Posterior Propriety and Computation for the Cox Regression Model with Applications to Missing Covariates

By Ming-Hui Chen, Joseph G. Ibrahim, and Qi-Man Shao*

University of Connecticut, University of North Carolina, and University of Oregon and HKUST

August 29, 2005

Summary

In this paper, we carry out an in-depth theoretical investigation for Bayesian inference for the Cox regression model (Cox, 1972, 1975). Specifically, we establish establish necessary and sufficient conditions for posterior propriety of the regression coefficients, $\boldsymbol{\beta}$, in Cox's partial likelihood, which can be obtained as the limiting marginal posterior distribution of $\boldsymbol{\beta}$ through the specification of a gamma process prior for the cumulative baseline hazard and a uniform improper prior for $\boldsymbol{\beta}$ (Kalbfleisch, 1978, Sinha, Ibrahim, and Chen, 2003). We also examine necessary and sufficient conditions for posterior propriety of the regression coefficients, $\boldsymbol{\beta}$ using full likelihood Bayesian approaches in which a gamma process prior is specified for the cumulative baseline hazard. We examine characterizations of posterior propriety under completely observed data settings as well as for missing covariates. Latent variables are introduced to facilitate a straightforward Gibbs sampling scheme in the Bayesian computation. A real dataset is presented to illustrate the proposed methodology.

Key words and phrases. Missing at random (MAR), Gamma process prior, Latent variable, Markov chain Monte Carlo, Necessary and sufficient conditions, Partial likelihood, Proportional hazards model, Propriety of posterior distribution.

*Ming-Hui Chen is Professor, Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, Email: mhchen@merlot.stat.uconn.edu. Joseph G. Ibrahim is Professor, Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, Chapel Hill, NC 27599, Email: ibrahim@bios.unc.edu. Qi-Man Shao is Professor, Department of Mathematics, University of Oregon, Eugene, OR 97403-1222, and Professor, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, Email: qmshao@darkwing.uoregon.edu.

1 Introduction

The study of specific conditions for propriety of the posterior distribution is a very important practical problem for the Cox model, since it is well known that certain data configurations and/or sample sizes lead to nonidentifiable models. Thus, it becomes critical in such situations to develop a theory that characterizes necessary and sufficient conditions for propriety of the posterior distribution of the Cox regression model under different settings. In this paper, we study this problem by considering two different but related approaches for obtaining the posterior distribution of $\boldsymbol{\beta}$ in the Cox model. The first approach is based on starting out with the Cox's partial likelihood itself and treating it as a likelihood, specifying an improper prior for $\boldsymbol{\beta}$, then examining necessary and sufficient conditions for propriety of the resulting posterior. Such a formulation of a posterior for the Cox model has been motivated and discussed by many including Carlin et al. (1993), Gustafson (1997), Volinsky et al. (1997), Sargent (1998), and Sinha, Ibrahim, and Chen (2003). Moreover, Cox's partial likelihood has been shown to have a Bayesian justification by Kalbfleisch (1978) and Sinha, Ibrahim, and Chen (2003), in which it is the limiting marginal posterior distribution based on a gamma process prior for the cumulative baseline hazard. The second approach we consider is based on a full likelihood approach rather than the partial likelihood, in which the baseline hazard is *not* eliminated from the likelihood at the likelihood stage and an authentic likelihood is used in the derivation of the posterior distribution of $\boldsymbol{\beta}$, in which and a gamma process prior is specified for the cumulative baseline hazard. The two approaches are related but different, and result in different conditions for posterior propriety as well as their computational implementation.

One of the most practical applications and driving motivations of our proposed methodology is in missing data settings, where models can easily be nonidentifiable and posterior propriety becomes a very important practical issue. Missing covariate data in the Cox model is a fundamentally important practical problem in biomedical research. In the presence of missing covariates, we only consider a full likelihood approach since a joint probability distribution must be specified for the failure time variable and the missing covariates, and hence a partial likelihood approach in this context is not as desirable. In the presence of missing covariates, we are led to very different theoretical characterizations for posterior propriety and more challenging computational implementation.

There has been some semiparametric Bayesian work done in the missing data context for the Cox model and cure rate models, including Chen, Ibrahim, and Lipsitz (2002), also discussed in the book by Ibrahim, Chen, and Sinha (2001). Previous frequentist work along those lines includes methods developed by Schluchter and Jackson (1989) and Lipsitz and Ibrahim (1996a, 1996b) for ignorably missing categorical covariates in fully parametric proportional hazards models. When the data are missing at random (MAR) and the pattern of missing data is monotone, estimating equations such as those proposed by Zhou and Pepe (1995) and Reilly and Pepe (1995) are useful. Although these estimating equation approaches may be used when missing covariates are categorical or continuous, they are restrictive because missing data often does not occur in a monotone fashion. Lin and Ying (1993) proposed approximate partial likelihood estimates that can accommodate any pattern of missing data but require the data to be MCAR. Paik (1997) proposes a multiple imputation method when only one covariate is missing, and Paik and Tsai (1997) propose an imputation method that can be used when data are MAR. Lipsitz and Ibrahim (1998) develop a Monte Carlo method for MAR categorical covariates in Cox's partial likelihood. Lawless, Kalbfleisch, and Wild (1999) examine missing covariate data in parametric regression models when missingness depends on a stratification variable that is always observed. Herring and Ibrahim (2001) develop a likelihood-based methodology for MAR covariates based on partial likelihood using an EM-type algorithm. However, as noted earlier, there has been virtually no literature on Bayesian (or frequentist) methods for theoretically characterizing posterior propriety (or existence of the MPLE) for the Cox model in the no missing data data setting or in the missing covariate data setting.

For a clear focus and ease of exposition, we only focus on the Bayesian paradigm in this paper rather than tackle both the frequentist and Bayesian paradigms simultaneously. In a certain sense, it suffices to only focus on the Bayesian paradigm for the Cox model since the partial likelihood does indeed have a Bayesian justification as noted earlier. Moreover, Bayesian methods of inference often lead to easier computational procedures than frequentist methods especially in missing data problems. Also, prior information from expert opinion or historical data can be more easily and naturally incorporated within the Bayesian paradigm, as discussed in Ibrahim and Chen (2000). We examine Bayesian methods of estimation when there is no missing data as well as when there are MAR covariates for Cox's partial likelihood. Specifically,

(a) we establish necessary and sufficient conditions for posterior propriety of regression coefficients, $\boldsymbol{\beta}$, based on partial likelihood with no missing data as well as the full likelihood with MAR covariate data using a gamma process prior on the baseline cumulative hazard; and

(b) develop a novel Bayesian computational scheme through the introduction of several latent variables for sampling from the posterior distribution of $\boldsymbol{\beta}$ in the presence of missing covariates.

The methodology proposed here is quite new and will shed light on the characterizations of posterior propriety for the Cox model with complete data as well as with missing covariate data. In (b), we devise a novel latent variable approach in which after the introduction of three latent variables, the resulting joint likelihood can be written as a product of independent observations so that the Gibbs sampling algorithm can be efficiently and easily implemented. The latent variable method for sampling from the posterior is quite new and useful in Bayesian inference for Cox's regression model with a gamma process on the cumulative baseline hazard function. In addition to this, we note that none of the literature cited above has examined theoretical necessary and sufficient conditions for the posterior propriety based on the partial likelihood for no missing data situation or the full likelihood with the gamma process prior and MAR covariates.

The significance of this work has several aspects: first, the proposed methodology will allow the data analyst to determine, for a given dataset, whether the posterior distribution is proper or not *before* carrying out the analysis. Such a methodology is critical since it is not always clear from the computer output in an analysis whether the posterior is proper or not. The methodology we propose here answers such questions with certainty given the dataset so that the analyst knows for sure whether the posterior is proper or not before an analysis is conducted. These necessary and sufficient conditions are straightforward to implement by the data analyst and these conditions will also be useful for determining suitable starting values for Gibbs samplers when fitting these models. Thus, the practical consequences of the proposed methodology is that we provide valuable tools for checking existence of propriety of the posterior as well as inferential and computational tools for Bayesian inference for the Cox model with MAR covariates. Such tools are critical in missing data problems and can easily be implemented in standard software packages.

We also note that the theory developed in Chen, Ibrahim, and Shao (2004) cannot be applied here, since (i) they fundamentally need to assume throughout that the likelihood function can be written as a product of independent observations, which is not the case for partial likelihood, and (ii) there are no unknown nuisance parameters in the response model, only the parameters of interest, $\boldsymbol{\beta}$. On the other hand, in the Cox model, we must deal with the baseline hazard (or cumulative hazard) as a nuisance parameter in the response model, which is not of primary inferential interest. In addition, we consider Bayesian approaches in this paper since Bayesian methods offer some advantages over maximum likelihood (ML) type methods. Specifically, (i) partial likelihood has a Bayesian justification, as shown in Kalbfleisch (1978) and Sinha, Ibrahim, and Chen (2003) and (ii) Bayesian methods of inference often lead to easier computational procedures than ML-type methods, especially when some of the missing covariates are continuous; (iii) prior information from expert opinion or historical data can be more easily and naturally incorporated within the Bayesian paradigm, as discussed in Ibrahim and Chen (2000).

The rest of this article is organized as follows. In Section 2, we consider Bayesian inference with partial likelihood and full likelihood through a gamma process prior for the cumulative baseline hazard with complete data and present theory in the presence of ties. Theoretical results for Bayesian inference with MAR covariates are given in Section 3. In Section 4, we develop a novel computational algorithm via latent variables to sample from the posterior distribution. Section 5 presents a melanoma dataset to illustrate the proposed methodology.

2 Bayesian Inference With No Missing Data

Let y_i denote the minimum of the censoring time C_i and the survival time T_i, and let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ be the $p \times 1$ vector of covariates associated with y_i for the i^{th} subject. Denote by $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ the $p \times 1$ vector of regression coefficients. Also, $\delta_i = 1\{T_i = y_i\}$ is the indicator for death for $i = 1, 2, \ldots, n$, where n is the total number of observations and $\mathcal{R}(t) = \{i : \ y_i \geq t\}$ is the set of subjects at risk at time t. Then, the partial likelihood of Cox (1975) is given by

$$L_p(\boldsymbol{\beta}|D_{obs}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{\sum_{j \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})} \right]^{\delta_i} , \tag{2.1}$$

where $D_{obs} = \{(y_i, \delta_i, \boldsymbol{x}_i) : i = 1, 2, \ldots, n\}$ is the observed univariate right censored data. As usual, we assume throughout that $\boldsymbol{x}_i$ does not include an intercept, since the intercept is not estimable in the Cox partial likelihood, and that given $\boldsymbol{x}_i$, T_i and C_i are independent.

Carrying out Bayesian inference for the Cox model is not an easy task since the baseline hazard function is typically left completely unspecified, and thus some type of nonparametric prior process is required for inference. In this section, we discuss how to characterize the conditions for propriety of the posterior distribution of $\boldsymbol{\beta}$ for the Cox model with an improper uniform prior on $\boldsymbol{\beta}$, namely, $\pi(\boldsymbol{\beta}) \propto 1$ in the no missing data situation.

We first establish a useful result, which is formally stated in the following lemma.

Lemma 2.1 *Let X^* be an $n^* \times p$ matrix ($p < n^*$). Also let R^{n^*} denote the n^*-dimensional Euclidean space. If there is no positive vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_{n^*})' \in R^{n^*}$ (denoted by $\boldsymbol{v} > 0$, i.e., $v_i > 0$ for $i = 1, 2, \ldots, n^*$) such that*

$$X^{*\prime}\boldsymbol{v} = 0, \tag{2.2}$$

then there exists a non-zero vector $\boldsymbol{b} \in R^p$ such that

$$\boldsymbol{b}'\boldsymbol{x}_i^* \leq 0, \tag{2.3}$$

where $\boldsymbol{x}_i^*$ is the i^{th} row of X^*.

Proof. Let

$$\mathcal{V} = \{X^{*\prime}\boldsymbol{v} : \quad \boldsymbol{v} > 0, \boldsymbol{v} \in R^{n^*}\}.$$

Then $\mathcal{V}$ is a convex cone in R^p (see Theorem 2.6 in Rockafellar (1970)). Since (2.2) does not hold, by Corollary 11.7.3 of Rockafellar (1970), there exists some non-zero vector $\boldsymbol{b}$ such that

$$\forall\ \boldsymbol{v} > 0,\ \boldsymbol{b}'X^{*\prime}\boldsymbol{v} \leq 0$$

and hence

$$\forall\ \boldsymbol{v} \geq 0,\ \boldsymbol{b}'X^{*\prime}\boldsymbol{v} \leq 0.$$

In particular, (2.3) holds. ■

We consider two approaches for carrying out Bayesian inference for $\boldsymbol{\beta}$ when there are no missing covariates. The first approach is to treat the partial likelihood $L_p(\boldsymbol{\beta}|D_{obs})$ in (2.1) as the "likelihood function" and take $\pi(\boldsymbol{\beta}) \propto 1$. In this development, the "posterior" distribution for $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|D_{obs}) \propto L_p(\boldsymbol{\beta}|D_{obs}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{\sum_{j \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})} \right]^{\delta_i}. \tag{2.4}$$

Treating (2.1) as the "likelihood function" has been considered by several authors in various contexts including Raftery et al. (1995), Volinksy et al. (1997), and Sargent (1998). We refer the reader to Ibrahim, Chen, and Sinha (2001, Chapter 4) for more details. The next theorem characterizes the necessary and sufficient conditions for propriety of the posterior distribution of $\boldsymbol{\beta}$ based on the partial likelihood.

Theorem 2.1 *Define X^* to be*

$$X^* = (\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i), j \in \mathcal{R}(y_i), 1 \leq i \leq n)'. \tag{2.5}$$

The posterior distribution $\pi(\boldsymbol{\beta}|D_{obs})$ in (2.4) is proper, i.e.,

$$\int_{R^p} L_p(\boldsymbol{\beta}|D_{obs})d\boldsymbol{\beta} < \infty,$$

if and only if the following conditions are satisfied:

(C1) X^* is of full rank; and

(C2) There exists a positive vector $\boldsymbol{v}$, i.e., each component of $\boldsymbol{v}$ is positive, such that

$$X^{*'}\boldsymbol{v} = 0. \tag{2.6}$$

Proof. For notational simplicity, we assume that $y_1, y_2, \ldots, y_d$ are the failure times. Let $\mathcal{R}_i = \mathcal{R}(y_i) - \{i\}$. Then, (2.1) reduces to

$$L_p(\boldsymbol{\beta}|D_{obs}) = \int_{R^{+d}} \exp(-\sum_{i=1}^{d} t_i) \prod_{1 \leq i \leq d, j \in \mathcal{R}_i} F((\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i} d\boldsymbol{t}, \tag{2.7}$$

where $\boldsymbol{t} = (t_1, t_2, \ldots, t_d)'$, $R^{+d} = R^+ \times \cdots \times R^+$ with $R^+ = (0, \infty)$, and $F(u) = \exp(-\exp(-u))$.

Sufficiency:

Let $1\{A\}$ denote the indicator function so that $1\{A\} = 1$ if A is true and 0 otherwise. Observe that

$$\begin{aligned} F((\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i} &= \int_{-\infty}^{\infty} 1\{u_{ij} \leq (\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\beta}\} dF^{t_i}(u_{ij}) \\ &= \int_{-\infty}^{\infty} 1\{u_{ij} \geq (\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta}\} d(-F^{t_i}(-u_{ij})). \end{aligned}$$

Let $k = \sum_{i=1}^d k_i$, where $k_i = \#\{\mathcal{R}_i\}$ is the cardinality of $\mathcal{R}_i$. Also let X^{**} denote the submatrix of X^* in (2.5) with rows corresponding to $\delta_i = 1$ and $j \in \mathcal{R}_i$ for $i = 1, 2, \ldots, d$. By the Fubini theorem, we get

$$\begin{aligned} &\int_{R^p} L_p(\boldsymbol{\beta}|D_{obs})d\boldsymbol{\beta} \\ &= \int_{R^{+d}} \int_{R^p} \int_{R^k} \exp(-\sum_{i=1}^{d} t_i) 1\{u_{ij} \geq (\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta}, 1 \leq i \leq d, j \in \mathcal{R}_i\} d\boldsymbol{F}(\boldsymbol{u}) d\boldsymbol{\beta} d\boldsymbol{t} \\ &= \int_{R^{+d}} \exp(-\sum_{i=1}^{d} t_i) \int_{R^k} \int_{R^p} 1\{X^{**}\boldsymbol{\beta} \leq \boldsymbol{u}\} d\boldsymbol{\beta} d\boldsymbol{F}(\boldsymbol{u}) d\boldsymbol{t}. \end{aligned}$$

where $d\boldsymbol{F}(\boldsymbol{u}) = \prod_{1\le i\le d, j\in\mathcal{R}_i} d(-F^{t_i}(-u_{ij}))$. Following Chen and Shao (2001), under condition $(C2)$ we obtain

$$\int_{R^p} 1\{X^{**}\boldsymbol{\beta} \le \boldsymbol{u}\} d\boldsymbol{\beta} \le K||\boldsymbol{u}||^p,$$

where $K > 0$ is a constant independent of $\boldsymbol{u}$. Note that

$$d\boldsymbol{F}(\boldsymbol{u}) = \prod_{1\le i\le d, j\in\mathcal{R}_i} d(-F^{t_i}(-u_{ij})) = \prod_{1\le i\le d, j\in\mathcal{R}_i} t_i \exp(u_{ij}) \exp\{-t_i \exp(u_{ij})\} du_{ij}.$$

Thus, we obtain

$$\begin{aligned}
&\int_{R^p} L_p(\boldsymbol{\beta}|D_{obs}) d\boldsymbol{\beta} \\
\le& K \int_{R^{+d}} \int_{R^k} \exp(-\sum_{i=1}^d t_i) ||\boldsymbol{u}||^p \prod_{1\le i\le d, j\in\mathcal{R}_i} t_i \exp(u_{ij}) \exp\{-t_i \exp(u_{ij})\} du_{ij} d\boldsymbol{t} \\
=& K \int_{R^k} ||\boldsymbol{u}||^p \Big[\prod_{1\le i\le d, j\in\mathcal{R}_i} \exp(u_{ij})\Big] \Big(\prod_{i=1}^d \int_0^\infty t_i^{k_i} \exp\Big\{ - t_i \Big[1 + \sum_{j\in\mathcal{R}_i} \exp(u_{ij})\Big]\Big\} dt_i\Big) d\boldsymbol{u} \\
=& K_1 \int_{R^k} ||\boldsymbol{u}||^p \Big[\prod_{1\le i\le d, j\in\mathcal{R}_i} \exp(u_{ij})\Big] \Big(\prod_{i=1}^d \Big[1 + \sum_{j\in\mathcal{R}_i} \exp(u_{ij})\Big]^{-(k_i+1)}\Big) d\boldsymbol{u} < \infty,
\end{aligned}$$

where $K_1 > 0$ is a constant.

Necessity: If X^* is not of full rank, then X^{**} is not of full rank as well. In this case, it is easy to see from (2.7) that the posterior distribution of $\boldsymbol{\beta}$ is not proper. Now assume that $(C2)$ is not satisfied. Since $(C2)$ does not hold, by Lemma 2.1, there exists a non-zero vector $\boldsymbol{b} = (b_1, b_2, \ldots, b_p)' \in R^p$ such that

$$\boldsymbol{b}'(\boldsymbol{x}_j - \boldsymbol{x}_i) \le 0 \text{ for } j \in \mathcal{R}_i,\ i = 1, 2, \ldots, d.$$

Without loss of generality, assume that $b_1 \neq 0$. Taking the transformation, $\boldsymbol{\beta} = s_1\boldsymbol{b} + (0, s_2, \ldots, s_p)'$, yields

$$\begin{aligned}
&\int_{R^p} L_p(\boldsymbol{\beta}|D_{obs}) d\boldsymbol{\beta} \\
=& \int_{R^p} \int_{R^{+d}} \exp(-\sum_{i=1}^d t_i) \prod_{1\le i\le d, j\in\mathcal{R}_i} F((\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i} d\boldsymbol{t} d\boldsymbol{\beta} \\
=& \int_{R^{+d}} \exp(-\sum_{i=1}^d t_i)|b_1| \int_{R^p} \prod_{1\le i\le d, j\in\mathcal{R}_i} F(s_1(\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{b} + (\boldsymbol{x}_i - \boldsymbol{x}_j)'(0, s_2, \ldots, s_p))^{t_i} d\boldsymbol{s} d\boldsymbol{t}.
\end{aligned}$$

Letting $\eta > 0$, we get

$$\begin{aligned}
&\int_{R^p} \prod_{1\le i\le d, j\in\mathcal{R}_i} F(s_1(\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{b} + (\boldsymbol{x}_i - \boldsymbol{x}_j)'(0, s_2, \ldots, s_p))^{t_i} d\boldsymbol{s} \\
\ge& \int_{s_1\ge 0,\ \ |s_l|\le\eta, 2\le l\le p} \left[\prod_{1\le i\le d, j\in\mathcal{R}_i} F(-p||\boldsymbol{x}_i - \boldsymbol{x}_j||\eta)^{t_i} \right] d\boldsymbol{s} \\
\ge& F(-\delta)^k \int_{s_1\ge 0,\ \ |s_l|\le\eta, 2\le l\le p} d\boldsymbol{s} = \infty,
\end{aligned}$$

where $\delta = p\eta \max_{1\le i\le d,\ j\in\mathcal{R}_i} ||\boldsymbol{x}_i - \boldsymbol{x}_j||$. Thus $\int_{R^p} L_p(\boldsymbol{\beta}|D_{obs})d\boldsymbol{\beta} = \infty$. ■

Remark 2.1: In X^* defined by (2.5), the rows corresponding to $\delta_i = 0$ or $\boldsymbol{x}_j = \boldsymbol{x}_i$ can be excluded. Thus, the effective numbers of rows in X^* can be reduced substantially.

Remark 2.2: When conditions $(C1)$ and $(C2)$ are satisfied for a subset of the data, the posterior is still proper. To see this, we assume that the subset consists of the first n^* observations. Then we have

$$L_p(\boldsymbol{\beta}|D_{obs}) \le \prod_{i=1}^{n^*} \left[\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{\sum_{j\in\mathcal{R}(y_i), j\le n^*} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})} \right]^{\delta_i}.$$

The propriety of the posterior can obtain by simply applying Theorem 2.1 to the above upper bound. These subset conditions are only sufficient but not necessary. However, this result is particularly useful for a large dataset, for which checking conditions $(C1)$ and $(C2)$ may not be computationally feasible.

Our second approach is to carry out Bayesian inference using the full likelihood function

$$L(\boldsymbol{\beta}, h_0|D) = \Big[\prod_{i=1}^{n} (h_0(y_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}))^{\delta_i} \Big] \exp\{-\sum_{j=1}^{n} H_0(y_j)\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\}. \tag{2.8}$$

In this case, Kalbfleisch (1978) and Sinha, Ibrahim, and Chen (2003) show that the partial likelihood in (2.1) can be obtained as a limiting case of the marginal posterior of $\boldsymbol{\beta}$ with continuous time survival data under a gamma process prior for the cumulative baseline hazard function using the likelihood in (2.8). The gamma process is defined as follows. Let $\mathcal{G}(a, b)$ denote the gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$. Let $\psi(t), t \ge 0$, be an increasing left continuous function such that $\psi(0) = 0$, and let $Z(t), t \ge 0$, be a stochastic process with the properties: (i) $Z(0) = 0$; (ii) $Z(t)$ has independent increments in disjoint intervals; and (iii) for $t > s$, $Z(t) - Z(s) \sim \mathcal{G}(c_0(\psi(t) - \psi(s)), c_0)$. Then the process $\{Z(t) : t \ge 0\}$ is called a gamma process and is denoted by $Z(t) \sim \mathcal{GP}(\psi(t), c_0)$, where $\psi(t)$ is the mean of $Z(t)$ and c_0 is a precision or confidence parameter about the prior mean $\psi(t)$.

Now assume the baseline cumulative hazard function $H_0(y) \sim \mathcal{GP}(H^*, c_0)$, where $H^*(y)$ is a known increasing differentiable function and $c_0 > 0$. We further assume that the observed failure times are all distinct and write $y_1 < y_2 < \cdots < y_n$. Assuming that the parameters of the distributions for the C_i's and $(\boldsymbol{\beta}, H_0)$ are distinct and independent a priori, then following Sinha, Ibrahim, and Chen (2003), the marginal posterior distribution of $\boldsymbol{\beta}$ using $\pi(\boldsymbol{\beta}) \propto 1$ is given by

$$\pi(\boldsymbol{\beta}|D_{obs}) \propto \prod_{i=1}^{n} \exp\left\{c_0 H^*(y_i)\log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} \times \left\{-c_0 h^*(y_i)\log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\}^{\delta_i}, \tag{2.9}$$

where $h^*(y) = \frac{dH^*(y)}{dy}$ and $A_i = \sum_{j \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})$. Under some regularity conditions, Sinha, Ibrahim, and Chen (2003) show that

$$\lim_{c_0 \to 0} K(c_0)\pi(\boldsymbol{\beta}|D_{obs}) = L_p(\boldsymbol{\beta}|D_{obs}), \tag{2.10}$$

where $K(c_0) > 0$ is a function of c_0 and independent of $\boldsymbol{\beta}$, and $L_p(\boldsymbol{\beta}|D)$ is the partial likelihood (2.1).

Next we introduce two useful results.

Lemma 2.2 *For $c > 0$, $\delta = 0$ or 1 and $0 < x < 1$, we have*

$$x^\delta (1-x)^c \le \exp\{c\log(1-x)\}(-\log(1-x))^\delta \le K_c x^\delta (1-x)^{c/2}, \tag{2.11}$$

where K_c is a finite positive constant depending only on c.

PROOF. The left hand side of the inequality is obvious because $-\log(1-x) \ge x$. To prove the right hand side, consider the following two cases:

Case 1. $0 < x \le 2/3$. Then

$$-\log(1-x) \le 2x$$

and hence

$$\exp(c\log(1-x))(-\log(1-x))^\delta \le (2x)^\delta (1-x)^c \le 2x^\delta (1-x)^{c/2}.$$

Case 2. $1/2 < x < 1$. Then

$$\begin{aligned} &\exp(c\log(1-x))(-\log(1-x))^\delta \\ &= \exp(.5c\log(1-x))\exp(-0.5c(-\log(1-x)))(-\log(1-x))^\delta \\ &\le \exp(.5c\log(1-x)) \sup_{t \ge \log 3} te^{-.5ct} \\ &\le 2 \sup_{t \ge \log 3} te^{-.5ct}(1-x)^{.5c}x^\delta. \end{aligned}$$

This proves (2.11) ■

Lemma 2.3 *For $c > 0$, we have*

$$\prod_{i=1}^{n} \exp\left\{cH^*(y_i)\log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} = \prod_{i=1}^{n} E[\exp(-h_i A_i)|ch_{0i}, c_0], \tag{2.12}$$

where $h_{0i} = H^*(y_i) - H^*(y_{i-1})$, $H^*(y_0) \equiv 0$, *and the expectation of* h_i *is taken with respect to the gamma distribution* $\mathcal{G}(ch_{0i}, c_0)$.

Proof. Observe that

$$E[\exp(-h_i A_i)|ch_{0i}, c_0] = \exp\left\{ch_{0i} \log\left(\frac{c_0}{c_0 + A_i}\right)\right\} \tag{2.13}$$

and

$$\prod_{i=1}^{n} \exp\left\{cH^*(y_i) \log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} = \prod_{i=1}^{n} \exp\left\{cH^*(y_i) \log\left(\frac{c_0 + A_{i+1}}{c_0 + A_i}\right)\right\}$$
$$= \prod_{i=1}^{n} \exp\left\{ch_{0i} \log\left(\frac{c_0}{c_0 + A_i}\right)\right\}. \tag{2.14}$$

Thus, (2.12) directly follows from (2.13) and (2.14). ■

Write

$$\pi^*(\boldsymbol{\beta}|D_{obs}) = \prod_{i=1}^{n} \exp\left\{c_0 H^*(y_i) \log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} \times \left\{-\log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\}^{\delta_i}, \tag{2.15}$$

which is the unnormalized posterior density of $\boldsymbol{\beta}$. Using Lemma 2.2, we have the inequalities:

$$\pi^*(\boldsymbol{\beta}|D_{obs}) \leq K_{c_0} \prod_{i=1}^{n} \exp\left\{\frac{c_0}{2} H^*(y_i) \log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} \times \left\{\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right\}^{\delta_i}, \tag{2.16}$$

where $K_{c_0} > 0$ is a constant, and

$$\pi^*(\boldsymbol{\beta}|D_{obs}) \geq \prod_{i=1}^{n} \exp\left\{c_0 H^*(y_i) \log\left(1 - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right\} \times \left\{\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0 + A_i}\right\}^{\delta_i}. \tag{2.17}$$

Let

$$G^* = (-\delta_i \boldsymbol{x}_i, \delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i), j \in \mathcal{R}(y_i), \boldsymbol{x}_i, i = 1, 2, \ldots, n)'. \tag{2.18}$$

We are led to the following theorem.

Theorem 2.2 *The posterior distribution* $\pi(\boldsymbol{\beta}|D_{obs})$ *in (2.9) is proper, i.e.,*

$$\int \pi^*(\boldsymbol{\beta}|D_{obs}) d\boldsymbol{\beta} < \infty,$$

where $\pi^*(\boldsymbol{\beta}|D_{obs})$ *is defined in (2.15), if and only if* ($C1^*$): G^* *is of full rank, and* ($C2^*$): *there exists a positive vector* $\boldsymbol{v}$ *such that* $(G^*)'\boldsymbol{v} = 0$.

Proof. Let $\mathcal{R}_i = \mathcal{R}(y_i) - \{i\}$ and $F(u) = \exp(-\exp(-u))$. Observing that for $\delta = 0$ or 1 and $x > -1$

$$\Big(\frac{1}{1+x}\Big)^{\delta} = \int_0^\infty e^{-t(1+\delta x)}dt, \tag{2.19}$$

we have

$$\begin{aligned}\Big(\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0+A_i}\Big)^{\delta_i} &= \left(\frac{1}{1+c_0\exp(-\boldsymbol{x}_i'\boldsymbol{\beta})+\sum_{j\in\mathcal{R}_i}\exp(-(\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})}\right)^{\delta_i} \\ &= \int_0^\infty \exp\Big(-t_i(1+\delta_i(c_0\exp(-\boldsymbol{x}_i'\boldsymbol{\beta})+\sum_{j\in\mathcal{R}_i}\exp(-(\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})))\Big)dt_i \\ &= \int_0^\infty e^{-t_i}F(\boldsymbol{x}_i'\boldsymbol{\beta})^{c_0t_i\delta_i}\prod_{j\in\mathcal{R}_i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i\delta_i}dt_i.\end{aligned} \tag{2.20}$$

By Lemma 2.3, (2.16), and (2.17), we obtain

$$\begin{aligned}\pi^*(\boldsymbol{\beta}|D_{obs}) \le & K_{c_0}\prod_{i=1}^n\Big[\int_0^\infty e^{-t_i}F(\boldsymbol{x}_i'\boldsymbol{\beta})^{c_0t_i\delta_i}\prod_{j\in\mathcal{R}_i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i\delta_i}dt_i\Big]E[\exp(-h_iA_i)|\frac{c_0}{2}h_{0i},c_0] \\ =&\prod_{i=1}^n\Big[\int_0^\infty e^{-t_i}F(\boldsymbol{x}_i'\boldsymbol{\beta})^{ct_i\delta_i}\prod_{j\in\mathcal{R}_i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i\delta_i}dt_i\Big]E[\prod_{j\in\mathcal{R}(y_i)}F(-\boldsymbol{x}_j)^{h_i}\Big|\frac{c_0}{2}h_{0i},c_0],\end{aligned} \tag{2.21}$$

where the expectation $E[.|\frac{c_0}{2}h_{0i},c_0]$ for h_i is taken with respect to the gamma distribution $\mathcal{G}((c_0/2)h_{0i},c_0)$, and

$$\begin{aligned}\pi^*(\boldsymbol{\beta}|D_{obs}) \ge &\prod_{i=1}^n\Big[\int_0^\infty e^{-t_i}F(\boldsymbol{x}_i'\boldsymbol{\beta})^{c_0t_i\delta_i}\prod_{j\in\mathcal{R}_i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i\delta_i}dt_i\Big]E[\exp(-h_iA_i)|c_0h_{0i},c_0] \\ =&\prod_{i=1}^n\Big[\int_0^\infty e^{-t_i}F(\boldsymbol{x}_i'\boldsymbol{\beta})^{c_0t_i\delta_i}\prod_{j\in\mathcal{R}_i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\boldsymbol{\beta})^{t_i\delta_i}dt_i\Big]E[\prod_{j\in\mathcal{R}(y_i)}F(-\boldsymbol{x}_j)^{h_i}\Big|c_0h_{0i},c_0].\end{aligned} \tag{2.22}$$

Thus, using (2.21) for sufficiency and (2.22) for necessity, the rest of the proof directly follows from the proof of Theorem 2.1, and thus the details are omitted for brevity. ■

Remark 2.3: It is easy to show that if conditions $(C1)$ and $(C2)$ in Theorem 2.1 hold, then conditions $(C1^*)$ and $(C2^*)$ in Theorem 2.2 are automatically satisfied. Thus, the necessary and sufficient conditions for propriety of the posterior distribution for $\boldsymbol{\beta}$ with an improper uniform prior based on the full likelihood function are weaker than those for propriety of the posterior distribution for $\boldsymbol{\beta}$ based on the partial likelihood. Thus a gamma process prior on the cumulative baseline hazard function, being inherently proper, brings additional information into the model resulting in weaker conditions.

When ties are present, as discussed in Klein and Moeschberger (1997, Chapter 8), the partial likelihood may be defined as

$$L_{pt}(\boldsymbol{\beta}|D_{obs}) = \prod_{i=1}^{d} \frac{\exp(\boldsymbol{z}_i'\boldsymbol{\beta})}{\left[\sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_i\boldsymbol{\beta})\right]^{d_i}}, \tag{2.23}$$

where $d = \sum_{i=1}^{n} \delta_i$, $\boldsymbol{z}_i = \sum_{j\in\mathcal{D}_i} \boldsymbol{x}_j$, d_i = the number of deaths at y_i, and $\mathcal{D}_i$ is the set of all individuals who die at time y_i. Note that the partial likelihood given by (2.23) is the likelihood of Breslow (1974), and the Breslow likelihood is the default choice in SAS to handle ties in the failure times.

Since L_{pt} can be rewritten as

$$L_{pt}(\boldsymbol{\beta}|D_{obs}) = \prod_{i=1}^{n} \frac{\exp(\delta_i\boldsymbol{x}_i'\boldsymbol{\beta})}{\left[\sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_i\boldsymbol{\beta})\right]^{\delta_i}},$$

Theorem 2.1 can be easily extended to the cases when ties are present. However, in the presence of missing covariates, there is no literature available defining the full likelihood with a gamma process prior on the baseline cumulative hazard. To circumvent the ties issue, we define

$$D_{obs}(\boldsymbol{\epsilon}) = \{y_i^* = y_i - \delta_i\epsilon_i, \delta_i, \boldsymbol{x}_i,\ i = 1, 2, \ldots, n\}. \tag{2.24}$$

We call $D_{obs}(\boldsymbol{\epsilon})$ the untied data of D_{obs}. Since $y_i^* = y_i$ when $\delta_i = 0$, the values of the censoring times remain unchanged. Also, it is easy to see that $\lim_{\boldsymbol{\epsilon}\to 0} D_{obs}(\boldsymbol{\epsilon}) = D_{obs}$. In (2.24) we choose the ϵ_i such that (a) $\epsilon_i > 0$, (b) the y_i^*'s associated with $\delta_i = 1$ are all distinct, and (c) the y_i^*'s associated with $\delta_i = 1$ are distinct from the y_i^*'s associated with $\delta_i = 0$. In this fashion, all failure times are distinct and the failure times are different from the censoring times in the untied data $D_{obs}(\boldsymbol{\epsilon})$.

Let $y_{(1)}^* < y_{(2)}^* < \cdots < y_{(n^*(\boldsymbol{\epsilon}))}^*$, where $n^*(\epsilon) \le n$, denotes the distinct ordered failure or censoring times in $D_{obs}(\boldsymbol{\epsilon})$. Therefore, if $h_j = H_0(y_{(j)}^*) - H_0(y_{(j-1)}^*)$, then $h_j \sim \mathcal{G}(c_0 h_{0j}, c_0)$, where $h_{0j} = H^*(y_{(j)}^*) - H^*(y_{(j-1)}^*)$. Suppose $y_{(j)}^* = y_{i_j} - \delta_{i_j}\epsilon_{i_j}$. Note that if $y_{(j-1)}^* < y_{(j)}^*$ are two failure times which correspond to two tied failure times $y_{i_j} = y_{i_{j-1}}$ in the original data D_{obs}, then $h_{0j} \to 0$ and $\mathcal{R}(y_{(j-1)}^*) = \mathcal{R}(y_{i_{j-1}})$ as $\epsilon_{i_j} \to 0$ and $\epsilon_{i_{j-1}} \to 0$. Let $A_j^* = \sum_{l\in\mathcal{R}(y_{(j)}^*)} \exp(\boldsymbol{x}_l'\boldsymbol{\beta})$ and let E_{GP} denote expectation with respect to the gamma process prior. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$ and $\boldsymbol{y}^* = (y_1^*, y_2^*, \ldots, y_n^*)'$ denote the random failure times and the observed times, respectively. Then, we have

$$\begin{aligned} P(\boldsymbol{Y} > \boldsymbol{y}^*|X, \boldsymbol{\beta}, H_0) &= K_C \exp\Big\{ -\sum_{i=1}^{n} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) H_0(y_i^*)\Big\} \\ &= K_C \exp\Big\{ -\sum_{j=1}^{n^*(\boldsymbol{\epsilon})} h_j \sum_{l\in\mathcal{R}(y_{(j)}^*)} \exp(\boldsymbol{x}_l'\boldsymbol{\beta})\Big\} = K_C \exp\Big(-\sum_{j=1}^{n^*(\boldsymbol{\epsilon})} h_j A_j^*\Big), \end{aligned}$$

where $X = (\boldsymbol{x}_i,\ i = 1, 2, \ldots, n)'$ and K_C is the part adhering to the distributions of the censoring times C_i's, and

$$E_{GP}\Big[P(\boldsymbol{Y} > \boldsymbol{y}^*|X, \boldsymbol{\beta}, H_0)|c_0, H^*\Big] = K_C \prod_{j=1}^{n^*(\boldsymbol{\epsilon})} \left(\frac{c_0}{c_0 + A_j^*}\right)^{c_0 h_{0j}}$$
$$= K_C \prod_{j=1}^{n^*(\boldsymbol{\epsilon})} \exp\left\{c_0 H^*(y^*_{(j)}) \log\left(\frac{c_0 + A^*_{j+1}}{c_0 + A^*_j}\right)\right\}.$$

As we assume that the parameters of the distributions for the C_i's and $(\boldsymbol{\beta}, H_0)$ are distinct and independent a priori, inference about the parameters of the distributions for the C_i's does not affect inference about $\boldsymbol{\beta}$. Thus, ignoring K_C, the likelihood function based on the data $D_{obs}(\boldsymbol{\epsilon})$ is given by

$$L_t(\boldsymbol{\beta}|D_{obs}(\boldsymbol{\epsilon})) = \prod_{j=1}^{n^*(\boldsymbol{\epsilon})} \Bigg\{ \exp\left[c_0 H^*(y^*_{(j)}) \log\left(\frac{c_0 + A^*_{j+1}}{c_0 + A^*_j}\right)\right]$$
$$\times \left[-c_0 h^*(y^*_{(j)}) \log\left(1 - \frac{\exp(x'_{i_j}\boldsymbol{\beta})}{c_0 + A^*_j}\right)\right]^{\delta_{i_j}}\Bigg\}. \tag{2.25}$$

Note that when $\delta_{i_j} = 1$, $A^*_j - A^*_{j+1} = \exp(x'_{i_j}\boldsymbol{\beta})$. By letting $\boldsymbol{\epsilon} \to 0$, we obtain

$$L_t(\boldsymbol{\beta}|D_{obs}) = \lim_{\boldsymbol{\epsilon}\to 0} L_t(\boldsymbol{\beta}|D_{obs}(\boldsymbol{\epsilon})) = \prod_{j=1}^{n^*} \Bigg\{ \exp\left[c_0 H^*(y_{(j)}) \log\left(\frac{c_0 + A_{j+1}}{c_0 + A_j}\right)\right]$$
$$\times \prod_{l \in \mathcal{D}(y_{(j)})} \left[-c_0 h^*(y_{(j)}) \log\left(1 - \frac{\exp(x'_l\boldsymbol{\beta})}{c_0 + A_j}\right)\right]^{\delta_{i_j}}\Bigg\}, \tag{2.26}$$

where $\mathcal{D}(y_{(j)}) = \{l: \ \ y_l = y_{(j)}, \delta_l = 1\}$ (i.e., the failure set at $y_{(j)}$, $y_{(1)} < y_{(2)} < \cdots < y_{(n^*)}$) are the n^* distinct failure and censoring times in D_{obs}, and $A_j = \sum_{l \in \mathcal{R}(y_{(j)})} \exp(\boldsymbol{x}'_l\boldsymbol{\beta})$. We call $L_t(\boldsymbol{\beta}|D_{obs})$ the *full likelihood when ties are present.* Similar to Sinha, Ibrahim, and Chen (2003), we can show that

$$\lim_{c_0 \to 0} K(c_0) L_t(\boldsymbol{\beta}|D_{obs}) = L_{pt}(\boldsymbol{\beta}|D_{obs}),$$

where $K(c_0) > 0$ is a constant and $L_{pt}(\boldsymbol{\beta}|D_{obs})$ is given by (2.23), which is the partial likelihood when ties are present. Using (2.25), the two inequalities for $L_t(\boldsymbol{\beta}|D_{obs}(\boldsymbol{\epsilon}))$ analogous to (2.16) and (2.17) are given by

$$L_t(\boldsymbol{\beta}|D_{obs}(\boldsymbol{\epsilon})) \le K_{c_0}(\boldsymbol{\epsilon}) \prod_{j=1}^{n^*(\boldsymbol{\epsilon})} \Bigg\{ \exp\left[(c_0/2) H^*(y^*_{(j)}) \log\left(\frac{c_0 + A^*_{j+1}}{c_0 + A^*_j}\right)\right]$$
$$\times \left[c_0 h^*(y^*_{(j)}) \left(\frac{\exp(x'_{i_j}\boldsymbol{\beta})}{c_0 + A^*_j}\right)\right]^{\delta_{i_j}}\Bigg\}, \tag{2.27}$$

where $K_{c_0}(\boldsymbol{\epsilon}) > 0$ is a constant, and

$$L_t(\boldsymbol{\beta}|D_{obs}(\boldsymbol{\epsilon})) \geq \prod_{j=1}^{n^*(\boldsymbol{\epsilon})} \left\{ \exp\left[c_0 H^*(y^*_{(j)}) \log\left(\frac{c_0 + A^*_{j+1}}{c_0 + A^*_j}\right)\right] \times \left[c_0 h^*(y^*_{(j)}) \left(\frac{\exp(x'_{i_j}\boldsymbol{\beta})}{c_0 + A^*_j}\right)\right]^{\delta_{i_j}} \right\}. \tag{2.28}$$

From the proof of Lemma 2.2, we can see that $\lim_{\epsilon \to 0} K_{c_0}(\boldsymbol{\epsilon}) = K_{c_0} > 0$. Thus, by taking $\boldsymbol{\epsilon} \to 0$, (2.27) and (2.28) reduce to

$$\begin{aligned} L_t(\boldsymbol{\beta}|D_{obs}) \leq & K_{c_0} \prod_{j=1}^{n^*} \left\{ \exp\left[\frac{c_0}{2} H^*(y_{(j)}) \log\left(\frac{c_0 + A_{j+1}}{c_0 + A_j}\right)\right] \right. \\ & \left. \times \prod_{l \in \mathcal{D}(y_{(j)})} \left[c_0 h^*(y_{(j)}) \left(\frac{\exp(x'_l\boldsymbol{\beta})}{c_0 + A_j}\right)\right]^{\delta_{i_j}} \right\} \end{aligned} \tag{2.29}$$

and

$$\begin{aligned} L_t(\boldsymbol{\beta}|D_{obs}) \geq & \prod_{j=1}^{n^*} \left\{ \exp\left[c_0 H^*(y_{(j)}) \log\left(\frac{c_0 + A_{j+1}}{c_0 + A_j}\right)\right] \right. \\ & \left. \times \prod_{l \in \mathcal{D}(y_{(j)})} \left[c_0 h^*(y_{(j)}) \left(\frac{\exp(x'_l\boldsymbol{\beta})}{c_0 + A_j}\right)\right]^{\delta_{i_j}} \right\}. \end{aligned} \tag{2.30}$$

Assuming $\pi(\boldsymbol{\beta}) \propto 1$, we have $\pi_t(\boldsymbol{\beta}|D_{obs}) \propto L_t(\boldsymbol{\beta}|D_{obs})$. Using (2.29) and (2.30), Theorem 2.2 can be easily extended to the case where ties are present.

3 Posterior Inference in the Presence of Missing Covariates

In the presence of missing covariates, we use the full likelihood based on the gamma process prior for the cumulative baseline hazard. We further assume that the distribution of the censoring time C_i does not depend on the missing covariates and the missingness is MAR. To unify the notation, we assume that ties may be present and $y_1 \leq y_2 \leq \cdots \leq y_n$. We write $\boldsymbol{x}_i = (\boldsymbol{x}'_{i,mis}, \boldsymbol{x}'_{i,obs})'$, $D_{obs} = (y_i, \delta_i, \boldsymbol{x}_{i,obs},\ i = 1, 2, \ldots, n)$, and $D = (y_i, \delta_i, \boldsymbol{x}_{i,obs}, \boldsymbol{x}_{i,mis},\ i = 1, 2, \ldots, n)$. Then the full likelihood function given in (2.26) can be rewritten as

$$L_t(\boldsymbol{\beta}|D) = \prod_{i=1}^{n} \left\{ \exp\left[c_0 H^*(y_i) \log\left(\frac{c_0 + A_{i+1}}{c_0 + A_i}\right)\right] \left[-c_0 h^*(y_i) \log\left(1 - \frac{\exp(x'_i\boldsymbol{\beta})}{c_0 + A_i}\right)\right]^{\delta_i} \right\}, \tag{3.1}$$

where $A_i = \sum_{l \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}'_i\boldsymbol{\beta})$. Note that in (3.1), when $y_i = y_{i+1}$, then $A_i = A_{i+1}$, which leads to $\log\left(\frac{c_0 + A_{i+1}}{c_0 + A_i}\right) = 0$. Let $f(\boldsymbol{x}_i|\boldsymbol{\alpha})$ denote the joint distribution for $\boldsymbol{x}_i$. Also let $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha})$ denote a joint prior distribution for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Then the joint posterior distribution for

$(\boldsymbol{\beta}, \boldsymbol{\alpha})$ based on the observed data is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs}) \propto \int L_t(\boldsymbol{\beta}|D)\Big[\prod_{i=1}^n f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})\Big] d\boldsymbol{x}_{mis} \times \pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha}), \tag{3.2}$$

where $\boldsymbol{x}_{mis} = (\boldsymbol{x}'_{i,mis}, i = 1, 2, \ldots, n)'$. Let

$$\pi(\boldsymbol{\alpha}|D_{obs}) \propto \pi^*(\boldsymbol{\alpha}|D_{obs}) \equiv \int \prod_{i=1}^n f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha}) d\boldsymbol{x}_{mis}\pi(\boldsymbol{\alpha}). \tag{3.3}$$

We are led to the following theorem.

Theorem 3.1 *Assume that $\pi(\boldsymbol{\beta}) \propto 1$. (i) If the x_{ij}'s are bounded, i.e., $a_i \le x_{ij} \le b_i$, define G^{**} to be $G^{**} = (-\delta_i \boldsymbol{x}_i^*, \delta_i(\boldsymbol{x}_j^* - \boldsymbol{x}_i^*), j \in \mathcal{R}(y_i), \delta_j = 0, (1-\delta_i)\boldsymbol{x}_i^*, 1 \le i \le n)'$, where $x_i^* = ((\boldsymbol{x}_{i,mis}^R)', \boldsymbol{x}'_{i,obs})'$ and each component of $\boldsymbol{x}_{i,mis}^R$ is equal to either $a_i^* = \delta_i a_i + (1-\delta_i) b_i$ or $b_i^* = (1-\delta_i)a_i + \delta_i b_i$ for all i. Then, the joint posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs})$ in (3.2) is proper if the following conditions are satisfied: (C1**) $\int \pi^*(\boldsymbol{\alpha}|D_{obs}) d\boldsymbol{\alpha} < \infty$; (C2**) G^{**} is of full rank; and (C3**) there exists a positive vector $\boldsymbol{v}$ such that ${G^{**}}'\boldsymbol{v} = 0$. (ii) If the x_{ij}'s are unbounded, the joint posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs})$ in (3.2) is proper if condition (C1**) in (i) and conditions (C1*) and (C2*) stated in Theorem 2.2 are satisfied for the completely observed cases.*

Proof. Rewrite (3.1) as

$$L_t(\boldsymbol{\beta}|D) = \prod_{i=1}^n \Bigg\{ \exp\Big[c_0 h_{0i} \log\Big(\frac{c_0}{c_0 + A_i}\Big)\Big] \Big[- c_0 h^*(y_i) \log\Big(1 - \frac{\exp(x_i'\boldsymbol{\beta})}{c_0 + A_i}\Big)\Big]^{\delta_i} \Bigg\}, \tag{3.4}$$

where $h_{0i} = H^*(y_i) - H^*(y_{i-1})$ for $i = 1, 2, \ldots, n$ and $H^*(y_0) = 0$. By (2.29), we have

$$\begin{aligned} L_t(\boldsymbol{\beta}|D) \le & K_{c_0} \prod_{i=1}^n \exp\Big[\frac{c_0}{2} h_{0i} \log\Big(\frac{c_0}{c_0 + A_i}\Big)\Big] \Big(\frac{\exp(x_i'\boldsymbol{\beta})}{c_0 + A_i}\Big)^{\delta_i} \\ \le & L_{tU}(\boldsymbol{\beta}|D) \equiv K_{c_0} \prod_{i=1}^n \exp\Big[\frac{c_0}{2} h_{0i} \log\Big(\frac{c_0}{c_0 + A_i^0}\Big)\Big] \Big(\frac{\exp(x_i'\boldsymbol{\beta})}{c_0 + \exp(x_i'\boldsymbol{\beta}) + A_i^0}\Big)^{\delta_i}, \end{aligned} \tag{3.5}$$

where $A_i^0 = \sum_{j \in \mathcal{R}(y_i), \delta_j = 0} \exp(x_j'\boldsymbol{\beta})$. Observe that $L_{tU}(\boldsymbol{\beta}|D)$ is an increasing function in $\boldsymbol{x}_i'\boldsymbol{\beta}$ for $\delta_i = 1$ and a decreasing function in $\boldsymbol{x}_j'\boldsymbol{\beta}$ for $\delta_j = 0$. For $1 \le l \le p$, let $x_{il}^R = b_i^* = (1-\delta_i)a_i + \delta_i b_i$ if $\beta_l \ge 0$ and $x_{il}^R = a_i^* = \delta_i a_i + (1-\delta_i) b_i$ if $\beta_l < 0$. Write $x_i^* = ((\boldsymbol{x}_{i,mis}^R)', \boldsymbol{x}'_{i,obs})'$ and $\boldsymbol{x}_{i,mis}^R = (x_{il}^R, r_{il} = 0, 1 \le l \le p)'$, where $r_{il} = 0$ if x_{il} is missing and $r_{il} = 1$ if x_{il} is observed. Then we have

$$\begin{aligned} & L_t(\boldsymbol{\beta}|D) \le L_{tU}(\boldsymbol{\beta}|D) \\ \le & L_{tU}(\boldsymbol{\beta}|D^R) \equiv K_{c_0} \prod_{i=1}^n \exp\Big[\frac{c_0}{2} h_{0i} \log\Big(\frac{c_0}{c_0 + A_i^{0*}}\Big)\Big] \Big(\frac{(\exp(x_i^*)'\boldsymbol{\beta})}{c_0 + \exp((x_i^*)'\boldsymbol{\beta}) + A_i^{0*}}\Big)^{\delta_i}, \end{aligned} \tag{3.6}$$

where $A_i^{0*} = \sum_{j \in \mathcal{R}(y_i), \delta_j = 0} \exp((x_j^*)'\boldsymbol{\beta})$ and $D^R = (y_i, \delta_i, \boldsymbol{x}_i^*, \ i = 1, 2, \ldots, n)$. Since D^R does not depend on $\boldsymbol{x}_{i,mis}$, (3.6) yields

$$\begin{aligned} &\int L_t(\boldsymbol{\beta}|D)\Big[\prod_{i=1}^n f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})\Big] d\boldsymbol{x}_{mis} \times \pi(\boldsymbol{\alpha}) \\ \le & L_{tU}(\boldsymbol{\beta}|D^R) \int \prod_{i=1}^n f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})\Big] d\boldsymbol{x}_{mis} \times \pi(\boldsymbol{\alpha}) = L_{tU}(\boldsymbol{\beta}|D^R)\pi^*(\boldsymbol{\alpha}|D_{obs}). \end{aligned} \tag{3.7}$$

Under conditions $(C1^{**})$, $(C2^{**})$, and $(C3^{**})$, the rest of the proof for (i) directly follows from (3.7) and the proof of Theorem 2.2. Observing that

$$L_t(\boldsymbol{\beta}|D) \le K_{c_0} \prod_{i=1}^n \exp\Big[\frac{c_0}{2} r_i h_{0i} \log\Big(\frac{c_0}{c_0 + A_i^1}\Big)\Big] \Big(\frac{(\exp(x_i)'\boldsymbol{\beta})}{c_0 + \exp((x_i)'\boldsymbol{\beta}) + A_i^1}\Big)^{r_i \delta_i},$$

where $r_i = 1$ if $\boldsymbol{x}_i$ is completely observed, $r_i = 0$ if at least one component of $\boldsymbol{x}_i$ is missing, and $A_i^1 = \sum_{j \in \mathcal{R}_i, r_j = 1} \exp(x_j'\boldsymbol{\beta})$, the proof for (ii) is straightforward. ■

Remark 3.1: If each missing component of $\boldsymbol{x}_i$ is discrete and bounded, then $(C2^{**})$ and $(C3^{**})$ are also necessary.

Remark 3.2: Chen and Ibrahim (2001) provide a comprehensive set of guidelines for specifying the joint distribution of the covariate vector $\boldsymbol{x}_i$ through a series of one dimensional conditional distributions. We can show that condition $(C1^{**})$ holds for various covariate distributions considered in Chen and Ibrahim (2001).

Remark 3.2: The main intuition behind Theorem 3.1 is that when the posterior is proper under conditions $(C2^{**})$ and $(C3^{**})$, for the most extreme possible values of the missing covariates, the posterior is also proper for any intermediate values of the missing covariates and averaging over the missing values will not affect the propriety of the posterior. In Theorem 3.1, the elements of the matrix G^{**} corresponding to missing covariates are "filled-in" by either $a_i^* = \delta_i a_i + (1-\delta_i) b_i$ or $b_i^* = (1-\delta_i) a_i + \delta_i b_i$, where a_i^* and b_i^* are in fact the two extreme possible values of the missing covariates when the x_{ij}'s are bounded. When x_{ij}'s are unbounded, we do not have finite extreme possible values. Thus, the portion of the data with missing covariates may not help to establish the propriety of the posterior.

4 Posterior Computation

In the presence of missing covariates, from (3.4), it does not appear possible to to carry out the posterior computation analytically. Even with modern sampling based-techniques such as Markov chain Monte Carlo (MCMC) sampling, the implementation of MCMC sampling can be still quite challenging. To overcome such difficulties, we propose a novel posterior sampling algorithm to sample from $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs})$ in (3.2) by introducing three sets of latent variables. Towards this goal, observe that

$$\left[-\log\left(1 - \frac{\exp(x_i'\boldsymbol{\beta})}{c_0 + A_i}\right)\right]^{\delta_i} = \int_0^1 \left[\frac{\frac{\exp(x_i'\boldsymbol{\beta})}{c_0+A_i}}{1 - w_i \frac{\exp(x_i'\boldsymbol{\beta})}{c_0+A_i}} \right]^{\delta_i} dw_i$$

$$= \int_0^1 \left[\frac{\exp(x_i'\boldsymbol{\beta})}{c_0 + (1-w_i)\exp(x_i'\boldsymbol{\beta}) + \sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}\right]^{\delta_i} dw_i$$

$$= \int_0^1 \int_0^\infty \exp(\delta_i \boldsymbol{x}_i'\boldsymbol{\beta}) \exp\Big\{ -t_i\Big[(1-\delta_i) + \delta_i\Big(c_0 + (1-w_i)\exp(x_i'\boldsymbol{\beta}) + \sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\Big)\Big]\Big\} dt_i dw_i. \tag{4.1}$$

Also, by letting $\nu_i = 0$ if $H^*(y_i) = H^*(y_{i-1})$ and $\nu_i = 1$ if $H^*(y_i) > H^*(y_{i-1})$, we have

$$\prod_{i=1}^n \exp\left[c_0 h_{0i}\log\left(\frac{c_0}{c_0 + A_i}\right)\right]$$
$$= \prod_{i=1}^n \int_0^\infty \exp(-\nu_i h_i A_i)\frac{c_0^{c_0 h_{0i} + (1-\nu_i)} h_i^{c_0 h_{0i} + (1-\nu_i) - 1}}{\Gamma(c_0 h_{0i} + (1-\nu_i))}\exp(-c_0 h_i) dh_i. \tag{4.2}$$

Let $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$, $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)'$, and $\boldsymbol{h} = (h_1, h_2, \ldots, h_n)'$ be three sets of latent variables. Assume that the joint distribution of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h})$ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}|D_{obs})$$
$$\propto \prod_{i=1}^n \Bigg[\exp(\delta_i \boldsymbol{x}_i'\boldsymbol{\beta}) \exp\Big\{ -t_i\Big[(1-\delta_i) + \delta_i\Big(c_0 + (1-w_i)\exp(x_i'\boldsymbol{\beta}) + \sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\Big)\Big]\Big\}$$
$$\times h_i^{c_0 h_{0i} + (1-\nu_i) - 1}\exp\{-h_i(c_0 + \nu_i A_i)\} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})\Bigg]\pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha}). \tag{4.3}$$

Let k denote the dimension of all of the missing covariates. By (4.1) and (4.2), it is easy to show that

$$\int_{R^k}\int_{R^{+n}}\int_{(0,1)^n}\int_{R^{+n}} \pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}|D_{obs}) d\boldsymbol{t} d\boldsymbol{w} d\boldsymbol{h} d\boldsymbol{x}_{mis} = \pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs}), \tag{4.4}$$

which is (3.2). In other words, $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs})$ is the marginal distribution of $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}|D_{obs})$.

Using $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}|D_{obs})$, an efficient Gibbs sampling algorithm can be developed. Specifically, we sample from the following conditional distributions in turn: (i) $\pi(\boldsymbol{\beta}|\boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}, D_{obs})$, (ii) $\pi(\boldsymbol{\alpha}|\boldsymbol{x}_{mis}, D_{obs})$, (iii) $\pi(\boldsymbol{x}_{mis}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}, D_{obs})$, (iv) $\pi(\boldsymbol{w}|\boldsymbol{\beta}, \boldsymbol{x}_{mis}, D_{obs})$, (v) $\pi(\boldsymbol{t}|\boldsymbol{\beta}, \boldsymbol{x}_{mis}, \boldsymbol{w}, D_{obs})$, and (vi) $\pi(\boldsymbol{h}|\boldsymbol{\beta}, \boldsymbol{x}_{mis}, D_{obs})$.

For (i), we have

$$\pi(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}, D_{obs})$$
$$\propto \left[\prod_{i=1}^{n} \exp(\delta_i \boldsymbol{x}_i'\boldsymbol{\beta}) \exp\Big\{ - t_i\delta_i\Big((1-w_i)\exp(x_i'\boldsymbol{\beta}) + \sum_{j\in\mathcal{R}_i} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\Big)\Big\} \exp\{-h_i\nu_i A_i\}\right]\pi(\boldsymbol{\beta}).$$

It can be shown that $\pi(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{x}_{mis}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}, D_{obs})$ is log-concave in $\boldsymbol{\beta}$ as long as $\pi(\boldsymbol{\beta})$ is log-concave, which is particularly true when an improper uniform prior, i.e., $\pi(\boldsymbol{\beta}) \propto 1$, is used. Hence we can sample $\boldsymbol{\beta}$ via the adaptive rejection algorithm of Gilks and Wild (1992). For (ii), $\pi(\boldsymbol{\alpha}|\boldsymbol{x}_{mis}, D_{obs}) \propto \Big[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})\Big]\pi(\boldsymbol{\alpha})$. For various covariate distributions specified through a series of one dimensional conditional distributions, sampling $\boldsymbol{\alpha}$ is straightforward. For (iii), given $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{w}$, $\boldsymbol{t}$, $\boldsymbol{h}$, and D_{obs}, the $\boldsymbol{x}_{i,mis}$'s are conditionally independent, and the conditional distribution for the i^{th} missing covariate is given by

$$\pi(\boldsymbol{x}_{i,mis}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{w}, \boldsymbol{t}, \boldsymbol{h}, D_{obs})$$
$$\propto \exp\Big\{\delta_i \boldsymbol{x}_i'\boldsymbol{\beta} - \Big[t_i\delta_i(1-w_i) + h_i\nu_i + \sum_{j:\, y_i\in\mathcal{R}_j} (t_j\delta_j + h_j\nu_j)\Big]\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\Big\} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha}).$$

This conditional independence of the missing covariates is a very attractive property which is totally facilitated by the introduction of the latent variables, and thus makes sampling $\boldsymbol{x}_{mis}$ easy and convenient. Without the latent variables, conditional independence is not obtained and the full conditional distribution of $\boldsymbol{x}_{mis}$ is quite unwieldy and computationally challenging.

For (iv), we use the modified collapsed Gibbs technique of Liu (1994) as discussed in Chen, Shao, and Ibrahim (2000). Specifically, we sample $\boldsymbol{w}$ after integrating out $\boldsymbol{t}$. It turns out that given $\boldsymbol{\beta}$, $\boldsymbol{x}_{mis}$, and D_{obs}, the w_i's are independent and the cumulative distribution function (CDF) for w_i is given by

$$P(w_i \le w|\boldsymbol{\beta}, \boldsymbol{x}_{mis}, D_{obs}) = \begin{cases} \dfrac{\log\left(1-\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0+A_i}w\right)}{\log\left(1-\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{c_0+A_i}\right)} & \text{if } \delta_i = 1, \\ w & \text{if } \delta_i = 0, \end{cases}$$

where $0 < w < 1$. Thus, the inverse CDF method can be directly applied for sampling w_i. For (v) and (vi), the t_i and h_i are conditionally independent. Given t_i given $(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{x}_{mis}, D_{obs})$ follows an exponential distribution with mean $\Big[(1-\delta_i) + \delta_i\Big(c_0 + (1-w_i)\exp(x_i'\boldsymbol{\beta}) + \sum_{j\in\mathcal{R}_i}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\Big)\Big]^{-1}$. h_i given $(\boldsymbol{\beta}, \boldsymbol{x}_{mis}, D_{obs})$ follows a gamma distribution $\mathcal{G}(c_0h_{0i} + (1-\nu_i), c_0 + \nu_i A_i)$. Therefore, sampling $\boldsymbol{t}$ and $\boldsymbol{h}$ is straightforward.

Remark 4.1: Note that if we take $\pi(\boldsymbol{\beta}) \propto 1$ and there are no missing covariates, then $\int_{R^{+n}} \pi(\boldsymbol{\beta}, \boldsymbol{t}|\boldsymbol{w} = 0, \boldsymbol{h} = 0, D_{obs})d\boldsymbol{t} = \pi(\boldsymbol{\beta}|D_{obs})$, where $\pi(\boldsymbol{\beta}|D_{obs})$ is given by (2.4). Thus, the above proposed Gibbs sampling algorithm can be easily modified for sampling $\boldsymbol{\beta}$ while treating (2.1) as the "likelihood function".

5 Melanoma Data

We present this example to illustrate how to check the conditions discussed in Sections 2 and 3 and to demonstrate the proposed methodology for analyzing real data from a cancer clinical trial. We consider data from a phase III melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG). The results from this study have been published by Kirkwood et al. (2000). This study was a clinical trial involving two treatment arms: high dose interferon (IFN) or observation. The results of this study suggested that IFN has a significant impact on disease free and overall survival, which led to FDA approval of this regimen as a standard adjuvant therapy for high risk melanoma patients. Here, disease free survival is defined as the time from randomization until progression of tumor or death, whichever comes first. The dataset used in this example had $n = 427$ patients. In this example, we consider three prognostic factors: x_1 = treatment (2 levels: observation, interferon, coded as 0 and 1), x_2 = type of primary (2 levels: nodular, other, coded as 1 and 0), and x_3 = Clark level (2 levels: Reticular dermis (IV), other, coded as 1 and 0). For these three prognostic factors, x_2 and x_3 had missing information and x_1 was completely observed for all cases. In this dataset, there is a total fraction of 16.55% missing covariate information on these two covariates. The outcome variable (y_i in years) was time to relapse, which is continuous and subject to right censoring, and δ_i denotes the censoring indicator which equals 1 if the i^{th} subject relapsed, and 0 otherwise. The median follow up time is 1.64 years.

We use the proposed methods to estimate the regression coefficients assuming the missing covariates are MAR. We consider a Cox regression model for $[y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, h_0]$ allowing for right censoring, and thus

$$f(y_i \mid \delta_i, \boldsymbol{x}_i, \boldsymbol{\beta}, h_0) = \Big[h_0(y_i)\exp\{\boldsymbol{x}_i'\boldsymbol{\beta})\}\Big]^{\delta_i} \exp\{-H_0(y_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}, \tag{5.1}$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3})'$ is a 3×1 vector of covariates, $i = 1, 2, \ldots, n$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ is the vector of the corresponding regression coefficients, $h_0(y_i)$ and $H_0(y_i)$ denote the baseline hazard function and the cumulative hazard function, respectively. Since only (x_2, x_3) have missing values, we model the covariate distribution as $f(x_{i2}, x_{i3} \mid x_{i1}, \boldsymbol{\alpha}) = f(x_{i3} \mid x_{i1}, x_{i1}, \boldsymbol{\alpha}_3) f(x_{i2} \mid x_{i1}, \boldsymbol{\alpha}_2)$, for $i = 1, 2, \ldots, n$. Since x_{i1} is always observed, it does not need to be modeled, and thus we condition on it throughout. We then use logistic regression models for x_{i2} and x_{i3} given by

$$f(x_{i3} \mid x_{i1}, x_{i2}, \boldsymbol{\alpha}_3) = \frac{\exp\{x_{i3}(\alpha_{30} + \alpha_{31}x_{i1} + \alpha_{32}x_{i2})\}}{1 + \exp(\alpha_{30} + \alpha_{31}x_{i1} + \alpha_{32}x_{i2})}, \tag{5.2}$$

where $\boldsymbol{\alpha}_3 = (\alpha_{30}, \alpha_{31}, \alpha_{32})'$, and

$$f(x_{i2} \mid x_{i1}, \boldsymbol{\alpha}_2) = \frac{\exp\{x_{i2}(\alpha_{20} + \alpha_{21}x_{i1})\}}{1 + \exp(\alpha_{20} + \alpha_{21}x_{i1})}, \tag{5.3}$$

where $\boldsymbol{\alpha}_2 = (\alpha_{20}, \alpha_{21})'$. We take $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}) \propto 1$ and independently, we take a gamma process prior for $H_0(y_i)$ with $H^*(y_i) = y_i$, which is the cumulative hazard function corresponding to a standard exponential distribution.

To illustrate how to apply the Theorems presented in Sections 2 and 3, we consider a subset of the melanoma data, which is given in Table 1.

Table 1: A Subset of the Melanoma Data

Obs (i)	y_i	δ_i	x_{i1}	x_{i2}	x_{i3}
1	0.016	1	1	1	1
2	0.077	1	1	1	0
3	0.156	1	0	0	0
4	0.197	1	0	1	1
5	2.535	0	1	0	1

Since all covariates are observed in this subset, using (2.5) after excluding the rows corresponding to $\delta_i = 0$ or $\boldsymbol{x}_j = \boldsymbol{x}_i$, we obtain

$$(X^*)' = \begin{pmatrix} 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & -1 & -1 & 0 & -1 & 1 & 0 & -1 \\ -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

The solution of $X^{*'}\boldsymbol{v} = 0$ is $v_1 = -k_2 + k_6 + k_7 + k_8 + k_9$, $v_2 = k_2$, $v_3 = -k_2 - k_5 - k_6 + k_9 + k_{10}$, $v_4 = -k_2 - k_5 - k_7 + k_8 - k_{10}$, $v_j = k_j$ for $j = 5, \ldots, 10$. It can be shown that if $k_8 \gg k_j > 0$, $k_9 \gg k_j$ for $j \neq 6, 8$, and $k_j > 0$ for $j \neq 8, 9$, then $v_j > 0$ for all j. Also, $|X^{*'}X^*| = 260 > 0$. Thus, conditions $(C1)$ and $(C2)$ given in Theorem 2.1 are met for this subset.

As discussed in Remark 2.2, when the conditions $(C1)$ and $(C2)$ are satisfied for a subset of the data, these two conditions hold for the entire dataset. Also, as discussed in Remark 2.3, the conditions $(C1^*)$ and $(C2^*)$ in Theorem 2.2 automatically hold when conditions $(C1)$ and $(C2)$ are met.

Since the posterior distributions $\pi(\boldsymbol{\beta}|D_{obs})$ and $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|D_{obs})$ given in (2.4) and (3.2), respectively, (using improper uniform priors for all parameters) are proper for this dataset, we can compute various estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_2$. The Gibbs sampler was used to sample from the posterior distribution and 50,000 Gibbs samples after a burn-in of 1,000 iterations were used to obtain all posterior estimates. We note that the computational algorithm developed in Section 4 performs well, and the autocorrelations for all model parameters disappear at lag 5, and the Gibbs sampler converges much earlier than 1,000 iterations. The resulting posterior estimates are shown in Tables 2 and 3. In both tables, posterior means, posterior standard deviations, and 95% highest posterior density (HPD) intervals are reported, where the HPD intervals were computed using the method of Chen and Shao (1999).

From Table 2, we can see that in the complete case (CC) analysis, the posterior estimates based the gamma process prior are closer to those based on the partial likelihood when c_0 becomes smaller, as expected. This result empirically confirms the findings of Sinha, Ibrahim, and Chen (2003) as stated in (2.10). From Table 3, we notice that the posterior estimates of $\boldsymbol{\alpha}$ are quite robust with respect to the choice of c_0. However, we also see some differences between the estimates in Table 2 and Table 3. In the complete case analysis, the 95% HPD interval for β_1 does include 0 in the the CC analysis but is below 0 when all cases are included, which indicates that the regression coefficient for treatment is not significant at the 0.05 level in the CC analysis, and that interferon treatment (IFN) may have a strong effect (i.e., more beneficial) compared to observation (OBS) with respect to time to relapse in the analysis incorporating all of the cases. Thus, we see the importance of incorporating all of the cases into the analysis.

Acknowledgment

Dr. Chen and Dr. Ibrahim's research was partially supported by NIH grants #CA 70101 and #CA 74015. Dr. Shao's research was partially supported by Research is partially supported by NSF grant DMS-0103487 and grant R-146-000-038-101 at the National University of Singapore. The work was done partially while Dr. Chen was visiting the Institute for Mathematical Sciences, National University of Singapore in 2005/2006. His visit was supported by the Institute.

Table 2: Posterior Estimates of β for Complete Case Analysis

Method	c_0	Parameter	Estimate	Std Dev	95% HPD Interval
Bayes Based on Partial Likelihood	-	β_1	-0.206	0.140	(-0.486, 0.064)
		β_2	0.056	0.146	(-0.231, 0.340)
		β_3	-0.151	0.146	(-0.436, 0.136)
Bayes Based on Gamma Process Prior	1.0	β_1	-0.256	0.140	(-0.526, 0.021)
		β_2	0.021	0.145	(-0.266, 0.303)
		β_3	-0.190	0.143	(-0.470, 0.087)
	0.5	β_1	-0.235	0.140	(-0.509, 0.039)
		β_2	0.035	0.146	(-0.250, 0.324)
		β_3	-0.171	0.144	(-0.444, 0.120)
	0.05	β_1	-0.210	0.142	(-0.487, 0.073)
		β_2	0.051	0.147	(-0.235, 0.336)
		β_3	-0.149	0.147	(-0.438, 0.137)

Table 3: Posterior Estimates of β and α Based On All Observed Data

c_0	Parameter	Estimate	Std Dev	95% HPD Interval
1.0	β_1	-0.301	0.128	(-0.558, -0.055)
	β_2	-0.023	0.144	(-0.302, 0.262)
	β_3	-0.157	0.140	(-0.430, 0.116)
	α_{20}	0.032	0.148	(-0.265, 0.316)
	α_{21}	0.365	0.211	(-0.056, 0.778)
	α_{30}	-0.207	0.183	(-0.558, 0.153)
	α_{31}	-0.040	0.214	(-0.458, 0.388)
	α_{32}	0.944	0.225	(0.514, 1.390)
0.5	β_1	-0.281	0.129	(-0.536, -0.032)
	β_2	-0.006	0.146	(-0.293, 0.283)
	β_3	-0.140	0.141	(-0.414, 0.140)
	α_{20}	0.034	0.147	(-0.254, 0.323)
	α_{21}	0.362	0.210	(-0.056, 0.769)
	α_{30}	-0.210	0.186	(-0.577, 0.153)
	α_{31}	-0.042	0.214	(-0.461, 0.376)
	α_{32}	0.949	0.224	(0.500, 1.374)
0.05	β_1	-0.262	0.130	(-0.518, -0.009)
	β_2	0.013	0.146	(-0.276, 0.293)
	β_3	-0.121	0.142	(-0.395, 0.159)
	α_{20}	0.032	0.148	(-0.262, 0.316)
	α_{21}	0.365	0.211	(-0.055, 0.768)
	α_{30}	-0.210	0.187	(-0.570, 0.165)
	α_{31}	-0.042	0.215	(-0.455, 0.393)
	α_{32}	0.949	0.227	(0.496, 1.384)

References

Breslow, N. E. (1974). Covariance Analysis of Censored Survival Data. *Biometrics, 30*, 89-99.

Carlin, B. P., Chaloner, K., Church, T., Louis, T. A., and Matts, J. P. (1993). Bayesian Approaches for Monitoring Clinical Trials with an Application to Toxoplasmic Encephalitis Prophylaxis. *The Statistician, 42*, 355-367.

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 Shao, Q.-M. (2004). On Propriety of the Posterior Distribution and Existence of the Maximum Likelihood Estimator for Regression models with Covariates Missing at Random. *Journal of the American Statistical Association, 99*, 421-438.

Chen, M.-H. and Shao, Q.-M. (2001). Propriety of Posterior Distribution for Dichotomous Quantal Response Models With General Link Functions. *Proceedings of the American Mathematical Society, 129*, 293-302.

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics, 8*, 69-92.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

Cox, D. R. (1972). Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society, Series B, 34*, 187-220.

Cox, D. R. (1975). Partial Likelihood. *Biometrika, 62*, 269-76.

Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics, 41*, 337-348.

Gustafson, P. (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics* **53**, 230–242.

Herring, A. H. and Ibrahim, J. G. (2001). Likelihood-based Methods for Missing Covariates in the Cox Proportional Hazards Model. *Journal of the American Statistical Association, 96*, 292-302.

Ibrahim, J. G. and Chen, M.-H. (2000). Power Prior Distributions for Regression Models. *Statistical Science, 15*, 46-60.

Ibrahim, J. G., Chen, M.-H., Sinha, D. (2001). *Bayesian Survival Analysis*, New York: Springer-Verlag.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian Analysis of Survival Time Data. *Journal of the Royal Statistical Society, Series B, 40*, 214-221.

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M., and Blum, R. H. (2000). The Role of High- and Low-dose Interferon Alfa-2b in High-risk Melanoma: First Analysis of Intergroup Trial E1690/S9111/C9190. *Journal of Clinical Oncology, 18*, 2444-2458.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis*. New York: Springer-Verlag.

Lawless, J. F., Kalbfleisch, J. D., and Wild, C. J., (1999). Semiparametric Methods for Response-selective and Missing Data Problems in Regression. *Journal of the Royal Statistical Society, Series B, 61*, 413-438.

Lin, D. Y. and Ying, Z. (1993). Cox Regression with Incomplete Covariate Measurements. *Journal of the American Statistical Association, 88*, 1341-1349.

Lipsitz, S. R. and Ibrahim, J. G. (1996a). A Conditional Model for Incomplete Covariates in Parametric Regression Models. *Biometrika, 83*, 916-922.

Lipsitz, S. R. and Ibrahim J. G. (1996b). Using the EM Algorithm for Survival Data with Incomplete Categorical Covariates. *Lifetime Data Analysis, 2*, 5-14.

Lipsitz, S. R. and Ibrahim, J. G. (1998). Estimating Equations with Incomplete Categorical Covariates in the Cox Model. *Biometrics, 54*, 1002-1013.

Liu, J. S. (1994). The Collapsed Gibbs Sampler in Bayesian Computations with Applications to A Gene Regulation Problem. *Journal of the American Statistical Association, 89*, 958-966.

Paik, M. C. (1997). Multiple Imputation for the Cox Proportional Hazards Model with Missing Covariates. *Lifetime Data Analysis, 3*, 289-298.

Paik, M. C. and Tsai, W.-Y. (1997). On Using the Cox Proportional Hazards Model with Missing Covariates. *Biometrika, 84*, 579-593.

Raftery, A. E., Madigan, D., and Volinsky, C. T. (1995). Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance. In *Bayesian Statistics 5* (Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith). Oxford: Oxford University Press, pp. 323-350.

Reilly, M. and Pepe, M. S. (1995). A Mean Score Method for Missing and Auxiliary Covariate Data in Regression Models. *Biometrika, 82*, 299-314.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton, N.J.: Princeton University Press.

Sargent, D. J. (1998). A General Framework for Random Effects Survival Analysis in the Cox Proportional Hazards Setting. *Biometrics, 54*, 1486-1497.

Schluchter, M. and Jackson, K. (1989). Log-linear Analysis of Censored Survival Data with Partially Observed Covariates. *Journal of the American Statistical Association, 84*, 42-52.

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

Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997). Bayesian Model Averaging in Proportional Hazards Models: Assessing the Risk of A Stroke. *Applied Statistics, 46*, 433-448.

Zhou, H. and Pepe, M. S. (1995). Auxiliary Covariate Data in Failure Time Regression. *Biometrika, 82*, 139-149.