Ann Inst Stat Math (2010) 62:465–485
DOI 10.1007/s10463-008-0187-z

Generalized time-dependent conditional linear models under left truncation and right censoring

Bianca Teodorescu · Ingrid Van Keilegom · Ricardo Cao

Received: 2 March 2007 / Revised: 24 January 2008 / Published online: 16 July 2008

Abstract Consider the model $\phi(S(z|X)) = \boldsymbol{\beta}(z)\vec{X}$, where ϕ is a known link function, $S(\cdot|X)$ is the survival function of a response Y given a covariate X, $\vec{X} = (1, X, X^2, \ldots, X^p)$ and $\boldsymbol{\beta}(z)$ is an unknown vector of time-dependent regression coefficients. The response is subject to left truncation and right censoring. Under this model, which reduces for special choices of ϕ to e.g. Cox proportional hazards model or the additive hazards model with time dependent coefficients, we study the estimation of the vector $\boldsymbol{\beta}(z)$. A least squares approach is proposed and the asymptotic properties of the proposed estimator are established. The estimator is also compared with a competing maximum likelihood based estimator by means of simulations. Finally, the method is applied to a larynx cancer data set.

Keywords Additive hazards model · Bootstrap · Least-squares estimator · Logistic model · Proportional hazards model · Semiparametric regression · Survival analysis · Time-dependent coefficients · U-statistics

B. Teodorescu (✉) · I. Van Keilegom
Institute of Statistics,
Université catholique de Louvain,
Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium
e-mail: teodorescu@stat.ucl.ac.be

I. Van Keilegom
e-mail: vankeilegom@stat.ucl.ac.be

R. Cao
Departamento de Matemáticas, Facultad de Informática,
Universidade da Coruña, Campus de Elviña,
15071 A Coruña, Spain
e-mail: rcao@udc.es

1 Introduction

In survival analysis interest often lies in the relationship between the survival function and a certain number of covariates. It usually happens that for some individuals we cannot observe the event of interest, due to the presence of right censoring and/or left truncation. A typical example is given by a retrospective medical study, in which one is interested in the time interval between birth and death due to a certain disease. Patients who die of the disease at early age will rarely have entered the study before death and are therefore left truncated. On the other hand, for patients who are alive at the end of the study, only a lower bound of the true survival time is known and these patients are hence right censored.

In the case of censored responses (in the absence of truncation), lots of models exist in the literature that describe the relationship between the survival function and the covariates (proportional hazards model or Cox model, log-logistic model, accelerated failure time model, etc.). These models are special cases of linear transformation models, that are of the following form (a): $\varphi(Y) = \boldsymbol{\beta}\vec{X} + \epsilon$, with Y the survival time, φ an unknown monotone increasing function, ϵ a random variable with known distribution function F and independent of the covariates $\vec{X}$, and $\boldsymbol{\beta}$ is a vector of constant coefficients or, equivalently, (b): $\varsigma\left(S_Y(z|\vec{X})\right) = \varphi(z) - \boldsymbol{\beta}\vec{X}$, where ς is a known increasing function, $\varsigma^{-1}(u) = 1 - F(u)$ and $S_Y(z|\vec{X}) = P(Y > z|\vec{X})$. We get the proportional hazards model and the proportional odds models by taking in (a) ϵ to follow the extreme-value distribution and the standard logistic distribution, respectively or, by taking in (b) $\varsigma(u) = \log(-\log(u))$ and $\varsigma(u) = \log(\frac{u}{1-u})$, respectively. For more details on transformation models see, for example, Doksum (1987), Cheng et al. (1995, 1997), Fine et al. (1998), Fine and Gray (1999) or Chen et al. (2002).

In these models, the regression coefficients are usually supposed to be constant over time. In practice, the structure of the data might however be more complex, and it might therefore be better to consider coefficients that can vary over time. In the previous example, certain covariates (e.g. sex, genetic indicators, smoking status, etc.) can have a relatively low impact on early age survival, but a higher influence at higher age. This motivated a number of authors to extend the Cox model to allow for time-dependent coefficients, see, for example Murphy and Sen (1991), Nan and Lin (2003), Cai and Sun (2003), among others. Also other time-dependent survival models have been considered, see, for example Lambert and Eilers (2004) and Kauermann (2005).

In this paper we go one step further. We consider a very general model, which includes as special cases the above mentioned models (Cox model, additive model, log-logistic model, linear transformation models, etc.) and study the estimation of the (time-dependent) regression coefficients by means of a least squares approach. The response is allowed to be subject to right censoring and/or left truncation.

More precisely, let Y denote the survival time, T the truncation time and C the censoring time. When data are left-truncated and right-censored we observe (T, Z, δ) only if $Z \geq T$, where $Z = \min\{Y, C\}$ and $\delta = 1_{\{Y \leq C\}}$. Let $(T_i, Z_i, \delta_i, X_i)$, $i = 1, \ldots, n$ be an iid sample from (T, Z, δ, X), where X is a (one-dimensional) covariate. A very common assumption that is made in this setup is that Y, T and C are independent given X. We are interested in the relationship between the survival function of Y, $S(z|X) = P(Y > z|X)$ and X. We suppose that this relationship is of polynomial type, via a known monotone transformation $\phi : [0, 1] \to I\!R$ of the survival function, i.e.,

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \cdots + \beta_p(z)X^p, \quad (1)$$

for some known p. No assumption is made on the form of the survival function $S(z|X)$, except for the usual smoothness assumptions. Particular choices of ϕ give well known models in survival analysis, but extended to time-dependent coefficients. The choice $\phi(u) = \log(\frac{u}{1-u})$ gives the logistic model, $\phi(u) = -\log(u)$ gives the additive risk model and $\phi(u) = \log(-\log(u))$ leads to a model close to the proportional hazards model.

In the absence of truncation, model (1) has been considered by Jung (1996), who proposed an estimator for the regression coefficients based on the maximum likelihood method, when the observations are censored and the covariate is discrete. His method is valid only in the case where the censoring is independent of the covariates. Using the same technique, Subramanian (2001) improved Jung's estimator by relaxing the hypothesis of independence between the censoring time and the covariates. Subramanian (2004) extended the estimator to the case of a one-dimensional continuous covariate.

All of these papers propose estimators that are based on a maximum likelihood approach, whereas the estimator we propose in this paper is based on a least squares principle. In comparison with the former, the latter approach has the advantage of being easier to compute, since it does not require any iterative computation. The method proposed in this paper is inspired by Cao and González-Manteiga (2008), who study a least squares procedure for the case where the coefficients are considered as being time-independent.

The paper is organized as follows. In the next section we introduce the proposed estimator and its asymptotic properties. In Sect. 3 we present a bootstrap based method for the selection of the smoothing parameter, while in Sect. 4 we give some numerical results. The analysis of larynx cancer data is conducted in Sect. 5. Finally, Section 6 contains the proofs.

2 Least squares estimator and its asymptotic properties

We need to introduce the following notations: $M(x) = P(X \le x)$, $F(y|x) = P(Y \le y|x)$, $G(y|x) = P(C \le y|x)$, $L(y|x) = P(T \le y|x)$, $H(y|x) = P(Z \le y|x)$, $H_1(y|x) = P(Z \le y, \delta = 1|x)$, $L(y) = P(T \le y)$, $H(y) = P(Z \le y)$, $H_1(y) = P(Z \le y, \delta = 1)$, $C(y|x) = P(T \le y \le Z|x, T \le Z)$, and $\alpha(x) = P(T \le Z|X = x)$, which is the probability of absence of truncation conditionally on $X = x$. For any distribution function $W(t) = P(\eta \le t)$, we denote the left and right support endpoints by $a_W = \inf\{t|W(t) > 0\}$ and $b_W = \sup\{t|W(t) < 1\}$, respectively. We define $W^*(t) = P(\eta \le t|T \le Z)$. Finally, let m denote the density of X and m^* the density of X conditionally on $T \le Z$.

The estimator of $\boldsymbol{\beta}(z) = (\beta_0(z), \ldots, \beta_p(z))^t$ we propose, is based on a least squares estimation procedure. More precisely, for a fixed value of z, we estimate $\boldsymbol{\beta}(z)$ by fitting a pth degree polynomial through the points $((1, X_i, \ldots, X_i^p), \phi(\hat{S}_n(z|X_i)))$ $(i = 1, \ldots, n)$, for some estimator $\hat{S}_n(z|X_i)$. We estimate the survival function $S(z|X_i)$ in a completely nonparametric way, by means of the estimator of the conditional distribution, proposed by Iglesias-Pérez and González-Manteiga (1999):

$$\hat{S}_n(z|x) = 1 - \hat{F}_n(z|x) = \prod_{i=1}^{n} \left(1 - \frac{1_{\{Z_i \le z, \delta_i = 1\}} B_{ni}(x)}{C_n(Z_i|x)}\right), \tag{2}$$

where

$$B_{ni}(x) = \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)}$$

are Nadaraya–Watson weights, K is a known probability density function (kernel), $h = h_n \to 0$ a bandwidth sequence, and $C_n(u|x) = \sum_{j=1}^{n} 1_{\{T_j \le u \le Z_j\}} B_{nj}(x)$.

Note that this estimator reduces to the estimator of Beran (1981) in the absence of truncation, to the one of Tsai et al. (1987) in the absence of covariates and to the classical Kaplan and Meier (1958) estimator when there is no truncation and there are no covariates.

Next, using the estimated responses $\phi(\hat{S}_n(z|X_i))$ $(i = 1, \ldots, n)$, apply the classical weighted least squares method to compute the estimators of $\beta_j(z)$ $(j = 0, \ldots, p)$:

$$\hat{\boldsymbol{\beta}}(z) = \begin{pmatrix} \hat{\beta}_0(z) \\ \hat{\beta}_1(z) \\ \vdots \\ \hat{\beta}_p(z) \end{pmatrix} = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \hat{\boldsymbol{\phi}}(z), \tag{3}$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 & \cdots & X_1^p \\ 1 & X_2 & \cdots & X_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \cdots & X_n^p \end{pmatrix}, \quad \hat{\boldsymbol{\phi}}(z) = \begin{pmatrix} \phi(\hat{S}_n(z|X_1)) \\ \phi(\hat{S}_n(z|X_2)) \\ \vdots \\ \phi(\hat{S}_n(z|X_n)) \end{pmatrix}$$

and $\boldsymbol{W} = \mathrm{diag}(w(X_1), \ldots, w(X_n))$, where $w(\cdot)$ is a trimmed function defined in terms of a proper weight function $\tilde{w}$, as precised in (H11).

The above procedure can be repeated for all possible z. In practice only the uncensored data need to be considered, since the estimator of the survival function, and hence the estimator of $\boldsymbol{\beta}(z)$, only changes at these points.

Note that the above procedure can be adapted in a straightforward way to the case where the covariate is discrete (or categorical). In fact, it suffices to estimate the survival function without using any smoothing in that case. We will not consider this case any further, as the results for continuous covariates can be reduced in an obvious way to discrete covariates. Also, combinations of several discrete covariates and a (one-dimensional) continuous covariate can be considered. An example is given in Sect. 5, where we analyse data containing one continuous and three binary covariates.

In order to obtain the asymptotic properties of $\hat{\boldsymbol{\beta}}(z)$ some conditions, (H1)–(H12), have to be assumed. They are collected in Sect. 6.

Let $\boldsymbol{\phi}(z) = (\phi(S(z|X_1)), \ldots, \phi(S(z|X_n)))^t$.

Model (1) implies that $\boldsymbol{\phi}(z) = \boldsymbol{X}\boldsymbol{\beta}(z)$, which leads to

$$\boldsymbol{\beta}(z) = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{\phi}(z), \tag{4}$$

and hence

$$\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} (\hat{\boldsymbol{\phi}}(z) - \boldsymbol{\phi}(z)).$$

The latter expression is the starting point for the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}(z)$, which is established in the next theorem.

Theorem 1 *Suppose that conditions (H1) through (H12) hold. Then, for any $a \leq z \leq b$,*

$$n^{1/2}(\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z)) \xrightarrow{d} N(0, \boldsymbol{A}^{-1} \boldsymbol{\Sigma}(z) (\boldsymbol{A}^{-1})^t),$$

where $\boldsymbol{\Sigma}(z) = (\sigma_{ij}(z))_{i,j=0}^p$, with

$$\sigma_{ij}(z) = \int_I x^{i+j} \tilde{w}^2(x) S^2(z|x) \phi'(S(z|x))^2 \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)} m^*(x) dx, \tag{5}$$

and $\boldsymbol{A} = (a_{ij})_{i,j=0}^p$, with $a_{ij} = E(X^{i+j} w(X))$.

Corollary 2 *Suppose that conditions (H1) through (H12) hold. Then*

$$\sup_{z \in [a,b]} |\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z)| = O_p(n^{-1/2})$$

Remark 1 In a similar way we can obtain the asymptotic properties of the estimator of the coefficients $\beta_j(z)$ when we have only discrete covariates or a combination of discrete covariates and a one-dimensional continuous covariate. Note that in the case where we have only discrete covariates, no smoothing is required, since the estimator of the survival function $\hat{S}(z|x)$ is the Kaplan–Meier estimator extended to the case when we also have truncation (Tsai et al. 1987).

Remark 2 As an immediate consequence of this result we can obtain the asymptotic normality of the estimator $\tilde{S}(z|x) = \phi^{-1}(\hat{\beta}_0(z) + \hat{\beta}_1(z)x + \ldots + \hat{\beta}_p(z)x^p)$ of the conditional survival function under model (1). Note that this estimator is in general non-monotone. A convenient and satisfactory solution is to keep the estimator constant until it starts decreasing again.

Remark 3 It is important to have at hand a procedure to check the validity of the assumed model (1), thus to test $H_0 : \exists\boldsymbol{\beta}(z)$ such that (1) holds against H_A :(1) does not hold for any $\boldsymbol{\beta}(z)$. The main idea is to compare a semiparametric estimator of the response, $\phi(\hat{S}(z|\boldsymbol{X}))$, with its parametric counterpart, $\hat{\boldsymbol{\beta}}(z)\boldsymbol{X}$, where $\hat{S}$ is the estimator in (2) and $\hat{\boldsymbol{\beta}}(z)$ is the least-squares estimator proposed in (3). A large deviation between them indicates the lack of fit of the parametric form and thus the rejection of the null (this topic will be the object of future work).

Remark 4 The choice of the weight matrix $\boldsymbol{W}$ can be done in such a way to accommodate heterocedasticity in the model.

Remark 5 The $\hat{\beta}_j(z)$ are step functions, but can be smoothed in z using, e.g., the kernel method.

Remark 6 Note that the above methodology is not only applicable to polynomial models. Consider for example a general nonlinear model of the form:

$$\phi(S(z|x)) = v(\boldsymbol{\beta}(z), x) \text{or equivalently} S(z|x) = \mu(\boldsymbol{\beta}(z), x),$$

where $\mu = \phi^{-1} \circ v$ and where the function $v : \mathbb{R}^{p+1} \times \mathbb{R} \to \mathbb{R}$ is known. Then, $\boldsymbol{\beta}(z)$ can be estimated by

$$\hat{\boldsymbol{\beta}}(z) = \underset{\boldsymbol{\beta}(z)}{\operatorname{argmin}} \sum_{i=1}^{n} w(X_i) \left[\phi(\hat{S}_n(z|X_i)) - v(\boldsymbol{\beta}(z), X_i)\right]^2$$

When $v(\boldsymbol{\beta}(z), x) \equiv \beta_0(z) + \beta_1(z)x + \cdots + \beta_p(z)x^p$, this estimator coincides with (3). Theorem 2.1 can be extended in a similar way to this more general estimator.

3 Bandwidth selection

The estimator $\hat{\boldsymbol{\beta}}(z)$ defined in Sect. 2, is based on a kernel estimator of the conditional survival function $S(z|X)$. Therefore, a bandwidth parameter h needs to be selected. We propose a bootstrap procedure which selects for a fixed z, the bandwidth for which the estimated mean squared error (MSE) of $\hat{\boldsymbol{\beta}}(z)$ is minimal. It suffices to consider the uncensored observations, since the estimator $\hat{\boldsymbol{\beta}}(z)$ only changes at these points. The procedure is as follows:

1. For fixed z consider values for $h \in \{h_1, \ldots, h_r\}$, a fine grid of bandwidths in the interval $(0, \mu(\text{supp}(X)))$, where μ is the Lebesgue measure.
2. For each h_j $(j = 1, \ldots, r)$:
   (a) Choose a pilot bandwidth, g_j, (usually larger than h_j) to estimate $S(z|X_i)$, $G(z|X_i)$ and $L(z|X_i)$ by $\hat{S}_{g_j}(z|X_i)$, $\hat{G}_{g_j}(z|X_i)$ and $\hat{L}_{g_j}(z|X_i)$, respectively $(i = 1, \ldots, n)$, where $\hat{S}_{g_j}(z|X_i)$ is the estimator in (2),

$$\hat{G}_{g_j}(z|x) = 1 - \prod_{i=1}^{n}\left(1 - \frac{1_{\{Z_i \le z, \delta_i = 0\}} B_{ni}(x)}{C_n(Z_i|x)}\right),$$

$$\hat{L}_{g_j}(z|x) = \prod_{i=1}^{n}\left(1 - \frac{1_{\{T_i > z\}} B_{ni}(x)}{C_n(T_i|x)}\right),$$

and the subscript g_j indicates the bandwidth we are working with.

(b) Replace $S(z|X_i)$ by $\hat{S}_{g_j}(z|X_i)$ in (1) and estimate $\beta_0(z), \ldots, \beta_p(z)$ by the least squares estimator in (3) to obtain $\hat{\beta}_{0,g_j}(z), \ldots, \hat{\beta}_{p,g_j}(z)$. Plug these estimators into (1) and re-estimate $S(z|X_i)$ by

$$\tilde{S}_{g_j}(z|X_i) = \phi^{-1}(\hat{\beta}_{0,g_j}(z) + \hat{\beta}_{1,g_j}(z)X_i + \cdots + \hat{\beta}_{p,g_j}(z)X_i^p).$$

(c) For every $i = 1, \ldots, n$ draw random observations Y_i^*, C_i^* and T_i^* from $\tilde{S}_{g_j}(\cdot|X_i)$, $\hat{G}_{g_j}(\cdot|X_i)$ and $\hat{L}_{g_j}(\cdot|X_i)$, respectively. Compute $Z_i^* = \min\{Y_i^*, C_i^*\}, \delta_i^* = 1_{\{Y_i^* \le C_i^*\}}$ and simulate new values Y_i^*, C_i^* and T_i^* if $T_i^* > Z_i^*$.

(d) Use this resample $\{(T_1^*, Z_1^*, \delta_1^*, X_1), \ldots, (T_n^*, Z_n^*, \delta_n^*, X_n)\}$ to estimate a bootstrap version of the conditional survival function, $\hat{S}_{h_j}^*(z|X_i)$ $(i = 1, \ldots, n)$ using the bandwidth h_j. This bootstrap version is used to obtain the bootstrap coefficients $\hat{\beta}_{0,h_j}^*(z), \ldots, \hat{\beta}_{p,h_j}^*(z)$ using the least squares estimator.

(e) Repeat the steps (c)–(d) B times and compute the bootstrap estimator of the mean squared error (MSE):

$$MSE^*(h_j) = \sum_{k=0}^{p}\left\{\frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}_{k,h_j,b}^*(z) - \hat{\beta}_{k,g_j}(z))^2\right\}$$

3. Choose the value h_j which leads to the smallest $MSE^*(h_j)$.
4. Repeat steps 1–3 for all the values of z considered.

Remark 7 A similar bootstrap procedure can be used to estimate the variance of $\hat{\boldsymbol{\beta}}(z)$, or to approximate the distribution of $\hat{\boldsymbol{\beta}}(z)$. For small samples, this might lead to better approximations than the normal limit established in Theorem 1.

Remark 8 The asymptotic validity of a slight variation of the above bootstrap procedure has been established by Iglesias-Pérez and González-Manteiga (2003). In fact, they resampled from $\hat{S}_g(z|X_i)$, $\hat{G}_g(z|X_i)$ and $\hat{L}_g(z|X_i)$ for each X_i $(i = 1, \ldots, n)$ in order to obtain Y_j^*, C_j^* and T_j^* respectively. Bootstrapping from $\tilde{S}$ instead of $\hat{S}$ allows us to actually mimic the model.

4 Numerical results

In this section, we will first conduct some simulations in order to compare the proposed least squares method (LS) with the maximum likelihood method (ML) proposed by

Jung (1996) and Subramanian (2001, 2004). We will deal with the cases of discrete covariates and of a one-dimensional continuous covariate, both under censoring. Next, we will study the performance of the new method in the case of a one-dimensional continuous covariate when truncation is also present. Finally, some simulations will illustrate the effect of the bootstrap bandwidth selector, proposed in Sect. 3.

Along the simulations, the following model is considered:

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x. \tag{6}$$

In the discrete case, model 1 considers that X is uniformly distributed in {0.1, 0.3, 0.5, 0.7, 0.9}, $Y|_{X=x} \sim \text{Logistic}(0, \frac{\pi^2}{3(4x)^2})$ (i.e. $F(y|x) = 1/(1+\exp(4xy))$, $E(Y|x) = 0$ and $\text{Var}(Y|x) = \pi^2/\{3(4x)^2\}$), $\exp(C)|_{X=x} \sim U[0, dx]$, where $d > 0$ will be chosen according to the desired censoring probability, and $\phi(u) = \log(\frac{u}{1-u})$ (logistic model), which gives us the true model $\phi(S(z|x)) = -4zx$. A similar model has also been considered by Subramanian (2001). The sample size is taken $n = 200$, the number of Monte Carlo simulations is $M = 10{,}000$ and d is taken to be 0.008, 0.022, 0.05 and 0.1 in order to give 10, 20, 30 and 40% of censoring, respectively. For estimating the survival function we use the Kaplan–Meier estimator, since there is no truncation and no smoothing is required. For $z = 0.1$ the results are given in Table 1. We notice that the results are very similar for the two methods in the case of censoring and in the presence of discrete covariates. Other simulations not reported here lead to similar conclusions: the difference between the two procedures is only very minor, regarding both bias and variance.

Model 2 deals with the continuous case, $X \sim U[0, 1]$, $Y|_{X=x} \sim \text{Exp}(4x)$, $C|_{X=x} \sim \text{Exp}(dx)$ with $d > 0$ that gives different censoring probabilities, and $\phi(u) = \log(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = -4zx$. Note that $\phi(S(z|X=0)) = \beta_0(z)$ and $\phi(S(z|X=1)) = \beta_0(z) + \beta_1(z)$. The sample size is taken $n = 100$, $M = 10{,}000$ Monte Carlo simulations are conducted and d is taken to be 1 and 8/3 in order to give 20 and 40% of censoring, respectively. Since

Table 1 Comparison between the ML and LS methods for model 1, at point $z = 0.1$

Censoring percentage	Method	$\beta_0(z) = 0$ Bias	$\beta_0(z) = 0$ MSE	$\beta_1(z) = -0.4$ Bias	$\beta_1(z) = -0.4$ MSE
10	LS	−0.0035	0.0917	−0.0071	0.2742
	ML	0.0025	0.0910	0.0056	0.2810
20	LS	−0.0044	0.1002	−0.0056	0.3017
	ML	0.0013	0.0993	0.0081	0.3065
30	LS	−0.0028	0.1246	−0.0053	0.3475
	ML	0.0026	0.1217	0.0126	0.3492
40	LS	0.0056	0.1696	−0.0083	0.4207
	ML	0.0040	0.1630	0.0113	0.4239

we have a one-dimensional continuous covariate, a bandwidth, h, is needed in order to estimate $S(z|x)$. We worked with $h \in \{0.2, 0.25, 0.3, 0.35, 0.4\}$. The Nadaraya–Watson weights are calculated based on the Epanechnikov kernel $K(u) = 1_{\{-1\leq u\leq 1\}} \cdot 3(1-u^2)/4$ and the weight function $\tilde{w}(x) = 1_{\{0.025\leq x\leq 0.975\}}$ has been chosen in order to avoid the boundary problems.

Table 2 shows the results for $z = 0.3$. The table shows the bias and MSE of the estimators of $\beta_0(z)$ and $\beta_1(z)$, and also of the estimators of the regression function $\beta_0(z) + \beta_1(z)x$, evaluated at the endpoints of the support of X, namely at 0 and 1. The values are computed with the LS and the ML methods and they correspond to the minimum in h of the total MSE, i.e. $TMSE(h) = MSE_h(\hat{\beta}_0(z)) + MSE_h(\hat{\beta}_1(z))$. The optimal bandwidth is also displayed. The results in Table 2 show that the LS method behaves better than the ML procedure. In the simulation studies, we also noticed that the ML algorithm had some serious convergence problems. In the case of 20% censoring we noticed 393 divergences on a total of 10,000, while for the 40% censoring case 1,323 on a total of 10,000.

Table 3 gives the results for model 2 computed only with the LS method and with the bandwidth estimated by means of the bootstrap procedure described in Sect. 3. The bandwidth is selected from the grid $\{0.2, 0.25, 0.3, 0.35, 0.4\}$. $B = 100$ bootstrap replications are generated each time in order to compute the bootstrap version of the MSE and $M = 1,000$ Monte Carlo simulations are conducted. We notice that the results are very comparable to those of Table 2.

Model 3 is a variation of model 2, where a truncation variable has been added: $T|_{X=x} \sim \text{Exp}(rx)$, where $r > 0$ controls the probability of truncation and is taken to be 45 and 22 in order to give 10 and 20% of truncation, respectively. The results are given in Tables 4 and 5. No comparison with other methods is possible here. The

Table 2 Comparison between the ML and LS methods for model 2, at point $z = 0.3$ ($x_1 = 0, x_2 = 1$)

Censoring	h_{opt}	Method	$\beta_0(z) = 0$		$\beta_1(z) = -1.2$		$\phi(S(z\|x_2)) = -1.2$	
percentage			Bias	MSE	Bias	MSE	Bias	MSE
20	0.25	LS	−0.0644	0.0181	0.1213	0.1217	0.0569	0.0644
	0.4	ML	−0.1333	0.0444	0.3085	0.1507	0.1752	0.0977
40	0.35	LS	−0.0863	0.0219	0.3282	0.1529	0.1041	0.0799
	0.3	ML	−0.0978	0.0449	0.2349	0.1636	0.1371	0.0839

Table 3 MSE of the LS estimator for model 2 using the bootstrap bandwidth selector, at point $z = 0.3$ ($x_1 = 0, x_2 = 1$)

Censoring	$\beta_0(z) = 0$		$\beta_1(z) = -1.2$		$\phi(S(z\|x_2)) = -1.2$	
percentage	Bias	MSE	Bias	MSE	Bias	MSE
20	−0.0974	0.0234	0.2799	0.1595	0.1826	0.0758
40	−0.1225	0.0281	0.4068	0.2343	0.2843	0.1176

Table 4 MSE of the LS estimator for model 3, at point $z = 0.3$ ($x_1 = 0, x_2 = 1$)

Censoring percentage	Truncation percentage	h_{opt}	$\beta_0(z) = 0$ Bias	MSE	$\beta_1(z) = -1.2$ Bias	MSE	$\phi(S(z\|x_2)) = -1.2$ Bias	MSE
20	10	0.25	−0.0851	0.0378	0.1335	0.1674	0.0484	0.0736
	20	0.2	−0.0581	0.0381	0.0727	0.1844	0.0146	0.0864
40	10	0.2	−0.0514	0.0358	0.0636	0.1999	0.0123	0.0968
	20	0.2	−0.0537	0.0424	0.0610	0.2129	0.0073	0.1003

Table 5 MSE of the LS estimator for model 3 using the bootstrap bandwidth selector, at point $z = 0.3$ ($x_1 = 0, x_2 = 1$)

Censoring percentage	Truncation percentage	$\beta_0(z) = 0$ Bias	MSE	$\beta_1(z) = -1.2$ Bias	MSE	$\phi(S(z\|x_2)) = -1.2$ Bias	MSE
20	10	−0.1391	0.0488	0.2708	0.2028	0.1317	0.0755
	20	−0.1524	0.0583	0.2923	0.2283	0.1399	0.0851
40	10	−0.1506	0.0494	0.3072	0.2178	0.1566	0.0859
	20	−0.1692	0.0662	0.3126	0.2370	0.1434	0.0853

tables show similar results to those obtained for model 2, when we had only censoring, thus the MSE increases with the percentage of censoring and truncation. For Table 4 we conducted $M = 10{,}000$ Monte Carlo simulations, while for Table 5, $M = 1{,}000$ and $B = 100$ for the bootstrap replications.

5 Data analysis

The methods presented in the previous sections have been applied to the larynx cancer data set previously studied by Klein and Moeschberger (1997). The data consist of 90 observations about males suffering from larynx cancer. No truncation is present, thus $T = 0$ with probability one. Patients are classified in four groups, according to the stage of their disease. For each individual i ($i = 1, \ldots, 90$) we observe the time-to-death or on-study, Z_i, the death indicator δ_i ($0 =$ alive, $1 =$ dead), the stage of the disease and the age at diagnosis. For these data the independece between Y and C given $\boldsymbol{X}$ seems a reasonable assumption.

The model considered by Klein and Moeschberger (1997) is the additive hazards model, which can be written in the following form:

$$\phi(S(z|\boldsymbol{X})) = \beta_0(z) + \beta_1(z)X_1 + \beta_2(z)X_2 + \beta_3(z)X_3 + \beta_4(z)X_4, \tag{7}$$

where $\phi(u) = -\log(u)$, X_i is the indicator of being at stage $i + 1$ ($i = 1, 2, 3$) and X_4 is the age at diagnosis minus its mean (64.11 years).

Klein and Moeschberger (1997) estimated the regression functions $\beta_i(z)$ ($i = 0, \dots, 4$) by means of the classical method for additive models (see Chap. 10 in their book for more details). They also verified that the assumptions for the additive hazards model hold. We apply the proposed least squares method to estimate the coefficients of this model and compare them to the results obtained by Klein and Moeschberger (1997). Denote $\tau = 4.4$ for the largest Z_i, for which at least one patient is still at risk in each of the four disease stages. The coefficients are estimated for time-points $z \in [0, \tau]$ and $\tilde{w}(x) = 1_{\{-18.66 \leq x \leq 19.44\}}$, where -18.66 and 19.44 are the percentiles 2.5 and 97.5% of X_4, respectively. For the new method, the bandwidth, h, that is needed for the estimation has been chosen by bootstrap among the values 20, 25, 30, 35, 40, 45. Its value was 25.

The 95% pointwise confidence intervals for $\beta_k(z)$ ($0 \leq k \leq 4$) have also been constructed. For the classical method they were found as:

$$\hat{\beta}_k(z) \pm 1.96\sqrt{V\hat{a}r[\hat{\beta}_k(z)]} \quad (0 \leq k \leq 4).$$

with the variance computed using the formulas presented in Chap. 10 of Klein and Moeschberger (1997), while for the new method they were estimated using percentile method, via the bootstrap procedure presented in Sect. 3.

As it can be seen in Fig. 1 the estimator of the cumulative baseline hazard rate, $\beta_0(z)$, is almost the same with both methods, as well as its confidence intervals. Similar things happen for the cumulative excess risk of stage 2, stage 3 and stage 4 of larynx cancer, as compared to stage 1, given by the functions $\beta_1(z)$, $\beta_2(z)$ and $\beta_3(z)$, respectively. As an example, we give the graphs of the estimators of $\beta_3(z)$, as well as their pointwise confidence intervals in Fig. 2. As for the coefficient corresponding to the continuous

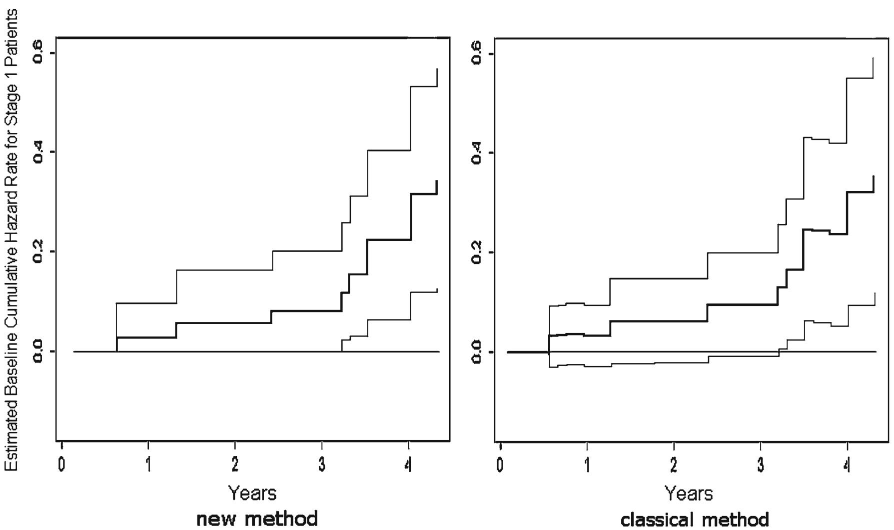

Fig. 1 Estimate of the cumulative baseline hazard rate ($\beta_0(z)$) and 95% pointwise confidence intervals

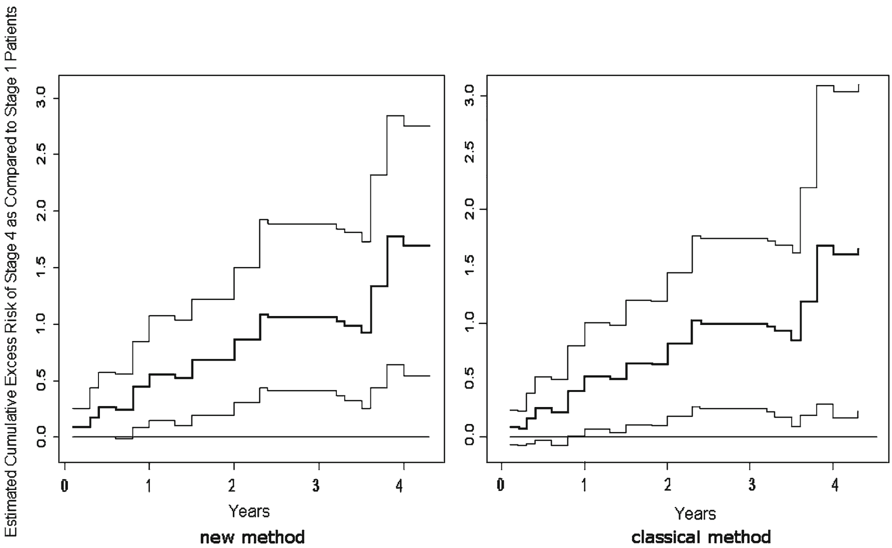

Fig. 2 Estimate of the cumulative excess risk of stage 4 cancer as compared to stage 1 cancer ($\beta_3(z)$) and 95% pointwise confidence intervals

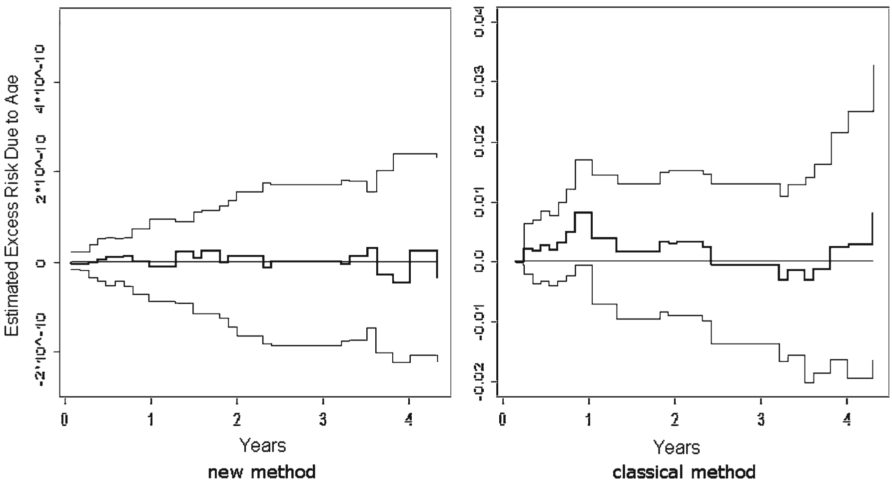

Fig. 3 Estimate of the cumulative effect of age ($\beta_4(z)$) and 95% pointwise confidence intervals

covariate, $\beta_4(z)$, we notice in Fig. 3 that the two curves are slightly different but both very close to zero.

Acknowledgements The first and second author acknowledge support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy). Research of the third author was supported in part by MEC Grant MTM2005-00429 (EU ERDF support included) and XUGA Grant PGIDIT03-PXIC10505PN.

Appendix

Conditions

We now state the conditions used in the result of Sect. 2. Conditions (H1)–(H6) are taken from Iglesias-Pérez and González-Manteiga (1999), on which our proof is based.

Remark 9 Condition (H2) comes from Dabrowska (1989) and comes from the fact that we stay away from the boundaries of the domain of the covariate while estimating the survival function $S(z|x)$, to avoid boundary effects.

(H1) X, Y, T, C are absolutely continuous random variables (r.v.).

(H2) (a) Let $I = [x_1, x_2]$ be an interval contained in the support of m^*, such that

$$0 < \gamma = \inf\{m^*(x) : x \in I_\delta\} < \sup\{m^*(x) : x \in I_\delta\} = \Gamma < \infty$$

for some $1_\delta = [x_1 - \delta, x_2 + \delta]$ with $\delta > 0$ and $0 < \delta\Gamma < 1$.

(b) For all $x \in I$ the r.v. Y, T, C are independent conditionally on $X = x$.

(c) $a_{L(\cdot|x)} \leq a_{H(\cdot|x)}$ and $b_{L(\cdot|x)} \leq b_{H(\cdot|x)}$ for all $x \in I_\delta$.

(d) There exist $a < b \in \mathbb{R}$ satisfying

$$\inf\{\alpha^{-1}(x)(1 - H(b|x))L(a|x) : x \in I_\delta\} \geq \theta > 0.$$

(H3) The first and second derivatives with respect to x of the functions $m(x)$ and $\alpha(x)$ exist and are continuous in I_δ.

(H4) All first and second derivatives with respect to x and y of the functions $L(y|x)$, $H(y|x)$ and $H_1(y|x)$ exist and are continuous and bounded in $(y, x) \in [0, \infty) \times I_\delta$.

(H5) The corresponding (improper) densities of the distribution (subdistribution) functions $L(y)$, $H(y)$ and $H_1(y)$ are bounded away from 0 in $[a, b]$.

(H6) The kernel function K is a symmetric density vanishing outside $(-1, 1)$ and the total variation of K is less than some $\lambda < +\infty$.

(H7) The function ϕ is twice continuously differentiable and its first and second derivatives are bounded by N_1 and N_2, respectively.

(H8) There exists some $N_3 < \infty$ such that $P(|X| \leq N_3) = 1$.

(H9) The matrix $\boldsymbol{A}$ is nonsingular.

(H10) $h \to 0$ as $n \to \infty$ and $\dfrac{\log^3 n}{nh^3} \to 0$, $nh^4 \to 0$.

(H11) The weights $w(x)$ are given by $w(x) = 1_{\{x\in I\}}\tilde{w}(x)$, with I as defined in condition (H2) and where $\tilde{w}(x)$ satisfies $\tilde{w}(x) \geq 0$ for all x, $\sup_x \tilde{w}(x) \leq B$ for some $B < \infty$ and $\int_I \tilde{w}(x) \int_0^\infty \frac{dH_1^*(u|x)}{C(u|x)} dx < \infty$.

(H12) $\det(\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X}) \neq 0$.

Proof of Theorem 1

From (3) and (4) we may write

$$\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} (\hat{\boldsymbol{\phi}}(z) - \boldsymbol{\phi}(z)) = \hat{\boldsymbol{A}}^{-1} \hat{\boldsymbol{b}}(z) \tag{8}$$

where $\hat{\boldsymbol{A}} = n^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X} = (\hat{a}_{ij})_{i,j=0}^{p}$, with $\hat{a}_{ij} = n^{-1} \sum_{l=1}^{n} X_l^{i+j} w(X_l)$ and

$$\hat{\boldsymbol{b}}(z) = n^{-1} \boldsymbol{X}^t \boldsymbol{W} (\hat{\boldsymbol{\phi}}(z) - \boldsymbol{\phi}(z)).$$

The strong law of large numbers implies that $\hat{\boldsymbol{A}} \to \boldsymbol{A}$ a.s., provided that $E(X^{i+j} w(X))$ is finite for all $i, j = 0, \ldots, p$. Using condition (H9) this implies that $\hat{\boldsymbol{A}}^{-1} \to \boldsymbol{A}^{-1}$. On the other hand, $\hat{\boldsymbol{b}}(z) = (\hat{b}_0(z), \hat{b}_1(z), \ldots, \hat{b}_p(z))^t$, with

$$\begin{aligned}
\hat{b}_i(z) &= \frac{1}{n} \sum_{j=1}^{n} X_j^i w(X_j)(\hat{\phi}_j(z) - \phi_j(z)) \\
&= \frac{1}{n} \sum_{j=1}^{n} X_j^i w(X_j)(\phi(\hat{S}_n(z|X_j)) - \phi(S(z|X_j))).
\end{aligned}$$

A Taylor expansion of ϕ around $S(z|X_j)$ gives $\hat{b}_i(z) = \hat{b}_i^{(1)}(z) + \hat{b}_i^{(2)}(z)$, where

$$\hat{b}_i^{(1)}(z) = \frac{1}{n} \sum_{j=1}^{n} X_j^i w(X_j) \phi^{'}(S(z|X_j))(\hat{S}_n(z|X_j) - S(z|X_j))$$

and

$$\hat{b}_i^{(2)}(z) = \frac{1}{2n} \sum_{j=1}^{n} X_j^i w(X_j) \phi^{''}(\Delta_j(z))(\hat{S}_n(z|X_j) - S(z|X_j))^2,$$

with some $\Delta_j(z)$ in between $S(z|X_j)$ and $\hat{S}_n(z|X_j)$.

First, we will prove that $\hat{b}_i^{(2)}(z) = o_p(n^{-1/2})$. Note that

$$|\hat{b}_i^{(2)}(z)| \leq \tfrac{1}{2n} \sup_{\substack{y \in [a,b] \\ x \in I}} |F(y|x) - \hat{F}_n(y|x)|^2 \sum_{j=1}^{n} |X_j^i| \, w(X_j) \, |\phi^{''}(\Delta_j(z))|.$$

Applying the uniform consistency of $\hat{F}_n(z|x)$, given by Lemma 5 in Iglesias-Pérez and González-Manteiga (1999), together with conditions (H7) and (H8) gives that $\hat{b}_i^{(2)}(z) = o_p\left(n^{-1/2}\right)$ uniformly in z.

Let us now concentrate on $\hat{b}_i^{(1)}(z)$. Using the iid representation for $\hat{S}_n(z|X)$ given in Iglesias-Pérez and González-Manteiga (1999), we have:

$$\hat{S}_n(z|X_j) - S(z|X_j) = \sum_{l=1}^{n} B_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z) + R_n(z|X_j), \quad (9)$$

where

$$\sup_{\substack{y\in[a,b]\\ x\in I}} |R_n(y|x)| = O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right), \quad (10)$$

and

$$\xi(Z, T, \delta, x, y) = \frac{1_{\{Z\le y,\delta=1\}}}{C(Z|x)} - \int_0^y \frac{1_{\{T\le u\le Z\}}}{C^2(u|x)} dH_1^*(u|x).$$

Observe that $E[\xi(Z, T, \delta, x, y)|X = x] = 0$. We plug (9) into $\hat{b}_i^{(1)}(z)$ to obtain:

$$\begin{aligned}\hat{b}_i^{(1)}(z) &= \frac{1}{n}\sum_{j=1}^{n} X_j^i w(X_j)\phi^{'}(S(z|X_j)) \sum_{l=1}^{n} B_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z)\\ &\quad +\frac{1}{n}\sum_{j=1}^{n} X_j^i w(X_j)\phi^{'}(S(z|X_j))R_n(z|X_j)\\ &= \hat{b}_i^{(11)}(z) + \hat{b}_i^{(R)}(z).\end{aligned}$$

Define $\tilde{B}_{nl}(X_j) = m^*(X_j)^{-1}(nh)^{-1}K(\frac{X_j-X_l}{h})$. Then,

$$\hat{b}_i^{(11)}(z) = \hat{b}_i^{(111)}(z) + \hat{b}_i^{(112)}(z) + \hat{b}_i^{(113)}(z),$$

with

$$\hat{b}_i^{(111)}(z) = \frac{1}{n}\sum_{j=1}^{n} X_j^i w(X_j)\phi^{'}(S(z|X_j))\sum_{\substack{l\ne j\\ l=1}}^{n} \tilde{B}_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z), \quad (11)$$

$$\hat{b}_i^{(112)}(z) = \frac{1}{n}\sum_{j=1}^{n} X_j^i w(X_j)\phi^{'}(S(z|X_j))\tilde{B}_{nj}(X_j)S(z|X_j)\xi(Z_j, T_j, \delta_j, X_j, z),$$

$$\begin{aligned}\hat{b}_i^{(113)}(z) &= \frac{1}{n}\sum_{j=1}^{n} X_j^i w(X_j)\phi^{'}(S(z|X_j)) \sum_{l=1}^{n}(B_{nl}(X_j)\\ &\quad -\tilde{B}_{nl}(X_j))S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z).\end{aligned}$$

We shall first prove that $\hat{b}_i^{(R)}(z)$, $\hat{b}_i^{(112)}(z)$ and $\hat{b}_i^{(113)}(z)$ are $o_p(n^{-1/2})$.

For $\hat{b}_i^{(R)}(z)$ we have from (10) and using condition (H10) that

$$|\hat{b}_i^{(R)}(z)| \leq N_3^i N_1 O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right) = o_p(n^{-1/2}),$$

uniformly in z.

For $\hat{b}_i^{(113)}(z)$, note that

$$B_{nl}(X_j) - \tilde{B}_{nl}(X_j) = B_{nl}(X_j)\frac{m^*(X_j) - \hat{m}^*(X_j)}{m^*(X_j)},$$

where $\hat{m}^*(x) = (nh)^{-1}\sum_{j=1}^n K(\frac{x-X_j}{h})$. This implies that

$$\begin{aligned}\hat{b}_i^{(113)}(z) &= \frac{1}{n}\sum_{j=1}^n X_j^i w(X_j)\phi^{'}(S(z|X_j))\frac{m^*(X_j) - \hat{m}^*(X_j)}{m^*(X_j)} \\ &\quad\times S(z|X_j)\sum_{l=1}^n B_{nl}(X_j)\xi(Z_l, T_l, \delta_l, X_j, z) \\ &= \frac{1}{n}\sum_{j=1}^n X_j^i w(X_j)\phi^{'}(S(z|X_j))\frac{m^*(X_j) - \hat{m}^*(X_j)}{m^*(X_j)} \\ &\quad\times\left\{F(z|X_j) - \hat{F}_n(z|X_j) + O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right)\right\}.\end{aligned}$$

Since

$$\sup_{x\in I}|m^*(x) - \hat{m}^*(x)| = O_p\left(\left(\frac{\log n}{nh}\right)^{1/2} + h^2\right)$$

(see e.g. Silverman 1978), it follows that

$$\begin{aligned}|\hat{b}_i^{(113)}(z)| &\leq \frac{1}{n}N_3^i N_1\left\{\sup_{x\in I}|m^*(x) - \hat{m}^*(x)|\right\} \\ &\quad\times\left\{\sup_{x\in I, y\in[a,b]}|\hat{F}_n(y|x) - F(y|x)| + O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right)\right\}\sum_{j=1}^n\frac{w(X_j)}{m^*(X_j)} \\ &= O_p\left(\frac{\log n}{nh}\right) = o_p(n^{-1/2}).\end{aligned}$$

As done for the term $\hat{b}_i^{(111)}(z)$ in the proof of Corollay 2 we can show that under condition (H11) $\hat{b}_i^{(112)}(z)$ is $O_p(n^{-1/2})$ uniformly in z.

So far we have proved that

$$\hat{b}_i(z) = \hat{b}_i^{(111)}(z) + o_p(n^{-1/2}),$$

uniformly in $z \in [a, b]$.

We will now prove the asymptotic normality of $\hat{b}_i^{(111)}(z)$ for a fixed $a \leq z \leq b$. Define

$$h_i(\boldsymbol{V}_j, \boldsymbol{V}_l) = X_j^i w(X_j)\phi^{'}(S(z|X_j))S(z|X_j)\tilde{B}_{nl}(X_j)\xi(Z_l, T_l, \delta_l, X_j, z),$$

where $\boldsymbol{V}_j = (Z_j, T_j, \delta_j, X_j)$. Let $\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_l) = \frac{1}{2}(h_i(\boldsymbol{V}_j, \boldsymbol{V}_l) + h_i(\boldsymbol{V}_l, \boldsymbol{V}_j))$. Then,

$$\hat{b}_i^{(111)}(z) = \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l \neq j \\ l=1}}^{n} \tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_l).$$

Thus, $\hat{b}_i^{(111)}(z)$ is a symmetric U statistic. Note however that its kernel $\tilde{h}_i$ depends on n. We use the Hájek projection [see Serfling (1980), p. 190] to decompose it into the following sum:

$$\hat{b}_i^{(111)}(z) = D_i^{(1)} + D_i^{(2)} + D_i^{(3)} + D_i^{(4)},$$

where

$$\begin{aligned} D_i^{(1)} &= \frac{2}{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ k>j}}^{n} h_i^{(1)}(\boldsymbol{V}_j, \boldsymbol{V}_k), \\ D_i^{(2)} &= \frac{n-1}{n}\sum_{j=1}^{n} h_i^{(2)}(\boldsymbol{V}_j), \\ D_i^{(3)} &= \frac{n-1}{n}\sum_{k=1}^{n} h_i^{(3)}(\boldsymbol{V}_k), \\ D_i^{(4)} &= (n-1)E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)], \end{aligned}$$

with

$$\begin{aligned} h_i^{(1)}(\boldsymbol{V}_j, \boldsymbol{V}_k) &= \tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k) - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_j] \\ &\quad - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_k] + E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)], \\ h_i^{(2)}(\boldsymbol{V}_j) &= E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_j] - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)], \\ h_i^{(3)}(\boldsymbol{V}_k) &= E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_k] - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)]. \end{aligned}$$

Note that $D_i^{(2)} = D_i^{(3)}$ because of the symmetry of $\tilde{h}_i$. Since $D_i^{(1)}$, $D_i^{(2)}$, $D_i^{(3)}$ and $D_i^{(4)}$ depend on n, standard results for U statistics cannot be applied, and so we need to compute directly the mean and the variance of each of the above terms. We will first prove that $D_i^{(1)} = o_p(n^{-1/2})$. It is easy to prove that $E(D_i^{(1)}) = 0$, while tedious but straightforward algebra show that

$$\mathrm{Var}\left(D_i^{(1)}\right) = \frac{2(n-1)}{n} E\left\{[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)]^2\right\}.$$

It can be easily proved that

$$E[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)^2] \leq E[\tilde{h_i}^2(\boldsymbol{V}_1, \boldsymbol{V}_2)],$$

with $E[\tilde{h_i}^2(\boldsymbol{V}_1, \boldsymbol{V}_2)] \leq E[h_i^2(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^{-1}n^{-2})$. This implies that $E[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)^2] = O(h^{-1}n^{-2})$ and, consequently, $\mathrm{Var}(D_i^{(1)}) = O(n^{-2}h^{-1})$, which gives

$$D_i^{(1)} = O_p(n^{-1}h^{-1/2}) = o_p(n^{-1/2}).$$

Now, $D_i^{(4)} = (n-1)E[\tilde{h_i}(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^2) = o_p(n^{-1/2})$.

It remains only to deal with $D_i^{(2)}$ and $D_i^{(3)}$, which are two sums of iid terms and will give the asymptotic normality of $\hat{b}_i^{(111)}(z)$. For $D_i^{(2)}$ it is easy to show that $E[D_i^{(2)}] = 0$ and that for any $0 \leq i, j \leq p$,

$$\begin{aligned}\mathrm{Cov}(D_i^{(2)}, D_j^{(2)}) = &\frac{(n-1)^2}{n} E[\tilde{h_i}(\boldsymbol{V}_1, \boldsymbol{V}_2)\tilde{h_j}(\boldsymbol{V}_3, \boldsymbol{V}_2)] \\ &-\frac{(n-1)^2}{n} E[\tilde{h_i}(\boldsymbol{V}_1, \boldsymbol{V}_2)]E[\tilde{h_j}(\boldsymbol{V}_1, \boldsymbol{V}_2)].\end{aligned}$$

On the other hand,

$$E[\tilde{h_i}(\boldsymbol{V}_3, \boldsymbol{V}_2)\tilde{h_j}(\boldsymbol{V}_1, \boldsymbol{V}_2)] = \Delta_{ij}^{(1)} + \Delta_{ij}^{(2)} + \Delta_{ij}^{(3)} + \Delta_{ij}^{(4)},$$

where $\Delta_{ij}^{(1)} = \frac{1}{4}E[h_i(V_3, V_2)h_j(V_1, V_2)]$, $\Delta_{ij}^{(2)} = \frac{1}{4}E[h_i(V_2, V_3)h_j(V_1, V_2)]$, $\Delta_{ij}^{(3)} = \frac{1}{4}E[h_i(V_3, V_2)h_j(V_2, V_1)]$ and $\Delta_{ij}^{(4)} = \frac{1}{4}E[h_i(V_2, V_3)h_j(V_2, V_1)]$. It can be easily seen that

$$\Delta_{ij}^{(1)} = \frac{1}{4n^2}\int_I x^{i+j}\tilde{w}^2(x)S^2(z|x)\phi^{'}(S(z|x))^2\int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)}m^*(x)dx + O(h^2n^{-2}),$$

$\Delta_{ij}^{(2)} = \Delta_{ij}^{(3)} = O(h^2n^{-2})$ and $\Delta_{ij}^{(4)} = O(h^4n^{-2})$, since $E[h_i(\boldsymbol{V}_1, \boldsymbol{V}_2)|\boldsymbol{V}_1] = O(h^2n^{-1})$.

As a consequence

$$\mathrm{Cov}(D_i^{(2)}, D_j^{(2)}) = \frac{(n-1)^2}{4n^3}[\sigma_{ij}(z) + O(h^2)],$$

with $\sigma_{ij}(z)$ defined in (5). It now follows from the central limit theorem for triangular arrays that for any $d \in \mathbb{R}^{p+1}$,

$$n^{1/2}d^t\hat{\boldsymbol{b}}(z) = 2n^{1/2}d^t(D_0^{(2)}, \ldots, D_p^{(2)})^t + o_p(n^{-1/2}) \xrightarrow{d} N(0, d^t\boldsymbol{\Sigma}(z)d).$$

Direct application of the Cramér-Wold device implies that

$$n^{1/2}\hat{\boldsymbol{b}}(z) \xrightarrow{d} N(0, \boldsymbol{\Sigma}(z)).$$

From this and the fact that $\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = \hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{b}}(z)$ (see (8)), we then get

$$n^{1/2}(\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z)) \xrightarrow{d} N\left(0, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}(z)(\boldsymbol{A}^{-1})^t\right),$$

which concludes the proof.

Proof of Corollary 2

From the proof of Theorem 1 it follows that we only have to prove that $\hat{b}_i^{(111)}(z)$ (defined in (11)) is is $O_p(n^{-1/2})$ uniformly in z. For this, partition the interval $[a, b]$ into $U_n \sim n^{1/2}$ subintervals $a = Z_0^{'} \leq Z_1^{'} \leq \ldots \leq Z_{U_n}^{'} = b$, of length $O(n^{-1/2})$. Then,

$$\sup_z |\hat{b}_i^{(111)}(z)| \leq \max_{1\leq r\leq U_n} \sup_{Z_{r-1}^{'}\leq z\leq Z_r^{'}} |\hat{b}_i^{(111)}(z) - \hat{b}_i^{(111)}(Z_r^{'})| + \max_{0\leq r\leq U_n} |\hat{b}_i^{(111)}(Z_r^{'})|$$
$$:= R_1 + R_2$$

We will first prove that $R_2 = o_p(n^{-1/2})$ and then that $R_1 = O_p(n^{-1/2})$. Consider for $\epsilon_n = cn^{-1/2}$ $(c > 0)$,

$$\begin{aligned}
&P(\max_{0\leq r\leq U_n} |\hat{b}_i^{(111)}(Z_r^{'})| > \epsilon_n) \\
&= P\left(\max_{0\leq r\leq U_n} \left|\frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} X_j^i w(X_j)\phi^{'}(S(Z_r^{'}|X_j))\right.\right. \\
&\quad \left.\left.\times S(Z_r^{'}|X_j)\tilde{B}_{nl}(X_j)\xi(Z_l, T_l, \delta_l, X_j, Z_r^{'})\right| > \epsilon_n\right) \\
&\leq \sum_{r=0}^{U_n} P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} X_j^i w(X_j)\phi^{'}(S(Z_r^{'}|X_j))\right.\right. \\
&\quad \left.\left.\times S(Z_r^{'}|X_j)\tilde{B}_{nl}(X_j)\xi(Z_l, T_l, \delta_l, X_j, Z_r^{'})\right| > \epsilon_n\right) \\
&\leq \frac{1}{\epsilon_n^2}\sum_{r=0}^{U_n} Var\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} X_j^i w(X_j)\phi^{'}(S(Z_r^{'}|X_j))\right. \\
&\quad \left.\times S(Z_r^{'}|X_j)\tilde{B}_{nl}(X_j)\xi(Z_l, T_l, \delta_l, X_j, Z_r^{'}) > \epsilon_n\right) \\
&= O(n^{1/2}\cdot n \cdot n^{-1}h^2) = o(1).
\end{aligned}$$

where we used Chebysev's inequality and the variance that appears in the expression above is proved to be $O(n^{-1}h^2)$ by Taylor developments and expectations computations. Next, consider R_1:

$$
\begin{aligned}
R_1 &= \max_{1\le r\le U_n} \sup_{Z'_{r-1}\le z\le Z'_r} |\hat{b}_i^{(111)}(z) - \hat{b}_i^{(111)}(Z'_r)| \\
&\le \max_{1\le r\le U_n} \sup_{Z'_{r-1}\le z\le Z'_r} \Bigg| \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} X_j^i w(X_j)\tilde{B}_{nl}(X_j) \\
&\quad \times \left[\phi'(S(z|X_j))S(z|X_j)\frac{1_{\{Z_l\le z,\delta_l=1\}}}{C(Z_l|X_j)} - \phi'(S(Z'_r|X_j))S(Z'_r|X_j)\frac{1_{\{Z_l\le Z'_r,\delta_l=1\}}}{C(Z_l|X_j)}\right]\Bigg| \\
&\qquad + \max_{1\le r\le U_n} \sup_{Z'_{r-1}\le z\le Z'_r} \Bigg| \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} X_j^i w(X_j)\tilde{B}_{nl}(X_j) \\
&\quad \times \Bigg[\phi'(S(z|X_j))S(z|X_j)\int_0^z \frac{1_{\{T_l\le u\le Z_l\}}}{C^2(u|X_j)} dH_1^*(u|X_j) \\
&\qquad - \phi'(S(Z'_r|X_j))S(Z'_r|X_j)\int_0^{Z'_r} \frac{1_{\{T_l\le u\le Z_l\}}}{C^2(u|X_j)} dH_1^*(u|X_j)\Bigg]\Bigg| \\
&\le A_1 + A_2.
\end{aligned}
$$

Write

$$
\begin{aligned}
A_1 &= \max_{1\le r\le U_n} \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} |X_j|^i \cdot w(X_j)\cdot \tilde{B}_{nl}(X_j) \cdot \sup_{Z'_{r-1}\le z\le Z'_r} \frac{1_{\{Z_l\le z,\delta_l=1\}}}{C(Z_l|X_j)} \\
&\qquad \cdot \sup_{Z'_{r-1}\le z\le Z'_r} |\phi'(S(z|X_j))S(z|X_j) - \phi'(S(Z'_r|X_j))S(Z'_r|X_j)| \\
&\qquad + \max_{1\le r\le U_n} \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\ l=1}}^{n} |X_j|^i \cdot w(X_j) \\
&\qquad \cdot \tilde{B}_{nl}(X_j)\frac{1}{C(Z_l|X_j)} |\phi'(S(Z'_r|X_j))S(Z'_r|X_j)| \\
&\qquad \cdot \sup_{Z'_{r-1}\le z\le Z'_r} |1_{\{Z_l\le z,\delta_l=1\}} - 1_{\{Z_l\le Z'_r,\delta_l=1\}}| \\
&= A_{11} + A_{12}.
\end{aligned}
$$

Using the fact that $\sup_{Z'_{r-1}\le z\le Z'_r} |\phi'(S(z|X_j))S(z|X_j) - \phi'(S(Z'_r|X_j))S(Z'_r|X_j)| = O(U_n^{-1})$ and condition (H11) it is easy to prove that $A_{11} = O(U_n^{-1}) = O_p(n^{-1/2})$, and in a similar way it can be shown that $A_2 = O(U_n^{-1}) = O_p(n^{-1/2})$.

Similar techniques as for R_2 combined with

$$\sup_{Z'_{r-1}\leq z\leq Z'_r} |1_{\{Z_l\leq z,\delta_l=1\}} - 1_{\{Z_l\leq Z'_r,\delta_l=1\}}| = 1_{\{Z'_{r-1}\leq Z_l\leq Z'_r,\delta_l=1\}},$$

give us that $A_{12} = O_p(n^{-1/2})$. Hence, $\hat{b}_i^{(111)}(z) = O_p(n^{-1/2})$ uniformly in z.

References

Beran, R. (1981). Nonparametric regression with randomly censored data, *Technical report*. Berkeley: University California.

Cai, Z., Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*, *30*, 93–111.

Cao, R., González-Manteiga, W. (2008). Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics*, *143*, 166–190.

Chen, K. N., Jin, Z. Z., Ying, Z. L. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, *89*, 659–668.

Cheng, S. C., Wei, L. J., Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, *83*, 835–846.

Cheng, S. C., Wei, L. J., Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association*, *92*(437), 227–235.

Dabrowska, D. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Annals of Statistics*, *17*, 1157–1167.

Doksum, K. A. (1987). An extention of partial likelihood methods for proportional hazard models to general transformation models. *Annals of Statistics*, *15*, 325–345.

Fine, J. P., Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *96*(454), 620–628.

Fine, J. P., Ying, Z., Wei, L. J. (1998). On the linear transformation model for censored data. *Biometrika*, *85*, 980–986.

Iglesias-Pérez, C., González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *Journal of Nonparametric Statistics*, *10*, 213–244.

Iglesias-Pérez, C., González-Manteiga, W. (2003). Bootstrap for the conditional distribution function with truncated and censored data. *The Annals of the Institute of Statistical Mathematics*, *55*, 331–357.

Jung, S. H. (1996). Regression analysis for long-term survival rate. *Biometrika*, *83*, 227–232.

Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

Kauermann, G. (2005). Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics and Data Analysis*, *49*, 169–186.

Klein, J. P., Moeschberger, M. L. (1997). *Survival analysis: Techniques for censored and truncated data*. New York: Springer.

Lambert, P., Eilers, P. H. C. (2004). Bayesian survival models with smooth time-varying coefficients using penalized Poisson regression. Technical Report TR0435. http://www.stat.ucl.ac.be/publications.

Murphy, S. A., Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications*, *39*, 153–180.

Nan, B., Lin, X. (2003). A varying-coefficient Cox model for the effect of age at a marker event on age at menopause. http://www.bepress.com/umichbiostat/paper16.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics*, *6*, 177–184.

Subramanian, S. (2001). Parameter estimation in regression for long-term survival rate from censored data. *Journal of Statistical Planning and Inference*, *99*, 211–222.

Subramanian, S. (2004). Survival-rate regression using kernel conditional Kaplan–Meier estimators. *Journal of Statistical Planning and Inference*, *123*, 187–205.

Tsai, W. Y., Jewell, N. P., Wang, M. C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika*, *74*, 883–886.