Ann Inst Stat Math (2016) 68:353–384
DOI 10.1007/s10463-014-0496-3

A semiparametric generalized proportional hazards model for right-censored data

M. L. Avendaño · M. C. Pardo

Received: 28 March 2013 / Revised: 30 September 2014 / Published online: 21 November 2014

Abstract We introduce a flexible family of semiparametric generalized logit-based regression models for survival analysis. Its hazard rates are proportional as the Cox model, but its relative risk related to a covariate is different for the values of the other covariates. The method of partial likelihood approach is applied to estimate its parameters in presence of right censoring and its asymptotic normality is established. We perform a simulation study to evaluate the finite-sample performance of these estimators. This new family of models is illustrated with lung cancer data and compared with Cox model. The importance of the conclusions obtained from the relative risk is pointed out.

Keywords Survival analysis · Proportional hazards · Type-I generalized logistic distribution · Semiparametric models · Profile likelihood · Partial likelihood

1 Introduction

Analysis of event times (also referred as survival analysis) deals with data representing the time to a well-defined event. These data arise in engineering, economy, reliability, public health, biomedicine and other areas. Data arising from survival analysis often

M. L. Avendaño (✉)
Faculty of Mathematics, University of Veracruz,
Circuito Gonzalo Aguirre Beltrán, s/n, Zona Universitaria, 91090 Xalapa, Mexico
e-mail: maravend@ucm.es

M. C. Pardo
Department of Statistics and Operational Research I,
Faculty of Mathematics, Complutense University of Madrid, Plaza de las Ciencias 3, Ciudad Universitaria, 28040 Madrid, Spain

consist of a response variable that measures the duration of time until the occurrence of a specific event and a set of variables (covariates) thought to be associated with the event-time variable, they have some features that present difficulties to traditional statistical methods. The first is that the data are generally asymmetrically distributed, while the second feature is that lifetimes are frequently censored (the end point of interest has not been observed for that individual). Regression models for survival data have traditionally been based on the Cox regression model (Cox 1972). One of the reasons for the popularity of this model is because the unknown parameter can be estimated by partial likelihood without putting a structure on baseline hazard. This model requires that the hazards between any two individuals are proportional across time. There are many works whose aim has been to extend the Cox model in different ways. Some of them appealing the necessity to allow non-proportional hazards such as Aranda-Ordaz (1983), Tibshirani and Ciampi (1983), Etezadi-Amoli and Ciampi (1987), Thomas (1986), Sasieni (1995), Clayton (1978), Hougaard (1984), Younes and Lachin (1997), MacKenzie (1996, 1997) and Devarajan and Ebrahimi (2011). However, when the assumption of proportional hazards is tenable, a Cox regression model is usually the preferred model. Therefore, proportional hazards models that maintain the good properties of the Cox model, but give flexibility to the model are welcome.

As mentioned above, regression models for survival data have traditionally been based on the proportional hazards model of Cox (1972) which is defined through the hazard function $\lambda(t \mid \mathbf{z})$ of the form

$$\lambda(t \mid \boldsymbol{\beta}, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}), \tag{1}$$

where $\lambda_0(\cdot)$ is an arbitrary function of time called baseline hazard function, $\mathbf{z}' = (z_1, \ldots, z_d)$ is a vector of covariates for the individual at time t, and $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_d)$ is a vector of unknown parameters to be estimated. In case that the baseline hazard function is treated non-parametrically, then this model becomes a semiparametric model.

In this paper, we introduce a flexible semiparametric family of proportional hazards models based on replacing the exponential link function in (1) by a generalization of the logistic distribution (see Balakrishnan 1992) called Type-I generalized logistic distribution in Sect. 2. The partial likelihood approach is proposed to estimate the covariate effects of the model considering right-censored data and an application is shown in Sects. 3 and 4, respectively. To prove the asymptotic normality of the partial likelihood estimator in our model, first we establish the equivalence of the partial likelihood estimator and the profile likelihood estimator for our model. Second, the efficiency of the profile likelihood estimator for our semiparametric model is proven. This proof is based on the method of an approximate least favorable submodel used by Murphy and van der Vaart (2000) with the Cox regression model and it is described in Sect. 5, but developed in an "Appendix". Finally, the small sample performance of the maximum partial likelihood estimators of the regression parameters in three- and four-covariate hazard function models is evaluated through a simulation study in Sect. 6.

2 The generalized proportional hazards model

The Cox model given in (1) can be generalized by replacing the exponential function in (1) by a generalization of the logistic distribution called Type-I generalized logistic distribution which is given by

$$F(y) = \left(\frac{1}{1+e^{-y}}\right)^{\alpha}, \qquad -\infty < y < \infty,\ \alpha > 0,$$

where the parameter α is a proportionality constant for the distribution. See Balakrishnan (1992).

Therefore, we propose a proportional hazards model defined through the hazard function

$$\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}) = \lambda_0(t) K(\boldsymbol{\theta}, \mathbf{z}), \tag{2}$$

with

$$K(\boldsymbol{\theta}, \mathbf{z}) = \left(\frac{1}{1+\exp(-\boldsymbol{\beta}'\mathbf{z})}\right)^{\alpha}, \quad \alpha > 0,$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)'$, which we call generalized logit-link proportional hazards model. In fact, if $\alpha = 0$ or $\boldsymbol{\beta} = 0$, the model reduces to a non-parametric form.

The cumulative hazard function for the model (2) is given by

$$\Lambda(t|\boldsymbol{\theta}, \mathbf{z}) = \Lambda_0(t) K(\boldsymbol{\theta}, \mathbf{z}),$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)\mathrm{d}u$ is the baseline cumulative hazard function. Thus, the survival function corresponding to the hazard model is

$$S(t \mid \boldsymbol{\theta}, \mathbf{z}) = \exp(-\Lambda(t|\boldsymbol{\theta}, \mathbf{z})),$$

and Eq. (2) characterizes the model with density given by

$$f(t \mid \boldsymbol{\theta}, \mathbf{z}) = \exp(-\Lambda(t \mid \boldsymbol{\theta}, \mathbf{z}))\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}).$$

Note that, in the particular case $\alpha = 1$, we obtain the proportional hazards model with a logit-link function

$$\lambda(t \mid \boldsymbol{\beta}, \mathbf{z}) = \lambda_0(t)\frac{\exp(\boldsymbol{\beta}'\mathbf{z})}{1+\exp(\boldsymbol{\beta}'\mathbf{z})}.$$

For two-covariate profiles $\mathbf{z}_i$ and $\mathbf{z}_j$, the hazard ratio is proportional and the relative risk does not depend on t as

$$\begin{aligned}\rho(t \mid \boldsymbol{\theta}, \mathbf{z}_i, \mathbf{z}_j) &= \frac{\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}_i)}{\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}_j)} \\ &= \left(\frac{1+\exp(-\boldsymbol{\beta}'\mathbf{z}_j)}{1+\exp(-\boldsymbol{\beta}'\mathbf{z}_i)}\right)^{\alpha} \\ &= \rho(t \mid \boldsymbol{\beta}, \alpha, \mathbf{z}_i, \mathbf{z}_j).\end{aligned}$$

The interpretation of the hazard ratio of model (2) is similar to the Cox model hazard ratio since if $\rho(t \mid \boldsymbol{\theta}, \mathbf{z}_i, \mathbf{z}_j) = 1$, then the individuals with covariates $\mathbf{z}_i$ and $\mathbf{z}_j$ have the same relative risk of death and if $\rho(t \mid \boldsymbol{\theta}, \mathbf{z}_i, \mathbf{z}_j) < 1 (>1)$, the individual with covariate $\mathbf{z}_i$ has lower (higher) relative risk of death than the individual with covariate $\mathbf{z}_j$.

As particular case, when the covariate z_m increases $z_m + 1$ and the rest of covariates remain equal the relative risk is equal (lower or higher) if and only if $\beta_m = 0$ (<0 or >0), respectively, and the parameter α is a proportionality parameter for the model since the relative risk increases or decreases when α increases depending if the relative risk is less or greater than 1. Moreover, this parameter gives flexibility to the model to adapt better to the data.

For illustrative purposes, we consider a two-covariate model, the first one a continuous covariate and the second one a 4-level factor covariate. The 4-level covariate is transformed in an indicator 3-dimensional vector. So, we obtain a full model given by

$$\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}) = \lambda_0(t)\left(\frac{1}{1+\exp(-(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \beta_4 z_4))}\right)^{\alpha}, \quad \alpha > 0.$$

In this case, if we take the relative risk of individuals with values $z_1 = z_{1i}$ and $z_1 = z_{1j}$, respectively, at the first covariate and the same level at the second covariate, we have that the relative risk depend on the level of the second covariate, that is:

$$\begin{aligned}\text{Level I:} &\ \left(\frac{1+\exp(-z_{1j}\beta_1)}{1+\exp(-z_{1i}\beta_1)}\right)^{\alpha}\\ \text{Level II:} &\ \left(\frac{1+\exp(-z_{1j}\beta_1-\beta_2)}{1+\exp(-z_{1i}\hat{\beta}_1-\beta_2)}\right)^{\alpha}\\ \text{Level III:} &\ \left(\frac{1+\exp(-z_{1j}\beta_1-\beta_3)}{1+\exp(-z_{1i}\beta_1-\beta_3)}\right)^{\alpha}\\ \text{Level IV:} &\ \left(\frac{1+\exp(-z_{1j}\beta_1-\beta_4)}{1+\exp(-z_{1i}\beta_1-\beta_4)}\right)^{\alpha}.\end{aligned}$$

Note that, in this case, the relative risk under Cox model (1) is given by

$$\rho_{\text{Cox}}(t|\boldsymbol{\beta}, \mathbf{z}_i, \mathbf{z}_j) = \exp(\beta_1(z_{1i} - z_{1j}))$$

which does not depend on the second covariate level. Therefore, the proposed model gives more flexibility to the relative risk.

Up to now, we have not made any assumption about the baseline hazard function $\lambda_0(\cdot)$, but in the following we assume an unknown functional form for the baseline hazard function of the model, so we have a semiparametric model.

3 Estimation procedure

Consider a right-censored data sample of size n where the observed data are the iid triplets $(X_i, \delta_i, \mathbf{Z}_i)$ where $X_i = \min(T_i, C_i)$ is the observed time with T_i and C_i the survival and censoring time, respectively, $\delta_i = 1_{\{T_i \leq C_i\}}$ is the indicator variable of censoring and $\mathbf{Z}_i$ is the vector of covariates for $i = 1, \ldots, n$.

In this case, analogously to Cox (1972), we can construct the partial likelihood function for the sample, and it can be written as

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\frac{K(\boldsymbol{\theta}, \mathbf{z}_i)}{\sum_{j \in R(x_i)} K(\boldsymbol{\theta}, \mathbf{z}_j)} \right]^{\delta_i},$$

where $R(x_i)$ is the risk set at time x_i, and the partial log-likelihood function is given by

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \ln(L_n(\boldsymbol{\theta})) \\ &= \sum_{i=1}^{n} \delta_i \left[\ln(K(\boldsymbol{\theta}, \mathbf{z}_i)) - \ln\left(\sum_{j \in R(x_i)} K(\boldsymbol{\theta}, \mathbf{z}_j) \right) \right]. \end{aligned} \tag{3}$$

Then, the maximum partial log-likelihood estimator of the unknown vector of parameters of model (2) is given by

$$\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) = \arg \max_{\boldsymbol{\theta}=(\boldsymbol{\beta}, \alpha>0)} l_n(\boldsymbol{\theta}), \tag{4}$$

and the variance of the parameter estimator $\hat{\boldsymbol{\theta}}_n$ is

$$\text{var}(\hat{\boldsymbol{\theta}}_n) = \text{diag}(\hat{\mathbf{I}}_n^{-1}(\hat{\boldsymbol{\theta}}_n)),$$

where $\hat{\mathbf{I}}_n$ is the observed information matrix.

To obtain (4), we would need to use numerical methods for carrying out the required constrained ($\alpha > 0$) optimization. However, the numerical algorithms are high sensitive to initial values of the parameters and they fall at local optimums. To avoid it, we consider $\{\alpha_1, \ldots, \alpha_l\}$ a grid for the value of the parameter α then we maximize the partial log-likelihood function so we obtain l estimations of $\boldsymbol{\beta}$, $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_l\}$ with partial log-likelihood function values $\{l_n(\alpha_1, \boldsymbol{\beta}_1), \ldots, l_n(\alpha_l, \boldsymbol{\beta}_l)\}$. Finally, we consider $\arg\max\{l_n(\alpha_1, \boldsymbol{\beta}_1), \ldots, l_n(\alpha_l, \boldsymbol{\beta}_l)\}$ as estimations of the parameters $(\alpha, \boldsymbol{\beta})$.

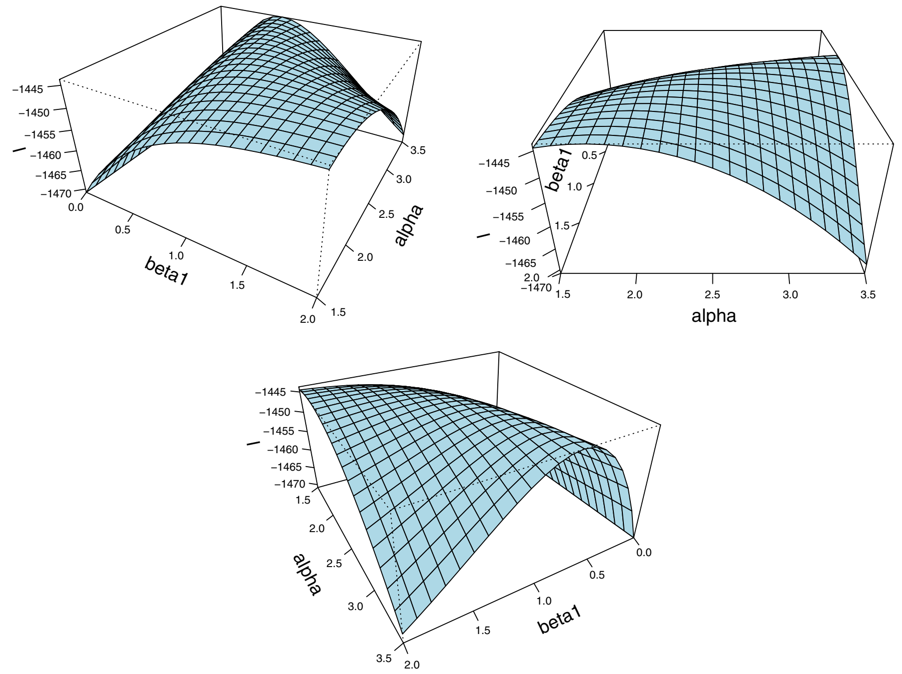

Fig. 1 Surface for partial log-likelihood function

To check how the method employed works and also if there exists an identifiable problem for $\boldsymbol{\beta}$ and α, we show in Fig. 1 a surface with the values of the partial log-likelihood function for different values of α and β_1 for a data set of size 500 generated from our model with the covariate z_1 a Bernoulli of parameter 0.45 and $z_2 = 0$ for $\alpha = 2.5$ and $\beta_1 = \beta_2 = 1$. It shows clearly that the method employed works very well. Furthermore, we graph the level curves for the partial log-likelihood function in Fig. 2. The true value of the parameters is the red square and the blue circle is that which achieves the maximum value for the partial log-likelihood function. It shows there is non-identifiability issue even if we have only one covariate.

4 Veteran's administration lung cancer data

We present results from fitting the generalized proportional hazards model to a dataset from the Veteran's Administration Lung Cancer Study Clinical Trial (Kalbfleisch and Prentice 2002). In this clinical trial, males with advanced inoperable lung cancer were randomized to either a standard or test chemotherapy. The primary end point for therapy comparison was time to death. Only nine out of the 137 survival times were censored, so the censoring rate is 6.6 %. The dataset includes six covariates: treatment (1 = standard and 2 = test), age at diagnosis (in years), Karnofsky score (from 0 to 100), diagnosis time, cell type (1 = Squamous, 2 = Small, 3 = Adenocarcinoma and 4 = Large) and prior therapy. The primary purpose of this clinical trial was to

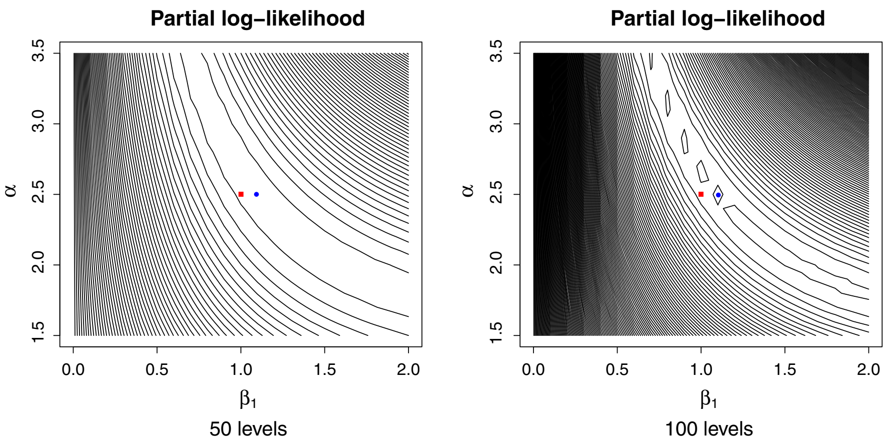

Fig. 2 Level curves for partial log-likelihood function

Table 1 Fit of models for veteran's administration lung cancer data

Covariate	PH		GLPH	
	Est	SE	Est	SE
Karnofsky	−0.0309	0.00518	−0.0083	0.00160
Cell type 2	0.7121	0.25274	0.1869	0.07310
Cell type 3	1.1508	0.29286	0.3100	0.09806
Cell type 4	0.3251	0.27669	0.0837	0.08006
α	–	–	6.5	0.98507
AIC	960.9972		962.8318	

investigate whether the new chemotherapy works better or worse than the standard chemotherapy after adjusting for other covariates. The Veteran data were used by Kalbfleisch and Prentice (2002) to illustrate the Cox proportional hazards model and the treatment effect was found to be non-significant. The veteran cancer data can be obtained from the `randomSurvivalForest` Package of the `R` statistical software, these data are called `veteran`.

In this example, we consider only two covariates: Karnofsky score and cell type (factor with four levels). We fit the Cox proportional hazards model (PH) and the generalized logit-link proportional hazards model (GLPH). In Table 1, we can see the parameter estimates (Est) and their standard errors (SE) for the two fitted models, for GLPH model we use parametric bootstrap with 500 replicas to obtain the SE of the parameters. We compare the fit of the two models with the the Akaike's entropy-based Information Criterion (AIC). From these results, we conclude that the fit of the two models is similar.

On the other hand, the interpretation of regression coefficients is similar in both models. However, the interpretation of their relative risks differs. We consider two individuals with a difference of 10 in the Karnofsky score value with the same cell type. Under PH model the relative risk is

$$\exp(10\hat{\beta}_1) \approx 0.73.$$

However, under the GLPH model the relative risk depends on the Cell type of the individuals. By way of illustration, we consider individuals with a difference of Karnofsky score of 10, for example, individuals with Karnofsky score of 20 and 10, respectively:

$$\text{Cell type 1}: \left(\frac{1+\exp(-10\hat{\beta}_1)}{1+\exp(-20\hat{\beta}_1)}\right)^{\hat{\alpha}} \approx 0.75$$

$$\text{Cell type 2}: \left(\frac{1+\exp(-10\hat{\beta}_1-\hat{\beta}_2)}{1+\exp(-20\hat{\beta}_1-\hat{\beta}_2)}\right)^{\hat{\alpha}} \approx 0.77$$

$$\text{Cell type 3}: \left(\frac{1+\exp(-10\hat{\beta}_1-\hat{\beta}_3)}{1+\exp(-20\hat{\beta}_1-\hat{\beta}_3)}\right)^{\hat{\alpha}} \approx 0.78$$

$$\text{Cell type 4}: \left(\frac{1+\exp(-10\hat{\beta}_1-\hat{\beta}_4)}{1+\exp(-20\hat{\beta}_1-\hat{\beta}_4)}\right)^{\hat{\alpha}} \approx 0.76.$$

Note that, an individual with a Karnofsky score of 20 compared with an individual with a Karnofsky score of 10 has lower relative risk, but now it is different depending on the cell type. The individuals with cell type 1 have the lowest risk. The individuals with cell type 3 have the highest risk.

On other hand, considering individuals with Karnofsky score of 100 and 90, respectively (also a difference of 10), the relative risks are:

$$\text{Cell type 1}: \left(\frac{1+\exp(-90\hat{\beta}_1)}{1+\exp(-100\hat{\beta}_1)}\right)^{\hat{\alpha}} \approx 0.69$$

$$\text{Cell type 2}: \left(\frac{1+\exp(-90\hat{\beta}_1-\hat{\beta}_2)}{1+\exp(-100\hat{\beta}_1-\hat{\beta}_2)}\right)^{\hat{\alpha}} \approx 0.71$$

$$\text{Cell type 3}: \left(\frac{1+\exp(-90\hat{\beta}_1-\hat{\beta}_3)}{1+\exp(-100\hat{\beta}_1-\hat{\beta}_3)}\right)^{\hat{\alpha}} \approx 0.72$$

$$\text{Cell type 4}: \left(\frac{1+\exp(-90\hat{\beta}_1-\hat{\beta}_4)}{1+\exp(-100\hat{\beta}_1-\hat{\beta}_4)}\right)^{\hat{\alpha}} \approx 0.70,$$

thus, an individual with a Karnofsky score of 100 compared with an individual with a Karnofsky score of 90 has lower relative risk to die than when we compare to individuals with Karnofsky scores of 10 and 20, respectively, and this risk is different depending on the cell type. The individuals with cell type 1 have the lowest risk. The individuals with cell type 3 have the highest risk.

To sum up, the relative risk under the Cox model is independent from the Cell type and also the concrete Karnofsky score since only it takes into account the difference between the score of the individuals. Therefore, the proposed model provides useful information from its relative risk.

5 Asymptotic normal distribution of the parameter estimators

To make inference, we establish the asymptotic normal distribution of the parameter estimator of the generalized proportional hazards model. To do it, we use the main Theorem of Murphy and van der Vaart (2000) and its Corollary 1, (see Kosorok 2008). Note that, this theorem considers a maximum profile log-likelihood estimator and we propose a maximum partial log-likelihood estimator, then it is necessary to verify that both estimators are equivalent.

Consider the semiparametric model

$$(\boldsymbol{\theta}, \Lambda_0) \mapsto P_{\boldsymbol{\theta},\Lambda_0},$$

where $\boldsymbol{\theta}$ is the regression parameter and Λ_0 the cumulative hazard function in the model (2) and $P_{\boldsymbol{\theta},\Lambda_0}$ is the data distribution function. Recall that an observation with right censoring is given by $\mathbf{Y} = (X, \delta, \mathbf{Z})$ where $X = \min(T, C)$, $\delta = 1_{\{T \leq C\}}$ and $\mathbf{Z} \in \mathbb{R}^d$ is the covariate vector. We assume that T is the survival time with cumulative hazard function given $\mathbf{Z}$

$$\Lambda(t|\mathbf{z}) = \Lambda_0(t)K(\boldsymbol{\theta}, \mathbf{z})$$

and C is a censoring time independent of T given $\mathbf{Z}$ and uninformative of $(\boldsymbol{\theta}, \Lambda_0)$.

In this case, the density function of $\mathbf{y} = (x, \delta, \mathbf{z})$ is given by

$$p_{\boldsymbol{\theta},\Lambda_0}(\mathbf{y}) = [\lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z})S_{T|\mathbf{Z}}(x|\mathbf{z})S_{C|\mathbf{Z}}(x|\mathbf{z})]^{\delta}[S_{T|\mathbf{Z}}(x|\mathbf{z})p_{C|\mathbf{Z}}(x|\mathbf{z})]^{1-\delta}p_{\mathbf{Z}}(\mathbf{z}).$$

Thus, the log-likelihood for the observation $\mathbf{y}$ is given by

$$\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \log p_{\boldsymbol{\theta},\Lambda_0}(\mathbf{y}) = \delta[\log \lambda_0(x) + \log K(\boldsymbol{\theta}, \mathbf{z})] - \Lambda_0(x)K(\boldsymbol{\theta}, \mathbf{z}) + c \quad (5)$$

where c is a constant. Note that, the log-likelihood function depends on the parameters $\boldsymbol{\theta}$ and Λ_0. Furthermore, in "Appendix A" it is shown that its efficient score function for $\boldsymbol{\theta}$ is

$$\tilde{\ell}_{\boldsymbol{\theta},\Lambda_0}(\mathbf{y}) = \delta(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - h_0(x)) - K(\boldsymbol{\theta}, \mathbf{z}) \int_0^x (\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - h_0(s))\mathrm{d}\Lambda_0(s), \quad (6)$$

where

$$\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) = ([K_1(\boldsymbol{\theta}, \mathbf{z})]', K_2(\boldsymbol{\theta}, \mathbf{z}))' \quad (7)$$

with

$$K_1(\boldsymbol{\theta}, \mathbf{z}) = \frac{\alpha \mathbf{z}}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})},$$
$$K_2(\boldsymbol{\theta}, \mathbf{z}) = \log[K(\boldsymbol{\theta}, \mathbf{z})]$$

and $h_0 : \mathbb{R} \to \mathbb{R}^{d+1}$ is the least favorable direction (see "Appendix A").

We consider the log-likelihood function based on a random sample of n observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$ which is given by

$$\ell_n(\boldsymbol{\theta}, \Lambda_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}_i),$$

then the profile log-likelihood for $\boldsymbol{\theta}$ is defined as

$$p\ell_n(\boldsymbol{\theta}) = \sup_{\Lambda_0} \ell_n(\boldsymbol{\theta}, \Lambda_0). \tag{8}$$

Note that, the maximum likelihood estimator for $\boldsymbol{\theta}$, the first component of the pair $(\hat{\boldsymbol{\theta}}_n, \hat{\Lambda}_0)$ that maximizes $\ell_n(\boldsymbol{\theta}, \Lambda_0)$, is the maximizer of the profile log-likelihood function $\boldsymbol{\theta} \mapsto p\ell_n(\boldsymbol{\theta})$. Then, $\tilde{\boldsymbol{\theta}}_n$ denote the maximum profile log-likelihood estimator.

To establish the asymptotic properties of the parameter estimates $\tilde{\boldsymbol{\theta}}_n$ we assume the following:

(i) Exist $\tau < \infty$ such that $P_0(C \geq \tau) = P_0(C = \tau) > 0$, where P_0 is the true probability measure.
(ii) The observed time $X \in [0, \tau]$.
(iii) Let H be all monotone increasing continuous functions Λ_0 on $[0, \tau]$ with $\Lambda_0(0) = 0$. We define $\hat{H}$ to be the set of all monotone increasing Cadlag functions Λ_0 on $[0, \tau]$ with $\Lambda_0(0) = 0$ and $\Lambda_0^0 \in \hat{H}$, where Λ_0^0 is the true value of the parameter.
(iv) The vector $\mathbf{Z} \in \mathbb{R}^d$ is bounded.
(v) Let $\boldsymbol{\Theta} \subset \mathbb{R}^d \times \mathbb{R}^+$ be a compact set such as $\boldsymbol{\theta}_0 \in \Theta$, where $\boldsymbol{\theta}_0$ is the true value of the parameter.
(vi) The efficient Fisher information matrix $\tilde{\mathbf{I}}_0$ is invertible.

Theorem 1 *Under regularity conditions (i)–(vi), the maximum profile log-likelihood estimator $\tilde{\boldsymbol{\theta}}_n$ for the generalized proportional hazards model* (2) *obtained maximizing* (8) *is asymptotically normal distributed as*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{d} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1})$$

where $\tilde{\mathbf{I}}_0 = P_0[(\tilde{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0})(\tilde{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0})']$ *with* $\tilde{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0}$ *given in* (6).

Proof To apply Corollary 1 of the main result of Murphy and van der Vaart (2000) we need to verify Conditions (8)–(11) of Murphy and van der Vaart (2000) and also conditions of its Theorem 1.

Let us define the least favorable submodel

$$\mathbf{t} \mapsto \Lambda_{\mathbf{t}}(\boldsymbol{\theta}, \Lambda_0)$$

with $\mathbf{t} \in \mathbb{R}^{d+1}$, where

$$\Lambda_{\mathbf{t}}(\boldsymbol{\theta}, \Lambda_0)(s) = \Lambda_{\mathbf{t}}(s) = \int_0^s (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(u)) \mathrm{d}\Lambda_0(u),$$

where $h_0 : \mathbb{R} \to \mathbb{R}^{d+1}$ is the least favorable direction at the true parameter $(\boldsymbol{\theta}_0, \Lambda_0^0)$. For $\mathbf{t}$ small enough $\Lambda_{\mathbf{t}}(\boldsymbol{\theta}, \Lambda_0) \in \hat{H}$. Moreover, we obtain

$$\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \Lambda_0)(\cdot) = \Lambda_{\boldsymbol{\theta}}(\cdot) = \Lambda_0(\cdot),$$

that is, Condition (8) of Murphy and van der Vaart (2000) is satisfied.

The least favorable direction $h_0(\cdot)$ is given by

$$h_0(x) = \frac{P_0[K(\boldsymbol{\theta}, \mathbf{z}) 1_{\{X \geq x\}}]}{P_0[\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z}) 1_{\{X \geq x\}}]},$$

with $\tilde{K}(\boldsymbol{\theta}, \mathbf{z})$ as (7). Details about the construction of $h_0(\cdot)$ is given in "Appendix A" at the end of this work.

On the other hand, we consider the map $\mathbf{t} \mapsto \ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(y)$ defined by

$$\ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \log p_{\mathbf{t}, \Lambda_{\mathbf{t}}}(\mathbf{y}), \tag{9}$$

thus the log-likelihood for the data $\mathbf{y}$ under the least favorable submodel is given by

$$\begin{aligned} \ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta[&\log((1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x)) \lambda_0(x)) \\ &+ \log K(\mathbf{t}, \mathbf{z})] - K(\mathbf{t}, \mathbf{z}) \int_0^x (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) \mathrm{d}\Lambda_0(s) \end{aligned}$$

and its score function for $\mathbf{t}$ is given by

$$\begin{aligned} \dot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) =& \frac{\partial}{\partial \mathbf{t}} \ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) && (10) \\ =& \delta \tilde{K}(\mathbf{t}, \mathbf{z}) - \int_0^\tau \tilde{K}(\mathbf{t}, \mathbf{z}) K(\mathbf{t}, \mathbf{z}) 1_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) \mathrm{d}\Lambda_0(s) \\ & - \frac{\delta h_0(x)}{(1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(x))} + \int_0^\tau K(\mathbf{t}, \mathbf{z}) 1_{\{X \geq s\}} h_0(s) \mathrm{d}\Lambda_0(s) && (11) \\ =& \int_0^\tau \left(\tilde{K}(\mathbf{t}, \mathbf{z}) - \frac{h_0(s)}{1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)} \right) \mathrm{d}M(s), && (12) \end{aligned}$$

with $\mathrm{d}M(s) = \mathrm{d}1_{\{X \leq s\}} \delta - K(\mathbf{t}, \mathbf{z}) 1_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \mathbf{t})' h_0(s)) \mathrm{d}\Lambda_0(s)$. Thus, we have

$$\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y}) = \tilde{\ell}_{\boldsymbol{\theta}_0, \Lambda_0^0}(\mathbf{y}),$$

this is, Condition (9) of Murphy and van der Vaart (2000) is satisfied.

In addition, considering the likelihood $\mathcal{L}(\boldsymbol{\theta}, \Lambda_0)$, the maximizer of $\Lambda_0 \mapsto \mathcal{L}(\boldsymbol{\theta}, \Lambda_0)$ has the form

$$\hat{\Lambda}_{\boldsymbol{\theta}}(t) = \int_0^t \frac{\mathbb{P}_n[\mathrm{d}N(u)]}{\mathbb{P}_n[1_{\{X \geq u\}} K(\boldsymbol{\theta}, \mathbf{z})]}, \tag{13}$$

with $\mathbb{P}_n$ the empirical distribution function and $N(u) = 1_{\{X \leq u\}}\delta$. This estimator is equivalent to the Breslow cumulative baseline hazard estimator in Cox model. Note that, the estimator (13) depends on the parameter $\boldsymbol{\theta}$. In Theorem 5 ("Appendix B") it is verified that $\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}$ is uniformly consistent for Λ_0 for any sequence $\tilde{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$. Thus, Condition (10) of Murphy and van der Vaart (2000) is satisfied.

The non-bias Condition (11) of Murphy and van der Vaart (2000) is verified in Lemma 6 in "Appendix B" of this work.

Finally, standard arguments allow to prove Donsker and Glivenko–Cantelli conditions of Theorem 1 of Murphy and van der Vaart (2000) (see "Appendix B").

Then, we can apply Theorem 1 of Murphy and van der Vaart (2000) and we obtain

$$\log p\ell_n(\tilde{\boldsymbol{\theta}}_n) = \log p\ell_n(\boldsymbol{\theta}_0) + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \sum_{i=1}^n \tilde{\ell}_0(\mathbf{y}_i)$$
$$- \frac{1}{2} n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \tilde{\mathbf{I}}_0 (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{P_0}(\sqrt{n}\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| + 1)^2$$

where $\tilde{\ell}_0$ is the efficient score function for $\boldsymbol{\theta}$ at $(\boldsymbol{\theta}_0, \Lambda_0^0)$ given in (6) and $\tilde{\mathbf{I}}_0$ is the efficient Fisher matrix for $\boldsymbol{\theta}$ at $(\boldsymbol{\theta}_0, \Lambda_0^0)$.

Now, assuming that $\tilde{\mathbf{I}}_0$ is invertible (Assumption iv)) and Theorem 5 of "Appendix B", we can apply Corollary 1 of Murphy and van der Vaart (2000) and we obtain

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n \to \infty]{d} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_0^{-1}),$$

where, the efficient Fisher matrix is given by

$$\tilde{\mathbf{I}}_0 = P_0[(\tilde{\ell}_{\boldsymbol{\theta}_0, \Lambda_0^0})(\tilde{\ell}_{\boldsymbol{\theta}_0, \Lambda_0^0})'].$$

□

Last theorem establishes the asymptotic distribution of the maximum profile log-likelihood estimator of the unknown parameters of model (2), but we propose to use maximum partial log-likelihood estimator. Nevertheless, next lemma shows equivalence of the partial likelihood and profile likelihood estimators.

Lemma 2 *The estimator $\tilde{\boldsymbol{\theta}}_n$ based on the profile log-likelihood for the model (2) under right censoring and the estimator $\hat{\boldsymbol{\theta}}_n$ based on the partial log-likelihood function (3) are equivalent.*

Proof We define the function $N_i(x) = 1_{\{X_i \le x\}}\delta_i$, thus the partial likelihood function for the sample is given by

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{0\le x\le\tau} \left(\frac{1_{\{X_i \ge x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)}{\sum_{j=1}^{n} 1_{\{X_j \ge x\}} K(\boldsymbol{\theta}, \mathbf{z}_j)} \right)^{\Delta N_i(x)},$$

the partial log-likelihood is

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \log(L_n(\boldsymbol{\theta})) \\ &= \sum_{i=1}^{n} \int_0^{\tau} \left(\log[1_{\{X_i \ge x\}} K(\boldsymbol{\theta}, \mathbf{z}_i)] - \log \left[\sum_{j=1}^{n} 1_{\{X_j \ge x\}} K(\boldsymbol{\theta}, \mathbf{z}_j) \right] \right) \mathrm{d}N_i(x), \end{aligned}$$

and its score function is given by

$$\frac{\partial}{\partial \boldsymbol{\theta}} l_n(\boldsymbol{\theta}) = n\mathbb{P}_n \left[\int_0^{\tau} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{\mathbb{P}_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{\mathbb{P}_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \mathrm{d}N(x) \right]. \quad (14)$$

On the other hand, if we consider the second term of the score efficient function (6) and the empirical density function we have

$$\begin{aligned} &\mathbb{P}_n \left[-K(\boldsymbol{\theta}, \mathbf{z}) \int_{[0,x]} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{\mathbb{P}_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{\mathbb{P}_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \mathrm{d}\hat{\Lambda}_0(s) \right] \\ &= - \int_0^{\tau} \Bigg(\mathbb{P}_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})] - \\ &\qquad \mathbb{P}_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})] \frac{\mathbb{P}_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{\mathbb{P}_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \Bigg) \mathrm{d}\hat{\Lambda}_0(s) \\ &= 0, \end{aligned}$$

with $\hat{\Lambda}_0(\cdot)$ the estimator of the cumulative baseline hazard function. Taking the first term of (6) we have

$$\begin{aligned} &\mathbb{P}_n \left[\delta \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \right] \\ &\quad = \mathbb{P}_n \left[\int_0^{\tau} \left(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \frac{E_n[1_{\{X \ge x\}} \tilde{K}(\boldsymbol{\theta}, \mathbf{z}) K(\boldsymbol{\theta}, \mathbf{z})]}{E_n[1_{\{X \ge x\}} K(\boldsymbol{\theta}, \mathbf{z})]} \right) \mathrm{d}N(x) \right]. \end{aligned}$$

Thus, the estimator $\hat{\boldsymbol{\theta}}_n$ and the estimator $\tilde{\boldsymbol{\theta}}_n$ are equivalent. □

6 Simulation study

The aim of this section is to check the normality of the estimators of the covariate effects of model (2) for finite-sample through a simulation study.

The survival times t were generated under the model

$$\lambda(t \mid \boldsymbol{\theta}, \mathbf{z}) = \lambda_0(t) K(\boldsymbol{\theta}, \mathbf{z}),$$

with

$$K(\boldsymbol{\theta}, \mathbf{z}) = \left(\frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})} \right)^{\alpha}, \quad \alpha > 0 \ \text{ fixed}$$

as

$$t = \Lambda_0^{-1} \left(\frac{-\log(u)}{K(\boldsymbol{\theta}, \mathbf{z})} \right),$$

where $u \sim \mathcal{U}(0, 1)$ (Bender et al. (2005)). We assume that the cumulative baseline hazard function Λ_0 is a Weibull distribution with shape parameter $\gamma = 3.2$ and scale parameter $\mu = 1.1$. We consider three different experiment designs:

First Design: We consider $\mathbf{Z}'\boldsymbol{\beta} = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4$ with true values of the parameters $\boldsymbol{\beta}_0 = (0.5, 0.2, 0.3, 0.1)$ and $\alpha = 8$, the covariates Z_1, Z_2 and Z_3 are binary variables that were generated from a factor of 4 levels and each level has a probability of $\{0.37, 0.19, 0.30, 0.14\}$, respectively, and Z_4 is generated from a $\mathcal{U}(-1, 1)$.

Second Design: We consider $\mathbf{Z}'\boldsymbol{\beta} = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3$ with true values of the parameters $\boldsymbol{\beta}_0 = (0.3, 0.6, 0.1)$ and $\alpha = 8$, the covariates Z_1, Z_2 and Z_3 were generated from a $\mathcal{B}(0.35)$, $\mathcal{U}(-1, 1)$ and $\mathcal{N}(1.5, 2)$, respectively.

Third Design: We consider $\mathbf{Z}'\boldsymbol{\beta} = Z_1\beta_1 + Z_2\beta_2$ with true values of the parameters $\boldsymbol{\beta}_0 = (2, -1)$ and $\alpha = 0.2$, the covariates Z_1 and Z_2 were generated from a $\mathcal{N}(0, 1)$ and $\mathcal{U}(-1, 1)$, respectively.

Several sample sizes $n = 100, 200, 400$ and different percent of censoring 15, 45 and 70 % for First and Second designs were considered and only 15 % of censoring for Third design was considered. A total of 500 simulated datasets were generated for each configuration of three designs.

In Sect. 5, we proved the asymptotic normality of the estimators of the covariate effects and now we investigate this property for small and moderate sample sizes. Therefore, we fit a generalized logit-based proportional hazard model for each one of the 500 simulate datasets. In Tables 2, 3 and 5, we present the average and the median of the estimators of the covariate effects, their estimated standard errors (SE), empirical standard errors (SEe), the p value for the Shapiro–Wilks test for testing normality by marginals and the coverage probabilities (CP) of a 90, 95 and 99 % confidence interval by marginals, respectively. In Table 4, the p values for the Henze–Zirkler multivariate normality test are presented for the First and Second Designs for each sample size

Table 2 Average and median of the parameter estimates, standard error (SE), empirical standard error (SEe), p value for Shapiro–Wilks test and coverage probabilities (CP) for a nominal 90, 95 and 99 %, respectively, for the first design

Censoring	Sample size	Parameter	Average	Median	SE	SEe	p value	CP for 0.90	CP for 0.95	CP for 0.99
15 %	100	β_1	0.52896	0.52070	0.12460	0.13066	0.00002	0.872	0.924	0.976
		β_2	0.21965	0.21197	0.08306	0.08491	0.00005	0.872	0.930	0.980
		β_3	0.32589	0.32140	0.11383	0.11941	0.00000	0.914	0.952	0.986
		β_4	0.10197	0.09954	0.05665	0.05979	0.01339	0.884	0.942	0.984
	200	β_1	0.51267	0.49862	0.08451	0.09725	0.00000	0.846	0.912	0.978
		β_2	0.20341	0.20339	0.05660	0.05903	0.17780	0.878	0.942	0.984
		β_3	0.31225	0.30518	0.07773	0.08479	0.00070	0.890	0.938	0.978
		β_4	0.09996	0.10109	0.03870	0.03866	0.19144	0.904	0.948	0.986
	400	β_1	0.50603	0.50043	0.05899	0.06881	0.00854	0.822	0.898	0.978
		β_2	0.20462	0.20295	0.03981	0.04011	0.04824	0.904	0.954	0.986
		β_3	0.30716	0.30274	0.05424	0.05973	0.00303	0.862	0.922	0.980
		β_4	0.10179	0.10170	0.02713	0.02740	0.78848	0.876	0.934	0.990
45 %	100	β_1	0.54146	0.52331	0.16010	0.17754	0.00000	0.858	0.900	0.956
		β_2	0.21324	0.20800	0.10441	0.11171	0.00018	0.870	0.932	0.982
		β_3	0.32191	0.30361	0.14705	0.15981	0.00000	0.878	0.916	0.974
		β_4	0.11144	0.10794	0.07221	0.07659	0.16895	0.866	0.934	0.976
	200	β_1	0.52802	0.51629	0.10798	0.11714	0.00066	0.856	0.918	0.972
		β_2	0.20995	0.20793	0.07161	0.07443	0.39918	0.886	0.938	0.986
		β_3	0.31986	0.31547	0.09907	0.10533	0.02710	0.862	0.928	0.978
		β_4	0.10801	0.10607	0.04909	0.04939	0.33180	0.902	0.946	0.984
	400	β_1	0.50671	0.50083	0.07331	0.08141	0.00150	0.860	0.912	0.978
		β_2	0.20492	0.20530	0.04959	0.05388	0.01724	0.896	0.938	0.982

Table 2 continued

Censoring	Sample size	Parameter	Average	Median	SE	SEe	p value	CP for 0.90	CP for 0.95	CP for 0.99
		β_3	0.30805	0.30543	0.06759	0.07702	0.00389	0.868	0.918	0.978
		β_4	0.10229	0.10410	0.03383	0.03402	0.79969	0.900	0.942	0.984
70 %	100	β_1	0.56340	0.52049	0.44202	0.51533	0.00000	0.834	0.884	0.938
		β_2	0.24258	0.20871	0.28117	0.47029	0.00000	0.864	0.908	0.962
		β_3	0.30960	0.29561	0.63577	0.62722	0.00000	0.860	0.914	0.956
		β_4	0.10420	0.09984	0.10412	0.16359	0.00000	0.866	0.912	0.964
	200	β_1	0.52015	0.50629	0.14801	0.15207	0.00002	0.866	0.926	0.984
		β_2	0.21102	0.20937	0.09808	0.10264	0.04140	0.894	0.940	0.974
		β_3	0.30352	0.31030	0.19855	0.19320	0.00000	0.900	0.940	0.972
		β_4	0.10836	0.10876	0.06750	0.06929	0.48689	0.874	0.926	0.982
	400	β_1	0.51422	0.50877	0.10090	0.11539	0.02883	0.842	0.902	0.962
		β_2	0.20797	0.20566	0.06760	0.07152	0.43157	0.872	0.922	0.982
		β_3	0.30877	0.31168	0.09271	0.09495	0.59518	0.886	0.942	0.980
		β_4	0.10019	0.09903	0.04614	0.04748	0.49890	0.874	0.946	0.992

Table 3 Average and median of the parameter estimates, standard error (SE), empirical standard error (SEe), *p* value for Shapiro–Wilks test and coverage probabilities (CP) for a nominal 90, 95 and 99 %, respectively, for the second design

Censoring	Sample size	Parameter	Average	Median	SE	SEe	*p* value	CP for 0.90	CP for 0.95	CP for 0.99
15 %	100	β_1	0.31225	0.30630	0.08559	0.09181	0.00021	0.864	0.938	0.972
		β_2	0.62663	0.61607	0.09405	0.10834	0.00000	0.840	0.894	0.960
		β_3	0.10628	0.10476	0.02225	0.02382	0.00003	0.864	0.914	0.970
	200	β_1	0.30951	0.30939	0.05781	0.06059	0.00454	0.884	0.940	0.978
		β_2	0.60836	0.60294	0.06303	0.07822	0.00003	0.802	0.912	0.974
		β_3	0.10175	0.10006	0.01491	0.01739	0.00001	0.854	0.910	0.958
	400	β_1	0.30166	0.29935	0.03979	0.04496	0.00013	0.874	0.928	0.978
		β_2	0.60265	0.59819	0.04342	0.05902	0.00092	0.766	0.846	0.940
		β_3	0.10010	0.09977	0.01017	0.01121	0.17298	0.870	0.934	0.980
45 %	100	β_1	0.31524	0.31323	0.10690	0.11154	0.00026	0.882	0.950	0.982
		β_2	0.63217	0.60955	0.11788	0.12870	0.00000	0.868	0.932	0.964
		β_3	0.10680	0.10288	0.02793	0.03007	0.00000	0.852	0.922	0.972
	200	β_1	0.30726	0.29930	0.07316	0.07522	0.00038	0.872	0.932	0.982
		β_2	0.62303	0.62235	0.08059	0.09249	0.01260	0.826	0.896	0.966
		β_3	0.10453	0.10253	0.01888	0.02072	0.00094	0.844	0.898	0.972
	400	β_1	0.30600	0.30154	0.05034	0.05477	0.08605	0.866	0.926	0.984
		β_2	0.61245	0.60473	0.05492	0.06542	0.00199	0.820	0.894	0.964
		β_3	0.10237	0.10105	0.01290	0.01421	0.00007	0.862	0.908	0.966

Table 3 continued

Censoring	Sample size	Parameter	Average	Median	SE	SEe	p value	CP for 0.90	CP for 0.95	CP for 0.99
70 %	100	β_1	0.34020	0.31814	0.16028	0.18905	0.00000	0.864	0.918	0.956
		β_2	0.67274	0.64338	0.17912	0.22422	0.00000	0.848	0.896	0.952
		β_3	0.11270	0.10424	0.04217	0.05451	0.00000	0.858	0.908	0.948
	200	β_1	0.31892	0.31371	0.10125	0.10926	0.00001	0.864	0.928	0.972
		β_2	0.62598	0.60825	0.11109	0.12647	0.00000	0.846	0.908	0.962
		β_3	0.10519	0.10257	0.02609	0.02842	0.00000	0.868	0.918	0.972
	400	β_1	0.30947	0.30308	0.06885	0.07721	0.00005	0.884	0.928	0.976
		β_2	0.62163	0.61481	0.07544	0.08963	0.00398	0.818	0.896	0.960
		β_3	0.10292	0.10166	0.01757	0.01942	0.00317	0.858	0.916	0.972

Table 4 *p* Values for the Henze–Zirkler test

Censoring	Sample size	*p* values Design 1	*p* values Design 2
15 %	100	0.000025	0.000016
	200	0.000008	0.000011
	400	0.183852	0.005988
45 %	100	0.000000	0.000000
	200	0.010884	0.008949
	400	0.125454	0.030476
70 %	100	0.000000	0.000000
	200	0.000000	0.000000
	400	0.118224	0.348330

and each censoring percentage. Moreover, Fig. 3 shows the estimated density plots of the standardized parameter estimates (by marginals) to check the normality of the estimator of $\boldsymbol{\beta}$, and Fig. 4 shows the multivariate qq-plots, for sample size 400 for First and Second designs and different censoring.

From Tables 2 and 3 it can be seen that the average of the estimates of the parameters is really close to the true value of the parameters for all configurations in both designs. Furthermore, the estimated standard errors are very near to the empirical ones. The empirical coverage probabilities appear to be quite close to the nominal levels. Although the p values of the Shapiro–Wilks normality test increase with n for most censoring, some p values are too small for not rejecting the normality of some estimations, even for $n = 400$. This can be caused by the optimization algorithm used that it favors one estimator over the others. However, the p values of the Henze–Zirkler multivariate normality test (Table 4) accept the multivariate normality for $n = 400$. Moreover, sometimes the results of the formal tests contradict with the impressions of the graphical analysis. In particular for large sample sizes the blind trust on the formal tests for normality can lead to erroneous results. Figure 3 shows that the empirical density plots (by marginals) improve when n increases and these plots are near to a normal density for moderate censoring, even for small sample size and Fig. 4 supports the approach to a multivariate normality distribution based on the qq-plots for $n = 400$.

From Table 5, we can see that the estimation of the parameters is bad, however, it is expected because for the Third design α is near to zero, so the model approaches to a non-parametric model. Therefore, we should prevent to use this model for α less than 0.5 since the estimations get worse cause the model approaches to non-parametric form.

Appendix A: The efficient score function and the least favorable direction

To find the efficient score function for the model (2), we follow the steps in the *Cox model Example* of Murphy and van der Vaart (2000).

Table 5 Average and median of the parameter estimates, standard error (SE), empirical standard error (SEe), *p* value for Shapiro–Wilks test and coverage probabilities (CP) for a nominal 90, 95 and 99 %, respectively, for the third design

Censoring	Sample size	Parameter	Average	Median	SE	SEe	*p* value	CP for 0.90	CP for 0.95	CP for 0.99
15 %	100	β_1	2.50437	1.63975	1.22515	2.26258	0.00000	0.666	0.716	0.786
		β_2	−1.37003	−0.81227	1.98859	2.92753	0.00000	0.762	0.810	0.866
	200	β_1	2.35903	1.67197	0.79797	1.79754	0.00000	0.438	0.608	0.766
		β_2	−1.16351	−0.80366	1.28217	1.77248	0.00000	0.796	0.838	0.892
	400	β_1	2.32239	1.88265	0.56460	1.49761	0.00000	0.294	0.376	0.604
		β_2	−1.10131	−0.82287	0.91745	1.22820	0.00000	0.790	0.840	0.910

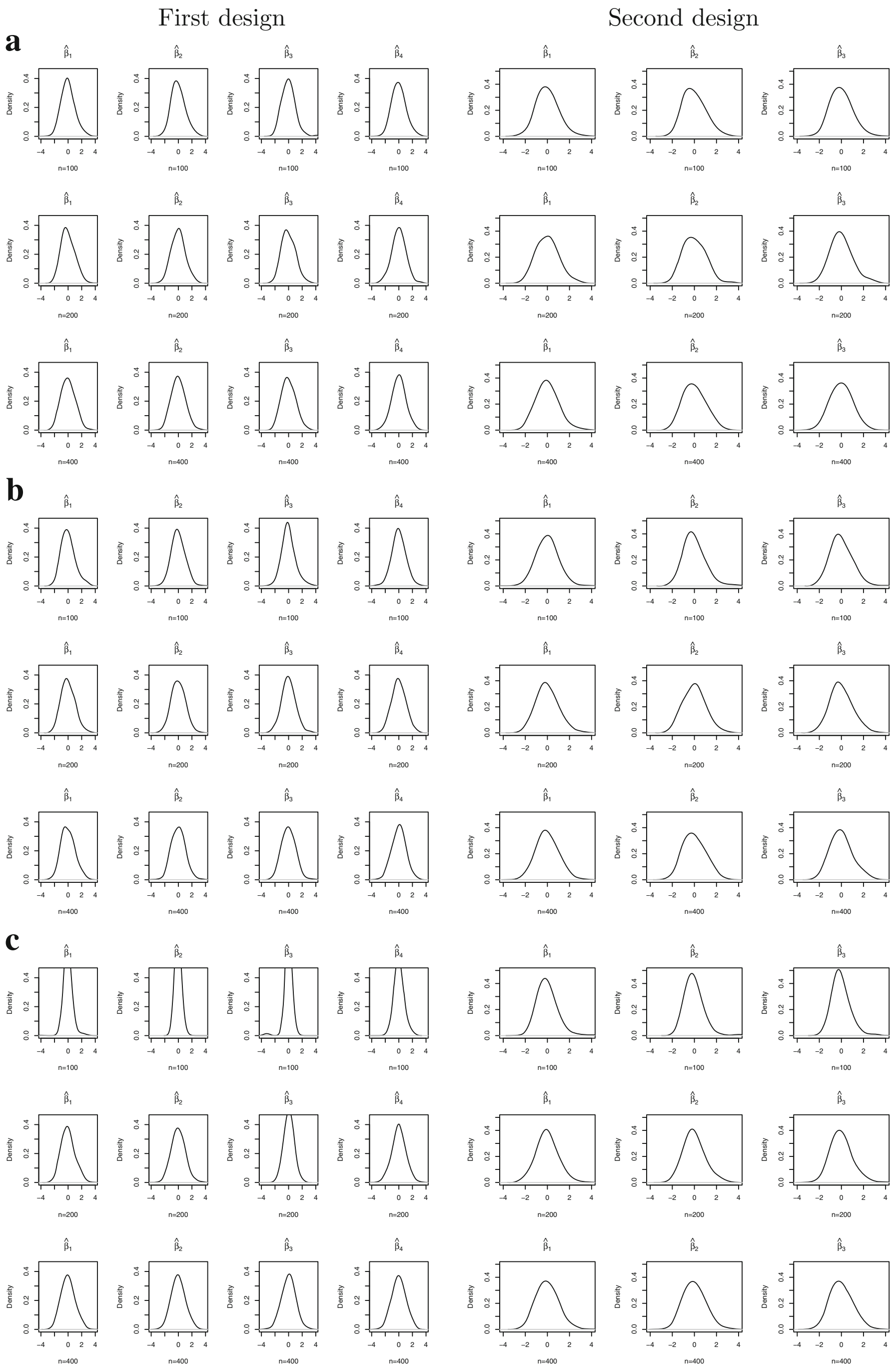

Fig. 3 Empirical density plots for the standardized parameter estimates for sample sizes $n = 100$, 200 and 400 for the First and Second design with **a** 15 % of censoring, **b** 45 % of censoring and **c** 70 % of censoring

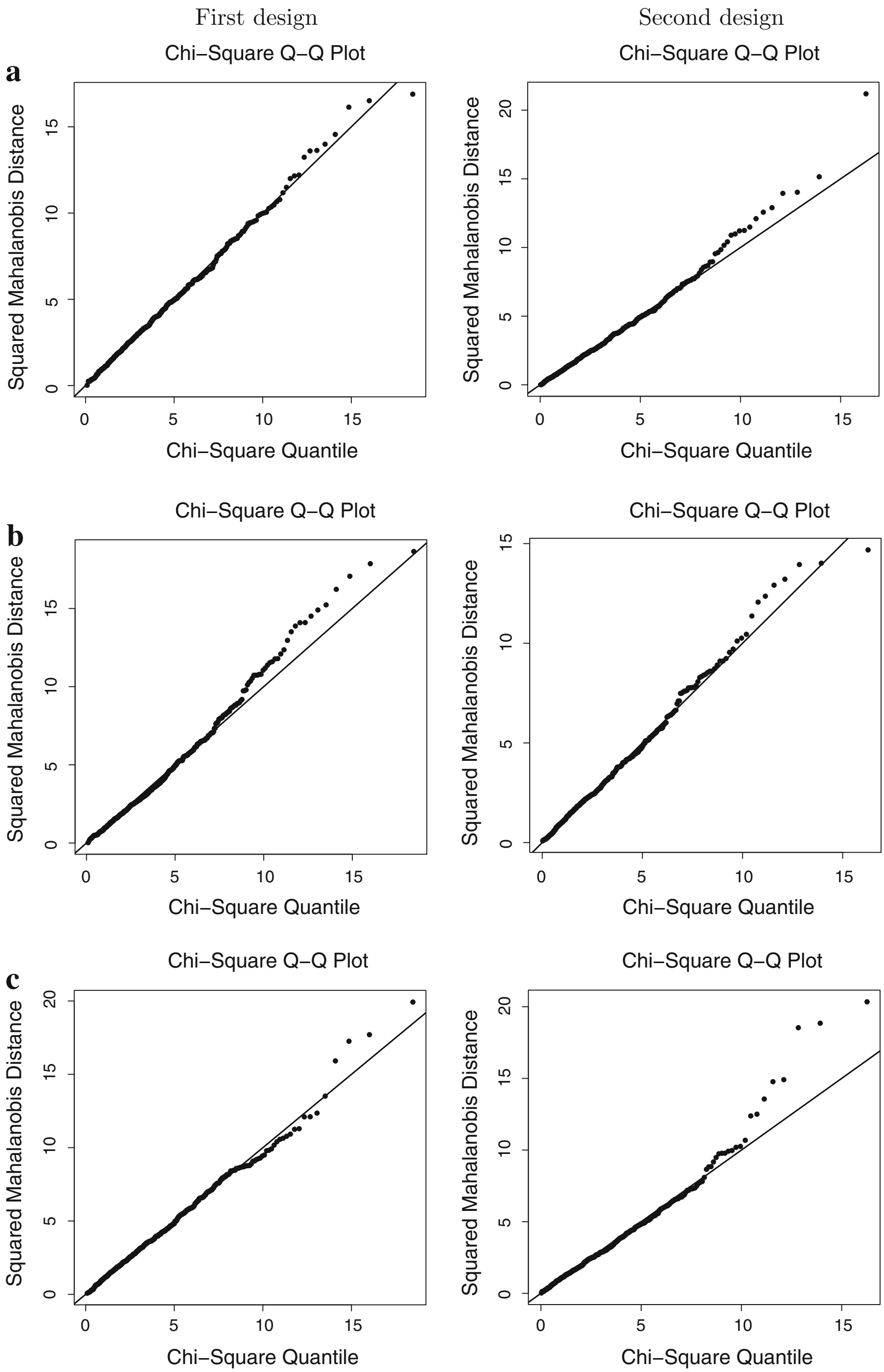

Fig. 4 Multivariate qq-plots for sample sizes $n = 400$ for the First and Second design with **a** 15 % of censoring, **b** 45 % of censoring and **c** 70 % of censoring

The efficient score function for $\boldsymbol{\theta}$ at $(\boldsymbol{\theta}, \Lambda_0^0)$ is defined as

$$\tilde{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0}(\mathbf{y}) = \dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0}(\mathbf{y}) - \Pi_{\boldsymbol{\theta}_0,\Lambda_0^0}\dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0}(\mathbf{y}) \tag{15}$$

where $\dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0}(\mathbf{y})$ is the score function for $\boldsymbol{\theta}$ at $(\boldsymbol{\theta}, \Lambda_0^0)$ and $\Pi_{\boldsymbol{\theta}_0,\Lambda_0^0}$ minimizes the squared distance $P_0(\dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0} - g)^2$ over all functions g in the closed linear span of the score functions for Λ_0.

In this case, as in the Cox model, we have

$$\Pi_{\boldsymbol{\theta}_0,\Lambda_0^0}\dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0} = A_{\boldsymbol{\theta}_0,\Lambda_0^0}h_0 \tag{16}$$

$$h_0 = (A^*_{\boldsymbol{\theta}_0,\Lambda_0^0}A_{\boldsymbol{\theta}_0,\Lambda_0^0})^{-1}A^*_{\boldsymbol{\theta}_0,\Lambda_0^0}\dot{\ell}_{\boldsymbol{\theta}_0,\Lambda_0^0} \tag{17}$$

where $A_{\boldsymbol{\theta}_0,\Lambda_0^0}h$ is the score function for Λ_0 at $(\boldsymbol{\theta}_0, \Lambda_0^0)$, $A^*_{\boldsymbol{\theta}_0,\Lambda_0^0}$ is the adjoint operator and h_0 is the least favorable direction.

Then, the score function for $\boldsymbol{\theta}$ of the model (2) is given by

$$\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \left(\left[\frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})\right]', \frac{\partial}{\partial\alpha}\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})\right)'.$$

From (5) we have

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta K_1(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x)K_1(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z}),$$
$$\frac{\partial}{\partial\alpha}\ell(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta K_2(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x)K_2(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z}),$$

with

$$K_1(\boldsymbol{\theta}, \mathbf{z}) = \frac{\alpha\mathbf{z}}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})}$$

and

$$K_2(\boldsymbol{\theta}, \mathbf{z}) = \ln[K(\boldsymbol{\theta}, \mathbf{z})].$$

Note that, the score function for $\boldsymbol{\theta}$ can be expressed as

$$\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y}) = \delta\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - \Lambda_0(x)\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z}), \tag{18}$$

with

$$\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) = (K_1(\boldsymbol{\theta}, \mathbf{z})', K_2(\boldsymbol{\theta}, \mathbf{z}))'.$$

To obtain the score function for Λ_0, we define the path

$$\Lambda_{\mathbf{t}}(x) = \int_0^x (1 + \mathbf{t}' h(s)) \mathrm{d}\Lambda_0(s) \tag{19}$$

for $\mathbf{t} \in \mathbb{R}^{d+1}$, $\Lambda_0(\cdot)$ given and a bounded function $h : [0, \tau] \to \mathbb{R}^{d+1}$. The path $\Lambda_{\mathbf{t}}(\cdot) \in \hat{H}$ if $\mathbf{t} \approx \mathbf{0}$. Replacing the path (19) in the log-likelihood (5) we obtain

$$\begin{aligned} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) = \delta[\log((1 + \mathbf{t}' h(x))\lambda_0(x)) + \log K(\boldsymbol{\theta}, \mathbf{z})] \\ -K(\boldsymbol{\theta}, \mathbf{z}) \int_0^x (1 + \mathbf{t}' h(s)) \mathrm{d}\Lambda_0(s) + c, \end{aligned}$$

thus, its derivative with respect to $\mathbf{t}$ is given by

$$\frac{\partial}{\partial \mathbf{t}} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) = \delta \frac{h(x)}{(1 + \mathbf{t}' h(x))} - K(\boldsymbol{\theta}, \mathbf{z}) \int_0^x h(s) \mathrm{d}\Lambda_0(s),$$

and taking $\mathbf{t} = \mathbf{0}$, we can get

$$\begin{aligned} \frac{\partial}{\partial \mathbf{t}} \ell(\boldsymbol{\theta}, \Lambda_{\mathbf{t}})(\mathbf{y}) \bigg|_{\mathbf{t}=\mathbf{0}} &= \delta h(x) - K(\boldsymbol{\theta}, \mathbf{z}) \int_0^x h(s) \mathrm{d}\Lambda_0(s) \\ &= A_{\boldsymbol{\theta}, \Lambda_0} h(\mathbf{y}). \end{aligned} \tag{20}$$

Then, if we consider that h_0 is the least favorable direction and by expressions (15), (16), (18) and (20) the efficient score function for $\boldsymbol{\theta}$ is given by

$$\tilde{\ell}_{\boldsymbol{\theta}, \Lambda_0}(\mathbf{y}) = \delta(\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - h_0(x)) - K(\boldsymbol{\theta}, \mathbf{z}) \int_0^x (\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) - h_0(s)) \mathrm{d}\Lambda_0(s).$$

In addition, to find the expression of the least favorable direction h_0 we use (17), so it is necessary to obtain the expressions of $A^*_{\boldsymbol{\theta}, \Lambda_0} \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(\mathbf{y})$ and $A^*_{\boldsymbol{\theta}, \Lambda_0} A_{\boldsymbol{\theta}, \Lambda_0}(x)$ where $A^*_{\boldsymbol{\theta}, \Lambda_0}$ is the adjoint operator characterized by

$$P_{\boldsymbol{\theta}, \Lambda_0}[(A_{\boldsymbol{\theta}, \Lambda_0} h) g] = \Lambda_0[h(A^*_{\boldsymbol{\theta}, \Lambda_0} g)].$$

Using the last expression we can get

$$P_{\boldsymbol{\theta}, \Lambda_0}[(A_{\boldsymbol{\theta}, \Lambda_0} g)(A_{\boldsymbol{\theta}, \Lambda_0} h)] = \Lambda_0[g(A^*_{\boldsymbol{\theta}, \Lambda_0} A_{\boldsymbol{\theta}, \Lambda_0} h)] \tag{21}$$

and

$$\Lambda_0[(A^*_{\boldsymbol{\theta}, \Lambda_0} \dot{\ell}(\boldsymbol{\theta}, \Lambda_0)) h] = P_{\boldsymbol{\theta}, \Lambda_0}[\dot{\ell}(\boldsymbol{\theta}, \Lambda_0)(A_{\boldsymbol{\theta}, \Lambda_0} h)]. \tag{22}$$

First, we calculate $(A_{\boldsymbol{\theta},\Lambda_0}g)(A_{\boldsymbol{\theta},\Lambda_0}h)$ that it is given by

$$(A_{\boldsymbol{\theta},\Lambda_0}g(\mathbf{y}))'(A_{\boldsymbol{\theta},\Lambda_0}h(\mathbf{y})) = \delta g(x)'h(x) - \delta K(\boldsymbol{\theta},\mathbf{z})\left[g(x)'\int_0^x h(s)\mathrm{d}\Lambda_0(s) + h(x)'\int_0^x g(s)\mathrm{d}\Lambda_0(s)\right] + (K(\boldsymbol{\theta},\mathbf{z}))^2\left(\int_0^x g(s)\mathrm{d}\Lambda_0(s)\right)'\int_0^x h(s)\mathrm{d}\Lambda_0(s).$$

We define

$$a(x) = g(x)'\int_0^x h(s)\mathrm{d}\Lambda_0(s) + h(x)'\int_0^x g(s)\mathrm{d}\Lambda_0(s)$$

and integrating partially we have

$$(A_{\boldsymbol{\theta},\Lambda_0}g(\mathbf{y}))'(A_{\boldsymbol{\theta},\Lambda_0}h(\mathbf{y})) = \delta g(x)'h(x) - \delta K(\boldsymbol{\theta},\mathbf{z})a(x) + (K(\boldsymbol{\theta},\mathbf{z}))^2\int_0^x a(s)\mathrm{d}\Lambda_0(s).$$

Furthermore, we can obtain

$$P_{\boldsymbol{\theta},\Lambda_0}[-\delta K(\boldsymbol{\theta},\mathbf{z})a(x)] = -\int_0^\tau a(x)E[(K(\boldsymbol{\theta},\mathbf{z}))^2 1_{\{X\geq x\}}]\mathrm{d}\Lambda_0(x).$$

Therefore, we have

$$\begin{aligned} P_{\boldsymbol{\theta},\Lambda_0}&\left[\delta g(x)'h(x) - \delta K(\boldsymbol{\theta},\mathbf{z})a(x) + (K(\boldsymbol{\theta},\mathbf{z}))^2\int_0^x a(s)\mathrm{d}\Lambda_0(s)\right] \\ &= P_{\boldsymbol{\theta},\Lambda_0}[\delta g(x)'h(x)] \\ &= \int_0^\tau g(x)'h(x)\left[\int_{\mathbb{R}^d} K(\boldsymbol{\theta},\mathbf{z})P[T\wedge C > x|\mathbf{z}]p_{\mathbf{Z}}(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathrm{d}\Lambda_0(x) \\ &= \int_0^\tau g(x)'E[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]h(x)\mathrm{d}\Lambda_0(x) \\ &= \Lambda_0[g'E[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]h]. \end{aligned}$$

Finally, by Eq. (21)

$$A^*_{\boldsymbol{\theta},\Lambda_0}A_{\boldsymbol{\theta},\Lambda_0}h(\mathbf{y}) = E[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]h(\mathbf{y}).$$

Analogously, taking the product $\dot{\ell}(\boldsymbol{\theta},\Lambda_0)A_{\boldsymbol{\theta},\Lambda_0}h$, integrating and using (22) we have

$$A^*_{\boldsymbol{\theta},\Lambda_0}\dot{\ell}(\boldsymbol{\theta},\Lambda_0)(\mathbf{y}) = E[\tilde{K}(\boldsymbol{\theta},\mathbf{z})K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}].$$

Then, we can define the least favorable direction at the true parameters $(\boldsymbol{\theta}_0, \Lambda_0^0)$ as

$$h_0(x) = \frac{E_0[\tilde{K}(\boldsymbol{\theta}, \mathbf{z})K(\boldsymbol{\theta}, \mathbf{z})1_{\{X\geq x\}}]}{E_0[K(\boldsymbol{\theta}, \mathbf{z})1_{\{X\geq x\}}]}.$$

Appendix B: Technical details for the proof of Theorem 1

We follow the procedure of Murphy (1994) to establish the consistency of the estimators of the generalized proportional hazards model. First, we prove the following lemmas.

Lemma 3 *The estimator* (13) *is consistent at* $\boldsymbol{\theta}_0$*, this is,*

$$\sup_{x\in[0,\tau]} |\hat{\Lambda}_{\boldsymbol{\theta}_0}(x) - \Lambda_0^0(x)| \overset{\text{a.s.}}{\to} 0,$$

when $n \to \infty$.

Proof We can obtain that

$$\sup_{x\in[0,\tau]} |\hat{\Lambda}_{\boldsymbol{\theta}_0}(x) - \Lambda_0^0(x)| \leq \sup_{x\in[0,\tau]} \left| \frac{1}{\mathbb{P}_n[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]} - \frac{1}{P_0[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]} \right| + \sup_{x\in[0,\tau]} \left| (\mathbb{P}_n - P_0)\left[\frac{N(x)}{P_0[1_{\{X\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]}\right]\right|.$$

On the other hand, the classes

$$\{1_{\{X\geq x\}}K(\boldsymbol{\theta}, \mathbf{z}) : x \in [0, \tau], \boldsymbol{\theta} \in \Theta\}$$

and

$$\left\{ \frac{N(x)}{P_0[1_{\{X\geq x\}}K(\boldsymbol{\theta}, \mathbf{z})]} : x \in [0, \tau], \boldsymbol{\theta} \in \Theta \right\} \tag{23}$$

are Donsker, and as consequence they are Glivenko–Cantelli classes. Thus, we have

$$\sup_{x\in[0,\tau]} |(\mathbb{P}_n - P_0)[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]| \overset{\text{a.s.}}{\to} 0,$$

additionally, as $P_0[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})] > 0$ for $x \in [0, \tau]$

$$\sup_{x\in[0,\tau]} \left| \frac{1}{\mathbb{P}_n[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]} - \frac{1}{P_0[1_{\{X_i\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]} \right| \overset{\text{a.s.}}{\to} 0.$$

Moreover, as the class given in (23) is Glivenko–Cantelli we obtain

$$\sup_{x\in[0,\tau]} \left| (\mathbb{P}_n - P_0)\left[\frac{N(x)}{P_0[1_{\{X\geq x\}}K(\boldsymbol{\theta}_0, \mathbf{z})]}\right]\right| \overset{\text{a.s.}}{\to} 0.$$

Then, the stated lemma follows. □

Lemma 4 $\limsup \hat{\Lambda}_{\boldsymbol{\theta}}(\tau) < \infty$ *almost surely.*

Proof Under assumption (iv), there exist $0 < c < \infty$ such as

$$||\mathbf{z}|| < c.$$

We define the constant c_2 such as

$$c_2 = \min_{||\mathbf{z}||<c} K(\boldsymbol{\theta}, \mathbf{z})$$

for $\boldsymbol{\theta}$ given. Thus, we obtain

$$\mathbb{P}_n[1_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] \geq c_2 \mathbb{P}_n[1_{\{X \geq x\}}],$$

and by the law of large numbers, almost surely we have

$$\mathbb{P}_n[1_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] \geq c_2 P_0[1_{\{X \geq x\}}] + o_{P_0}(1).$$

We assume that $P_0[1_{\{X \geq x\}}] > 0$ for $x \in [0, \tau]$, thus,

$$\mathbb{P}_n[1_{\{X \geq x\}} K(\boldsymbol{\theta}, \mathbf{z})] > 0$$

almost surely when $n \to \infty$. This is, the jumps of $\hat{\Lambda}_{\boldsymbol{\theta}}$ in τ are bounded by $1/c_2$, thus,

$$0 \leq \hat{\Lambda}_{\boldsymbol{\theta}}(\tau) \leq O(1)\mathbb{P}_n N(\tau)/c_2$$

almost surely when $n \to \infty$, with $N(x) = 1_{\{X \leq x\}}\delta$. □

Now, we can prove the following consistency theorem, where $\tilde{\boldsymbol{\theta}}_n$ is the maximum profile log-likelihood estimator of model (2).

Theorem 5 *The estimators of the generalized proportional hazards model* (2) *considering n right-censored data are consistent, this is*

$$\sup_{x \in [0,\tau]} |\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(x) - \Lambda_0^0(x)| \text{ and } ||\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0||$$

converge to 0 almost surely when $n \to \infty$.

Proof By Lemma 4 and Helly's selection Theorem, we have that along of a sequence

$$\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(x) \to \Lambda^*(x) \quad \text{for } x \in [0, \tau],$$

where $\Lambda^*(\cdot)$ is a continuous non-decreasing function and $\tilde{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^*$ with $\boldsymbol{\theta}^* \in \Theta$.

On the other hand, because $\mathbb{P}_n[1_{\{X\geq x\}}K(\boldsymbol{\theta},\mathbf{z})] > \epsilon > 0$, we have that $\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(\cdot)$ is absolutely continuous with respect to $\hat{\Lambda}_{\boldsymbol{\theta}_0}(\cdot)$ and $\frac{d\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(x)}{d\hat{\Lambda}_{\boldsymbol{\theta}_0}(x)}$ converges to a bounded measurable function $\zeta(x)$, this is

$$\Lambda^*(x) = \int_0^x \zeta(s)\mathrm{d}\Lambda_0^0(s).$$

Thus, $\Lambda^*(\cdot)$ is absolutely continuous with respect to $\Lambda_0^0(\cdot)$ and its derivative is denoted by $\lambda^*(\cdot)$, moreover $\zeta(x) = \frac{\lambda^*(x)}{\lambda_0^0(x)}$.

In addition, as $(\tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})$ maximize $\ell_n(\boldsymbol{\theta}, \Lambda_0)$ we have

$$0 \leq \frac{1}{n}\sum_{i=1}^{n}[\ell(\tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})(\mathbf{y}) - \ell(\boldsymbol{\theta}_0, \hat{\Lambda}_{\boldsymbol{\theta}_0})(\mathbf{y})],$$

if $n \to \infty$,

$$0 \leq P_0[\ell(\boldsymbol{\theta}^*, \Lambda^*)(\mathbf{y}) - \ell(\boldsymbol{\theta}_0, \Lambda_0^0)(\mathbf{y})].$$

By the Glivenko–Cantelli Theorem and as $\frac{\mathrm{d}\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(x)}{\mathrm{d}\hat{\Lambda}_{\boldsymbol{\theta}_0}(x)}$ converges uniformly to $\frac{\lambda^*(x)}{\lambda_0^0(x)}$, we can conclude that the Kullback–Leibler information between the density given by the parameters $(\boldsymbol{\theta}^*,\Lambda^*)$ and the density given by the parameters $(\boldsymbol{\theta}_0,\Lambda_0^0)$ is negative, thus two densities are equal almost surely. This implies that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\Lambda^*(\cdot) = \Lambda_0^0(\cdot)$ in $[0, \tau]$. Then, $\tilde{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ and $\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(x) \to \Lambda_0^0(x)$ almost surely with $x \in [0, \tau]$. □

Lemma 6 *Under the consistency of the estimators $\tilde{\boldsymbol{\theta}}_n$ and $\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}$ we have*

$$P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] = o_{P_0}(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|).$$

Proof From expression (11) we have

$$\begin{aligned}P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] =& P_0\Bigg[\int_0^\tau K(\boldsymbol{\theta}_0,\mathbf{z})1_{\{X\geq s\}}(h_0(s) - \tilde{K}(\boldsymbol{\theta}_0,\mathbf{z}))\mathrm{d}\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s)\\ &-\int_0^\tau \tilde{K}(\boldsymbol{\theta}_0,\mathbf{z})K(\boldsymbol{\theta}_0,\mathbf{z})1_{\{X\geq s\}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'h_0(s)\mathrm{d}\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s)\\ &+\frac{\delta(\tilde{K}(\boldsymbol{\theta}_0,\mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'h_0(x)} + \frac{\delta\tilde{K}(\boldsymbol{\theta}_0,\mathbf{z})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'h_0(x)}\Bigg].\end{aligned}$$

Moreover, we can obtain the following equalities

$$P_0[K(\boldsymbol{\theta}_0, \mathbf{z})1_{\{X \geq x\}}(h_0(x) - \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}))] = 0,$$

$$P_0\left[\frac{\delta(\tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) - h_0(x))}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\right] = 0$$

and

$$P_0\left[\frac{\delta \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\right] = P_0\left[\frac{\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\right].$$

Thus,

$$\begin{aligned} P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] = P_0\Bigg[& -\int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) 1_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) \mathrm{d}\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\ & + \frac{\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\Bigg]. \end{aligned} \tag{24}$$

As

$$M(s) = N(s) - \int_0^s K(\boldsymbol{\theta}_0, \mathbf{z}) 1_{\{X \geq s\}} (1 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' h_0(s)) \mathrm{d}\Lambda_0^0(s)$$

is a martingale with media 0, we have

$$\begin{aligned} P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \Lambda_0^0)] &= 0 \\ &= P_0\Bigg[-\int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) 1_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) \mathrm{d}\Lambda_0^0(s) \\ &\qquad + \delta \frac{h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\Bigg]. \end{aligned} \tag{25}$$

Finally, from expressions (24) and (25) we have

$$\begin{aligned} P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] &= P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n})] - P_0[\dot{\ell}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_n, \Lambda_0^0)] \\ &= P_0\Bigg[-\int_0^\tau \tilde{K}(\boldsymbol{\theta}_0, \mathbf{z}) K(\boldsymbol{\theta}_0, \mathbf{z}) 1_{\{X \geq s\}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(s) \mathrm{d}(\hat{\Lambda}_{\tilde{\boldsymbol{\theta}}_n}(s) \\ &\qquad - \Lambda_0^0(s)) + \frac{2\delta h_0(x)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}{1 + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' h_0(x)}\Bigg] \\ &= o_{P_0}(\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|). \end{aligned}$$

□

Lemma 7 *Let V be a neighborhood of $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0, \Lambda_0^0)$ where $\boldsymbol{\theta}_0$ and Λ_0^0 are the true values of the parameter of the generalized proportional hazards model (2) considering right-censored data and $\ell(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0)$ the log-likelihood function for the submodel as (9).*

(i) *The class of functions*

$$\{\dot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) : (\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) \in V\}$$

is Donsker with squared-integrable envelope function.

(ii) *The class of functions*

$$\{\ddot{\ell}(\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) : (\mathbf{t}, \boldsymbol{\theta}, \Lambda_0) \in V\}$$

is Glivenko–Cantelli and bounded in $L_1(P_0)$.

Proof Let $\mathcal{F}$ be the set of continuous distribution functions. For $\rho > 0$, we define

$$C_\rho = \{P \in \mathcal{F} : \|P - P_0\|_\infty \le \rho\},$$

where P_0 is the true distribution function.

On the other hand, as the vector $\mathbf{z}$ is bounded (assumption iv), the class

$$\{\boldsymbol{\beta}'\mathbf{z} : \boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}\}$$

is Donsker, with $\Theta_{\boldsymbol{\beta}}$ such as $\Theta = \Theta_{\boldsymbol{\beta}} \times \Theta_\alpha$. As the function $\exp(\boldsymbol{\beta}'\mathbf{z})$ is differentiable and its derivatives are bounded, the class

$$\{\exp(\boldsymbol{\beta}'\mathbf{z}) : \boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}\}$$

is Donsker. Moreover, as $1 + \exp(\boldsymbol{\beta}'\mathbf{z}) > 0$ we have that

$$\left\{K(\boldsymbol{\theta}, \mathbf{z}) = \left(\frac{1}{1 + \exp(-\boldsymbol{\beta}'\mathbf{z})}\right)^\alpha : \boldsymbol{\theta} \in \Theta\right\}, \tag{26}$$

and

$$\left\{K_1(\boldsymbol{\theta}, \mathbf{z}) = \frac{\alpha\mathbf{z}}{1 + \exp(\boldsymbol{\beta}'\mathbf{z})} : \boldsymbol{\theta} \in \Theta\right\}$$

are Donsker. In addition, as $f(x) = \ln(x)$ with domain in $[c, \infty)$ where $c > 0$ is Lipschitz we obtain that the class

$$\{K_2(\boldsymbol{\theta}, \mathbf{z}) = \ln[K(\boldsymbol{\theta}, \mathbf{z})] : \boldsymbol{\theta} \in \Theta\}$$

is Donsker. Thus,

$$\{\tilde{K}(\boldsymbol{\theta}, \mathbf{z}) : \boldsymbol{\theta} \in \Theta\}$$

is Donsker. As the class $\{1_{\{X\geq x\}} : x \in [0,\tau]\}$ is Donsker, we have that the classes

$$\{K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}} : x \in [0,\tau], \boldsymbol{\theta} \in \Theta\}$$

and

$$\{\tilde{K}(\boldsymbol{\theta},\mathbf{z})K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}} : x \in [0,\tau], \boldsymbol{\theta} \in \Theta\}$$

are Donsker. For $P \in C_\rho$, as the function $P \to E_P[f]$ is Lipschitz, the classes

$$\{E_P[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}] : x \in [0,\tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\}$$

and

$$\{E_P[\tilde{K}(\boldsymbol{\theta},\mathbf{z})K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}] : x \in [0,\tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\}$$

are Donsker. On the other hand, as $\mathbf{z}$ is bounded as $\boldsymbol{\theta} \in \Theta$, there exist m and M such as

$$0 < m < K(\boldsymbol{\theta},\mathbf{z}) < M < \infty. \tag{27}$$

As the function $P \to E_P[f]$ is continuous, there exist $\rho_1 > 0$ such as for each $P \in C_\rho$ we have

$$E_P[1_{\{X\geq x\}}] \geq \rho_1 > 0. \tag{28}$$

From (27) and (28) we obtain

$$0 < \rho_1 m \leq mE_P[1_{\{X\geq x\}}] \leq E_P[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}] \leq ME_P[1_{\{X\geq x\}}] < \infty$$

thus, the class

$$\left\{\frac{1}{E_P[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]} : x \in [0,\tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\right\}$$

is Donsker and in consequence the class

$$\left\{\tilde{K}(\boldsymbol{\theta},\mathbf{z}) - \frac{E_P[\tilde{K}(\boldsymbol{\theta},\mathbf{z})K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]}{E_P[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]} : x \in [0,\tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\right\} \tag{29}$$

is it. As the function $\Lambda_0 \to \int f \, d\Lambda_0$ is Lipschitz, the class

$$\left\{\int_{[0,x]}\left(\tilde{K}(\boldsymbol{\theta},\mathbf{z}) - \frac{E_P[\tilde{K}(\boldsymbol{\theta},\mathbf{z})K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]}{E_P[K(\boldsymbol{\theta},\mathbf{z})1_{\{X\geq x\}}]}\right)d\Lambda_0 : x \in [0,\tau], \boldsymbol{\theta}\in\Theta, P\in C_\rho\right\} \tag{30}$$

is Donsker. Finally, from (26), (29) and (30) we have that

$$\{\tilde{\ell}(\boldsymbol{\theta}, \Lambda_0) : x \in [0, \tau], \boldsymbol{\theta} \in \Theta, P \in C_\rho\}$$

is Donsker with squared-integrable envelope function, this proves (i). Analogously, (ii) could be proved. □

Acknowledgments This work was supported by research grant MTM2013-40778-R and MAEC-AECID. The authors are grateful to the referees for their valuable comments and proposals which have improved the contents of this paper.

References

Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics*, *39*, 109–117.

Balakrishnan, N. (1992). *Handbook of the logistic distribution*. New York: Marcel Dekker.

Bender, R., Augustin, T., Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards model. *Statistics in Medicine*, *24*(11), 1713–1723.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*, 141–151.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, *34*(2), 187–220.

Devarajan, K., Ebrahimi, N. (2011). A semi-parametric generalization of the cox proportional hazards regression model: inference and applications. *Computational Statistics and Data Analysis*, *65*(1), 667–676.

Etezadi-Amoli, J., Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: a spline approximation for the background hazard function. *Biometrics*, *43*, 181–192.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing heterogeneity. *Biometrika*, *71*, 75–83.

Kalbfleisch, J. D., Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). New York: Wiley.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. New York: Springer.

MacKenzie, G. (1996). Regression models for survival data: the generalized time-logistic family. *Statistician*, *45*, 21–34.

MacKenzie, G. (1997). On a non-proportional hazards regression model for repeated medical random counts. *Statistics in Medicine*, *16*, 1831–1843.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics*, *22*(2), 712–731.

Murphy, S. A., van der Vaart, W. (2000). On profile likelihood. *Journal of the American Statistical Association, Theory and Methods*, *95*(450), 449–465.

Sasieni, P. D. (1995). Efficiently weighted estimating equations with application to proportional excess hazards. *Lifetime data analysis*, *1*, 49–57.

Thomas, D. C. (1986). Use of auxiliary information in fitting non-proportional hazards model. In S. H. Moolgavkar R. L. Prentice (Eds.), *Modern statistical methods in chronic disease epidemiology* (pp. 197–210). New-York: Wiley.

Tibshirani, R., Ciampi, A. (1983). A family of additive and proportional hazard models for survival data. *Biometrics*, *39*(1), 141–147.

Younes, N., Lachin, J. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, *53*, 1199–1211.