T E C H N I C A L

R E P O R T

0557

GENERALIZED TIME-DEPENDENT CONDITIONAL LINEAR MODELS UNDER LEFT TRUNCATION AND RIGHT CENSORING

TEODORESCU B., VAN KEILEGOM I., and R. CAO

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Generalized Time-dependent Conditional Linear Models under Left Truncation and Right Censoring

Bianca Teodorescu[1] Ingrid Van Keilegom[1] Ricardo Cao[2]

November 14, 2005

Abstract

Consider the model $\phi(S(z|X)) = \boldsymbol{\beta}(z)^t \vec{X}$, 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 $\boldsymbol{\beta}(z) = (\beta_0(z), \ldots, \beta_p(z))^t$ 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's 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.

Key words: Additive hazards model, bootstrap, least-squares estimator, logistic model, proportional hazards model, semiparametric regression, survival analysis, time-dependent coefficients, U-statistics.

[1] Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: teodorescu@stat.ucl.ac.be, vankeilegom@stat.ucl.ac.be.

[2] 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's model, log-logistic model, accelerated failure time model, etc.). 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 e.g., 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 e.g. Murphy & Sen (1991), Nan & Lin (2003), Cai & Sun (2003), among others. Also other time-dependent survival models have been considered, see for example Lambert & 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, 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 (Z, T, δ) only if $Z \geq T$, where $Z = \min\{Y, C\}$ and $\delta = I_{\{Y \leq C\}}$. Let $(Z_i, T_i, \delta_i, X_i)$, $i = 1, \ldots, n$ be an iid sample from (Z, T, δ, X), where X is a (one-dimensional) covariate. 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.:

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \ldots + \beta_p(z)X^p, \tag{1.1}$$

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 version of the proportional hazards model.

In the absence of truncation, model (1.1) 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 & González-Manteiga (2003), 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 Section 3 we present a bootstrap based method for the selection of the smoothing parameter, while in Section 4 we give some numerical results. The analysis of larynx cancer data is conducted in Section 5. Finally, Section 6 contains the proofs.

2 Least squares estimator and its asymptotic properties

We need to introduce the following notations: $M(x) = P(X \leq x)$, $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $L(y|x) = P(T \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H_1(y|x) = P(Z \leq y, \delta = 1|x)$, $L(y) = P(T \leq y)$, $H(y) = P(Z \leq y)$, $H_1(y) = P(Z \leq y, \delta = 1)$, $C(y|x) = P(T \leq y \leq Z|x, T \leq Z)$, and $\alpha(x) = P(T \leq Z|X = x)$, 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 p-th 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 & González-Manteiga (1999):

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

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 $C_n(u|x) = \sum_{j=1}^{n} I_{\{T_j \le u \le Z_j\}} B_{nj}(x)$.

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-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 $\beta_j(z)$ $(j = 0, \ldots, p)$:

$$\hat{\boldsymbol{\beta}}(z) = \begin{pmatrix} \hat{\beta}_0(z) \\ \hat{\beta}_1(z) \\ \vdots \\ \hat{\beta}_p(z) \end{pmatrix} = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \hat{\boldsymbol{\phi}}(z), \tag{2.1}$$

where

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

and $\boldsymbol{W} = \mathrm{diag}(w(X_1), \dots, w(X_n))$, 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 Section 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 Section 6.

Let $\boldsymbol{\phi}(z) = (\phi(S(z|X_1)), \dots, \phi(S(z|X_n)))^t$. Model (1.1) implies that $\boldsymbol{\phi}(z) = \boldsymbol{X}\boldsymbol{\beta}(z)$, which leads to

$$\boldsymbol{\beta}(z) = (\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{\phi}(z), \tag{2.2}$$

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 2.1 *Suppose that conditions (H1) through (H12) hold. Then, for $a \le z \le b$,*

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

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

$$\sigma_{ij}(z) = \int_I x^{i+j}\tilde{w}^2(x)S^2(z|x)\phi'(S(z|x))^2 \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)} m^*(x)dx, \tag{2.3}$$

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

Remark 2.1 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.

Remark 2.2 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 (1.1). Note that this estimator can in certain cases be non-monotone. A convenient and satisfactory solution is to keep the estimator constant until it starts decreasing again.

Remark 2.3 It is important to have at hand a procedure to test the validity of the assumed model (1.1). This can be done by measuring the distance between $\phi(\hat{S}_n(z|x))$ and $\hat{\beta}_0(z) + \hat{\beta}_1(z)x + \ldots + \hat{\beta}_p(z)x^p$ uniformly over all z and x.

3 Bandwidth selection

The estimator $\hat{\boldsymbol{\beta}}(z)$ defined in Section 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

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

   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 (1.1) and estimate $\beta_0(z), \ldots, \beta_p(z)$ by the least squares estimator in (2.1) to obtain $\hat{\beta}_{0,g_j}(z), \ldots, \hat{\beta}_{p,g_j}(z)$. Plug these estimators into (1.1) and re-estimate $S(z|X_i)$ by

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

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 $Z_i^* = \min\{Y_i^*, C_i^*\}$, $\delta_i^* = I_{\{Y_i^* \leq C_i^*\}}$ and simulate new values Y_i^*, C_i^* and T_i^* if $T_i^* > Z_i^*$.

d) Use this resample $\{(Z_1^*, T_1^*, \delta_1^*, X_1), \ldots, (Z_n^*, T_n^*, \delta_n^*, X_n)\}$ 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 3.1 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 2.1.

Remark 3.2 The asymptotic validity of a slight variation of the above bootstrap procedure has been established by Iglesias-Pérez & 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.

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 Section 3.

Along the simulations, the following model is considered:

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x. \quad (4.1)$$

In the discrete case, model 1 considers that X is uniformly distributed in $\{1.1, 1.3, 1.5, 1.7, 1.9\}$, $Y|_{X=x} \sim$ 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 $n = 200$ and the number of Monte Carlo simulations is $M = 10000$. 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.

Censoring		$\beta_0(z) = 0$		$\beta_1(z) = -0.4$	
percentage	Method	Bias	MSE	Bias	MSE
10	LS	0.0075	0.7842	-0.0188	0.3408
	ML	0.0161	0.7607	0.0059	0.3495
20	LS	-0.0104	0.9107	-0.0072	0.3921
	ML	-0.0003	0.8733	0.0213	0.3999
30	LS	-0.0123	1.0267	-0.0099	0.4336
	ML	-0.0009	0.9761	0.0191	0.4358
40	LS	-0.0125	1.2383	-0.0119	0.5116
	ML	0.0043	1.1607	0.0153	0.5133

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

Model 2 deals with the continuous case, $\exp(X) \sim U[2, 3]$, $Y|_{X=x} \sim \text{Exp}(4x)$, $C|_{X=x} \sim \text{Exp}(dx)$ with $d > 0$ that gives different censoring probabilities, and $\phi(u) = \log(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = -4zx$. The sample size is taken $n = 100$ and $M = 10000$ Monte Carlo simulations are conducted. Since we have a one-dimensional continuous covariate, a bandwidth, h, is needed in order to estimate $S(z|x)$. We worked with $h = 0.15$ and $h = 0.30$. The Nadaraya-Watson weights are calculated based on the uniform kernel $K(u) = I_{\{-1 \leq u \leq 1\}} \cdot 1/2$.

Censoring percentage	h	Method	$\beta_0(z) = 0$ Bias	$\beta_0(z) = 0$ MSE	$\beta_1(z) = -1$ Bias	$\beta_1(z) = -1$ MSE
	0.15	LS	-0.2954	0.6807	0.2583	0.7867
20		ML	-0.2985	0.6449	0.2712	0.7572
	0.3	LS	-0.7617	0.6275	0.8315	0.7355
		ML	-0.7623	0.6282	0.8334	0.7386
	0.15	LS	-0.2767	0.8175	0.2848	1.0122
40		ML	-0.2801	0.7484	0.3005	0.9485
	0.3	LS	-0.7501	0.6258	0.8247	0.7377
		ML	-0.7515	0.6274	0.8278	0.7423

Censoring percentage	h	Method	$\phi(S(z\|x_1)) = -0.6931$ Bias	$\phi(S(z\|x_1)) = -0.6931$ MSE	$\phi(S(z\|x_2)) = -1.0986$ Bias	$\phi(S(z\|x_2)) = -1.0986$ MSE
	0.15	LS	-0.1163	0.0511	-0.0115	0.0657
20		ML	-0.1106	0.0538	-0.0122	0.0762
	0.3	LS	-0.1853	0.0438	0.1517	0.0412
		ML	-0.1846	0.0477	0.1534	0.0434
	0.15	LS	-0.0793	0.0611	0.0361	0.0694
40		ML	-0.0718	0.0629	0.0499	0.0844
	0.3	LS	-0.1784	0.0554	0.1559	0.0504
		ML	-0.1777	0.0552	0.1579	0.0512

Table 2: Comparison between the ML and LS methods for model 2, at point $z = 0.25$ ($x_1 = \log(2)$, $x_2 = \log(3)$).

Table 2 shows the results for $z = 0.25$. 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 $\log(2)$ and $\log(3)$. The results in Table 2 are very similar for the two methods. We can also notice from Table 2, that the choice of h has quite a big influence on the results. Table 3 gives the results for the bandwidth estimated by means of the bootstrap procedure described in Section 3. The bandwidth is selected from the grid $\{0.1, 0.15, 0.2, 0.25, 0.3\}$. $B = 100$ bootstrap replications are constructed each time in order to compute the bootstrap version of the MSE and $M = 1000$ Monte Carlo simulations are conducted.

Censoring	$\beta_0(z) = 0$		$\beta_1(z) = -1$	
percentage	Bias	MSE	Bias	MSE
20	-0.7601	0.6123	0.8493	0.7666
40	-0.7016	0.6250	0.8605	0.7816

Censoring	$\phi(S(z\|x_1)) = -0.6931$		$\phi(S(z\|x_2)) = -1.0986$	
percentage	Bias	MSE	Bias	MSE
20	-0.1738	0.0412	0.1525	0.0432
40	-0.1891	0.0527	0.1537	0.0545

Table 3: MSE of the LS estimator for model 2 using the bootstrap bandwidth selector, at point $z = 0.25$ ($x_1 = \log(2)$, $x_2 = \log(3)$).

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. The results are given in Tables 4 and 5. No comparison with other methods is possible here. The tables show similar results to those obtained for model 2, when we had only censoring.

Censoring percentage	Truncation percentage	h	$\beta_0(z)=0$ Bias	MSE	$\beta_1(z)=-1$ Bias	MSE
	10	0.15	-0.2857	0.7001	0.3039	0.8565
20		0.3	-0.7646	0.6417	0.8359	0.7503
	20	0.15	-0.2977	0.7747	0.3178	0.9378
		0.3	-0.7620	0.6436	0.8325	0.7494
	10	0.15	-0.2782	0.8252	0.2876	1.0167
40		0.3	-0.7621	0.6478	0.8338	0.7564
	20	0.15	-0.2675	0.8493	0.2832	1.0435
		0.3	-0.7578	0.6452	0.8273	0.7507

Censoring percentage	Truncation percentage	h	$\phi(S(z\|x_1))=-0.6931$ Bias	MSE	$\phi(S(z\|x_2))=-1.0986$ Bias	MSE
	10	0.15	-0.0750	0.0535	0.0481	0.0570
20		0.3	-0.1852	0.0554	0.1537	0.0467
	20	0.15	-0.0774	0.0604	0.0514	0.0604
		0.3	-0.1849	0.0573	0.1525	0.0483
	10	0.15	-0.0788	0.0626	0.0377	0.0695
40		0.3	-0.1841	0.0586	0.1539	0.0510
	20	0.15	-0.0712	0.0643	0.0436	0.0699
		0.3	-0.1843	0.0601	0.1511	0.0521

Table 4: MSE of the LS estimator for model 3, at point $z = 0.25$ ($x_1 = \log(2)$, $x_2 = \log(3)$).

Censoring percentage	Truncation percentage	$\beta_0(z) = 0$		$\beta_1(z) = -1$	
		Bias	MSE	Bias	MSE
20	10	-0.7610	0.6362	0.8347	0.7483
	20	-0.7712	0.6410	0.8466	0.7491
40	10	-0.7632	0.6398	0.8366	0.7534
	20	-0.7748	0.6421	0.8276	0.7551

Censoring percentage	Truncation percentage	$\phi(S(z\|x_1)) = -0.6931$		$\phi(S(z\|x_2)) = -1.0986$	
		Bias	MSE	Bias	MSE
20	10	-0.1823	0.0527	0.1560	0.0461
	20	-0.1879	0.0562	0.1553	0.0473
40	10	-0.1845	0.0549	0.1592	0.0498
	20	-0.1860	0.0597	0.1613	0.0517

Table 5: MSE of the LS estimator for model 3 using the bootstrap bandwidth selector, at point $z = 0.25$ ($x_1 = \log(2)$, $x_2 = \log(3)$).

5 Data Analysis

The methods presented in the previous sections have been applied to the larynx cancer data set previously studied by Klein & Moeschberger (1997). The data consist of 90 observations about males suffering from larynx cancer. 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.

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

$$\phi(S(z|\boldsymbol{X})) = \beta_0(z) + \beta_1(z)X_1 + \beta_2(z)X_2 + \beta_3(z)X_3 + \beta_4(z)X_4, \tag{5.1}$$

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 & Moeschberger (1997) estimated the regression functions $\beta_i(z)$ $(i = 0, \dots, 4)$ by means of the classical method for additive models (see Chapter 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 & Moeschberger (1997). Denote $\omega = 4.4$ 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 $z \in [0; 4.4]$. 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 Chapter 10 of Klein & Moeschberger (1997), while for the new method they were estimated using percentile bootstrap, via the bootstrap procedure presented in Section 3.

As it can be seen in Figure 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 $\beta_1(z)$, as well as their pointwise confidence intervals in Figure 2. As for the coefficient corresponding to the continuous covariate, $\beta_4(z)$, we notice in Figure 3 that both curves are very close to zero.

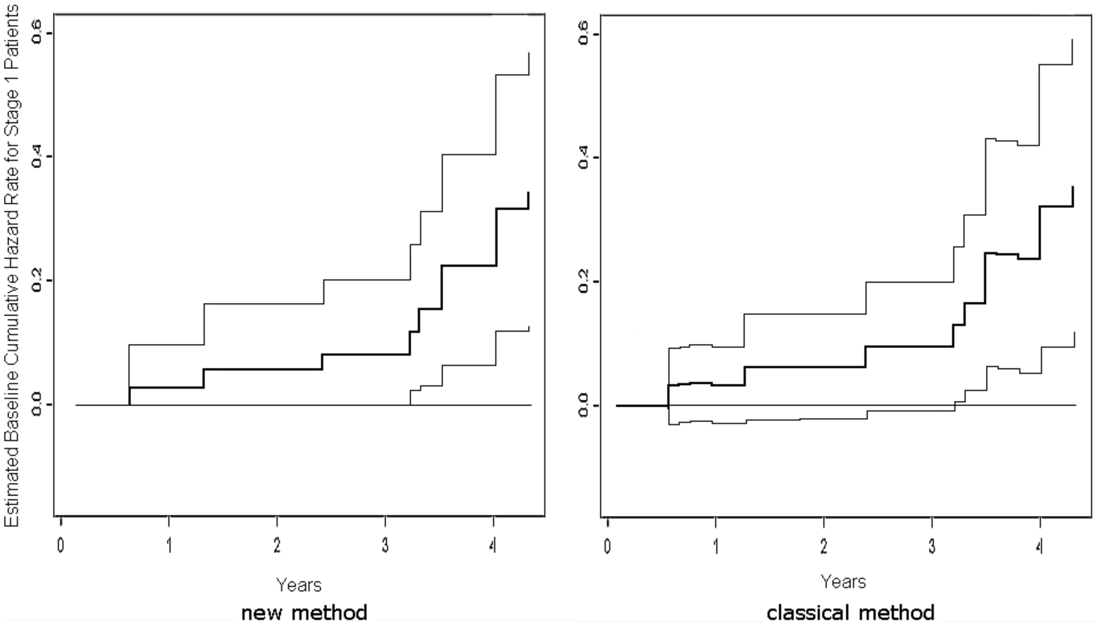

Figure 1: Estimate of the cumulative baseline hazard rate ($\beta_0(z)$) and 95% pointwise confidence intervals.

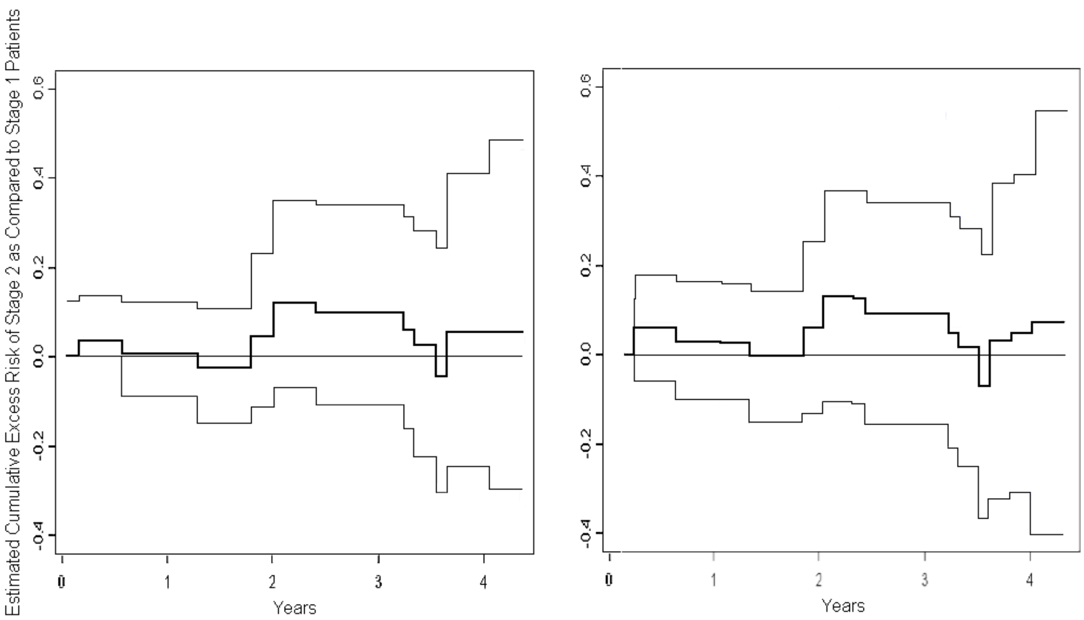

Figure 2: Estimate of the cumulative excess risk of stage 2 cancer as compared to stage 1 cancer ($\beta_1(z)$) and 95% pointwise confidence intervals.

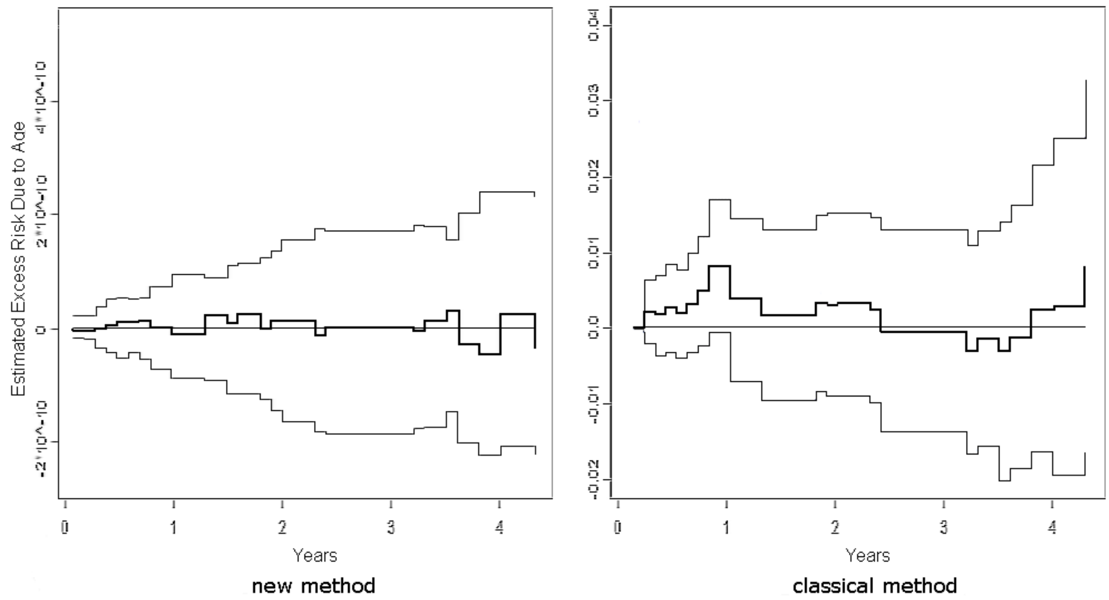

Figure 3: Estimate of the cumulative effect of age ($\beta_4(z)$) and 95% pointwise confidence intervals.

6 Appendix

6.1 Conditions

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

(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 $I_\delta = [x_1 - \delta, x_2 + \delta]$ 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 $w(x) = I_{\{x \in I\}} \tilde{w}(x)$, 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) $\det(\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X}) \neq 0$.

6.2 Proof of Theorem 2.1

From (2.1) and (2.2) we may write

$$\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} (\hat{\boldsymbol{\phi}}(z) - \boldsymbol{\phi}(z)) = \hat{\boldsymbol{A}}^{-1} \hat{\boldsymbol{b}} \tag{6.1}$$

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

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

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, $\hat{\boldsymbol{b}} = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p)^t$, with

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

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

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

and

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

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

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

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

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

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

$$\hat{S}_n(z|X_j) - S(z|X_j) = \sum_{l=1}^{n} B_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z) + R_n(z|X_j), \tag{6.2}$$

where

$$\sup_{\substack{y\in[a,b]\\ x\in I}} |R_n(y|x)| = O_p\left((\frac{\log n}{nh})^{3/4} \right), \tag{6.3}$$

and

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

Observe that $E[\xi(Z, T, \delta, x, y)|X = x] = 0$. We plug (6.2) into $\hat{b}_i^{(1)}$ to obtain:

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

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

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

with

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

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

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

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

For $\hat{b}_i^{(R)}$ we have from (6.3) and using condition (H10) that

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

For $\hat{b}_i^{(112)}$ we use Markov's inequality. Consider

$$E(|\hat{b}_i^{(112)}|) \leq \frac{1}{n} N_3^i N_1 \sum_{j=1}^{n} E[w(X_j) \tilde{B}_{nj}(X_j) E(|\xi(Z_j, T_j, \delta_j, X_j, z)| \ |X_j)].$$

On the other hand,

$$|\xi(Z_j, T_j, \delta_j, X_j, z)| \leq \frac{\delta_j}{C(Z_j|X_j)} + \int_0^\infty \frac{I_{\{T_j \leq u \leq Z_j\}}}{C^2(u|X_j)} dH_1^*(u|X_j),$$

which implies

$$E[|\xi(Z_j, T_j, \delta_j, X_j, z)| \ |X_j] \leq 2 \int_0^\infty \frac{1}{C(u|X_j)} dH_1^*(u|X_j).$$

Consequently (with $\|K\| = \sup_u |K(u)|$),

$$\begin{aligned} E(|\hat{b}_i^{(112)}|) &\leq \frac{2}{n} N_3^i N_1 \sum_{j=1}^{n} E\left[w(X_j) \tilde{B}_{nj}(X_j) \int_0^\infty \frac{dH_1^*(u|X_j)}{C(u|X_j)}\right] \\ &\leq 2\|K\| N_3^i N_1 \frac{1}{nh} E\left[\frac{w(X_1)}{m^*(X_1)} \int_0^\infty \frac{dH_1^*(u|X_1)}{C(u|X_1)}\right] \\ &= 2\|K\| N_3^i N_1 \frac{1}{nh} \int_I \tilde{w}(x) \int_0^\infty \frac{dH_1^*(u|x)}{C(u|x)} dx. \end{aligned}$$

Since the latter integral is bounded, it follows that $E(|\hat{b}_i^{(112)}|) = O((nh)^{-1})$, which, using (H10) implies that $\hat{b}_i^{(112)} = O_p((nh)^{-1}) = o_p(n^{-1/2})$.

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

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

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

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

So far we have proved that

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

We will now prove the asymptotic normality of $\hat{b}_i^{(111)}$. 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,

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

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

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

where

$$D_i^{(1)} = \frac{2}{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k>j}}^{n} h_i^{(1)}(\boldsymbol{V}_j, \boldsymbol{V}_k),$$

$$D_i^{(2)} = \frac{n-1}{n}\sum_{j=1}^{n} h_i^{(2)}(\boldsymbol{V}_j),$$

$$D_i^{(3)} = \frac{n-1}{n}\sum_{k=1}^{n} h_i^{(3)}(\boldsymbol{V}_k),$$

$$D_i^{(4)} = (n-1)E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)],$$

with

$$h_i^{(1)}(\boldsymbol{V}_j, \boldsymbol{V}_k) = \tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k) - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_j] - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_k] + E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)],$$

$$h_i^{(2)}(\boldsymbol{V}_j) = E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_j] - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)],$$

$$h_i^{(3)}(\boldsymbol{V}_k) = E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)|\boldsymbol{V}_k] - E[\tilde{h}_i(\boldsymbol{V}_j, \boldsymbol{V}_k)].$$

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

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

It can be easily proved that

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

with $E[\tilde{h}_i^{\,2}(\boldsymbol{V}_1, \boldsymbol{V}_2)] \le E[h_i^2(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^{-1}n^{-2})$. 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 $\hat{b}_i^{(111)}$. For $D_i^{(2)}$ it is easy to show that $E[D_i^{(2)}] = 0$ and that for any $0 \le i, j \le p$,

$$\mathrm{Cov}(D_i^{(2)}, D_j^{(2)}) = \frac{(n-1)^2}{n}E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)\tilde{h}_j(\boldsymbol{V}_3, \boldsymbol{V}_2)] - \frac{(n-1)^2}{n}E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)]E[\tilde{h}_j(\boldsymbol{V}_1, \boldsymbol{V}_2)].$$

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 (2.3). It now follows from the central limit theorem for triangular arrays that for any $a \in \mathbb{R}^{p+1}$,

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

Direct application of the Cramér-Wold device implies that

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

From this and the fact that $\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = \hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{b}}$ (see (6.1)), 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.

Acknowledgments. 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 PGIDIT03PXIC10505PN.

References

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

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

Cao, R. and González-Manteiga, W. (2003). Goodness-of-fit tests for conditional models under censoring and truncation (unpublished manuscript available at www.udc.es/dep/mate/ricardo/preprints.html).

Iglesias-Pérez, C. and 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. and 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. and 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. and Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York.

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

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

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

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. and Wang, M.C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika*, 74, 883-886.