) \propto \exp\left\{\sum_{i=1}^{n}\left[\delta_i(\mathbf{x}_i'\boldsymbol{\beta}) - \exp(\mathbf{x}_i'\boldsymbol{\beta})\left(1 - \exp\left\{-\sum_{j=1}^{b}\lambda_\rho^{(j)}(t_{ij} - a_{j-1})\right\}\right)\right]\right\}. \quad (22)$$

and

$$(N_i|\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs}) \sim Poisson\left(\exp\left\{\mathbf{x}_i'\boldsymbol{\beta} - \sum_{j=1}^{b}\lambda_\rho^{(j)}(t_{ij} - a_{j-1})\right\}\right) + \delta_i, \quad (23)$$

for $i = 1, ..., n$.

It can be shown that $\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}_\rho, \lambda_0, \rho, D_{obs})$ is log-concave in each component of $\boldsymbol{\beta}$. Thus, the adaptive rejection sampling algorithm (Gilks and Wild (1992)) can be used to sample from $(\boldsymbol{\beta}|\boldsymbol{\lambda}_\rho, \rho, D_{obs})$. Sampling from $(\mathbf{N}|\beta, \lambda_\rho, \rho, D_{obs})$ is straightforward.

Similarly, samples from $(\boldsymbol{\lambda}_\rho, \lambda_0, \rho|\boldsymbol{\beta}, \mathbf{N}, D_{obs})$ are obtained from the f.c.p. distributions $(\lambda_0, \rho|\boldsymbol{\beta}, \mathbf{N}, D_{obs})$ and $(\boldsymbol{\lambda}_\rho|\boldsymbol{\beta}, \lambda_0, \rho, \mathbf{N}, D_{obs})$, respectively, given by

$$\pi(\lambda_0|\rho, \boldsymbol{\beta}, \mathbf{N}, D_{obs}) \propto \left(\prod_{j=1}^{b}\frac{(1/\kappa^j)^{\mu_j\kappa^{-j}}}{\Gamma(\mu_j\kappa^{-j})}\frac{\Gamma(\mu_j\kappa^{-j}+\eta_j)}{(\kappa^{-j}+\xi_j)^{\mu_j\kappa^{-j}+\eta_j}}\right)\pi(\lambda_0) \quad (24)$$

and

$$\pi(\rho|\lambda_0, \boldsymbol{\beta}, \mathbf{N}, D_{obs}) \propto \prod_{j=1}^{b}\left(\frac{(1/\kappa^j)^{\mu_j\kappa^{-j}}}{\Gamma(\mu_j\kappa^{-j})}\frac{\Gamma(\mu_j\kappa^{-j}+\eta_j)}{(\kappa^{-j}+\xi_j)^{\mu_j\kappa^{-j}+\eta_j}}\right)c_\rho^{(j)}. \quad (25)$$

The Metropolis-Hasting (M-H) algorithm (see Gilks et al. (1996) and Casella and Robert (2005)) or the adaptive rejection Metropolis sampling (ARMS) algorithm proposed by Gilks et al. (1995) can be used to sample from (24). The product form of the joint prior distribution given in (19) enables us to use the Gibbs sampling algorithm proposed by Barry and Hartigan (1993) to sample from (25).

Finally, the joint prior distribution for $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho)$ given in (20) implies that the complete conditional posterior of each common failure rate $\lambda_\rho^{(j)}$ is:

$$(\lambda_\rho^{(j)}|\boldsymbol{\beta}, \lambda_0, \rho, \mathbf{N}, D_{obs}) \sim Gamma(\mu_j\kappa^{-j} + \eta_j; \kappa^{-j} + \xi_j), \tag{26}$$

As a consequence, the f.c.p. distribution for λ_k, $k = 1, ..., m'$, can be recovered by taking:

$$\pi(\lambda_k|\boldsymbol{\beta}, \lambda_0, \mathbf{N}, D_{obs}) = \sum_{i_{j-1}<k\leq i_j} \pi(\lambda_\rho^{(j)}|\boldsymbol{\beta}, \mathbf{N}, D_{obs}) R(I_\rho^{(j)}|\boldsymbol{\beta}, \mathbf{N}, D_{obs}), \tag{27}$$

where $R(I_\rho^{(j)}|\boldsymbol{\beta}, \mathbf{N}, D_{obs})$ is called posterior relevance, and denotes the probability of each clustered interval $I_\rho^{(j)}$ belongs to the random partition ρ, and $\pi(\lambda_\rho^{(j)}|\boldsymbol{\beta}, \mathbf{N}, D_{obs})$ denotes the posterior distribution of the common parameter $\lambda_\rho^{(j)}$, for $j = 1, ..., b$ and $b \in \{1, ..., m'\}$.

Under the framework we are proposing, the posterior populational survival function and posterior cure fraction are given, respectively, by

$$\begin{aligned} S_{pop}(t_{n+1}|D_{obs}) &= \sum_\rho \int_{\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0} \exp\{-\exp(\mathbf{x}_{n+1}\boldsymbol{\beta}) F(t_{n+1}|\boldsymbol{\lambda}_\rho, \rho)\} \\ &\times \pi(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, |D_{obs}) d\boldsymbol{\beta} d\boldsymbol{\lambda}_\rho d\lambda_0; \end{aligned} \tag{28}$$

and

$$\widehat{\pi_{n+1}} = \sum_\rho \int_{\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0} \exp\left\{-\exp(\mathbf{x}'_{n+1}\boldsymbol{\beta})\right\} \pi(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \lambda_0, \rho, |D_{obs}) d\boldsymbol{\beta} d\boldsymbol{\lambda}_\rho d\lambda_0. \tag{29}$$

Since (28) and (29) are not available in closed form expressions, inferences regarding such quantities are made numerically through the posterior sample of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho, \lambda_0)$.

4 Model selection

In this section we present a brief overview of conditional predictive ordinate (CPO) statistics. The CPO is a useful tool for model selection which allows us to assess goodness of fit as well as to perform model comparisons. For a more detailed discussions about CPO we refer the reader to Dey et al. (1997) and Ibrahim et al. (2001b), and references therein.

Let $D_{obs}^{(-i)}$ denote the observed data set excluding the i-th observation, $D_{obs}^{(i)} = (t_i, \delta_i, \mathbf{x}_i)$. Denote by $f(t_i|\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, D_{obs}^{(-i)}) = [\theta_i f(t_i|\boldsymbol{\lambda}_\rho)]^{\delta_i} \exp\{-\theta_i[1 - S(t_i|\boldsymbol{\lambda}_\rho)]\}$, where $\theta_i = \exp(\mathbf{x}'_i\boldsymbol{\beta})$.

The CPO statistic associated with the i-th observation is defined as the predictive density of t_i, conditionally on $D_{obs}^{(-i)}$, that is:

$$\begin{aligned} CPO_i &= f(t_i|D_{obs}^{(-i)}) \\ &= \sum_{\rho} \int_{\boldsymbol{\beta},\boldsymbol{\lambda}_\rho,\lambda_0} f(t_i|\boldsymbol{\beta},\boldsymbol{\lambda}_\rho, D_{obs}^{(-i)}) \pi(\boldsymbol{\beta},\boldsymbol{\lambda}_\rho,\rho|D_{obs}^{(-i)}) d\boldsymbol{\beta} d\boldsymbol{\lambda}_\rho. \end{aligned} \tag{30}$$

The CPO_i in (30) cannot be obtained in close form. However, it is possible to estimate (30) by using the MCMC samples from the posterior $\pi(\boldsymbol{\beta},\boldsymbol{\lambda}_\rho,\rho|D_{obs})$. It can be shown (Dey et al. (1997)) that a Monte Carlo estimate for CPO_i , $i = 1, ..., n$, is given by:

$$\widehat{CPO}_i = L \left(\sum_{l=1}^{M} \left[f(t_i|\boldsymbol{\beta}^l, \boldsymbol{\lambda}_\rho^l, D_{obs}) \right]^{-1} \right)^{-1}, \tag{31}$$

where M denotes the MCMC sample size of the posterior $\pi(\lambda_\rho, \beta, \rho|D_{obs})$. The posterior distribution of $(\lambda_\rho, \beta, \rho)$ is obtained by conditioning on the complete observed data set D_{obs}. Therefore, the computation of the CPO_i's is straightforward when MCMC samples of the posteriors of the quantities of interest are available.

A visual criterion for evaluating the performance of two or more competing models consists in plotting in the same graph their respective CPO_i's against the observations. Then, that model whose trajectory of CPO_i's is predominantly higher than the trajectories of the remaining competing models should be preferred.

Another useful measure for comparing competing models, which can be derived directly from the CPO_i's, is the logarithm of the pseudo-marginal likelihood (LPML), which is defined as:

$$LPML = \sum_{i=1}^{n} \log\left(CPO_i\right), \tag{32}$$

where n denotes the sample size. Hence, the higher the LPML is, the better is the fit of the model under investigation.

As a final remark on the the CPO_i statistics (and consequently the LPML statistic), we emphasize that it is well defined as long as $\pi(\boldsymbol{\beta},\boldsymbol{\lambda}_\rho,\rho|D_{obs}^{(-i)})$ is proper, $i = 1, ..., n$. In addition, it is also very computational stable (Ibrahim et al. (2001b)), and it can be easily computed when the number of parameters is not fixed, which turns the CPO particularly attractive for the approach we are proposing.

5 Case Study

In this section we analyze the E1673 data from a melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG) reported in Kim et al. (2006) using the proposed methodology. The E1673 data consists of the overall survival (in years), defined as

the time from randomization until death, and three covariates, namely age in years, gender (0 = male, 1 = female), and performance status (0 = fully active, 1 = other). The E1673 study lasted over 20 years and included n = 650 patients. By the time the study was closed, it was observed that 257 patients remained alive, which corresponds to approximately 40% of the patients involved in the trial. We refer the reader to Ibrahim et al. (2001b) and Kim et al. (2006) for more details regarding this data set.

Our aim here is to carry out a sensitivity analysis for different choices of m' and κ when the Weibull distribution with shape and scale parameters α and γ, respectively, is assumed for $F_0(\cdot|\lambda_0)$. For comparison purposes, we further compare the results provided by our approach with those yielded by the competing approach proposed by Kim et al. (2006). With these regards, the LPML measure discussed in the previous section is used to evaluate the goodness of fit of the fitted models.

Under the assumption that there is no prior information available regarding the time grid, we put equal mass onto the $2^{m'-1}$ possible partitions $\rho = \{i_0, i_1, ..., i_b\}$ by setting $c_{I_\rho^{(j)}} = 1,\ \forall\ (i_{j-1}, i_j) \in \mathcal{I}$. This choice for the prior cohesions $c_{I_\rho^{(j)}}$ yields

$$\pi(\rho = \{i_0, i_1, ..., i_b\}) = \frac{1}{2^{m'-1}},$$

which turns to be the Bayes-Laplace prior distribution for ρ.

We consider the Weibull distribution as a parametric model for $F_0(\cdot|\lambda_0)$, along with the cumulative hazard function in the following parametrization: $H(t|\lambda_0) = \exp(\gamma)t^\alpha$, where $\lambda_0 = (\alpha, \gamma)'$, so that $\mu_j = \exp(\gamma)(a_j^\alpha - a_{j-1}^\alpha)/(a_j - a_{j-1})$ and $\sigma_j^2 = \kappa^j \exp(\gamma)(a_j^\alpha - a_{j-1}^\alpha)/(a_j - a_{j-1})$. We also assume independence between α and γ, and as prior distributions we elicit $\alpha \sim Ga(0.001; 0.001)$ and $\gamma \sim N(0; 1000)$.

To carry out our Bayesian analysis using the collapsed Gibbs sampler discussed in the previous section, we considered posterior samples of size 1,000 for all scenarios analyzed by running single long chains, discarding the initial iterations needed to reach convergence and assuming lag 100 to eliminate correlations. All computational procedures were implemented in the Object-Oriented Matrix Programming Language Ox version 5.1 (Doornik, 2007), and are available from the author upon request.

We evaluated the effect of the maximum number of intervals intervals admitted a *priori* on the goodness of fit of the proposed model by considering values of m' ranging from 10 to 60. We further examined the role played by the smoothing parameter κ on the model fitting. For this purpose, we chose κ such that $\kappa^{m'}$ was equal to 1×10^{-01} and 1×10^{-05}. The results obtained are displayed in Table 1 and Figure 1.

As it can be observed in Figure 1, the proposed model yielded similar performances for both choices of $\kappa^{m'}$, regardless of the choice of m'. We further observed that $m' = 10$ led to the poorest fits whereas the best fits were obtained when $m' = 20$. For the fixed time grid model, both choices of $\kappa^{m'}$ also provided close results for $m' \leq 30$, and better fits when

Table 1: LPMLs for the melanoma data according to different specifications for m' and κ.

τ	$\kappa^{m'}$	10	20	30	40	50	60
random	1×10^{-01}	-1339.879	-1329.780	-1332.310	-1332.576	-1331.657	-1335.453
	1×10^{-05}	-1338.355	-1327.910	-1331.356	-1331.690	-1332.511	-1335.214
fixed	1×10^{-01}	-1339.178	-1332.954	-1337.905	-1344.310	-1350.325	-1356.122
	1×10^{-05}	-1339.407	-1334.567	-1336.502	-1337.340	-1340.383	-1344.551

$m' = 20$. Nevertheless, a decrease in the goodness of fit for $m' > 30$ could be observed. In particular, the performance of the model was substantially affected when $\kappa^{m'} = 1 \times 10^{-01}$.

Based on the results obtained, the introduction of a random time grid in the modeling seems to improve the goodness of fit of the model. It is also noteworthy that the LPML measures provided in Table 1 were, in general, greater than those yielded by the competing model presented in Kim et al. (2006), regardless of the assumption of a fixed or random time grid for the PEM, which suggests the superiority of the proposed model.

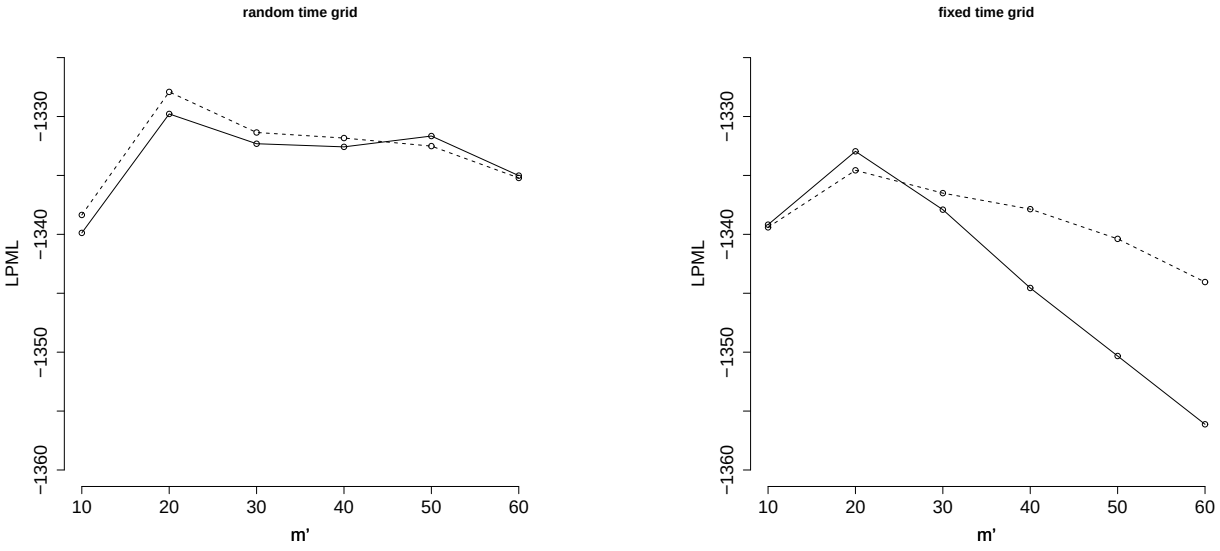

Figure 1: LPML associated with the fitted models according to different choices of m', for $\kappa^{m'} = 1 \times 10^{-01}$ (solid line) and $\kappa^{m'} = 1 \times 10^{-05}$ (dashed line).

According to Table 1, the best fitted model was obtained when a random time grid for the PEM was assumed along with $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-05}$. For this model, it was observed that the posterior most probable number of intervals was 11, with probability 0.238, and the corresponding 95% HPD intervals was $[8, 14]$. It is remarkable that there is a significant reduction on the number of intervals when compared with the maximum number of intervals assumed a *priori* ($m' = 20$). Other posterior summaries related to the number of intervals of the PEM are given in Table 2 and Figure 2. Finally, the most probable time grid a posteriori was $\tau = \{0, 0.476, 0.668, 0.794, 1.533, 5.629, 11.077, 24.208\}$, with probability 0.004.

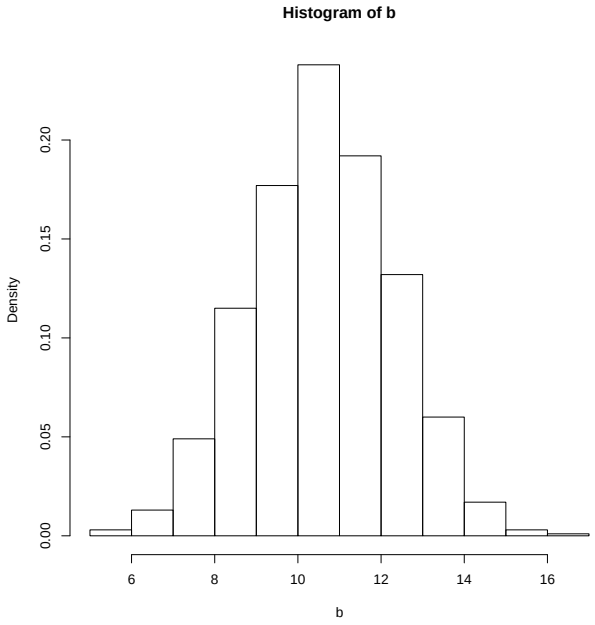

Figure 2: Posterior summaries for the number of intervals - PEM with random time grid, $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-05}$.

Table 2: Summary of the sample of the posterior distribution of the number of intervals - PEM with random time grid, $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-05}$.

	mean	sd	min	max
	11.103	1.734	5	17
$P_{2.50}$	$P_{25.0}$	$P_{50.0}$	$P_{75.0}$	$P_{97.50}$
8	10	11	12	14

The parameter estimates associated with the best fitted models considering the PEM with random and fixed time grids are displayed in Tables 3 and 4, respectively. The estimated regression coefficients associated with age, gender and PS were quite similar when fixed and random were assumed for the PEM. However, different estimates for the hyper-parameters α and γ, as well as the intercept, were obtained when fixed and random time grids were considered. The analysis of Tables 3 and 4 also shows that the status performance not be not significant to explain the survival time of patients with melanoma tumor, whereas gender and age are statistically significant prognostic variables. In this spirit, one my conclude that older patients are more likely to die from melanoma cancer than younger patients, and the risk of a man to die from melanoma cancer is approximately 1.5 times a woman's risk.

Figure 3 shows the estimated baseline hazard functions associated with the best fitted models with random and fixed time grids respectively, along with their corresponding 95% HPD intervals. As expected, the PEM with random time grid provides a smoother estimate for the baseline hazard function when compared with the PEM with fixed time grid. We

Table 3: Summary of the posterior distributuions - PEM with random time grid, $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-05}$.

			95% HPD interval	
variable	mean	sd	lower	upper
intercept	1.0374	0.3347	0.4784	1.7304
age	0.2053	0.0545	0.0910	0.3018
gender	-0.3902	0.1063	-0.6208	-0.2065
PS	0.2585	0.1445	-0.0124	0.5557
α	0.4981	0.1452	0.2405	0.7737
γ	-1.3167	0.4424	-2.1053	-0.4474

Table 4: Summary of the posterior distributions - PEM with fixed time grid, $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-01}$.

			95% HPD interval	
variable	mean	sd	lower	upper
intercept	0.6136	0.1831	0.2684	0.9582
age	0.2037	0.0521	0.1027	0.3010
gender	-0.3830	0.1058	-0.5774	-0.1815
PS	0.2424	0.1492	-0.0717	0.5143
α	0.6966	0.1389	0.4473	0.9860
γ	-0.7233	0.3844	-1.5272	-0.0370

further observed that both baseline hazard functions displayed in Figure 3 are smoother than those provided by the competing model given in Kim et al. (2006).

The boxplot for the overall cure rate estimates associated with the best fitted models considering the PEM with random and fixed time grids are shown in Figure 4, and the estimated populational survival functions (solid lines) along with their corresponding 95% HPD intervals (dashed lines) for average fully active men and women are displayed in Figure 5. The PEM with random time grid provided smaller estimates for the cure rate of the patients included in the E1673 study than the PEM with fixed time grid, as well as a shorter variability among estimates associated with different patients. This is in agreement with the estimated populational survival functions plotted in Figure 5, which indicates slightly lower populational survival curves associated with the PEM with random time grid when compared with the PEM with fixed time grid.

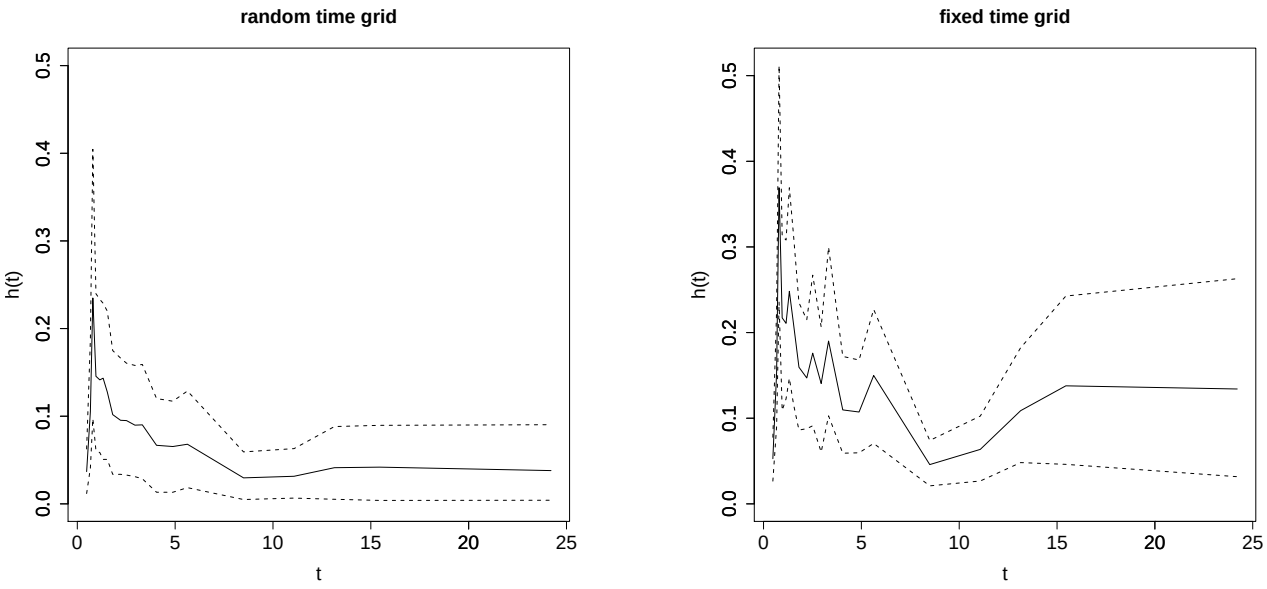

Figure 3: Estimated baseline hazard functions assuming $m' = 20$, $\kappa^{m'} = 1 \times 10^{-05}$ (random time grid), and $\kappa^{m'} = 1 \times 10^{-01}$ (fixed time grid).

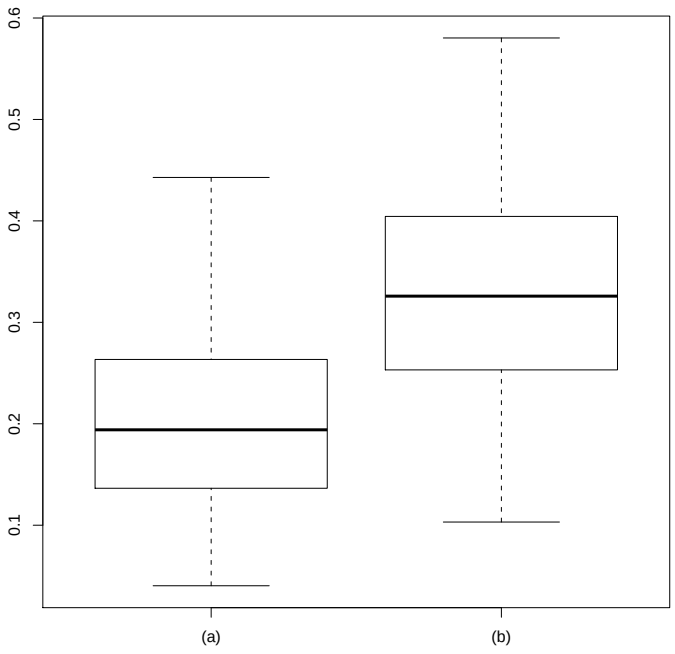

Figure 4: Overall cure rate estimates - PEM with random time grid, $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-05}$ (a), and PEM with fixed time grid, $m' = 20$, $\kappa^{m'} = 1 \times 10^{-01}$ (b).

6 Conclusions

In this paper we have proposed a novel semiparametric full Bayesian approach for modeling survival data with cure fraction using the PEM with random time grid and a structural prior

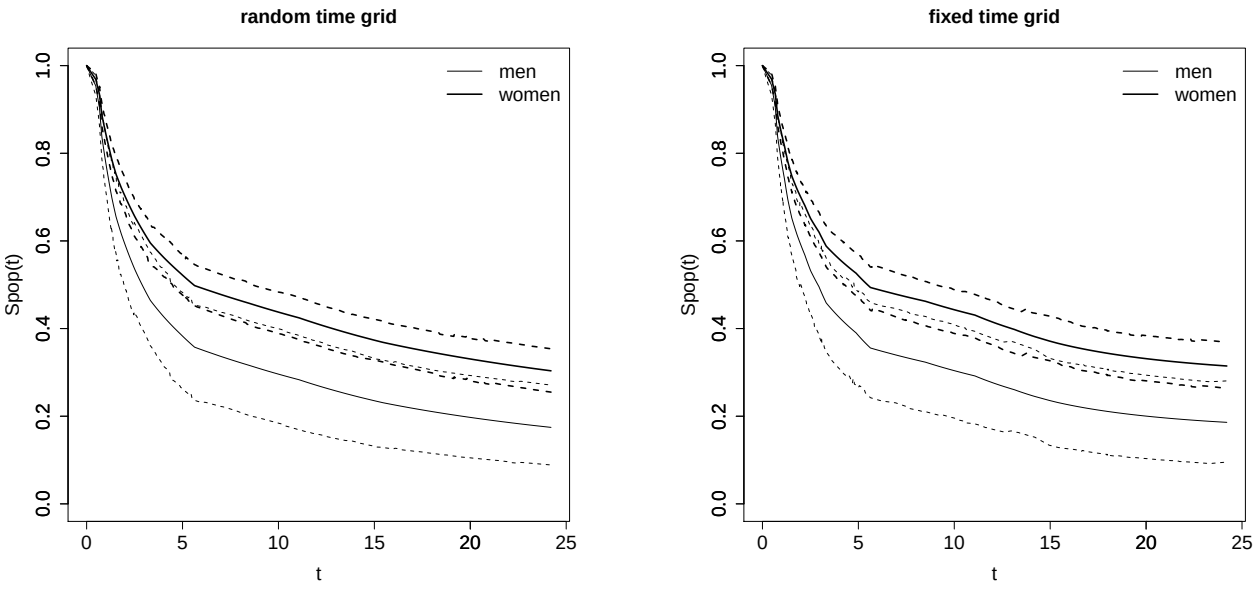

Figure 5: Estimated populational survival function for average fully active men and women - $m' = 20$, $\kappa^{m'} = 1 \times 10^{-05}$ (random time grid), and $m' = 20$ and $\kappa^{m'} = 1 \times 10^{-01}$ (fixed time grid).

distributions for the failure rates. We discussed in detail properties of the proposed model and provided an efficient and easy-to-implement Gibbs sampler algorithm for carrying out Bayesian analysis using the proposed approach. The usefulness of the proposed model was illustrated through an extensive analysis of the E1673 data from a melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG). A sensitivity analysis taking into account different prior specifications was performed. Finally, comparisons between the results yielded by the proposed model and those reported in Kim et al. (2006) indicates the superiority of the proposed approach over the competing model.

Our approach is fully Bayesian since all quantities involved in the model can be estimated after direct prior elicitation. Given the random time grid, prior elicitation for the failure rates of the PEM are based on local approximations to a continuous hazard function associated with a parametric model which is expected to describe the behavior of the right tail of the survival curve. The resulting model corresponds to an extension of the model proposed by Demarqui et al. (2008) and includes the semiparametric model introduced by Ibrahim et al. (2001a) as a special case. Another nice characteristic of the proposed model is that the constrains imposed on the set of possible time grids are based on solid arguments which permits, in a very elegant way, to model jointly the time grid and failure rates of the PEM, adding more flexibility in survival data modeling. Moreover, it makes better use of the data information in modeling the time grid of the PEM by taking into account the exact position in which the failure times have occurred, so that the endpoints of the intervals which defines the PEM are more likely to provide a better approximation for the unknown true baseline hazard function.

Future research includes the extension of the proposed methodology to account for multivariate survival data such as time to first and second infection, time to relapse to cancer and death, and so forth. Another possible extension comprehends the inclusion of spatial frailies into the model to accounting for spatial correlated survival data.

Acknowledgements

Fabio N. Demarqui's research has been sponsored by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) of the Ministry for Science and Technology of Brazil, and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais). Rosangela H. Loschi's research has been partially funded by CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 306085/2009-7, 304505/2006-4. Enrico A. Colosimo's research has been partially funded by the CNPq, grants 150472/2008-0 and 306652/2008-0. Part of this work was made during the period the first author was visiting the Department of Statistics of the University of Connecticut, Storrs-CT, USA.

7 Appendix

We provide the proof for the theorem given in Section 3.3. The following result, which is given in Chen et al. (1999), will be helpful in our proof.

Lemma: *Let* $g_1(z) = \frac{ze^{-z}}{1-e^{-z}}$ *and* $g_2(z) = ze^{-z}$. *Then, there exist a common constant* M_0 *such that* $g_1(z) \leq M_0$ *and* $g_2(z) \leq M_0$, *for all* $z > 0$;

Proof of theorem:

In order to prove the result, we need to show that:

$$\sum_{\rho} \int_{\lambda_0} \int_{\boldsymbol{\lambda}_\rho} \int_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D_{obs}) \pi(\boldsymbol{\lambda}_\rho | \lambda_0, \rho) \pi(\lambda_0) \pi(\rho) d\boldsymbol{\beta} d\boldsymbol{\lambda}_\rho d\lambda_0 < \infty. \tag{33}$$

Since the sum with respect to ρ in (33) is finite and $\pi(\rho)$ is a genuine probability mass function, it is sufficient to show that $\int_{\lambda_0} \int_{b\lambda_\rho} \int_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D_{obs}) \pi(\boldsymbol{\lambda}_\rho | \lambda_0, \rho) \pi(\lambda_0) d\boldsymbol{\beta} d\boldsymbol{\lambda}_\rho d\lambda_0 < \infty$ for a given ρ.

Remember that

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D_{obs}) &= \sum_{\mathbf{N}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_\rho, \rho | D) \\ &= \prod_{i=1}^{n} \left[\exp(\mathbf{x}_i' \boldsymbol{\beta}) f(t_i | \boldsymbol{\lambda}_\rho, \rho) \right]^{\delta_i} \exp\left\{ - \exp(\mathbf{x}_i' \boldsymbol{\beta}) [1 - S(t_i | \boldsymbol{\lambda}_\rho, \rho)] \right\}, \end{aligned} \tag{34}$$

where $f(t_i|\boldsymbol{\lambda}_\rho,\rho)=\prod_{j=1}^{b}\left(\lambda_\rho^{(j)}\right)^{\delta_{ij}}\exp\left\{-\lambda_\rho^{(j)}(t_{ij}-a_{j-1})\right\}$ and $S(t_i|\boldsymbol{\lambda}_\rho,\rho)=\exp\left\{-\sum_{j=1}^{b}\boldsymbol{\lambda}_\rho^{(j)}(t_{ij}-a_{j-1})\right\}$

We first note that, if $\delta_i=0$, then

$$\exp\left\{-\exp(\mathbf{x}_i'\boldsymbol{\beta})[1-S(t_i|\boldsymbol{\lambda}_\rho,\rho)]\right\}\leq 1, \tag{35}$$

Hence, it follows from (34) and (35) that

$$\begin{aligned}&\int_\beta\prod_{i=1}^{n}\left[\exp(\mathbf{x}_i'\boldsymbol{\beta})f(t_i|\boldsymbol{\lambda}_\rho,\rho)\right]^{\delta_i}\exp\left\{-\exp(\mathbf{x}_i'\boldsymbol{\beta})[1-S(t_i|\boldsymbol{\lambda}_\rho,\rho)]\right\}d\boldsymbol{\beta}\\ \leq&\int_\beta\prod_{k=1}^{m}\left[\exp(\mathbf{x}_{l_k}'\boldsymbol{\beta})f(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)\right]^{\nu_{l_k}}\exp\left\{-\exp(\mathbf{x}_{l_k}'\boldsymbol{\beta})[1-S(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)]\right\}d\boldsymbol{\beta},\end{aligned} \tag{36}$$

where $\{l_k: l_k=i,\delta_i=1,k=1,...,m,i=1,...,n\}$.

Using the lemma, it can be shown that

$$\theta_i f(t_i|\boldsymbol{\lambda}_\rho,\rho)\exp\left\{\theta_i[1-S(t_i|\boldsymbol{\lambda}_\rho,\rho)]\right\}\leq M_0\tfrac{f(t_i|\boldsymbol{\lambda}_\rho,\rho)}{1-S(t_i|\boldsymbol{\lambda}_\rho,\rho)}. \tag{37}$$

Furthermore, since X^* is of full rank and $m\geq p$, there must exist $l_1^*,l_2^*,...,l_p^*$ such that $\delta_{l_1^*}=\delta_{l_2^*}=...=\delta_{l_p^*}=1$ and $\mathbf{X}_p^*=(\mathbf{x}_{l_1^*},\mathbf{x}_{l_2^*},...,\mathbf{x}_{l_p^*})'$ and $\mathbf{X}_p^*$ is of full rank. Thus, using (37) in (36) we have that:

$$\begin{aligned}&\int_\beta\prod_{k=1}^{m}\left[\exp(\mathbf{x}_i'\boldsymbol{\beta})f(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)\right]^{\nu_{l_k}}\exp\left\{-\exp(\mathbf{x}_{l_k}'\boldsymbol{\beta})[1-S(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)]\right\}d\boldsymbol{\beta}\\ =&\left(\int_{\boldsymbol{\beta}}\prod_{k=1}^{p}f(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)\exp\left\{x_{l_k^*}'\boldsymbol{\beta}-[1-S(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)]\exp(x_{l_k^*}'\boldsymbol{\beta})\right\}d\boldsymbol{\beta}\right)\\ \times&\prod_{k=p+1}^{m}M_0\frac{f(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)}{1-S(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)}.\end{aligned} \tag{38}$$

To integrate out $\boldsymbol{\beta}$ in expression (38), we take the one-to-one transformation $\mathbf{u}=(u_1,...,u_p)'=\exp(\mathbf{X}_p^*\boldsymbol{\beta})$. Then, we have

$$\begin{aligned}&\int_{\boldsymbol{\beta}}\prod_{k=1}^{p}f(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)\exp\left\{\mathbf{x}_{l_k^*}'\boldsymbol{\beta}-[1-S(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)]\exp(\mathbf{x}_{l_k^*}'\boldsymbol{\beta})\right\}d\boldsymbol{\beta}\\ =&\frac{1}{|X_p^*|}\prod_{k=1}^{p}\int_{-\infty}^{\infty}f(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)\exp\left\{u_k-[1-S(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)]\exp(u_k)\right\}du_k\\ =&\prod_{k=1}^{p}\frac{f(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)}{1-S(t_{l_k^*}|\boldsymbol{\lambda}_\rho,\rho)}.\end{aligned} \tag{39}$$

Moreover, for each pair $\delta_{l_k}=1$ and $t_{l_k}\in I_\rho^{(j)}=(a_{j-1},a_j]$, $j=1,...,b$, and using the

lemma again, we have

$$\begin{aligned}
\frac{f(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)}{1-S(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)} &= \left[\prod_{j=1}^{b}\left(\boldsymbol{\lambda}_\rho^{(j)}\right)^{\delta_{l_kj}}\right]\frac{\exp\{-\sum_{j=1}^{b}\lambda_\rho^{(j)}(t_{l_kj}-a_{j-1})\}}{1-\exp\{-\sum_{j=1}^{b}\lambda_\rho^{(j)}(t_{l_kj}-a_{j-1})\}} \\
&\leq \left[\prod_{j=1}^{b}\left(\boldsymbol{\lambda}_\rho^{(j)}\right)^{\delta_{l_kj}}\right]\frac{\exp\{-\lambda_\rho^{(j)}(t_{l_kj}-a_{j-1})\}}{1-\exp\{-\lambda_\rho^{(j)}(t_{l_kj}-a_{j-1})\}} \\
&\leq M_0\prod_{j=1}^{b}\left(\boldsymbol{\lambda}_\rho^{(j)}\right)^{\delta_{l_kj}}. \qquad (40)
\end{aligned}$$

Therefore, using the lemma again, the right-hand of expression (38) becomes:

$$M_0^{m-p}\int_{\lambda_0}\int_{\boldsymbol{\lambda}_\rho}\prod_{k=1}^{m}\frac{f(t_{l_k}|\lambda_\rho,\rho)}{1-S(t_{l_k}|\boldsymbol{\lambda}_\rho,\rho)}\pi(\boldsymbol{\lambda}_\rho|\boldsymbol{\lambda}_0,\rho)\pi(\lambda_0)d\boldsymbol{\lambda}_\rho d\lambda_0 \qquad (41)$$

$$\leq M_0^{2m-p}\int_{\lambda_0}\int_{\boldsymbol{\lambda}_\rho}\prod_{k=1}^{m}\prod_{j=1}^{b}\left(\boldsymbol{\lambda}_\rho^{(j)}\right)^{\delta_{l_kj}}\pi(\boldsymbol{\lambda}_\rho|\boldsymbol{\lambda}_0,\rho)\pi(\lambda_0)d\boldsymbol{\lambda}_\rho d\lambda_0$$

$$= M_0^{2m-p}\int_{\lambda_0}\int_{\boldsymbol{\lambda}_\rho}\prod_{j=1}^{b}\left(\boldsymbol{\lambda}_\rho^{(j)}\right)^{\sum_{k=1}^{m}\delta_{l_kj}}\pi(\boldsymbol{\lambda}_\rho|\boldsymbol{\lambda}_0,\rho)\pi(\lambda_0)d\boldsymbol{\lambda}_\rho d\lambda_0, \qquad (42)$$

which is obviously finite, since $\sum_{k=1}^{m}\delta_{lkj}\geq 1\ \forall\ j$, and each component of $\boldsymbol{\lambda}_\rho$ has a gamma prior distribution. This completes the proof.

References

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data. *Statistica Sinica* **4,** 505–524.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems. *Ann Statist* **20,** 260–279.

Barry, D. and Hartigan, J. A. (1993). Bayesian analysis for change point problems. *Journal of the American Statistical Association* **88,** 309–319.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47,** 501–515.

Casella, G. and Robert, C. P. (2005). *Monte Carlo Statistical Methods.* Springer-Verlag, New York.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (1999). A new Bayesian model for survival data with a survival fraction. *Journal of the American Statistical Association* **94,** 909–919.

Chen, M. H., Shao, Q.-M., and Ibrahim, J. G. (1999). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.

Demarqui, F. N., Loschi, R. H., and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis* **14,** 333–356.

Dey, D. K., Chen, M. H., and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53,** 1239–1252.

Doornik, J. A. (2007). *Ox 5 - An Object-oriented Matrix Programming Language*. Timberlake Consultants Ltd.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38,** 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *The Canadian Journal of Statistics* **14,** 257–262.

Gamerman, D. (1994). Bayes estimation of the piece-wise exponential distribution. *IEEE Transactions on Reliability* **43,** 128–131.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **44,** 455–472.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Mont Carlo in Practice: Interdisciplinary Statistics*. Chapman & Hall.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41,** 337–348.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001a). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* **57,** 383–388.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001b). *Bayesian survival analysis*. Springer-Verlag, New York.

Kim, S., Chen, M. H., Dey, D. K., and Gamerman, D. (2006). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis* **13,** 17–35.

Kuk, A. Y. C. and Chen, M. H. (1992). A mixture models combinig logistic regression with proportional razards regression. *Biometrika* **79,** 731–739.

Maller, R. and Zhou, X. (1996). *Survival analysis with long-term survivors*. John Wiley and Sons, New York.

McKeague, I. W. and Tighiouart, M. (2000). Bayesian estimators for conditional hazard functions. *Biometrics* **56,** 1007–1015.

Peng, Y. and Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56,** 237–243.

Rodrigues, J., Cancho, V. G., de Castro, M., and Louzada-Neto, F. (2009a). On the unification of the long-term survival models. *Statistics and Probability Letters* **79,** 753–759.

Rodrigues, J., de Castro, M., Cancho, V. G., and Balakrishnan, N. (2009b). Com-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference* **139,** 3605 –3611.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56,** 227–236.

Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixuture models. *Journal of the American Statistical Association* **98,** 1063–1078.

Tsodikov, A. D., Loeffer, M., and Yakovlev, A. Y. (1998). A cure model with time-changing risk factor: An application to the analysis of secondary leukemia. a report from the international database on hodgkin's disease. *Statistics in Medicine* **17,** 27–40.

Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A., and Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer. *Biometrie et Analyse de Donnees Statio-Temporales* **12,** 66–82. B. Asselain, M. Boniface, C. Duby, C. Lopes, J. P. Masson, and J. Tranchefort, France, Rennes: Societe Francaise de Biometrie, ENSA.

Yin, G. and Ibrahim, J. G. (2005). Cure cate models: an unified approach. *The Canadian Journal of Statistics* **33,** 559–570.

Conclusões

Nesta tese de doutorado foram apresentadas extensões da abordagem introduzida por Demarqui *et al.* (2008) para a modelagem da grade do MEP. As contribuições mais importantes desta tese dizem respeito à generalização e unificação da abordagem inicialmente proposta por Demarqui *et al.* (2008) com outras abordagens já consagradas na literatura. A metodologia proposta se encaixa em diferentes contextos de análise de sobrevivência, tais como a estimação da função risco, modelos de regressão e modelos com fração de cura, e pode ser aplicada para dados do tipo tempo até a ocorrência de um evento de interesse oriundos de qualquer área do conhecimento.

Na obordagem originalmente introduzida por Demarqui *et al.* (2008), uma grade com tantos intevalos quanto tempos distintos de falha observados é assumida como a grade mais fina possível para o MEP, e a estrutura de agrupamento do MPP é introduzida sobre os intervalos resultantes. Esta suposição garante que cada intervalo agrupado induzido pelas grades aleatórias contenha pelo menos um tempo de falha observado. Na prática, um número muito grande de intervalos, no entanto, é desnecessário para uma boa aproximação da função risco, e esta condição pode ser ralaxada. Neste contexto, foi proposto um novo procedimento para a especificação da grade mais fina associada ao MEP. Especificamente, assumiu-se que esta grade é formada por um subconjunto do conjunto de tempos distintos de falhas observados. Tal suposição permite-nos controlar o número máximo de intervalos, bem como a suavidade das estimativas das taxas de falha. Ademais, a especificação de um número máximo de intervalos *a priori* evita o crescimento, sem limites, do número de intervalos quando o número de falhas tende a infinito, o que tornaria a abordagem proposta inviável, tanto do ponto de vista teórico quanto computacional. Na prática, uma análise de sensibilidade considerando-se valores pequenos à moderados para o número máximo de intervalos é recomendada.

O mecanismo considerado neste trabalho para modelar a grade do PEM apresenta claras vantagens sobre outras abordagens existentes na literatura. Primeiro, a restrição imposta sobre a grade mais fina possível para o PEM permite a modelagem direta da mesma após apropriadas especificações *a priori*, enquanto a utilização de um PPH para a modelagem do MEP com grade aleatória não possibilita tal estimação. Neste contexto,

a abordagem proposta enriquece a análise possibilitando inferências sobre a grade, bem como o número de intervalos aleatórios associados ao MEP. Além disso, a abordagem proposta apresenta outras características interessantes que não são compartilhadas pelas abordagens que assumem um PPH para a modelagem do MEP com grade aleatória, como por exemplo, assegurar que todos os intervalos aleatórios contenham pelo menos um tempo de falha, além de considerar a disposição dos tempos de falha observados ao longo do eixo do tempo. Finalmente, os modelos modelos discutidos nesta tese abrigam uma ampla classe de distribuições *a priori*, tanto para a grade de tempo quanto para as taxas de falha do PEM (nem todas consideradas neste trabalho), e inclui outros modelos consagrados na literatura como casos especiais, tornando, desta forma, a abordagem proposta bastante flexivel e atraente.

A seguir são discutidos, resumidamente, as principais contribuições e resultados associados a cada artigo apresentado nesta tese.

Artigo 1

Neste artigo foi proposta uma análise Bayesiana objetiva assumindo-se a distribuição de Bayes-Laplace para a grade de tempos e a distribuição *a priori* de Jeffreys para as taxas de falha do MEP. A escolha desta distribuição *a priori* conjunta para a modelagem do MEP é atrativa em situações nas quais não se tem informação *a priori* disponível. Além disso, tal especificação *a priori* possibilita uma comparação mais justa entre as estimativas Bayesianas e as de máxima verossimilhança para as taxas de falha e função de sobrevivência do MEP.

Os resultados obtidos a partir da análise dos tempos de sobrevivência de paciêntes diagnosticados com câncer no cérebro mostraram que as estimativas fornecidas pelo modelo proposto são comparáveis com as estimativas de máxima verossimilhança obtidas para o mesmo conjunto de dados. Entretanto, para a abordagem frequentista encontram-se disponíveis apenas intervalos de confiança, baseados em aproximações assintóticas, para a função de sobrevivência, enquanto a abordagem Bayesiana proposta possibilita a estimação intervalar de ambas as funções risco e de sobrevivência, de maneira direta, e sem depender de resultados assintóticos.

Artigo 2

Uma classe de modelos Bayesianos semiparamétricos para a modelagem de dados de sobrevivência considerando-se o MEP com grade aleatória foi introduzida neste artigo. A estrutura de agrupamento do MPP foi utilizada para modelar a incerteza acerca da grade que define o MEP. Condicionando-se na grade aleatória, quatro diferentes especificações *a priori* para as taxas de falha do MEP foram apresentadas. Especificamente, a distribuição *a priori* de Bayes-Laplace foi assumida para a grade aleatória do MEP. Uma abordagem Bayesiana objetiva foi proposta considerando-se a distribuição *a priori* de Jeffreys para as taxas de falha do MEP. Distribuições *a priori* subjetivas para as taxas de falha do MEP foram propostas através de diferentes escolhas de distribuições gama, incluindo distribuições gama independentes não-informativas, distribuições gama condicionalmente independentes com hiper-parâmetros correlacionados (abordagem dinâmica) e distribuições gama como aquelas definidas em Ibrahim *et al.* (2001a), chamadas aqui de distribuições gamma estruturadas.

O modelo proposto inclui alguns importantes modelos existentes na literatura como casos particulares, e pode ser facilmente generalizado para acomodar uma classe ainda mais abrangente de distribuições *a priori* para as taxas de falha do MEP. Como ilustração das abordagens propostas neste artigo, o banco de dados referente aos tempos de sobrevivência de pacientes diagnosticados com câncer no cérebro no condado de Windham-CT, EUA, foi novamente considerado para ilustração. Comparações entre os modelos ajustados considerando-se grades fixas e aleatórias, bem como diferentes especificações *a priori* para a taxa de falha, realizadas utilizando-se como medida de comparação a média do logaritmo da pseudo-verossimilhança marginal, mostraram que o MEP com grade aleatória, na maioria dos casos considerados, forneceu melhores ajustes para o conjunto de dados estudado.

Artigo 3

Neste artigo uma abordagem dinâmica para o MEP com grade aleatória foi apresentada. O modelo proposto combina as abordagens propostas por Gamerman (1991) e Demarqui *et al.* (2008). Para a modelagem da grade de tempos do MEP, foi assumida a distribuição *a priori* de Bayes Laplace. Dada a grade aleatória, um processo correlacionado foi assumido para as taxas de falhas e os coeficientes da regressão associados aos sucessivos intervalos agrupados induzidos pela grade aleatória. A análise dos tem-

pos de sobrevivência de pacientes diagnosticados com câncer no cérebro no condado de Windham-CT, EUA, considerando-se a variável explicativa idade, foi apresentada. Os resultados obtidos indicaram que a idade dos pacientes é uma variável estatisticamente significativa para explicar o tempo de vida dos pacientes.

O desempenho da abordagem dinâmica para o MEP com grade aleatória foi avaliado através uma análise de sensibilidade realizada com base em diferentes especificações *a priori*, utilizando-se medidas como o fator de Bayes e probabilidades *a posteriori* dos modelos ajustados. Os resultados obtidos também foram comparados aos resultados fornecidos pelo mesma abordagem dinâmica considerando-se o MEP com grade fixa, e indicaram a superioridade do modelo com grade aleatória. Também foi observado que, em todos os cenários considerados, o MEP com grade aleatória forneceu estimativas mais suaves para os coeficientes da regressão.

O modelo proposto é uma alternativa bastante flexível para situações nas quais a suposição de riscos proporcionais não pode ser verificada pelos dados, tais como a presença de covariáveis depedentes do tempo e / ou cujos efeitos das mesmas variam ao longo do tempo, como foi ilustrado na análise dos tempos de sobrevivência de pacientes com câncer no cérebro, onde observou-se que o feito da idade dos pacientes diminui ao longo do tempo.

Artigo 4

Foi apresentada neste artigo uma nova abordagem Bayesiana semiparamétrica para o modelo de tempo de promoção introduzido por Yakovlev *et al.* (1993) e Chen *et al.* (1999). O MEP com grade aleatória foi assumido para modelar a distribuição dos tempos de promoção. Para a modelagem da grade, a estrutura de agrupamento do MPP foi utilizada. A distribuição *a priori* de Bayes-Laplace foi assumida, neste caso, para a grade aleatória. Dada a grade aleatória, distribuição *a priori* estrutural originalmente proposta por Ibrahim *et al.* (2001a) foi considerada para modelar a incerteza acerca das taxas de falha do MEP. Uma distribuição *a priori* uniforme imprópria foi assumida para os coeficientes da regressão, e condições garantindo que a distribuição *a posteriori* conjunta do modelo proposto é própria foram fornecidas. O modelo resultante é uma generalização do Modelo 4 apresentado no segundo artigo que compõe esta tese (o modelo proposto reduz-se ao Modelo 4 se $P(N_i = 1) = 1, i = 1, ..., n$), e inclui como caso particular o modelo de fração de cura proposto por Ibrahim *et al.* (2001a), que considera uma grade fixa de tempo para a modelagem do MEP.

A nova abordagem foi utilizada para modelar os tempos de sobrevivência de indivíduos diagnosticados com melanoma, que fizeram parte do estudo clínico E1673 conduzido pela ECOG. Com o objetivo de avaliar o efeito de diferentes especificações *a priori*, uma análise de sensibilidade do modelo, considerando-se a soma do logaritmo da pseudo-verossimilhança marginal como medida de comparação, foi conduzida. Foi observado que o modelo proposto é sensível à diferentes escolhas *a priori*. Finalmente, comparações realizadas entre os ajustes fornecidos pelo modelo proposto e o modelo competidor introduzido, (ver resultados disponíveis em Kim *et al.* (2006)) indicaram que o modelo proposto apresentou melhor desempenho quando ajustado ao conjunto de dados considerado.

Conclusões Gerais

Em linhas gerais, foi observado que modelagem do MEP com grade aleatória proposta nesta tese fornece estimativas mais suaves para a função risco do que aquelas em que uma grade fixa é assumida para o MEP. A modelagem proposta também enriquece a análise ao permitir, por exemplo, que inferências sobre a grade do eixo do tempo e o número de intervalos que define o MEP, sejam realizadas. Por fim, os resultados obtidos indicam que as abordagens consideradas fornecem resultados mais satisfatórios para a modelagem de dados de sobrevivência.

Trabalhos futuros incluem a introdução de fragilidades (efeitos aleatórios) na modelagem de dados de sobrevivência dentro dos contextos aqui considerados. Modelos de fragilidade tem recebido crescente atenção na literatura estatística, sobretudo nas duas últimas décadas, com especial destaque às abordagens Bayesianas. Os modelos discutidos nesta tese podem ser diretamente estendidos para acomodar dados de sobrevivência multivariados, tais como a modelagem de eventos recorrentes e ou tempos de sobrevivência espacialmente correlacionados. O tratamento de dados ausentes também pode ser facilmente incorporado às abordagens propostas. Outra possível proposta de trabalho é a elaboração de um artigo de revisão cobrindo as principais abordagens existentes para o MEP, e o desenvolvimento de um pacote estatístico para ajustar esses modelos, para ser disponibilizado em R.

Referências Bibliográficas

AITKIN, M.; LAIRD, N.; FRANCIS, B. A reanalysis of the Stanford heart transplant data (with discussion). *Journal of the American Statistical Association*, v. 78, p. 264–292, 1983.

ARJAS, E.; GASBARRA, D. Nonparametric Bayesian inference from right censored survival data. *Statistica Sinica*, v. 4, p. 505–524, 1994.

BARBOSA, E. P.; COLOSIMO, E. A.; LOUZADA-NETO, F. Accelerated life tests analyzed by a piecewise exponential distribution via generalized linear models. *IEEE Transactions on Reliability*, v. 45, p. 619–623, 1996.

BARRY, D.; HARTIGAN, J. A. Product partition models for change point problems. *Ann Statist*, v. 20, p. 260–279, 1992.

BASTOS, L. S.; GAMERMAN, D. Dynamic survival models with spatial frailty. *Lifetime Data Analysis*, v. 12, p. 441–460, 2006.

BRESLOW, N. E. Covariance analysis of censored survival data. *Biometrics*, v. 30, p. 89–99, 1974.

CHEN, M. H.; IBRAHIM, J. G. Maximum likelihood methods for cure rate models with missing covariates. *Biometrics*, v. 57, p. 43–52, 2001.

CHEN, M. H.; IBRAHIM, J. G.; SINHA, D. A new Bayesian model for survival data with a survival fraction. *Journal of the American Statistical Association*, v. 94, p. 909–919, 1999.

CLARK, D. E.; RYAN, L. M. Concurrent prediction of hospital mortality and length of stay from risk factors on admission. *Health Services Research*, v. 37, p. 631–645, 2002.

COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 34, p. 187–220, 1971. With discussion.

DEMARQUI, F. N.; LOSCHI, R. H.; COLOSIMO, E. A. Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis*, v. 14, p. 333–356, 2008.

FRIEDMAN, M. Piecewise exponential models for survival data with covariates. v. 10, p. 101–103, 1982.

GAMERMAN, D. Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 40, p. 63–79, 1991.

GAMERMAN, D. Bayes estimation of the piece-wise exponential distribution. *IEEE Transactions on Reliability*, v. 43, p. 128–131, 1994.

IBRAHIM, J. G.; CHEN, M. H.; SINHA, D. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, v. 57, p. 383–388, 2001.

IBRAHIM, J. G.; CHEN, M. H.; SINHA, D. *Bayesian survival analysis*. [S.l.]: Springer-Verlag, New York, 2001.

KALBFLEISCH, J. D.; PRENTICE, R. L. Marginal likelihoods based on Cox's regression and life models. *Biometrika*, v. 60, p. 267–278, 1973.

KASS, R. E.; RAFTERY, A. E. Bayes factor. *Journal of the American Statistical Association*, v. 90, p. 773–795, 1995.

KIM, J. S.; PROSCHAN, F. Piecewise exponential estimator of the survival function. *IEEE Transactions on Reliability*, v. 40, p. 134–139, 1991.

KIM, S. *et al.* Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, v. 13, p. 17–35, 2006.

MCKEAGUE, I. W.; TIGHIOUART, M. Bayesian estimators for conditional hazard functions. *Biometrics*, v. 56, p. 1007–1015, 2000.

QIOU, Z.; RAVISHANKER, N.; DEY, D. K. Multivariate survival analysis with positive stable frailties. *Biometrics*, v. 55, p. 637–644, 1999.

SAHU, S. K. *et al.* A weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis*, v. 3, p. 123–137, 1997.

SINHA, D.; CHEN, M. H.; GOSH, S. K. Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, v. 55, p. 585–590, 1999.

YAKOVLEV, A. Y. *et al.* A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer. *Biometrie et Analyse de Donnees Statio-Temporales*, v. 12, p. 66–82, 1993. B. Asselain, M. Boniface, C. Duby, C. Lopes, J. P. Masson, and J. Tranchefort, France, Rennes: Societe Francaise de Biometrie, ENSA.