Statistics and Its Interface Volume 11 (2018) 687–698

Prior specifications to handle the monotone likelihood problem in the Cox regression model

Frederico M. Almeida, Enrico A. Colosimo, and Vinícius D. Mayrink*

The monotone likelihood is a phenomenon that may affect the fitting process of well-established regression models such as the Cox proportional hazards model. In short, the problem occurs when the likelihood converges to a finite value, while at least one parameter estimate diverges to ±infinity. In survival analysis, monotone likelihood primarily appears in samples with substantial censored times and containing many categorical covariates; it is often observed when one level of a categorical covariate has not experienced any failure. A solution suggested in the literature (known as Firth correction) is an adaptation of a method originally created to reduce the bias of maximum likelihood estimates. The method leads to a finite estimate by means of a penalized maximum likelihood procedure. In this case, the penalty might be interpreted as a Jeffreys type of prior widely used in the Bayesian context; however, this approach has some drawbacks, especially biased estimators and high standard errors. The present paper explores other penalties for the partial likelihood function in the flavor of Bayesian prior distributions. A simulation study is developed, based on Monte Carlo replications and distinct sample sizes, to evaluate the impact of the suggested priors in terms of inference. Results show that a greater bias reduction can be achieved with respect to the Firth correction; however, this performance depends on the uncertainty level of the prior (vague priors do not manage well the monotone shape). A real application is also presented to illustrate the analysis using a melanoma skin data set.

AMS 2000 subject classifications: 62N02.
Keywords and phrases: Firth correction, MCMC, Proportional hazards, Partial likelihood, Survival analysis.

1. INTRODUCTION

The proportional hazards model (PHM) [6] is probably one of the most important statistical methods for the analysis of time-to-event data, perhaps because of its flexibility to explore the association of covariates with failure rates. Over the past decades, the PHM has been applied to several practical situations ranging from medical studies [2, 12, 34, 35] to the analysis of economic data on employment/unemployment cycle duration or bank failure [27, 24].

When fitting the PHM to some data sets, one may observe a phenomenon known as monotone likelihood or separation [1, 4, 29]. This issue occurs when the likelihood function is configured with a flat monotone shape indicating convergence to a finite value while one (or more) parameter estimate diverges to ±infinity. The monotone likelihood tends to occur in situations involving small sample sizes and containing unbalanced (highly predictive) covariates; an in-depth examination of the conditions of the monotone likelihood for the PHM is presented in [42]. In particular for survival data, a large number of censored observations can lead to this problem; for example, it occurs when only censored responses are associated with a category of a categorical covariate. In this sense, the larger the number of dichotomous regressors included in the model, the higher is the chance of monotone likelihood. The problem is rarely experienced when using only continuous covariates.

This paper is motivated by the analysis of a melanoma data set [5] obtained in a study developed between 1995-2012 at Hospital das Clínicas/UFMG (Brazilian public university hospital). Melanoma is a neoplasm that shows high mortality when diagnosed in advanced stages; therefore, the premature identification of patients in the high-risk group to develop metastasis is a key strategy to reduce mortality. The main aim of the mentioned study in Brazil is to assess the influence of five epidemiological and histopathologic features on the development of metastasis in patients diagnosed with invasive primary cutaneous melanoma. The PHM analysis is used here to determine the factors associated with the follow-up time until metastasis occurrence. However, a very important binary covariate (presence of mitosis) is clearly related to the monotone likelihood issue since, according to the histologic exam, the main event metastasis is not associated with those tumors in the level "without mitosis". The data set is composed by 221 patients and contains 33 metastasis detections; i.e., the censoring percentage is approximately 85%.

A solution to the problem is suggested in [23] and it is based on a procedure reported in [8], originally developed to reduce the bias of maximum likelihood estimates. The method (hereafter called "Firth correction/method")

*Corresponding author.

produces finite parameter estimates through a penalized maximum likelihood estimation. It is incorporated in computational packages available for different softwares; e.g., `SAS` [20, 22, 19] and `R` [41] (`logistf` in [21], `coxphf` in [37], `brglm` in [26]).

Penalization is a very general method for stabilizing estimates, which has both frequentist and Bayesian rationales. Firth method is a well known example of a penalization, which can be derived from a Jeffreys type of prior [25] in the Bayesian inference; however, this approach is not perfect and may suffer with biased estimators and high standard errors [16, 36, 28]. Naturally, one might consider other prior distributions being easier to compute and also more convenient to handle the monotone likelihood issue. The main goal of this paper is to evaluate and compare the impact of different prior distributions on the inferences developed for the Cox PHM. These priors can be interpreted as penalties for the partial likelihood function in a frequentist analysis. In summary, the main contributions of the present study are:

- Development of a broad comparative analysis evaluating the impact (in terms of bias) of different prior distributions for the regression coefficients (Normal, log-F, Jeffreys and the non-penalized case without prior).
- Investigation of the monotone likelihood issue in distinct simulation scenarios for Monte Carlo replications. Different sample sizes and percentages of monotone likelihood are considered.
- Examination of the effect of different levels of prior information to shine some light on the relationship between this factor and the level of observed bias in the estimation process.

The outline of this paper is as follows: Section 2 presents the notation related to the PHM. Section 3 introduces the elements of the Bayesian inference for the PHM, including the prior distributions seen as penalizations for the monotone partial likelihood. Section 4 describes the aspects of a simulation study, involving Monte Carlo replications, to investigate both Firth and non-Jeffreys approaches under different data scenarios. Section 5 shows the results for the real data analysis using the aforementioned melanoma data set. Finally, Section 6 presents the main conclusions and final remarks.

2. NOTATION AND THE COX REGRESSION

The most popular form of the Cox regression model has an exponential formulation for the hazard function:

$$\lambda(t) = \lambda_0(t)\exp(\boldsymbol{\beta}^\top \mathbf{x}), \tag{1}$$

in which $\lambda_0(t)$ is a baseline hazard (an unknown nonnegative function of time), $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters (to be estimated), and $\mathbf{x} = (x_1, x_2, \ldots, x_p)^\top$ is a vector of covariates.

The estimation of $\boldsymbol{\beta}$ in (1) is based on the partial log-likelihood function denoted by $\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$. In the absence of ties and in a sample with n individuals, where $k \le n$ failures are observed at $t_1 < t_2 < \cdots < t_k$, this function can be written for (1) as:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i - \log \left(\sum_{j \in R_{(t_i)}} \exp(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_j) \right) \right], \tag{2}$$

where $R_{(t_i)} = \{k : \ t_k \ge t_i\}$ is the risk set at time t_i, δ_i is the failure binary indicator and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ corresponds to the covariate column vector for the i-th individual. As an additional notation, consider $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^\top$. The Maximum Partial Likelihood Estimator (MPLE) of $\boldsymbol{\beta}$, notation $\widehat{\boldsymbol{\beta}}$, is obtained through the maximization of (2). Inference for $\boldsymbol{\beta}$ is usually based on: (i) Wald statistic $W = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top I(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ with $I(\boldsymbol{\beta})$ being the negative Hessian $-\ell''(\boldsymbol{\beta})$ and (ii) the Likelihood Ratio (LR) statistic given by $-2\log[L(\boldsymbol{\beta})/L(\widehat{\boldsymbol{\beta}})]$. Both statistics have a chi-squared asymptotic distribution.

As mentioned in Section 1, $L(\boldsymbol{\beta})$ might be monotone and so the MPLE cannot be found in this case. The penalty considered in [8] is a well known solution proposed in the literature for this problem; see [23]. The Firth method, initially designed for bias reduction, estimates $\boldsymbol{\beta}$ through the maximization of $\ell(\boldsymbol{\beta})$ penalized by $\log |I(\beta)|^{1/2}$. In other words, it maximizes $\ell^*(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) + (1/2)\log|I(\boldsymbol{\beta})|$. Here, $|A|$ denotes the determinant of the matrix A.

Note that the extra term $(1/2)\log|I(\boldsymbol{\beta})|$ is the log of a Jeffreys prior (apart from a constant); therefore, the maximizer of $\ell^*(\boldsymbol{\beta})$ can be seen as the mode of the posterior distribution under this prior. There is a large theoretical literature on the Jeffreys prior; see, for example [25, 3, 38, 10].

Other penalties structures have been recently proposed and explored in [16]. In this reference, the main interest is to evaluate the properties of the Jeffreys prior and to study other priors in the monotone partial likelihood situation. These interests are also incorporated in our paper, however, the simulated data analyses developed ahead are more complete, allowing the study of these and related points still unclear in the literature.

3. PRIOR DISTRIBUTIONS AND THE NON-JEFFREYS APPROACH

The essential characteristic of the Bayesian approach for inference is the explicit use of probability distributions to specify our uncertainty about unknown quantities. Here, the researcher expresses the initial information, available before observing the data, about the unknown elements of the model. This can be done by choosing a proper probability distribution (prior distribution highly informative or vague) or using an improper specification, which is considered non-informative given some criterion. In summary, a Bayesian data analysis has three mains steps: (i) choosing a full probability model, i.e., a joint probability distribution for all observable and unobservable quantities in the problem, (ii) calculating via Bayes' rule the posterior distribution – conditional distribution given the observed data – for the unobserved quantities of interest and (iii) evaluating and interpreting the information from the posterior distribution.

The partial likelihood modeling version based on (2) can be used in a Bayesian analysis. The Bayesian inference taking into account the Cox's partial likelihood is a central topic explored in [40] using the context of time-dependent covariates and time-varying regression parameters. According to the authors the regression coefficients can be well estimated in this situation.

The main focus of our study is to investigate the impact of different prior specifications, for the coefficients associated with the covariates included in the Cox regression model, used to fit a data set configured with the monotone likelihood scenario. The following cases are examined ($k = 1, \ldots, p$):

- $\beta_k \sim \mathrm{N}(m, v)$ with $m = 0$ and $v > 0$ small/large being typical choices to explore informative/vague initial uncertainty.
- $\beta_k \sim$ log-F$(l_1/2, l_2/2)$ with l_1 and l_2 being the degrees of freedom of the original F distribution. In this case, $E(\beta_k) = \Psi(l_1/2) - \Psi(l_2/2) + \log(l_2/l_1)$ and $Var(\beta_k) = \Psi'(l_1/2) + \Psi'(l_2/2)$, with $\Psi(.)$ and $\Psi'(.)$ being the digamma and trigamma functions, respectively. Higher variability is associated with small values of (l_1, l_2). The distribution is symmetric around 0 for $l_1 = l_2$. See [7] for more details.

Naturally, other prior distributions might be evaluated in this study; however, we choose to follow the steps of [16], which have investigated the previous options to deal with the monotone likelihood issue in both the logistic and survival regression modelings. The authors compare weakly informative configurations of these priors as penalizations for the likelihood to circumvent the monotone shape issue compromising the inference through maximization. The main conclusions of the paper are drawn from a real data analysis in a neonatal death study. They do not explore simulated data to evaluate and compare the behavior of the model fit, especially in terms of bias, when increasing the sample size or choosing distinct levels of prior information. The authors conclude that the log-F$(1, 1)$ is easier to interpret and it provides simpler implementation for applications in health sciences than the Jeffreys or the t-Student penalizations. In [17], the log-F$(2, 2)$ prior is recommended for typical noninfectious-disease epidemiology applications. In logistic regression, see [31] for a discussion of other possibilities of likelihood penalties and their relative advantage and disadvantages.

The joint posterior distribution of $\boldsymbol{\beta}$, given the observed time points and covariates, does not have a closed form, i.e., a proper probability density cannot be identified via Bayes' rule in this case. Indirect methods are required to sample from this unknown target distribution. The Markov Chain Monte Carlo (MCMC) algorithm called Gibbs Sampler is widely used for this task; see [11] and [9].

Depending on the model – true for those investigated in this paper – we may not be able to identify some of the full conditional posterior distributions required in the Gibbs Sampler. In this case, other methods must be considered to allow indirect sampling within the Gibbs Sampler structure itself. Some options are: Metropolis-Hastings [32, 18], Adaptive Rejection Sampling [15, 13], Adaptive Rejection Metropolis Sampling [14] and Slice Sampling [33]. In our work, we choose the random walk Metropolis-Hasting MCMC algorithm to sample from the full conditional $\pi(\beta_k|\boldsymbol{\beta}_{-k}, \mathbf{x}, \mathbf{t}, \boldsymbol{\delta})$; notation: $\boldsymbol{\beta}_{-k}$ is the vector $\boldsymbol{\beta}$ without β_k. The algorithm must be appropriately tuned for a good performance with reasonable acceptance rates related to the step generating and testing candidate values.

4. MONTE CARLO SIMULATION STUDY

In this section, a Monte Carlo (MC) simulation study is considered to compare the performances of the standard (Firth method) and the non-Jeffreys approaches to handle the monotone likelihood. This analysis is based on simulated data and is primarily designed to evaluate the properties of the estimators in each context.

Two covariates are included in the regression structure. The first one $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})^\top$ is binary with exactly 5 values (subjects) set to be 1 (the remaining cases are zeros); n is the sample size. This configuration is chosen to force the monotone likelihood occurrence in the MC samples, since the problem will appear when all subjects related to a category of a categorical covariate are not associated with failure times. The quantity 5 is good enough for this purpose and the difficulty to control the percentage of samples with monotone likelihood tend to increase when choosing larger values. The second covariate $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})^\top$ is continuous being generated from the standard normal distribution. This structure of covariates is considered for all sample sizes explored here.

In order to generate the data, the true values of the parameters are chosen in such a way to determine around 75% or 25% of the samples with monotone likelihood. They also provide a censoring fraction of approximately 75% (recall that this fraction is $\approx$ 85% in the melanoma data set). The failure times are generated via $T_i \sim \mathrm{Weibull}(\alpha, \varphi_i)$ for $i = 1, \ldots, n$. Here, $\varphi_i = \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)$ is the scale and

$\alpha > 0$ is the shape parameter. The censored observations are generated assuming $C_i \sim \text{Exp}(\lambda_c)$. As a consequence of this procedure, the corresponding baseline hazard can be written as $\lambda_0(t) = \alpha\,\varphi^*\,t_i^{\alpha-1}$, with $\varphi^* = e^{\beta_0}$. In order to simplify the analysis, we chose to work with a simpler version of the model without the intercept. This implies that $\varphi^* = 1$.

Two scenarios are considered in the simulations (assume $\beta_1 = -2.00$ and $\alpha = 1.50$ for both):

- Scenario 1: ≈ 75% of samples with monotone likelihood; $\beta_2 = -0.74$ and $\lambda_c = 2.10$.
- Scenario 2: ≈ 25% of samples with monotone likelihood; $\beta_2 = 4.97$ and $\lambda_c = 4.70$.

In terms of prior distributions expressing different levels of initial information for β_k, $k = 1$ or 2, consider:

$$\text{Case 1: } \beta_k \sim N(0,5); \quad \text{Case 2: } \beta_k \sim N(0,1);$$
$$\text{Case 3: } \beta_k \sim \text{log-F}(1,1); \quad \text{Case 4: } \beta_k \sim \text{log-F}(2,2).$$

For simplicity, we chose the log-F hyperparameters $l_1 = l_2$ to have a prior distribution symmetric around zero. The mean 0 is set for the regression coefficients in all cases, which suggests lack of information about the sign of these parameters. Figure 1 presents the shapes of the proposed priors. Comparing the Gaussian priors, note that Case 1 expresses higher uncertainty (larger variance for β_k) than Case 2. Similarly for the log-F distribution, Case 3 is vaguer than Case 4. It can also be seen that the N(0,5) has the highest uncertainty (among all cases) and the tail of the log-F(2,2) is slightly heavier than that of the N(0,1).

This study is based on the analysis of 1,000 samples generated in a Monte Carlo scheme for both standard and non-Jeffreys approaches. In the Bayesian context, the MCMC is applied to each sample leading to a total of 1,000 chains for each parameter of the model. The posterior inferences are focused on the 1,000 posterior means computed from these chains.

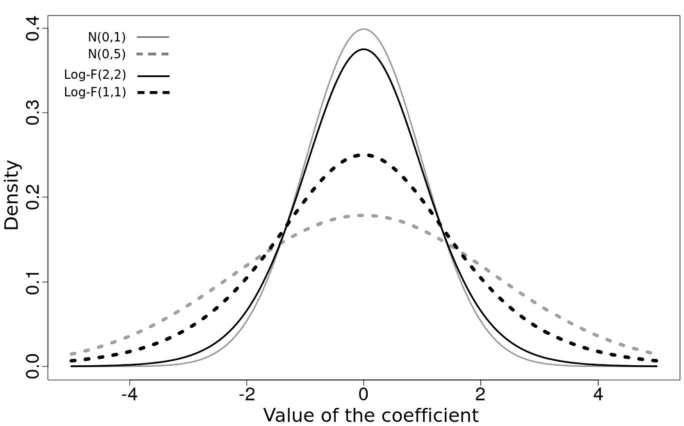

Figure 1. Comparing the shapes of the densities (Normal and log-F) chosen as prior distributions in the present study.

The MCMC algorithm is set to perform 7,000 iterations with the first 1,000 assumed as the burn-in period (discarded from the analysis). The Gaussian proposal distributions in the Metropolis-Hastings step of the Gibbs Sampler were adjusted to determine acceptance rates between $40 - 50\%$, as recommended in the literature [39].

In each scenario, different sample sizes are considered: $n = 50$, 200, 600 and 1,000. The nominal value of the coverage probability is assumed to be 0.95. The performance of the estimators were evaluated by the mean square error (MSE), relative bias (RB) and coverage rate (CR). The coverage rate provides the proportion of times (in the 1,000 MC replications) that the 95% confidence or credible intervals, built for each parameter, contain the corresponding true value. In addition, consider: $\text{MSE}_{\hat{\beta}_k} = Var(\hat{\beta}_k) + [E(\hat{\beta}_k - \beta_k)]^2$ and $\%\text{RB}_{\hat{\beta}_k} = 100\,E(\hat{\beta}_k - \beta_k)/|\beta_k|$, where β_k is the true value and $\hat{\beta}_k$ is the corresponding estimate (posterior mean or maximum likelihood estimate).

The simulation results discussed in the next sections are obtained for the standard (Firth correction) and the non-Jeffreys approaches. Additional results, involving the square root of the MSE and the hazard ratio $\exp(\hat{\beta}_k)$ estimators, are shown in Appendix A.

4.1 Standard approach analysis

Table 1 summarizes the results of the so-called standard approach analysis, which is based on the current classical solution (Firth correction) to deal with the monotone likelihood problem. The reported quantities are: the MSE, relative bias (in %) and confidence intervals based on Wald (CR.w_{β_k}) and likelihood ratio (CR.p_{β_k}) statistics. As can be seen, a huge relative bias is observed for the maximum partial likelihood estimates of β_1, in both scenarios representing 75% and 25% of monotone likelihood. In addition, the relative bias of β_1 does not decrease with the sample size increasing. Note that, for the different sample sizes, the $\text{RB}_{\hat{\beta}_1}$, $\text{MSE}_{\hat{\beta}_1}$, and the standard error $\text{SE}_{\hat{\beta}_1}$ are around 52%, 1.92 and 1.24 in Scenario 1 and around 49%, 3.10 and 1.02 in Scenario 2, respectively. The coverage rates for β_1 (Wald and likelihood ratio) are smaller than the nominal value. As an example, the Wald coverage rates for β_1 are approximately: 85% (Scenario 1) and 78% (Scenario 2). The profile coverage rate for β_1 is close to the nominal value when the sample size is small, but it decreases when the sample size increases for both scenarios. The estimates of β_2, which is not directly related to the monotone likelihood issue, show good asymptotic properties in this analysis.

4.2 Non-Jeffreys approach analysis

This section is devoted to the analysis of the Cox PHM based on the Bayesian approach assuming the Gaussian and log-F prior distributions (i.e., non-Jeffreys priors) previously specified as Cases 1–4. The main results are presented in Tables 2 and 3, for 75% and 25% of monotone likelihood, respectively.

Table 1. Simulation results based on the Firth correction. Scenarios 1 and 2 are indicated by the percentage of MC samples with monotone likelihood (%ML)

n	%ML	%RB$_{\hat{\beta}_1}$	%RB$_{\hat{\beta}_2}$	MSE$_{\hat{\beta}_1}$	MSE$_{\hat{\beta}_2}$	SE$_{\hat{\beta}_1}$	SE$_{\hat{\beta}_2}$	CR.w$_{\beta_1}$	CR.p$_{\beta_1}$	CR.w$_{\beta_2}$	CR.p$_{\beta_2}$
50	75	53.930	-2.400	2.200	0.150	1.430	0.360	86.900	91.680	95.100	94.100
	25	46.910	2.750	3.840	2.650	1.350	1.260	85.900	91.780	93.800	94.680
200	75	52.280	-1.660	1.860	0.030	1.280	0.160	85.100	94.500	94.500	94.100
	25	50.450	0.870	3.050	0.260	1.040	0.490	78.900	90.130	94.800	95.800
600	75	53.770	-0.700	1.920	0.010	1.240	0.090	83.200	74.090	95.100	95.100
	25	52.100	0.280	3.130	0.080	1.000	0.270	76.400	84.200	93.600	93.800
1,000	75	49.400	-0.540	1.720	0.001	1.270	0.070	86.800	72.730	95.400	95.300
	25	48.980	0.480	2.880	0.040	1.000	0.200	77.900	82.810	95.600	95.700

Table 2. Simulation results based on the non-Jeffreys priors for β_k. Case 1 = N(0,5), Case 2 = N(0,1), Case 3 = log-F(1,1) and Case 4 = log-F(2,2). Scenario 1: $\approx 75\%$ of MC samples with monotone likelihood

n	Cases	RB$_{\hat{\beta}_1}$	RB$_{\hat{\beta}_2}$	MSE$_{\hat{\beta}_1}$	MSE$_{\hat{\beta}_2}$	SE$_{\hat{\beta}_1}$	SE$_{\hat{\beta}_2}$	CR$_{\beta_1}$	CR$_{\beta_2}$
50	1	16.740	-9.640	0.821	0.171	0.840	0.400	96.700	95.400
	2	66.210	1.760	1.910	0.110	0.390	0.320	77.500	96.200
200	1	14.430	-0.960	0.670	0.030	0.770	0.170	96.400	94.000
	2	61.090	0.210	1.650	0.020	0.410	0.150	79.700	95.400
600	1	12.411	-0.860	0.660	0.010	0.790	0.090	95.900	94.600
	2	60.890	0.200	1.650	0.011	0.410	0.090	78.200	94.500
1,000	1	9.550	-0.640	0.660	0.004	0.790	0.060	96.600	95.300
	2	60.400	0.180	1.630	0.004	0.410	0.060	79.000	95.100
50	3	-8.890	-13.260	1.270	0.180	1.120	0.410	96.800	94.100
	4	34.560	-8.110	0.970	0.140	0.700	0.380	93.500	95.400
200	3	-14.890	-2.040	1.170	0.020	1.040	0.160	97.000	94.100
	4	28.640	-1.380	0.760	0.022	0.660	0.151	93.300	95.300
600	3	-15.100	-0.750	1.190	0.011	1.050	0.091	96.200	93.300
	4	28.550	-0.570	0.771	0.010	0.670	0.090	93.700	94.300
1,000	3	-16.190	0.150	1.240	0.004	1.060	0.071	96.100	94.300
	4	27.730	0.311	0.770	0.005	0.681	0.060	92.800	95.300

Table 3. Simulation results based on the non-Jeffreys priors for β_k. Case 1 = N(0,5), Case 2 = N(0,1), Case 3 = log-F(1,1) and Case 4 = log-F(2,2). Scenario 2: $\approx 25\%$ of MC samples with monotone likelihood

n	Cases	RB$_{\hat{\beta}_1}$	RB$_{\hat{\beta}_2}$	MSE$_{\hat{\beta}_1}$	MSE$_{\hat{\beta}_2}$	SE$_{\hat{\beta}_1}$	SE$_{\hat{\beta}_2}$	CR$_{\beta_1}$	CR$_{\beta_2}$
50	1	26.920	-11.370	0.995	0.711	0.840	0.620	95.400	92.500
	2	68.420	-41.290	2.009	4.278	0.370	0.260	58.400	0.200
200	1	17.070	-2.160	0.822	0.214	0.841	0.450	95.800	94.800
	2	56.050	-15.920	1.517	0.704	0.510	0.280	70.800	41.300
600	1	17.210	-0.720	0.841	0.064	0.850	0.250	94.900	94.300
	2	52.920	-5.930	1.457	0.135	0.580	0.220	70.200	77.500
1,000	1	14.690	-0.460	0.826	0.041	0.860	0.200	96.000	93.800
	2	50.900	-3.650	1.408	0.065	0.610	0.181	72.100	84.100
50	3	-14.610	9.901	1.026	1.088	0.971	0.922	97.900	98.300
	4	40.001	-12.252	1.188	0.918	0.740	0.742	92.500	90.500
200	3	-13.802	2.670	0.868	0.268	0.890	0.501	97.800	94.700
	4	28.650	-2.311	0.981	0.234	0.801	0.471	92.200	94.600
600	3	-12.451	0.731	0.854	0.069	0.891	0.260	97.400	94.900
	4	28.530	-0.770	0.948	0.069	0.810	0.261	91.900	94.000
1,000	3	3.056	0.220	1.078	0.041	1.037	0.203	95.200	94.600
	4	26.260	-0.491	0.950	0.041	0.821	0.201	92.600	93.500

Table 2 shows that the N(0,1) and log-F(2,2) (more informative cases) have much larger relative bias (%RB) for β_1 than the N(0,5) and log-F(1,1) (less informative cases). In terms of coverage rates, note that the values of CR_{β_1}, for the N(0,5) and the log-F(1,1), are close to the nominal level of 95%. This behavior is not observed (especially) for the N(0,1). A possible reason for these results is the fact that the distributions N(0,1) and log-F(2,2) have low uncertainty, but in the wrong direction. The true value is $\beta_1 = -2$ and these priors are centered at 0, with a small variability. This idea also explains the better results for N(0,5) and log-F(1,1). These options are also centered at 0, but with higher uncertainty. A justification for the better coverage rate and the lower relative bias obtained for the log-F(2,2), in a comparison with the N(0,1), is the fact that this log-F has variance larger than 1. This means that the probability mass is more concentrated around 0 for the Gaussian prior, making the PHM with log-F(2,2) less affected by the incorrect informative specification.

As expected, the analysis of β_2 in Table 2 shows that this coefficient is well estimated. The quality of the estimation improves with the sample size increasing. Recall that this parameter is not associated with the binary covariate responsible for the monotone likelihood problem.

The analysis of Table 3 leads to similar conclusions regarding β_1 in the scenario with 25% of monotone likelihood; i.e., better results are obtained for the less informative priors and an increase in the sample size does not seem to improve significantly the estimates. In terms of β_2, a poor result can be detected for the prior N(0,1) when the sample size is $n = 50$. This behavior can be explained by the fact that the true value of β_2 is 4.97, which is clearly distant from the prior mean 0. In other words, the prior distribution N(0,1) is concentrating too much mass of probability in a region that does not include the true value. This strong prior information dominates the likelihood due to the small sample size. Naturally, better results are observed as the sample size increases.

Comparing Table 1 with Tables 2 and 3, note that the vaguer priors (Cases 1 and 3) provide better results for β_1 than the standard approach in the literature based on the Firth correction. As an example, the smallest $\%RB_{\beta_1}$ obtained through Firth correction (Scenario 1: 75%) is 49.40, whereas, the largest absolute $\%RB_{\beta_1}$ in Table 2 (for Cases 1 and 3) is 16.74. This is one of the main conclusions in the present paper, the previous analyses clearly indicate that the quality of the estimates under monotone likelihood can be significantly improved by using non-Jeffreys priors.

The reader should also have in mind that assuming an informative prior for a coefficient (usually centered at 0) can be problematic if the true value is far from 0. This is the situation observed in Table 3 for the N(0,1) prior (Scenario: $\approx$ 25% of monotone likelihood). It seems safer to choose a vaguer prior; however, the bias may increase again if the prior is too vague, especially for the coefficient associated with the monotone likelihood. A key question is: how vague should be the prior? This discussion motivated the next section with some extra simulations to approach this topic.

4.3 Prior uncertainty against monotone likelihood

This section explores extra results focused on the most severe scenario (in terms of bias) exhibited in Table 3. Here, recall that the MC replications are configured with: $\approx$ 25% of monotone likelihood (Scenario 2), $n = 50$ (the smallest tested sample) and true values of coefficients away from 0 ($\beta_1 = -2.00$ and $\beta_2 = 4.97$). The main goal is to evaluate the behavior of RB, MSE, CR and standard errors, when the prior variance increases (the mean is fixed at 0). Nine different specifications are used to study the Gaussian prior; variance ranging from 1 to the much vaguer case of 200. In addition, five specifications of variability are tested for the log-F prior; the smallest is 0.24 and the largest is 9.87.

It is important to emphasize that using samples with $n = 50$ is a central strategy for the analysis developed in this section. The goal here is to investigate how the prior distribution affects a not too strong likelihood (based on only 50 observations). The uncertainty level, expressed in the prior, is certainly a key factor controlling the influence of this distribution on the likelihood. The uncertainty level also represents the strength of the penalty (in the classical interpretation) and this strength seems (as indicated ahead) crucial to handle the monotone shape.

Table 4 summarizes the results for the Gaussian prior distribution. Note that the relative bias for β_1 decreases from 68.41% to 7.16%, when the variance increases from 1 to 10. The RB closest to 0 is registered for variance 10. When moving from variance 10 towards 200, the absolute RB clearly increases reaching 158.3% with the vaguer specification. A similar reaction is observed for the RB of β_2; however, the magnitudes of the bias are lower than those of the β_1 case. Note that the N(0,10) specification is also the one providing the best estimation for β_2 (% RB = -1.14). As expected for both coefficients, when increasing the uncertainty level: a decreasing-increasing trend is also obtained for the MSE values and an increasing-only trend is reported for the standard errors. In terms of coverage rates, only the small variances specifications determine values below the nominal level of 95%. Intermediate and large variances lead to better approximations.

Table 5 shows the results based on the log-F prior specifications. The conclusions are quite similar to those drawn from the Gaussian prior analysis evaluating the impact of the increasing uncertainty level. In brief, the absolute RBs and the MSEs are configured with a decreasing-increasing trend, where the best estimation (for both coefficients) is related to the log-F(1,1), which is the option having the second largest variance in the tested group. The standard errors show again an increasing-only behavior and the coverage rates reach the 95% level as the prior variance increases.

Table 4. Analysis using the Normal prior with mean 0; the variances are indicated in the rows. Estimates based on Monte Carlo replications (≈ 25% ML and $n = 50$)

σ^2_{prior}	$\text{RB}_{\hat\beta_1}$	$\text{RB}_{\hat\beta_2}$	$\text{MSE}_{\hat\beta_1}$	$\text{MSE}_{\hat\beta_2}$	$\text{SE}_{\hat\beta_1}$	$\text{SE}_{\hat\beta_2}$	CR_{β_1}	CR_{β_2}
1	68.410	-41.302	2.011	4.282	0.373	0.261	60.320	0.100
3	41.538	-20.628	1.147	1.264	0.676	0.461	92.704	79.310
5	26.335	-11.692	0.954	0.719	0.823	0.618	96.200	92.102
10	7.158	-1.140	1.274	0.780	1.120	0.882	95.140	97.000
15	-9.812	4.971	1.701	1.139	1.290	1.039	96.510	97.501
20	-22.342	7.943	2.211	1.544	1.419	1.179	96.700	96.604
50	-59.393	16.821	5.912	2.938	2.123	1.497	96.201	95.700
100	-102.143	18.174	13.153	4.052	2.998	1.800	95.310	94.210
200	-158.301	23.610	30.031	5.562	4.481	2.050	95.501	94.501

Table 5. Analysis using the log-F prior centered at 0; $l_1 = l_2$ and the variances are indicated in the rows. Estimates based on Monte Carlo replications (≈ 25% ML and $n = 50$)

$l_k/2$	σ^2_{prior}	$\text{RB}_{\hat\beta_1}$	$\text{RB}_{\hat\beta_2}$	$\text{MSE}_{\hat\beta_1}$	$\text{MSE}_{\hat\beta_2}$	$\text{SE}_{\hat\beta_1}$	$\text{SE}_{\hat\beta_2}$	CR_{β_1}	CR_{β_2}
9.0	0.240	80.247	-48.388	2.640	5.859	0.253	0.275	2.500	0.000
5.0	0.441	68.797	-34.243	2.062	3.065	0.411	0.410	60.601	26.930
2.0	1.290	39.347	-12.602	1.163	0.924	0.738	0.730	93.600	90.240
1.0	3.291	-14.610	9.901	1.026	1.088	0.971	0.922	97.900	98.300
0.5	9.870	-33.607	11.044	3.567	2.289	1.766	1.411	95.701	96.805

Figure 2 compares the shapes of the penalized profile likelihoods constructed using the Gaussian and log-F priors (as penalties) centered at 0 and with different variances (different penalization strengths). This analysis is based on a simulated data set ($n = 50$) affected by the monotone likelihood issue. The profile curve for each coefficient is obtained by fixing the other coefficient at its true value and assuming the prior variance as specified at the top of each graph. Note that the vertical axes are not the same in these graphs, since the distinct penalized profile likelihoods have different scales. This is a necessary action to allow the visual inspection. The reader should also note that the scale presented in the horizontal axes are the same for each parameter.

Looking at the curves for β_1, exhibited in Figure 2, one can clearly see that the shape of the penalized profile likelihood seems to modify in the direction of a monotone shape as the variance increases. For the small variance cases, the maximum is centered between -2 and 0. This suggests that the likelihood is not strong enough to correct the prior information centered at 0. Assuming higher uncertainty *a priori* reduces the strength of the prior to dominate the likelihood. The transformation towards the typical flat shape of a monotone likelihood is not evident for the curves related to β_2 (coefficient of the continuous covariate unrelated to the monotone likelihood issue). Again, the small variance prior has a strong influence on the likelihood determining a curve centered between 0 and 4.97. This domination seems to reduce as the prior uncertainty increases.

The discussion presented in this section highlight some key aspects to be considered when fitting the Cox PHM to a data set affected by the monotone likelihood issue. In many practical situations, the researcher does have initial information about the coefficients; therefore, centering the prior distribution at 0 is a very usual choice. In this case, setting a small prior variance can be extremely dangerous (especially when n is small). Vague priors are safer; however, a prior expressing a high uncertainty can be viewed as a weak penalization unable to change the undesirable monotone shape, thus implying in biased estimates.

5. REAL DATA ANALYSIS

In this section, the standard (non-penalized or Firth method) and the non-Jeffreys approaches are applied to the melanoma data set mentioned in Section 1. The prognostics factors under consideration are: (i) Gender ("female" with coefficient β_2); (ii) Histological type with levels "nodular" (β_{31}), "acral lentiginous" (β_{32}) and the reference category "extensive superficial + lentigo maligna"; (iii) Breslow index with levels "$1 - 4$ mm" (β_{41}), "> 4 mm" (β_{42}) and the reference "< 1 mm"; (iv) Ulceration ("yes" with coefficient β_5). The fifth factor, "mitosis" (β_1), is the one associated with the monotone likelihood issue. This binary factor is coded as 1 when the mitotic rate is high for the patient. Metastasis is the response event under study and the follow-up time goes from the diagnosis of melanoma to the date of

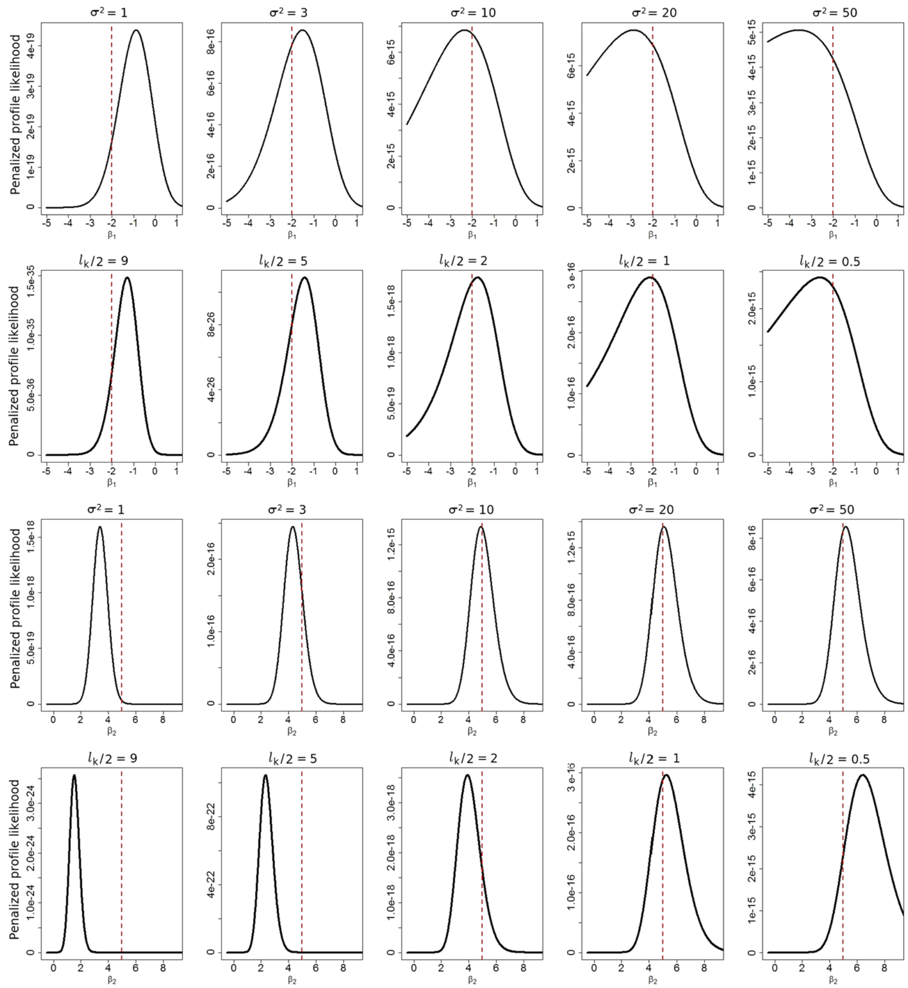

Figure 2. Penalized profile likelihoods built with the Normal and log-F priors (as penalties) for β_1 (rows 1 and 2) and β_2 (rows 3 and 4). The priors are centered at 0 with variance (Normal) or degree of freedom (log-F) indicated at the top of the graphs. The true value of the coefficient is identified by the vertical dashed line. For both cases, the variability expressed in the prior distribution increases from the left to right panel. Analysis based on a simulated data set ($n = 50$) detected with monotone likelihood.

the last visit (censoring) or the date when metastasis is detected (failure).

In each covariate, the observations are distributed as follows: gender (male 37.56%, female 62.44%), histological type (reference 72.40%, nodular 17.19%, acral lentiginous 10.41%), Breslow index ("< 1 mm" 60.63%, "$1 - 4$ mm" 29.41%, "> 4 mm" 9.96%), ulceration (no 79.64%, yes 20.36%), mitosis (low 35.29%, high 64.71%). Ignoring the censored cases, the median time to metastasis is 4.6 months (interquartile range is 11.14).

Table 6 presents the estimates of the coefficients for eight different model fits. The standard approach columns show the results for the ordinary Cox PHM analysis (without penalty) and for the Firth method. As expected, the coefficients not connected with the monotone likelihood issue have similar estimates across the rows of the table. A minor difference can be detected for the standard errors of $\hat{\beta}_1$ using the priors N(0,1) and log-F(9,9); these values are the smallest ones in that row. This result is probably due to a highly informative configuration mistakenly centered in the wrong part (around zero) of the parametric space. The large values reported for $\hat{\beta}_1$, in the non-penalized Cox regression fit, are exactly those found in the output of the software `R`. These estimates are not valid for inference, since their calculations are affected by the monotone likelihood issue.

Figure 3 (a) shows estimates for β_1 obtained through the classical approach (with Firth correction) and the non-Jeffreys approach. Some important aspects to be noted: (i) all Bayesian credible intervals include the classical penalized maximum likelihood estimate, (ii) the amplitude of the intervals are not the same and some Bayesian intervals are

Table 6. *Parameter estimates based on the standard and non-Jeffreys approaches for the melanoma data set; standard errors in parentheses. The column "Cox" refers to the Cox PHM model fit without penalization*

	Standard approach		Non-Jeffreys approach					
	Cox	Firth	N(0,1)	N(0,5)	N(0,10)	log-F(1,1)	log-F(2,2)	log-F(9,9)
$\hat{\beta}_1$	18.52 (5131)	2.221 (1.50)	1.403 (0.70)	2.772 (1.43)	3.664 (1.94)	3.691 (2.32)	2.431 (1.42)	1.080 (0.56)
$\hat{\beta}_2$	-0.651 (0.36)	-0.641 (0.36)	-0.589 (0.33)	-0.633 (0.35)	-0.643 (0.36)	-0.630 (0.35)	-0.620 (0.35)	-0.531 (0.31)
$\hat{\beta}_{31}$	0.984 (0.45)	0.958 (0.45)	1.038 (0.39)	1.077 (0.45)	1.052 (0.45)	1.049 (0.44)	1.047 (0.44)	0.994 (0.37)
$\hat{\beta}_{32}$	0.236 (0.63)	0.298 (0.62)	0.172 (0.53)	0.233 (0.63)	0.207 (0.64)	0.215 (0.61)	0.211 (0.59)	0.149 (0.45)
$\hat{\beta}_{41}$	0.110 (0.69)	1.036 (0.67)	0.731 (0.49)	1.061 (0.66)	1.115 (0.68)	1.060 (0.61)	0.922 (0.59)	0.526 (0.40)
$\hat{\beta}_{42}$	0.184 (0.73)	1.760 (0.71)	1.423 (0.53)	1.782 (0.70)	1.849 (0.72)	1.798 (0.65)	1.649 (0.63)	1.207 (0.43)
$\hat{\beta}_5$	0.699 (0.38)	0.685 (0.39)	0.793 (0.36)	0.759 (0.39)	0.744 (0.39)	0.733 (0.39)	0.766 (0.38)	0.784 (0.34)

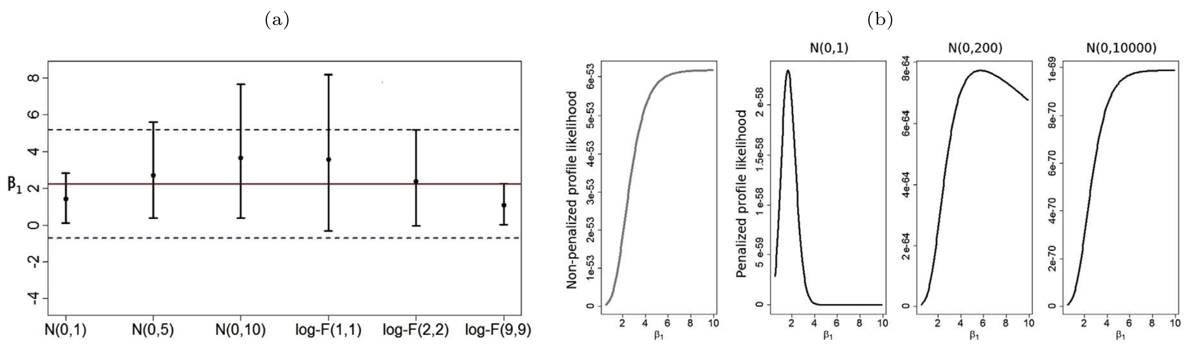

Figure 3. Panel (a) compares estimates for β_1 (covariate mitosis): 95% confidence intervals (dashed lines), 95% credible intervals (vertical solid segments), penalized maximum likelihood estimate (horizontal solid line) and the posterior mean (dot mark). Panel (b) shows the shapes of the non-penalized and penalized profile likelihoods based on the real data set (melanoma study) - the Gaussian prior (centered at 0) is used here as the penalty.

much smaller than the classical one, (*iii*) the largest amplitude corresponds to the log-F(1,1) prior (the least informative option in the tested log-F group), (*iv*) small amplitudes are found for the N(0,1) and log-F(9,9) (more informative cases), (*v*) the interval built with Firth correction includes the value 0 ($z = 2.221/1.50$) suggesting that mitosis is not significant, (*vi*) some Bayesian intervals indicate that mitosis is indeed an important factor affecting the time-to-metastasis.

Results for the covariates not associated with the monotone likelihood issue are presented in Appendix B. In summary, small differences are observed when comparing the amplitude of the intervals and only the conclusions regarding the significance of β_5 (ulceration) change between methods/priors.

Figure 3 (b) compares the shapes of the non-penalized (first panel) and the penalized profile likelihoods using the Normal prior centered at zero. The curves related to three different variances (1, 200 and 10,000) are presented in this illustration. The main conclusion is similar to the one obtained from Figure 2; the monotone shape (shown in the first panel) is clearly recovered when increasing the prior variance.

6. CONCLUSIONS

Monotone likelihood is a recurrent condition that can happen very often in situations involving: rare events, small sample sizes, large percentages of censored observations and the presence of many categorical covariates. The standard classical procedure, available in the literature to deal with this problem in a Cox regression model, is based on the so-called Firth correction originally created to reduce bias of maximum likelihood estimates. This method can be seen, under the Bayesian point of view, as the maximization of a penalized partial likelihood function, with the penalty being the Jeffreys prior specification for the regression coefficients. This vision motivated the present paper, where the main goal was to study the impact and benefit (if any) of other prior choices in the analysis of the Cox PHM.

An extensive simulated data analysis (using Monte Carlo replications) was developed to understand the behavior of the estimates in different configurations of data and levels of initial information expressed by the tested prior distributions. This study was focused on prior specifications centered at 0, with the level of information being controlled through its variance magnitude. The analyses have shown

Table 7. Simulation results related to the coefficients. Analysis involving: Firth, Case 1 = N(0,5) and Case 3 = log-F(1,1). Comparing standard errors and the square root of the MSE (RMSE). Scenario with $\approx 75\%$ of MC samples affected by the monotone likelihood

n	Cases	$\text{RMSE}_{\hat{\beta}_1}$	$\text{RMSE}_{\hat{\beta}_2}$	$\text{SE}_{\hat{\beta}_1}$	$\text{SE}_{\hat{\beta}_2}$
50	Firth	1.483	0.388	1.434	0.360
	1	0.886	0.412	0.819	0.401
	3	1.127	0.418	1.113	0.407
1,000	Firth	1.390	0.066	1.243	0.067
	1	0.809	0.064	0.787	0.064
	3	1.134	0.064	1.084	0.064

Table 8. Simulation results related to the exponential of the coefficients (hazard ratios). Analysis involving: Firth, Case 1 = N(0,5) and Case 3 = log-F(1,1). Scenario with $\approx 75\%$ of the MC samples affected by the monotone likelihood

n	Cases	$\text{RB}_{\exp(\hat{\beta}_1)}$	$\text{RB}_{\exp(\hat{\beta}_2)}$	$\text{MSE}_{\exp(\hat{\beta}_1)}$	$\text{MSE}_{\exp(\hat{\beta}_2)}$	$\text{RMSE}_{\exp(\hat{\beta}_1)}$	$\text{RMSE}_{\exp(\hat{\beta}_2)}$	$\text{SE}_{\exp(\hat{\beta}_1)}$	$\text{SE}_{\exp(\hat{\beta}_2)}$
50	Firth	430.149	5.468	3.296	0.038	1.815	0.195	0.194	0.172
	1	113.919	-2.126	0.208	0.031	0.456	0.177	0.429	0.177
	3	75.797	-2.116	0.233	0.032	0.483	0.178	0.472	0.178
1,000	Firth	342.858	-0.025	0.721	0.001	0.849	0.031	0.168	0.032
	1	76.332	-0.267	0.094	0.001	0.306	0.030	0.288	0.030
	3	41.611	-0.246	0.093	0.001	0.305	0.030	0.300	0.030

that, depending on the size of the variance, the bias of the estimates can be significantly reduced with respect to the Firth method. As expected, poor results were observed when assuming small variance *a priori* for a parameter having a true value far from 0. The performance improves as this variance increases, but it deteriorates again when the prior becomes too vague. The reason for this decay can be the fact that a vague prior is not strong enough to impose an effective penalization to handle the monotone shape. A possible strategy that may be used to guide the choice of the prior variance is summarized in the steps listed below:

1. Fit the Cox PHM using the standard approach based on the Firth correction in the context of the Jeffreys prior;
2. Identify the estimate of the coefficient affected by the monotone likelihood issue (denote it by $\hat{\beta}_k^{\text{Firth}}$);
3. Choose a prior variance (Gaussian or log-F) such that the credible *a priori* interval (say 95%) contains the value of $\hat{\beta}_k^{\text{Firth}}$ close to one of its borders.

In a real application, the Firth correction and the tested priors were confronted in a comparative study evaluating the relationship between some epidemiological and histopathologic factors and the time-to-metastasis for patients with melanoma. The main conclusions were drawn from the analysis of the binary covariate "mitosis" (directly associated with the monotone likelihood issue). The effect of this covariate is not significant via Firth method and this result changes for some of the prior specifications. This alteration illustrates how crucial can be the use of alternative priors in order to reach greater bias reduction with respect to the standard Jeffreys option implemented for the Firth correction.

APPENDIX A. ADDITIONAL RESULTS FOR THE SIMULATION STUDY

Table 7 compares the MSE and the square root of the MSE (RMSE) obtained in an additional MC simulation developed for the sample sizes 50 and 1,000. The number of MC replications is 1,000 and approximately 75% of them are affected by the monotone likelihood issue. This analysis evaluates the Firth correction and the less informative specifications N(0,5) and log-F(1,1). Note that the RMSE is quite close to the corresponding SE. The difference RMSE-SE are slightly larger for β_1 in all cases.

Table 8 reports results for hazard ratio estimators given by $\exp(\hat{\beta}_k)$. This option differs from $\hat{\beta}_k$ in terms of bias and it has been regarded in the literature as relevant for clinical decisions; see [30]. This analysis is also focused on the Firth method, N(0,5) and log-F(1,1). As can be seen, $\text{RB}_{\exp(\hat{\beta}_1)}$ is much bigger than $\text{RB}_{\exp(\hat{\beta}_2)}$. The highest bias for $\exp(\hat{\beta}_1)$ is related to the Firth correction for both samples sizes. Looking at the differences RMSE-SE for each coefficient, note that $\exp(\hat{\beta}_1)$ is the one associated with the largest values, especially for the Firth correction case.

APPENDIX B. ADDITIONAL RESULTS FOR THE REAL DATA ANALYSIS

Figure 4 compares estimates for the coefficients not associated with the monotone likelihood problem in the real data analysis presented in Section 5.

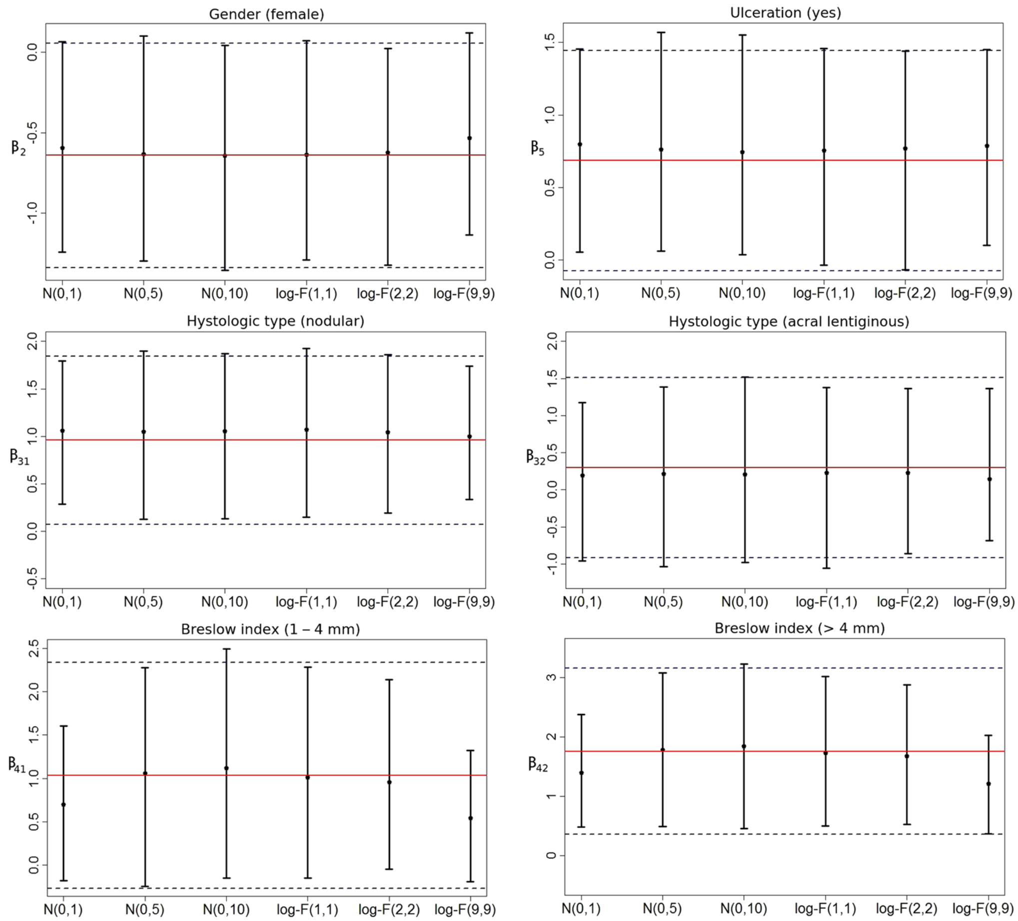

Figure 4. Results for the coefficients unrelated to the monotone likelihood issue: 95% *confidence intervals (dashed lines),* 95% *credible intervals (vertical solid segments), penalized maximum likelihood estimate (horizontal solid line) and the posterior mean (dot mark).*

ACKNOWLEDGEMENTS

The authors would like to thank the Associate Editor and two anonymous referees for their constructive comments. This research is partially supported by grants from Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Received 5 October 2017

REFERENCES

[1] Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10. MR0738319

[2] Andersen, P. K. (1991). Survival analysis 1982–1991: The second decade of the proportional hazards regression model. *Statistics in Medicine* **10** 1931–1941.

[3] Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory.* John Wiley & Sons, Hoboken.

[4] Bryson, M. C. and Johnson, M. E. (1991). The incidence of monotone likelihood in the Cox model. *Technometrics* **23** 381–383.

[5] Cherobin, A. C. F. P., Wainstein, A. J. A., Colosimo, E. A., Goulart, E. M. A. and Bittencourt, F. V. (2017). Fatores Prognosticos para a metastase no melanoma cutaneo. *Anais Brasileiros de Dermatologia*. (in press) http://www.anaisdedermatologia.org.br.

[6] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B* **34** 187–220. MR0341758

[7] Dupuis, D. J. (2001). Fitting log-F models robustly, with an application to the analysis of extreme values. *Computational Statistics and Data Analysis* **35** 321–333. MR1819043

[8] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38. MR1225212

[9] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85** 398–409. MR1141740

[10] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis*, 3 ed. CRC Press, Boca Raton. MR3235677

[11] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**

721–741.

[12] Genser, B. and Wernecke, K. D. (2005). Joint modelling of repeated transitions in follow-up data - a case study on breast cancer data. *Biometrical Journal* **47** 388–401. MR2142729

[13] Gilks, W. (1992). *Bayesian Statistics 4* Derivative-free adaptive rejection sampling for Gibbs sampling, 641–649. Oxford University Press.

[14] Gilks, W., Best, N. and Tan, K. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of Applied Statistics: Series C* **44** 455–472.

[15] Gilks, W. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41** 337–348.

[16] Greenland, S. and Mansournia, M. A. (2015). Penalization, bias reduction and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine* **34** 3133–3143. MR3402580

[17] Greenland, S., Mansournia, M. A. and Altman, D. G. (2016). Sparse data bias: a problem hiding in plain sight. *British Medical Journal* **352**. i1981.

[18] Hastings, W. (1970). Monte Carlo sampling using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

[19] Heinze, G. and Ploner, M. (2002). SAS and SPLUS programs to perform Cox without convergence problems. *Computer Methods and Programs in Biomedicine* **67** 217–223.

[20] Heinze, G. and Ploner, M. (2003). Fixing the nonconvergence bug in logistic regression with SAS and SPLUS. *Computer Methods and Programs in Biomedicine* **71** 181–187.

[21] Heinze, G., Ploner, M., Dunkler, D. and Southworth, H. (2013). logistf: Firth's bias reduced logistic regression R package version 1.21.

[22] Heinze, G. and Puhr, R. (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. *Statistics in Medicine* **29** 770–777. MR2752041

[23] Heinze, G. and Schemper, M. (2001). A Solution to the Problem of Monotone Likelihood. *Biometrics* **57** 114–119. MR1833296

[24] Henebry, K. L. (1996). Do cash flow variables improve the predictive accuracy of a Cox proportional hazards model for bank failure? *The Quartely Review of Economics and Finance* **36** 395–409.

[25] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A* **186** 453–461. MR0017504

[26] Kosmidis, I. (2013). brglm: bias reduction in binary-response generalized linear models R package version 0.5-9.

[27] Lane, W. R., Looney, S. W. and Wansley, J. W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance* **10** 511–531.

[28] Lin, I. F., Chang, W. P. and Liao, Y. N. (2013). Shrinkage methods enhanced the accuracy of parameter estimation using Cox models with small number of events. *Jounal of Clinical Epidemiology* **66** 743–751.

[29] Loughin, T. M. (1998). On the bootstrap and monotone likelihood in the Cox proportional hazards regression model. *Lifetime Data Analysis* **4** 393–403.

[30] Lyles, R. H., Guo, Y. and Greenland, S. (2012). Reducing bias and mean squared error associated with regression-based odds ratio estimators. *Journal of Statistical Planning and Inference* **142** 3235–3241. MR2956808

[31] Mansournia, M. A., Geroldinger, A., Greenland, S. and Heinze, G. (2017). Separation in logistic regression - causes, consequences and control. *American Journal of Epidemiology* **0** kwx299.

[32] Metropolis, N., Rosenbluth, A., Teller, M. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemistry and Physics* **21** 1087–1091.

[33] Neal, R. (2003). Slice sampling. *Annals of Statistics* **31** 705–767. MR1994729

[34] Oakes, D. (2013). An introduction to survival models: in honor of Ross Prentice. *Lifetime Data Analysis* **19** 442–462. MR3119992

[35] Pan, Q. and Schaubel, D. E. (2014). Proportional hazards regression in the presence of missing study eligibility information. *Lifetime Data Analysis* **20** 424–443. MR3217546

[36] Pianto, D. M. and Cribari-Neto, F. (2011). Dealing with monotone likelihood in a model for speckled data. *Computational Statistics and Data Analysis* **55** 1394–1409. MR2741423

[37] Ploner, M. and Heinze, G. (2015). coxphf: Cox regression with Firth's penalized likelihood R package version 1.11.

[38] Robert, C. P. (2007). *The Bayesian choice*, 2 ed. Springer, New York. MR2723361

[39] Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* **7** 110–120. MR1428751

[40] Sinha, D., Ibrahim, J. G. and Chen, M. H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika* **90** 629–641. MR2006840

[41] R Core Team (2017). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria.

[42] Wu, J., Castro, M., Schifano, E. D. and Chen, M. H. (2017). Assessing covariate effects using Jeffreys-type prior in the Cox model in the presence of a monotone partial likelihood. *Journal of Statistical Theory and Practice* **0** 1–19. http://dx.doi.org/10.1080/15598608.2017.1299058. MR3766710

Frederico M. Almeida
Av. Antônio Carlos, 6627
Departamento de Estatística
ICEx
Universidade Federal de Minas Gerais
Belo Horizonte, MG, 31270-901
Brazil
E-mail address: falmeida856@gmail.com

Enrico A. Colosimo
Av. Antônio Carlos, 6627
Departamento de Estatística
ICEx
Universidade Federal de Minas Gerais
Belo Horizonte, MG, 31270-901
Brazil
E-mail address: enricoc@est.ufmg.br

Vinícius D. Mayrink
Av. Antônio Carlos, 6627
Departamento de Estatística
ICEx
Universidade Federal de Minas Gerais
Belo Horizonte, MG, 31270-901
Brazil
E-mail address: vdm@est.ufmg.br