

I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

0823

**GOODNESS-OF-FIT TEST FOR GENERALIZED
CONDITIONAL LINEAR MODELS UNDER
LEFT TRUNCATION AND RIGHT CENSORING**

TEODORESCU, B. and I. VAN KEILEGOM

This file can be downloaded from
<http://www.stat.ucl.ac.be/ISpub>

Goodness-of-fit Test for Generalized Conditional Linear Models under Left Truncation and Right Censoring

Bianca Teodorescu¹

Ingrid Van Keilegom¹

July 14, 2008

Abstract

Consider a semiparametric time-varying coefficients regression model of the following form: $\phi(S(z|X)) = \boldsymbol{\beta}(z)^t \mathbf{X}$, where ϕ is a known link function, $S(\cdot|X)$ is the survival function of a response Y given a covariate X , $\mathbf{X} = (1, X, X^2, \dots, X^p)$ and $\boldsymbol{\beta}(z) = (\beta_0(z), \dots, \beta_p(z))^t$ is the unknown vector of regression coefficients. This model reduces for special choices of ϕ to e.g. the additive hazards model or the Cox proportional hazards model with time dependent coefficients. The response is subject to left truncation and right censoring. An omnibus goodness-of-fit test is developed to test whether the model fits the data. A bootstrap version, to approximate the critical values of the test, is proposed and proved to work from a practical point of view as well. The test is also applied to real data.

Key words: Additive hazards model, bootstrap, goodness-of-fit, least-squares estimator, proportional hazards model, semiparametric regression, survival analysis, time-varying coefficients.

¹ Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: bianca.teodorescu@uclouvain.be, ingrid.vankeilegom@uclouvain.be.

1 Introduction

In survival analysis, many models exist that account for the relationship between the survival function and a certain number of covariates, e.g. the Cox proportional hazards model, the log-logistic model or the accelerated failure time model. Cao and González-Manteiga (2007) considered a very general model, which includes as special cases the above mentioned models and where the response can also be subject to truncation, not only to censoring as in the previous models. In all these models the influence of the covariates is assumed to be constant over time except for variations in the baseline hazard function. This is an unrealistic assumption in many applications (for instance the prognostic value of a covariate measured at the beginning of a study may decline over the follow-up period), so it is important to develop methods that take into account this fluctuation. In the case of censoring, many authors extended the Cox model to allow for time-dependent coefficients, see e.g. Zucker and Karr (1990), Murphy and Sen (1991), Nan and Lin (2003), Cai and Sun (2003), Lambert and Eilers (2004) and Kauermann (2005), among others. Also, other time-dependent survival models have been considered, like the additive hazards model, see for example Aalen (1980), Huffer and McKeague (1991) or McKeague and Sasieni (1998). Jung (1996) extended the general model in Cao and González-Manteiga (2008) to allow for time-varying coefficients, but only in the case where the observations are censored, the covariates are discrete and the censoring is independent of the covariates. Subramanian (2001) improved Jung's model by relaxing the hypothesis of independence between the censoring time and the covariates and Subramanian (2004) extended the model, to allow for a one-dimensional continuous covariate. Teodorescu et al. (2008) extended this general model, by allowing the response to be subject to right censoring and/or left truncation and they used a least squares procedure instead of the maximum likelihood method used in the above models.

It is of great use to have at hand a method to check the validity of the above mentioned models. When the coefficients are time-independent, some work is available, like Andersen (1982), Arjas (1988), Schoenfeld (1980), Gill and Schumacher (1987), Parzen and Lipsitz (1999) for the Cox model, Yang (1998), Rossini, Wei and Ying (1996) and Chen (2001) for the accelerated failure time model, Akritas (1996) for a general polynomial regression model, Grigoletto and Akritas (1999) with a unifying methodology for the Cox model, the additive risk model and the proportional odds model and Cao and González-Manteiga (2008) for an extension of the work in Akritas (1996) and Grigoletto and Akritas (1999).

When the coefficients are time-dependent, little work is available, see e.g. Marzec and

Marzec (1997) who proposed a goodness-of-fit test for the Cox model with time-dependent coefficients and Kim and Lee (1998), Yuen and Burke (1997), Amir and McKeague (2000) for the additive hazards model. Klein and Moeschberger (1997) proposed a graphical method for assessing the fit of the additive hazards model.

We extend the methodology of Cao and González-Manteiga (2008) to the case when the coefficients are time-dependent. We study this problem in the framework of the generalized linear model presented in Teodorescu et al. (2008). We develop formal tests of hypothesis, where previously only ad-hoc graphical methods were available. We show that our approach works well on simulated data and apply it to data from a study on larynx cancer.

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 $(Z_i, T_i, \delta_i, X_i)$, $i = 1, \dots, n$ 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 like to test whether this relationship is of polynomial type, via a known monotone transformation $\phi : [0, 1] \rightarrow \mathbb{R}$ of the survival function, i.e.:

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \dots + \beta_p(z)X^p, \quad (1.1)$$

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 the proportional hazards model.

The appropriateness of the parametric modelling of regression data may be judged by comparison with a semi-parametric estimator of the response. For this purpose one may use a squared deviation measure between the two fits. The sum of the squared deviation over all the values of the covariates may be used as a test statistic for testing the parametric model where the critical value is determined from the asymptotic distribution of this statistic. The convergence to the distribution may be slow, so a bootstrap method is proposed in order to estimate the critical values.

The paper is organized as follows. In the next section we introduce the test statistic and its asymptotic distribution, while in Section 3 we present a bootstrap based method for the approximation of the critical values of the test. Section 4 shows some numerical

results, while the analysis of data on cancer of the larynx is conducted in Section 5. Finally, Section 6 contains the proofs.

2 The test statistic

Let us first introduce the following notations: $M(x) = P(X \leq x)$, $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $L(y|x) = P(T \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H_1(y|x) = P(Z \leq y, \delta = 1|x)$, $L(y) = P(T \leq y)$, $H(y) = P(Z \leq y)$, $H_1(y) = P(Z \leq y, \delta = 1)$, $C(y|x) = P(T \leq y \leq Z|x)$, 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 $W(t) = P(\eta \leq t)$, 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 $W^*(t) = P(\eta \leq t|T \leq Z)$. Finally, let m denote the density of X and m^* the density of X conditionally on $T \leq Z$.

If model (1.1) holds, then the coefficients $\boldsymbol{\beta}(z)$ can be estimated via a weighted least squares procedure presented in Teodorescu et al. (2008). More precisely,

$$\hat{\boldsymbol{\beta}}(z) = (\hat{\beta}_0(z), \hat{\beta}_1(z), \dots, \hat{\beta}_p(z))^t = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \hat{\boldsymbol{\phi}}(z), \quad (2.1)$$

where

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

$\mathbf{W} = \text{diag}(w(X_1), \dots, w(X_n))$ is a trimmed function defined in terms of a proper weight function \tilde{w} , as precised in condition (H11) in the Appendix and $\hat{S}(z|x)$ is the estimator of the conditional distribution, proposed by Iglesias-Pérez and González-Manteiga (1999):

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

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), $h = h_n \rightarrow 0$ a bandwidth sequence, and $C_n(u|x) = \sum_{j=1}^n 1_{\{T_j \leq u \leq Z_j\}} B_{nj}(x)$. See Teodorescu et al. (2008) for further details and asymptotic properties.

Now we want to check the appropriateness of the semi-parametric model (1.1). For a given ϕ and p we would like to test

H_0 : for all z there exists a vector $\beta(z) \in \mathbb{R}^{p+1}$ such that (1.1) holds,

against H_a : there exists a z such that (1.1) does not hold for any $\beta(z) \in \mathbb{R}^{p+1}$.

A natural way to proceed is to measure the distance between $\hat{\phi}(z)$ and the hypothesized model (1.1) and to use this distance as test statistic. Here we study a kind of L_2 - distance between these two:

$$\hat{\Phi}_n(\hat{\beta}(z)) = \frac{1}{n} \sum_{r=1}^n \left(\phi(\hat{S}(z|X_r)) - (\hat{\beta}_0(z) + \hat{\beta}_1(z)X_r + \dots + \hat{\beta}_p(z)X_r^p) \right)^2$$

Hence, it is reasonable to reject H_0 when $\hat{\Phi}_n(\hat{\beta}(z))$ is large. We measure this distance for all values of z lying in an interval $[a, b]$, where a and b are defined in condition (H2)(d) in the Appendix, and define :

$$T_n = \int_a^b \hat{\Phi}_n(\hat{\beta}(z)) dz \quad (2.2)$$

We should also multiply this quantity by a normalizing sequence in order to have a limiting distribution. This leads to the following test statistic: $nh^{1/2}T_n$. In order to obtain the limiting distribution of this statistic under the null hypothesis, some conditions are to be imposed. The conditions needed, (H1)-(H12), are collected in the Appendix.

Let us state our main result:

Theorem 2.1 *Suppose that conditions (H1) through (H12) hold. Then, under H_0 ,*

$$nh^{1/2}T_n - b_{0h} \xrightarrow{d} N(0, V),$$

$$\text{where } b_{0h} = h^{-1/2}K^{(2)}(0) \int_a^b \int_x g(z, z, x) dx dz,$$

$$V = 2K^{(4)}(0) \int_a^b \int_a^b \int_x g^2(z_1, z_2, x) dx dz_1 dz_2, \quad (2.3)$$

$$g(z_1, z_2, x) = \phi'(S(z_1|x))\phi'(S(z_2|x))S(z_1|x)S(z_2|x) \int_a^{z_1 \wedge z_2} \frac{dH_1^*(u|x)}{C^2(u|x)}. \quad (2.4)$$

$K^{(4)}$ is the convolution of $K^{(2)}$, and $K^{(2)}$ is the convolution of K , that is $K^{(2)}(u) = \int K(v)K(u+v) dv$.

Remark 1 In a similar way we can obtain the limiting distribution of the test statistic $nh^{1/2}T_n$ 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)).

3 Bootstrap version

The convergence of the distribution of the test statistic $nh^{1/2}T_n$ to a normal distribution is quite slow, so that it seems more appropriate not to use the asymptotic critical values in practice. We therefore compute the critical values based on a bootstrap method. The procedure is as follows:

1. Choose a bandwidth h in the interval $(0, \mu(\text{supp}(X)))$ and a pilot bandwidth g (larger than h), where μ is the Lebesgue measure.

2. For all $z \in \{Z_1, \dots, Z_n\}$ and $x \in \{X_1, \dots, X_n\}$:

- a) Estimate $S(z|x)$, $G(z|x)$ and $L(z|x)$ by $\hat{S}_g(z|x)$, $\hat{G}_g(z|x)$ and $\hat{L}_g(z|x)$, respectively, where

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

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

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

$$\tilde{S}_g(z|x) = \phi^{-1}(\hat{\beta}_{0,g}(z) + \hat{\beta}_{1,g}(z)x + \dots + \hat{\beta}_{p,g}(z)x^p). \quad (3.1)$$

3. For $b = 1, \dots, B$:

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

- b) Use this resample $\{(T_1^*, Z_1^*, \delta_1^*, X_1), \dots, (T_n^*, Z_n^*, \delta_n^*, X_n)\}$ to estimate a bootstrap version of the conditional survival function, $\hat{S}_h^*(z|X_i)$ ($i = 1, \dots, n$) using the bandwidth h . This bootstrap version is used to obtain the bootstrap vector of coefficients $\hat{\beta}_h^{*(b)} = \left(\hat{\beta}_{0,h}^{*(b)}(z), \dots, \hat{\beta}_{p,h}^{*(b)}(z)\right)$ using the least squares estimator, and to obtain the bootstrap version $\hat{\Phi}_n^* \left(\hat{\beta}_h^{*(b)}(z)\right)$ of $\hat{\Phi}_n \left(\hat{\beta}(z)\right)$.
- c) Compute the bootstrap version of the test statistic $nh^{1/2}T_n$, which is given by:

$$nh^{1/2}T_{n,b}^* = nh^{1/2} \int_a^b \hat{\Phi}_n^* \left(\hat{\beta}_h^{*(b)}(z)\right) dz.$$

4. Order the obtained test statistics and take $nh^{1/2}T_{n,[(1-\alpha)B]}^*$ which approximates the $(1 - \alpha)$ -quantile of the distribution of $nh^{1/2}T_n$ under H_0 .
5. If $nh^{1/2}T_n > nh^{1/2}T_{n,[(1-\alpha)B]}^*$, then reject H_0 , otherwise do not reject H_0 .

Remark 2 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 ($i = 1, \dots, n$) 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.

Remark 3 Note that the estimator $\tilde{S}(z|x)$ of the conditional survival function in (3.1) is in general non-monotone. A convenient and satisfactory solution is to keep the estimator constant until it starts decreasing again.

4 Numerical results

In this section, we study the finite sample properties of the proposed test. We will first deal with the case of a one-dimensional continuous covariate, under censoring. Next, we will study the performance of the test when truncation is also present.

Along the simulations, the following two models are considered:

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x + \beta_2(z) \sin\left(\frac{\pi x}{2}\right) \quad (4.1)$$

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x + \beta_2(z) \exp\left(\frac{2x}{3}\right). \quad (4.2)$$

For model (4.1), $X \sim U[4, 10]$, $Y|_{X=x} \sim \text{Exp}(4x + a_1 \sin(\frac{\pi x}{2}))$, $C|_{X=x} \sim \text{Exp}(d_1 x)$, where $d_1 > 0$ determines the censoring probability, $T|_{X=x} \sim \text{Exp}(r_1 x)$, where $r_1 > 0$ controls the probability of truncation and $\phi(u) = -\log(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = 4zx + a_1 \sin(\frac{\pi x}{2})$. The sample size is taken $n = 100$, $M = 1000$ Monte Carlo simulations are conducted, $B = 500$ bootstrap resamples are being generated, d_1 is taken in order to give 20% and 40% of censoring, respectively, r_1 is taken in order to give 10% and 20% of truncation, respectively. Since we have a one-dimensional continuous covariate, a bandwidth, h , is needed in order to estimate $S(z|x)$. We worked with $h \in \{1.2, 1.5, 1.8, 2.1, 2.4\}$ and the pilot bandwidth in the bootstrap procedure was taken to be $g = 2h$. This is a reasonable choice since g has to be asymptotically larger than h^2 (see Cao and González-Manteiga (2008) for more details and Härdle and Mammen (1993) for some insight about possible choices of g in the regression case). 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 $\tilde{w}(x) = 1_{\{4.35 \leq x \leq 9.65\}}$ has been chosen in order to avoid boundary problems. The level of the test was taken to be $\alpha = 0.05$, H_0 is model (4.1) for $a_1 = 0$, while as alternatives to H_0 we have considered model (4.1) for $a_1 \in \{4, 8, 12, 16, 20\}$. As a_1 becomes larger, the departure from H_0 also increases.

Table 1 displays the results for model (4.1) under censoring. We see that, as expected, the results for 20% censoring are better than those for 40% censoring and that as a_1 increases (we go further away from H_0), the power of the test also increases, to get to 0.945 when $a_1 = 20$ under 20% censoring. We also notice that the choice of h has an influence on the results. Under H_1 there are some cases where the power of the test varies with 0.2 (see $a_1 = 20$, 40% censoring), while generally we have a variation of 0.1 or less. Under H_0 we have smaller fluctuations (maximum of 0.05) and we see that usually the power is either underestimated or overestimated.

Table 2 contains the results for model (4.1) under both censoring and truncation. The results are similar to those of Table 1. We notice that the results for 20% censoring are better than those for 40% censoring and also that 10% truncation displays better results than 20% truncation.

Cens perc	h	H_0	H_1				
		$a_1 = 0$	$a_1 = 4$	$a_1 = 8$	$a_1 = 12$	$a_1 = 16$	$a_1 = 20$
20	1.2	0.031	0.077	0.200	0.445	0.752	0.931
	1.5	0.029	0.114	0.261	0.531	0.788	0.945
	1.8	0.039	0.113	0.305	0.534	0.820	0.945
	2.1	0.075	0.122	0.315	0.525	0.801	0.943
	2.4	0.080	0.140	0.275	0.422	0.650	0.870
40	1.2	0.024	0.051	0.107	0.225	0.420	0.660
	1.5	0.021	0.050	0.151	0.294	0.519	0.762
	1.8	0.038	0.065	0.180	0.325	0.535	0.757
	2.1	0.051	0.082	0.157	0.296	0.506	0.701
	2.4	0.068	0.106	0.134	0.231	0.376	0.562

Table 1: Power of the test for model (4.1) under censoring.

Cens perc	Trunc perc	h	H_0	H_1				
			$a_1 = 0$	$a_1 = 4$	$a_1 = 8$	$a_1 = 12$	$a_1 = 16$	$a_1 = 20$
20	10	1.2	0.051	0.067	0.191	0.385	0.662	0.893
		1.5	0.056	0.104	0.238	0.432	0.733	0.939
		1.8	0.057	0.107	0.227	0.445	0.768	0.960
		2.1	0.064	0.116	0.244	0.441	0.804	0.923
		2.4	0.068	0.139	0.269	0.395	0.693	0.857
	20	1.2	0.061	0.076	0.167	0.373	0.651	0.872
		1.5	0.026	0.092	0.199	0.421	0.702	0.931
		1.8	0.037	0.071	0.233	0.433	0.794	0.901
		2.1	0.036	0.111	0.269	0.415	0.692	0.912
		2.4	0.038	0.091	0.270	0.372	0.833	0.842
40	10	1.2	0.035	0.059	0.081	0.213	0.398	0.676
		1.5	0.047	0.063	0.141	0.286	0.461	0.691
		1.8	0.023	0.054	0.157	0.316	0.536	0.734
		2.1	0.024	0.073	0.156	0.278	0.497	0.706
		2.4	0.025	0.095	0.143	0.234	0.366	0.535
	20	1.2	0.041	0.060	0.090	0.204	0.412	0.645
		1.5	0.046	0.059	0.135	0.273	0.458	0.692
		1.8	0.031	0.053	0.141	0.312	0.556	0.725
		2.1	0.028	0.071	0.122	0.226	0.492	0.699
		2.4	0.032	0.092	0.127	0.330	0.359	0.522

Table 2: Power of the test for model (4.1) under censoring and truncation.

For model (4.2), $X \sim U[4, 10]$, $Y|_{X=x} \sim \text{Exp}(x + a_1 \exp(\frac{2x}{3}))$, $C|_{X=x} \sim \text{Exp}(d_2x)$, where $d_2 > 0$ determines the censoring probability, $T|_{X=x} \sim \text{Exp}(r_2x)$, where $r_2 > 0$ controls the probability of truncation and $\phi(u) = -\log(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = zx + a_2 \exp(\frac{2x}{3})$. The sample size is taken

$n = 100$, $M = 1000$ Monte Carlo simulations are conducted, $B = 500$ bootstrap resamples are being generated, d_2 is taken in order to give 20% and 40% of censoring, respectively, r_2 is taken in order to give 10% and 20% of truncation, respectively. We worked with $h \in \{1.2, 1.5, 1.8, 2.1, 2.4\}$, $g = 2h$, the Nadaraya-Watson weights are computed as before, based on the Epanechnikov kernel and the weight function $\tilde{w}(x)$ is the same as for model (4.1). The level of the test was taken to be $\alpha = 0.05$, H_0 is model (4.2) for $a_2 = 0$, while the alternative, H_1 , is model (4.2) for values of a_2 in $\{0.01, 0.1, 1, 10, 100\}$. As a_2 increases, we go further away from H_0 .

Cens perc	h	H_0	H_1				
		$a_2 = 0$	$a_2 = 0.01$	$a_2 = 0.1$	$a_2 = 1$	$a_2 = 10$	$a_2 = 100$
20	1.2	0.025	0.088	0.412	0.594	0.630	0.624
	1.5	0.024	0.089	0.538	0.770	0.811	0.821
	1.8	0.036	0.140	0.683	0.914	0.934	0.927
	2.1	0.065	0.172	0.821	0.961	0.970	0.966
	2.4	0.077	0.240	0.852	0.970	0.978	0.978
40	1.2	0.020	0.050	0.390	0.588	0.645	0.602
	1.5	0.025	0.073	0.490	0.748	0.780	0.771
	1.8	0.037	0.092	0.607	0.861	0.935	0.896
	2.1	0.054	0.146	0.739	0.919	0.957	0.951
	2.4	0.068	0.169	0.817	0.953	0.896	0.946

Table 3: Power of the test for model (4.2) under censoring.

Table 3 shows the results for model (4.2) under censoring. We remark more or less the same features as for model (4.1) under censoring.

Cens perc	Trunc perc	h	H_0	H_1				
			$a_2 = 0$	$a_2 = 0.01$	$a_2 = 0.1$	$a_2 = 1$	$a_2 = 10$	$a_2 = 100$
20	10	1.2	0.056	0.104	0.417	0.621	0.662	0.644
		1.5	0.062	0.126	0.495	0.723	0.749	0.780
		1.8	0.043	0.157	0.580	0.829	0.862	0.829
		2.1	0.074	0.153	0.653	0.859	0.890	0.893
		2.4	0.080	0.175	0.733	0.908	0.931	0.928
	20	1.2	0.054	0.091	0.456	0.595	0.637	0.691
		1.5	0.069	0.088	0.528	0.702	0.736	0.759
		1.8	0.078	0.145	0.615	0.789	0.845	0.860
		2.1	0.080	0.143	0.695	0.835	0.885	0.885
		2.4	0.076	0.169	0.721	0.886	0.900	0.915
40	10	1.2	0.043	0.074	0.378	0.598	0.603	0.601
		1.5	0.041	0.097	0.442	0.669	0.702	0.719
		1.8	0.055	0.118	0.525	0.732	0.780	0.814
		2.1	0.069	0.129	0.568	0.821	0.844	0.854
		2.4	0.068	0.155	0.637	0.896	0.888	0.906
	20	1.2	0.061	0.088	0.405	0.578	0.607	0.622
		1.5	0.059	0.123	0.470	0.682	0.709	0.712
		1.8	0.075	0.140	0.550	0.765	0.777	0.800
		2.1	0.062	0.144	0.573	0.792	0.829	0.848
		2.4	0.098	0.153	0.627	0.794	0.870	0.871

Table 4: Power of the test for model (4.2) under censoring and truncation.

Table 4 displays the results for model (4.2) under both censoring and truncation and they look similar to those of Table 2, for model (4.1) under censoring and truncation: the power of the test decreases with the augmentation of the censoring and the truncation percentage. We also notice that there is little difference between the power of the test

for H_1 with $a_2 = 10$ and $a_2 = 100$. That is because the multiplication factor for a_2 is exponential, and that as a_2 becomes larger, the models (curves) become similar, in the sense that while comparing them with a line, the difference is almost the same.

5 Data analysis

The methods presented in the previous sections have been applied to data on larynx cancer previously studied by Klein and 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 ($i = 1, \dots, 90$) 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 and Moeschberger (1997) is the additive hazards model, which can be written in the following form:

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

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).

They verified if the assumptions for the additive hazards model hold by two types of diagnostic plots suggested by Aalen (1993) (see Chapter 11.7 in Klein and Moeschberger (1997) for more details). Both plots showed no indication of incorrect modeling.

By using our proposed method, we will test

$$H_0 : \text{for all } z \text{ there exists a vector } \boldsymbol{\beta}(z) \in \mathbb{R}^5 \text{ such that (5.1) holds,}$$

against H_a : there exists a z such that (5.1) does not hold for any $\boldsymbol{\beta}(z) \in \mathbb{R}^5$.

Under H_0 , Teodorescu et al. (2008) calculated the estimator (2.1) of the coefficients $\beta_i(z)$ ($i = 0, \dots, 4$) and selected the optimal bandwidth $h = 25$ by means of a bootstrap method (see Teodorescu et al. (2008) for more details on how to select the bandwidth based on the bootstrap). In order to test H_0 against H_a , we took $h = 25$, $g = 1.5h$, $\alpha = 0.05$ and we conducted 10000 bootstrap simulations. The estimated p-value was found to be 0.1918. Hence, we do not reject H_0 , i.e. we conclude that the additive regression model is appropriate for these data, which agrees with the graphical tests carried out by Klein and Moeschberger (1997).

6 Appendix

6.1 Conditions

We now state the conditions mentioned in Theorem 2.1. Conditions (H1)–(H6) below are taken from Iglesias-Pérez and González-Manteiga (1999), on which our proof is based.

Condition (H2) comes from Dabrowska (1989) and is needed in order to 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 $I_\delta = [x_1 - \delta, x_2 + \delta]$ 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 $m(x)$, $m^*(x)$ and $\alpha(x)$ exist and are continuous in I_δ and $m^*(x)$ has bounded second derivative.

(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 thrice continuously differentiable and its first, second and third derivatives are bounded by $N_1 < \infty$, $N_2 < \infty$ and $N_3 < \infty$, respectively.

(H8) There exists some $N_4 < \infty$ such that $P(|X| \leq N_4) = 1$.

(H9) The matrix $\mathbf{A} = (a_{ij})_{i,j=0}^p$, with $a_{ij} = E(X^{i+j}w(X))$, is nonsingular.

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

(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 $B < \infty$.

(H12) $g(z_1, z_2, x)$ defined in (2.4) has bounded second derivative with respect to x .

6.2 Proof of Theorem 2.1

Using (2.2), we can write

$$T_n = T_{n,1} + R_{n,1} + R_{n,2}, \quad (6.1)$$

with

$$T_{n,1} = \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r)) \right]^2 dz$$

$$R_{n,1} = \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \dots + \hat{\beta}_p(z)X_r^p) \right]^2 dz \quad (6.2)$$

$$R_{n,2} = \int_a^b \frac{2}{n} \sum_{r=1}^n \left[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r)) \right] \left[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \dots + \hat{\beta}_p(z)X_r^p) \right] dz \quad (6.3)$$

where $R_{n,1}$ and $R_{n,2}$ are $o_p(n^{-1}h^{-1/2})$ by Lemma 6.6 below. The leading term $T_{n,1}$ of equation (6.1) can be easily handled using a Taylor expansion of ϕ around $S(z|X_r)$:

$$T_{n,1} = \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\phi'(S(z|X_r))(\hat{S}(z|X_r) - S(z|X_r)) + \frac{1}{2} \phi''(\Delta_r(z))(\hat{S}(z|X_r) - S(z|X_r))^2 \right]^2 dz$$

$$= T_{n,2} + R_{n,3} + R_{n,4} \quad (6.4)$$

with $\Delta_r(z)$ in between $S(z|X_r)$ and $\hat{S}(z|X_r)$,

$$T_{n,2} = \int_a^b \frac{1}{n} \sum_{r=1}^n \phi'(S(z|X_r))^2 (\hat{S}(z|X_r) - S(z|X_r))^2 dz$$

$$R_{n,3} = \int_a^b \frac{1}{n} \sum_{r=1}^n \frac{1}{4} \phi''(\Delta_r(z))^2 (\hat{S}(z|X_r) - S(z|X_r))^4 dz = o_p(n^{-1}h^{-1/2}) \quad (6.5)$$

$$R_{n,4} = \int_a^b \frac{1}{n} \sum_{r=1}^n \phi'(S(z|X_r)) \phi''(\Delta_r(z)) (\hat{S}(z|X_r) - S(z|X_r))^3 dz = o_p(n^{-1}h^{-1/2}), \quad (6.6)$$

where the order of $R_{n,3}$ and $R_{n,4}$ follows from the fact that $\sup_{x \in I} \sup_{a \leq z \leq b} |\hat{S}(z|x) - S(z|x)| = O_p((\ln n)^{1/2} (nh)^{-1/2})$ (see Lemma 5 in Iglésias-Pérez and Gonzalez-Manteiga (1999)).

By the iid representation for $\hat{S}(z|X_r)$, given in Iglésias-Pérez and Gonzalez-Manteiga (1999),

$$T_{n,2} = \int_a^b \frac{1}{n} \sum_{r=1}^n \phi'(S(z|X_r))^2 \left[S(z|X_r) \sum_{i=1}^n B_{ni}(X_r) \xi(Z_i, T_i, \delta_i, X_r, z) + R'_n(z|X_r) \right]^2 dz$$

$$= T_{n,3} + R_{n,5} + R_{n,6}, \quad (6.7)$$

where

$$\xi(Z, T, \delta, x, z) = \frac{1_{\{Z \leq z, \delta=1\}}}{C(Z|x)} - \int_0^z \frac{1_{\{T \leq u \leq Z\}}}{C^2(u|x)} dH_1^*(u|x) \quad (6.8)$$

and

$$\sup_{\substack{z \in [a,b] \\ x \in I}} |R'_n(z|x)| = O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right). \quad (6.9)$$

$$T_{n,3} = \int_a^b \frac{1}{n} \sum_{r=1}^n \phi'(S(z|X_r))^2 S(z|X_r)^2 \left[\sum_{i=1}^n B_{ni}(X_r) \xi(Z_i, T_i, \delta_i, X_r, z) \right]^2 dz$$

$$R_{n,5} = o_p(n^{-1}h^{-1/2})$$

$$R_{n,6} = O_p\left(\left(T_{n,3}\right)^{1/2} \cdot \left(\frac{\ln n}{nh}\right)^{3/4}\right). \quad (6.10)$$

By Lemma 6.5 below, we get that $R_{n,6} = o_p(n^{-1}h^{-1/2})$.

The term $T_{n,3}$ can be decomposed in the following way:

$$\begin{aligned}
T_{n,3} &= \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) + \Delta_r^{(7)}(z) \right]^2 dz \\
&= \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \right]^2 dz + \int_a^b \frac{1}{n} \sum_{r=1}^n [\Delta_r^{(7)}(z)]^2 dz \\
&\quad + \int_a^b \frac{1}{n} \sum_{r=1}^n \sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \Delta_r^{(7)}(z) dz = T_{n,4} + R_{n,7} + R_{n,8},
\end{aligned} \tag{6.11}$$

with

$$\begin{aligned}
\epsilon_i(z) &= \phi'(S(z|X_i)) S(z|X_i) \xi(Z_i, T_i, \delta_i, X_i, z) \\
\Delta_r^{(7)}(z) &= \sum_{i=1}^n B_{ni}(X_r) \left(\phi'(S(z|X_r)) S(z|X_r) \xi(Z_i, T_i, \delta_i, X_r, z) - \right. \\
&\quad \left. - \phi'(S(z|X_i)) S(z|X_i) \xi(Z_i, T_i, \delta_i, X_i, z) \right).
\end{aligned}$$

The terms $R_{n,7}$ and $R_{n,8}$ are $o_p(n^{-1}h^{-1/2})$ by Lemma 6.4 below.

By defining $b_{nij} = n^{-1} \sum_{r=1}^n B_{ni}(X_r) B_{nj}(X_r)$, $T_{n,4}$ can be decomposed into two new terms:

$$T_{n,4} = \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \right]^2 dz = T_{n,5} + R_{n,9} \tag{6.12}$$

with $T_{n,5} = \int_a^b 2 \sum_{i < j} b_{nij} \epsilon_i(z) \epsilon_j(z) dz$ and

$$R_{n,9} = \int_a^b \sum_{i=1}^n b_{nii} \epsilon_i^2(z) dz. \tag{6.13}$$

In Lemma 6.3 it is shown that

$$R_{n,9} = \frac{1}{nh} \left(\int_a^b \int_x g(z, z, x) dx dz \right) K^{(2)}(0) + o_p(n^{-1}h^{-1/2}). \tag{6.14}$$

To analyse $T_{n,5}$, define $\tilde{b}_{nij} = n^{-1} \sum_{r=1}^n \tilde{B}_{ni}(X_r) \tilde{B}_{nj}(X_r)$ with $\tilde{B}_{ni}(X_r) = (nh)^{-1} K\left(\frac{X_r - X_i}{h}\right) m^*(X_r)^{-1}$. Then,

$$B_{ni}(X_r) = \frac{K\left(\frac{X_r - X_i}{h}\right)}{nh} \cdot \frac{1}{\hat{m}^*(X_r)} = \tilde{B}_{ni}(X_r) + B_{ni}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)},$$

where $\hat{m}^*(x)$ is the Parzen-Rosenblatt estimator of $m(x)$:

$$\hat{m}^*(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

which implies that

$$\begin{aligned} b_{nij} &= \tilde{b}_{nij} + \frac{1}{n} \sum_{r=1}^n \tilde{B}_{ni}(X_r) B_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \\ &\quad + \frac{1}{n} \sum_{r=1}^n B_{ni}(X_r) \tilde{B}_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \\ &\quad + \frac{1}{n} \sum_{r=1}^n B_{ni}(X_r) B_{nj}(X_r) \frac{(m^*(X_r) - \hat{m}^*(X_r))^2}{m^*(X_r)^2}, \end{aligned}$$

and this leads to $T_{n,5} = T_{n,6} + R_{n,10} + R_{n,11} + R_{n,12}$, with

$$T_{n,6} = \int_a^b 2 \sum_{i < j} \tilde{b}_{nij} \epsilon_i(z) \epsilon_j(z) dz \quad (6.15)$$

$$R_{n,10} = \int_a^b \frac{2}{n} \sum_{i < j} \sum_{r=1}^n \tilde{B}_{ni}(X_r) B_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \epsilon_i(z) \epsilon_j(z) dz \quad (6.16)$$

$$R_{n,11} = \int_a^b \frac{2}{n} \sum_{i < j} \sum_{r=1}^n B_{ni}(X_r) \tilde{B}_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \epsilon_i(z) \epsilon_j(z) dz \quad (6.17)$$

$$R_{n,12} = \int_a^b \frac{2}{n} \sum_{i < j} \sum_{r=1}^n B_{ni}(X_r) B_{nj}(X_r) \frac{[m^*(X_r) - \hat{m}^*(X_r)]^2}{m^*(X_r)^2} \epsilon_i(z) \epsilon_j(z) dz. \quad (6.18)$$

From Lemma 6.2 we have that $R_{n,10}$, $R_{n,11}$ and $R_{n,12}$ are $o_p(n^{-1}h^{-1/2})$, while from Lemma 6.1 we get that

$$\sqrt{n^2 h} T_{n,6} \xrightarrow{d} N(0, V).$$

The result now follows, since we have shown that

$$T_n = T_{n,6} + \sum_{k=1}^{12} R_{n,k},$$

where all $R_{n,k}$ ($k = 1, \dots, 12$) are $o_p(n^{-1}h^{-1/2})$, except $R_{n,9}$, whose asymptotic expression is given in (6.14).

6.3 Lemmas and proofs

For more details on the proofs of the following lemmas, see the PhD thesis of Teodorescu (2008).

Lemma 6.1

$$\sqrt{n^2 h} T_{n,6} \xrightarrow{d} N(0, V),$$

with $T_{n,6}$ and V defined in (6.15) and (2.3), respectively.

Proof: Write

$$\begin{aligned} \text{Var}(T_{n,6}) &= \text{Var} \left(2 \sum_{i < j} \tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz \right) \\ &= 4 \sum_{i < j} \sum_{k < l} \text{Cov} \left(\tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz, \tilde{b}_{nkl} \int_a^b \epsilon_k(z) \epsilon_l(z) dz \right). \end{aligned}$$

Using that $i < j, k < l$ and $E(\epsilon_i(z)|X_i) = 0$, we obtain (where $\tilde{\mathbf{X}} = (X_1, \dots, X_n)$)

$$\begin{aligned} &\text{Cov} \left(\tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz, \tilde{b}_{nkl} \int_a^b \epsilon_k(z) \epsilon_l(z) dz \right) \\ &= E \left(\text{Cov} \left(\tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz, \tilde{b}_{nkl} \int_a^b \epsilon_k(z) \epsilon_l(z) dz \mid \tilde{\mathbf{X}} \right) \right) + \\ &\quad + \text{Cov} \left(E \left(\tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz \mid \tilde{\mathbf{X}} \right), E \left(\tilde{b}_{nkl} \int_a^b \epsilon_k(z) \epsilon_l(z) dz \mid \tilde{\mathbf{X}} \right) \right) \\ &= E \left(\tilde{b}_{nij} \tilde{b}_{nkl} \text{Cov} \left(\int_a^b \epsilon_i(z) \epsilon_j(z) dz, \int_a^b \epsilon_k(z) \epsilon_l(z) dz \mid \tilde{\mathbf{X}} \right) \right) \\ &= \begin{cases} E \left(\tilde{b}_{nij}^2 \text{Var} \left(\int_a^b \epsilon_i(z) \epsilon_j(z) dz \mid X_i, X_j \right) \right) = E \left(\tilde{b}_{n12}^2 g_2(X_1, X_2) \right), & \text{if } i = k < j = l \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where

$$g_2(X_1, X_2) = \int_a^b \int_a^b g(z_1, z_2, X_1) g(z_1, z_2, X_2) dz_1 dz_2,$$

with $g(z_1, z_2, x)$ defined in (2.4). Hence,

$$\begin{aligned} \text{Var}(T_{n,6}) &= 4 \cdot \frac{n(n-1)}{2} E \left(\tilde{b}_{n12}^2 g_2(X_1, X_2) \right) \\ &= \frac{2(n-1)}{n(nh)^4} \left[(n-2)(n-3) E \left(\frac{K \left(\frac{X_3-X_1}{h} \right) K \left(\frac{X_3-X_2}{h} \right)}{m^*(X_3)^2} \cdot \frac{K \left(\frac{X_4-X_1}{h} \right) K \left(\frac{X_4-X_2}{h} \right)}{m^*(X_4)^2} g_2(X_1, X_2) \right) \right. \\ &\quad + (n-2) E \left(\frac{K \left(\frac{X_3-X_1}{h} \right)^2 K \left(\frac{X_3-X_2}{h} \right)^2}{m^*(X_3)^4} g_2(X_1, X_2) \right) \\ &\quad \left. + 4(n-2) E \left(\frac{K(0) K \left(\frac{X_1-X_2}{h} \right)}{m^*(X_1)^2} \cdot \frac{K \left(\frac{X_4-X_1}{h} \right) K \left(\frac{X_4-X_2}{h} \right)}{m^*(X_4)^2} g_2(X_1, X_2) \right) \right] \end{aligned}$$

$$\begin{aligned} & +2E \left(\frac{K(0)^2 K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^2 m^*(X_2)^2} g_2(X_1, X_2) \right) \\ & + E \left(\frac{K(0)^2 K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^4} g_2(X_1, X_2) \right) \end{aligned} \Bigg]$$

Standard calculations lead to

$$\begin{aligned} E \left(\frac{K\left(\frac{X_3-X_1}{h}\right) K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)^2} \cdot \frac{K\left(\frac{X_4-X_1}{h}\right) K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)^2} g_2(X_1, X_2) \right) \\ = h^3 \int_v (K * K(v))^2 dv \int_a^b \int_a^b \int_x g(z_1, z_2, x)^2 dx dz_1 dz_2 + O(h^4) \end{aligned}$$

$$E \left(\frac{K\left(\frac{X_3-X_1}{h}\right) K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)^4} g_2(X_1, X_2) \right) = O(h^2)$$

$$E \left(\frac{K(0) K\left(\frac{X_1-X_2}{h}\right) K\left(\frac{X_4-X_1}{h}\right) K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_1)^2 m^*(X_1)^2} g_2(X_1, X_2) \right) = O(h^2)$$

$$E \left(\frac{K(0)^2 K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^2 m^*(X_2)^2} g_2(X_1, X_2) \right) = O(h)$$

$$E \left(\frac{K(0)^2 K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^4} g_2(X_1, X_2) \right) = O(h)$$

So, by using (H10), we get

$$\begin{aligned} \text{Var}(T_{n,6}) &= \frac{2(n-1)}{n(nh)^4} \left[(n-2)(n-3) \left[h^3 \int_v K * K(v)^2 dv \int_a^b \int_a^b \int_x g_1^2(z_1, z_2, x) dx dz_1 dz_2 \right. \right. \\ & \quad \left. \left. + O(h^4) \right] + O(n^2 h^4) + O(nh^2) + O(h) \right] \\ &= \frac{2(n-1)(n-2)(n-3)}{n^5 h} \left(K^{(4)}(0) \int_a^b \int_a^b \int_x g_1^2(z_1, z_2, x) dx dz_1 dz_2 \right) \\ & \quad + O(n^{-2}) + O(n^{-3} h^{-2}) + O(n^{-4} h^{-3}) + O(n^{-4} h^{-4}) \\ &= \frac{2}{n^2 h} K^{(4)}(0) \int_a^b \int_a^b \int_x g_1^2(z_1, z_2, x) dx dz_1 dz_2 + o(n^{-2} h^{-1}) \end{aligned}$$

In order to prove the asymptotic convergence, we will use the central limit theorem for double sums given by de Jong (1987). Using the notations of that theorem, let $W(n) = T_{n,6}$, with $W_{nij}(\mathbf{V}_i, \mathbf{V}_j) = 2\tilde{b}_{nij} \int_a^b \epsilon_i(z) \epsilon_j(z) dz$ and $\mathbf{V}_i = (X_i, Z_i, T_i, \delta_i)$.

Note that $E(W_{nij}(\mathbf{V}_i, \mathbf{V}_j)|\mathbf{V}_i) = 0$ and that

$$\begin{aligned}\sigma_{ij}^2 &= E(W_{nij}^2) = 4E\left\{\tilde{b}_{nij}^2 \text{Var}\left[\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i, X_j\right]\right\} \\ &= \frac{4}{n^4h}\left(\int_v K * K(v)^2dv \int_a^b \int_a^b \int_x g^2(z_1, z_2, x)dx dz_1 dz_2\right) + o(n^{-4}h^{-1})\end{aligned}$$

so $\frac{1}{\sigma^2(n)}\max_i \sum_j \sigma_{ij}^2 = O(n^2hn^{-3}h^{-1}) = O(n^{-1}) \rightarrow 0$

Hence, the first condition in de Jong (1987) is satisfied.

In order to check the second condition, we will use the same terminology as in de Jong (1987), page 266:

$$E(W(n)^4) = E((T_{n,6})^4) = G_I + 6G_{II} + 12G_{III} + 24G_{IV} + 6G_V \quad (6.19)$$

with

$$\begin{aligned}G_I &= \sum_{i < j} E(W_{ij}^4) \\ G_{II} &= \sum_{i < j < k} (E(W_{ij}^2 W_{ik}^2) + E(W_{ji}^2 W_{jk}^2) + E(W_{ki}^2 W_{kj}^2)) \\ G_{III} &= \sum_{i < j < k} (E(W_{ij}^2 W_{ki} W_{kj}) + E(W_{ik}^2 W_{ji} W_{jk}) + E(W_{kj}^2 W_{ij} W_{ik})) \\ G_{IV} &= \sum_{i < j < k < l} (E(W_{ij} W_{ik} W_{lj} W_{lk}) + E(W_{ij} W_{il} W_{kj} W_{kl}) + E(W_{ik} W_{il} W_{jk} W_{jl})) \\ G_V &= \sum_{i < j < k < l} (E(W_{ij}^2 W_{kl}^2) + E(W_{ik}^2 W_{jl}^2) + E(W_{il}^2 W_{jk}^2)).\end{aligned}$$

It is easy to show that $G_I = O(n^{-6}h^{-3})$, $G_{II} = O(n^{-5}h^{-2})$, $G_{III} = O(n^{-5}h^{-2})$, $G_{IV} = O(n^{-4}h^{-1})$, $G_V = C_V n^{-4}h^{-2} + o(n^{-4}h^{-2})$, for some $C_V > 0$ and that $\sigma^4(n) = 2G_V + o(\sigma^4(n))$. Hence,

$$\frac{E((T_{n,6})^4)}{\sigma^4(n)} = \frac{6G_V + o(n^{-4}h^{-2})}{2G_V + o(\sigma(n)^4)} \rightarrow 3$$

So we get the desired result.

Lemma 6.2 $R_{n,10}$, $R_{n,11}$ and $R_{n,12}$ defined in (6.16), (6.17) and (6.18), respectively, are $o_p(n^{-1}h^{-1/2})$.

Proof:

$$\begin{aligned}
R_{n,10} &= \int_a^b \frac{2}{n} \sum_{i < j} \sum_{r=1}^n \tilde{B}_{ni}(X_r) B_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \epsilon_i(z) \epsilon_j(z) dz \\
&= \frac{2}{n} \sum_{i < j} \sum_{r=1}^n \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \tilde{B}_{ni}(X_r) B_{nj}(X_r) \int_a^b \epsilon_i(z) \epsilon_j(z) dz \\
&= 2 \sum_{i < j} \tilde{b}_{nij}^{(1)} \int_a^b \epsilon_i \epsilon_j dz
\end{aligned}$$

with $\tilde{b}_{nij}^{(1)} = n^{-1} \sum_{r=1}^n \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \tilde{B}_{ni}(X_r) B_{nj}(X_r)$. For any $\epsilon > 0$ we have that $P(|R_{n,10}| > Kn^{-1}h^{-1/2}) = A_1 \cdot B_1 + A_2 \cdot B_2$, with

$$A_1 = P(R_{n,10} > Kn^{-1}h^{-1/2} | \inf_x \hat{m}^*(x) > \epsilon), \quad B_1 = P(\inf_x \hat{m}^*(x) > \epsilon)$$

$$A_2 = P(|R_{n,10}| > Kn^{-1}h^{-1/2} | \inf_x \hat{m}^*(x) \leq \epsilon), \quad B_2 = P(\inf_x \hat{m}^*(x) \leq \epsilon)$$

Note that $B_1 \leq 1$ and $A_2 \leq 1$. Moreover, $E(R_{n,10} | \inf_x \hat{m}^*(x) > \epsilon) = 0$ and similar calculations as those done in the proof of Lemma 6.1 lead to $\text{Var}(R_{n,10} | \inf_x \hat{m}^*(x) > \epsilon) = o(n^{-2}h^{-1})$, provided that $\sup_x |\hat{m}^*(x) - m^*(x)| = o_p(1)$. This proves that $A_1 = o(1)$.

Next we will prove that $B_2 = o_p(1)$.

It is clear that

$$m^*(x) \leq \hat{m}^*(x) + \sup_x |\hat{m}^*(x) - m^*(x)|,$$

thus

$$\inf_x \hat{m}^*(x) \geq \inf_x m^*(x) - \sup_x |\hat{m}^*(x) - m^*(x)|,$$

which implies that

$$B_2 \leq P\left(\sup_x |\hat{m}^*(x) - m^*(x)| \geq \inf_x m^*(x) - \epsilon\right) \rightarrow 0, \quad \text{if } \epsilon < \inf_x m^*(x).$$

So we get $R_{n,10} = o_p(n^{-1}h^{-1/2})$.

In a similar way $R_{n,11}$ and $R_{n,12}$ are proved to be $o_p(n^{-1}h^{-1/2})$.

Lemma 6.3

$$R_{n,9} = \frac{1}{nh} \left(\int_a^b \int_x g(z, z, x) dx dz \right) K^{(2)}(0) + o_p(n^{-1}h^{-1/2})$$

where $R_{n,9}$ and $g(z, z, x)$ are defined in (6.13) and (2.4), respectively.

Proof: This term can be dealt with in the same way as the term Δ_{121} in the proof of Theorem 2.1 in González-Manteiga and Cao (1993). Indeed,

$$\begin{aligned}
R_{n,9} &= \int_a^b \sum_{i=1}^n b_{nii} \epsilon_j^2(z) dz \\
&= \int_a^b \frac{1}{n^3 h^2} \sum_{r=1}^n \frac{1}{\hat{m}^*(X_r)^2} \sum_{i=1}^n K\left(\frac{X_r - X_i}{h}\right)^2 \epsilon_i^2(z) dz \\
&= R_{n,9}^{(1)} + R_{n,9}^{(2)},
\end{aligned} \tag{6.20}$$

where the terms $R_{n,9}^{(1)}$ and $R_{n,9}^{(2)}$ are given by

$$\begin{aligned}
R_{n,9}^{(1)} &= \int_a^b \frac{1}{n^3 h^2} \sum_{r=1}^n \frac{1}{m^*(X_r)^2} \sum_{i=1}^n K\left(\frac{X_r - X_i}{h}\right)^2 \epsilon_i^2(z) dz \\
R_{n,9}^{(2)} &= \int_a^b \frac{1}{n^3 h^2} \sum_{r=1}^n \frac{m^*(X_r)^2 - \hat{m}^*(X_r)^2}{\hat{m}^*(X_r)^2 m^*(X_r)^2} \sum_{i=1}^n K\left(\frac{X_r - X_i}{h}\right)^2 \epsilon_i^2(z) dz = \int_a^b R_{n,9}^{(3)}(z) dz
\end{aligned}$$

Standard arguments, condition (H10) and Theorem B in Silverman (1978) imply that

$$R_{n,9}^{(2)} \leq \int_a^b |R_{n,9}^{(3)}(z)| dz \leq \frac{\Gamma(1+D)D}{\gamma^2} R_{n,9}, \tag{6.21}$$

where

$$D = \sup_x |m^*(x) - \hat{m}^*(x)| = O_p\left(\left(\frac{\ln h^{-1}}{nh}\right)^{1/2} + h^2\right)$$

and

$$R_{n,9}^{(2)} = O_p\left\{\left[\left(\frac{\ln h^{-1}}{nh}\right)^{1/2} + h^2\right] \frac{1}{nh}\right\} = O_p\{n^{-3/2} h^{-3/2} (\ln h^{-1})^{1/2} + n^{-1} h\} = o_p(n^{-1} h^{-1/2}).$$

The mean and the variance of $R_{n,9}^{(1)}$ can be studied by using Taylor expansions:

$$\begin{aligned}
E(R_{n,9}^{(1)}) &= \frac{1}{nh} \left(\int_a^b \int_x g(z, z, x) dx dz \right) \left(\int_t K^2(t) dt \right) + O(n^{-1} h), \\
\text{Var}(R_{n,9}^{(1)}) &= O(n^{-4} h^{-3}),
\end{aligned}$$

which implies that

$$R_{n,9}^{(1)} = \frac{1}{nh} \left(\int_a^b \int_x g(z, z, x) dx dz \right) \left(\int_t K^2(t) dt \right) + O(n^{-1} h + n^{-2} h^{-1.5}) \tag{6.22}$$

We now use (6.20), (6.21) and (6.22) to conclude.

Lemma 6.4 $R_{n,7}$ and $R_{n,8}$ defined in (6.11) are $o_p(n^{-1}h^{-1/2})$.

Proof: The term $R_{n,7}$ can be handled in a similar way as T_{n5} in the proof of Theorem 2.1 in order to get rid of the random denominator in $B_{ni}(X_r)$. Thus, we get the following:

$$R_{n,7} = \int_a^b \frac{1}{n} \sum_{r=1}^n (\Delta_r^{(7)}(z))^2 dz = \int_a^b (1 + o_p(1)) \frac{1}{n} \sum_{r=1}^n \sum_{i,j=1}^n (\tilde{\Delta}_{rij}^{(7)}(z))^2 dz \quad (6.23)$$

with $\tilde{\Delta}_{rij}^{(7)}(z) = \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)\eta_{ri}(z)\eta_{rj}(z)$ and $\eta_{ri}(z) = \phi'(S(z|X_r))S(z|X_r)\xi(Z_i, T_i, \delta_i, X_r, z) - \phi'(S(z|X_i))S(z|X_i)\xi(Z_i, T_i, \delta_i, X_i, z)$.

It is clear that $\tilde{\Delta}_{rij}^{(7)} = 0$ when $r = i$ or $r = j$. On the other hand, using (6.8), we have

$$E(\xi(Z, T, \delta, y, z)|X = x) = \int_0^z \frac{dH_1^*(u|x)}{C(u|y)} - \int_0^z \frac{C(u|x)}{C(u|y)^2} dH_1^*(u|y)$$

and

$$E(\xi(Z, T, \delta, y, z)^2|X = x) = \int_0^z \frac{dH_1^*(u|y)}{C(u|x)^2}$$

It is straightforward, but long and tedious to compute the order of $E(\tilde{\Delta}_{rij}^{(7)})$. Standard arguments such as changes of variables and Taylor expansions lead to

$$E(\tilde{\Delta}_{rij}^{(7)}(z)) = \begin{cases} 0, & \text{if } r = i \text{ or } r = j \\ E(\tilde{\Delta}_{123}^{(7)}(z)) = O(n^{-2}h^4), & \text{if } i \neq j \neq r \\ E(\tilde{\Delta}_{122}^{(7)}(z)) = O(n^{-2}h^{-1}), & \text{if } i = j \neq r \end{cases}$$

uniformly in z . These results imply that

$$E\left(\frac{1}{n} \sum_{r=1}^n \sum_{i,j=1}^n \tilde{\Delta}_{rij}^{(7)}(z)\right) = O(h^4 + n^{-1})$$

which together with $n^{-1} \sum_{r=1}^n \sum_{i,j=1}^n \tilde{\Delta}_{rij}^{(7)} \geq 0$ and by Markov inequality, leads to

$$\int_a^b \frac{1}{n} \sum_{r=1}^n \sum_{i,j=1}^n \tilde{\Delta}_{rij}^{(7)}(z) dz = O_p(h^4 + n^{-1}).$$

Now (6.23) implies that

$$R_{n,7} = \int_a^b \frac{1}{n} \sum_{r=1}^n (\Delta_r^{(7)}(z))^2 dz = O_p(h^4 + n^{-1}). \quad (6.24)$$

$R_{n,8}$ may be bounded by means of the Cauchy-Schwarz inequality:

$$\begin{aligned} |R_{n,8}| &= \left| \int_a^b 2 \frac{1}{n} \sum_{r=1}^n \sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \Delta_r^{(7)}(z) dz \right| \\ &\leq 2 \int_a^b \left[\frac{1}{n} \sum_{r=1}^n \left(\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \right)^2 \right]^{1/2} \cdot \left[\frac{1}{n} \sum_{r=1}^n (\Delta_r^{(7)}(z))^2 \right]^{1/2} dz = o_p(n^{-1}h^{-1/2}), \end{aligned}$$

where $n^{-1} \sum_{r=1}^n \left(\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \right)^2 = O_p(n^{-1}h^{-1})$.

Lemma 6.5 $R_{n,6}$ defined in (6.10) is $o_p(n^{-1}h^{-1/2})$.

Proof: Note that

$$R_{n,6} = O_p\left((T_{n,3})^{1/2} \cdot \left(\frac{\ln n}{nh} \right)^{3/4} \right).$$

From Lemma 6.1, 6.2, 6.3, condition (H10) and the fact that $T_{n,4} = R_{n,9} + T_{n,5}$ we get

$$\begin{aligned} \sqrt{n^2 h} \left(\int_a^b \frac{1}{n} \sum_{r=1}^n \left(\sum_{i=1}^n B_{ni}(X_r) \epsilon_i(z) \right)^2 dz \right. \\ \left. - \frac{1}{nh} \left(\int_a^b \int_x g(z, z, x) dx dz \right) K^{(2)}(0) \right) \xrightarrow{d} N(0, V). \end{aligned}$$

This limit distribution together with Lemma 6.4 gives that

$$\sqrt{n^2 h} T_{n,3} - h^{-1/2} K^{(2)}(0) \int_a^b \int_x g(z, z, x) dx dz \xrightarrow{d} N(0, V). \quad (6.25)$$

As a consequence, $T_{n,3} = O_p(n^{-1}h^{-1})$, thus $R_{n,6} = O_p(n^{-6/5}h^{-1/2}) = o_p(n^{-1}h^{-1/2})$.

Lemma 6.6 $R_{n,1}$ and $R_{n,2}$ defined in (6.2) and (6.3), respectively, are $o_p(n^{-1}h^{-1/2})$.

Proof: Write

$$\begin{aligned} R_{n,1} &= \int_a^b \frac{1}{n} \sum_{r=1}^n \left[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \dots + \hat{\beta}_p(z)X_r^p) \right]^2 dz \\ &\leq \frac{1}{n} \sum_{r=1}^n \int_a^b \left[\sup_z \max_r |\phi(S(z|X_r)) - \hat{\beta}(z)X_r| \right]^2 dz = o_p(n^{-1}h^{-1/2}), \end{aligned}$$

and

$$\begin{aligned}
R_{n,2} &= \int_a^b \frac{2}{n} \sum_{r=1}^n \left[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r)) \right] \left[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \dots + \hat{\beta}_p(z)X_r^p) \right] dz \\
&= \int_a^b \frac{2}{n} \sum_{r=1}^n \left[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r)) \right] (1, X_r, \dots, X_r^p) (\boldsymbol{\beta}(z) - \hat{\boldsymbol{\beta}}(z)) dz \\
&= O_p(n^{-1}) = o_p(n^{-1}h^{-1/2}),
\end{aligned}$$

where $\sup_{z \in [a,b]} |\boldsymbol{\beta}(z) - \hat{\boldsymbol{\beta}}(z)| = O_p(n^{-1/2})$ by Corollary 2.2 in Teodorescu et al. (2008), and

where $\int_a^b 2n^{-1} \sum_{r=1}^n [\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r))] (1, X_r, \dots, X_r^p) = O_p(n^{-1/2})$ uniformly in z .

Acknowledgments. The authors acknowledge support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy). The second author also acknowledges support from ERC grant nr. 203650.

References

- Aalen, O.O. (1980). A model for non-parametric regression analysis of counting processes. In *Lecture Notes on Mathematical Statistics and Probability*, 2, W. Klonecki, A. Kozek and J. Rosiski, eds. New York: Springer-Verlag, 1-25.
- Akritas, M.G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, 52, 1342-1362.
- Amir, C. and McKeague, I.W. (2000). Identifying a time-dependent covariate effect in the additive risk model. *Unpublished manuscript available at <http://biostat.columbia.edu/mckeague/ps/index.html>*.
- Andersen, P.K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics*, 38, 67-77.
- Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association*, 83, 204-212.
- Beran, R. (1981). Nonparametric regression with randomly censored data. *Technical report, University of 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. (2008). Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics*, 143, 166-190.
- Chen, Y.Q. (2001). Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics*, 57, 853-860.
- Dabrowska, D. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics*, 17, 1157-1167.
- de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Prob. Th. Rel. Fields*, 75, 261-277.
- Gill, R.D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika*, 74, 289-300.
- Grigoletto, M. and Akritas, M.G. (1999). Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics*, 55, 1177-1187.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21, 1926-1947.
- Huffer, F.W. and McKeague, I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, 86, 114-129.

- 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.
- Kim, J. and Lee, S.-Y. (1998). Two-sample goodness-of-fit tests for additive risk models with censored observations. *Biometrika*, 85, 593-603.
- 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>).
- Marzec, L. and Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika*, 84, 901-908.
- McKeague, I.W. and Sasieni, P.D. (1994). A partly parametric additive risk model. *Biometrika*, 81, 501-514.
- 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>).
- Parzen, M. and Lipsitz, S.R. (1999). A global goodness-of-fit statistic for Cox regression models. *Biometrics*, 55, 580-584.
- Rossini, A.J., Wei, L.J. and Ying, Z. (1996). Checking adequacy of the semiparametric location shift model with censored data. *Lifetime Data Analysis*, 2, 145-157.
- Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67, 145-153.

- 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.
- Teodorescu, B., Van Keilegom, I. and Cao, R. (2008). Generalized conditional linear models under left truncation and right censoring. *The Annals of the Institute of Statistical Mathematics* (to appear).
- Teodorescu, B. (2008). *Generalized conditional linear models with time-varying coefficients under left truncation and right censoring*, PhD thesis, Université catholique de Louvain, Belgium.
- 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.
- Yang, S. (1998). Some scale estimators and lack-of-fit tests for the censored two-sample accelerated life model. *Biometrics*, 54, 1040-1052.
- Yuen, K.C. and Burke, M.D. (1997). A test of fit for a semiparametric additive risk model. *Biometrika*, 84, 631-639.
- Zucker, D.M. and Karr, A.F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Annals of Statistics*, 18, 329-353.