UNIVERSITE CATHOLIQUE DE LOUVAIN

INSTITUT DE STATISTIQUE

General Conditional Linear Models with Time-Dependent Coefficients under Censoring and Truncation

Membres du jury :

Prof. Ricardo Cao
Prof. Paul Janssen
Prof. Leopold Simar
Prof. Sébastien Van Bellegem
Prof. Ingrid Van Keilegom *(Promoteur)*

Thèse présentée en vue de
l'obtention du grade de
Docteur en Sciences
(orientation statistique) par :

Bianca Teodorescu

Louvain-la-Neuve

December 2008

Acknowledgements

J'aimerais tout d'abord remercier mon promoteur, Prof. Ingrid Van Keilegom pour son encadrement continu, ses encouragements, son dynamisme et sa disponibilité. Je tiens aussi à la remercier pour son amitié et toutes les bons moments passés en dehors du travail. Tambíen quiero agradecer el Prof. Ricardo Cao por su tiempo dedicado a la realisacion de este trabajo, por sus buenos consejos y por la patiencia de haberlo leído todo en detalles. Je tiens aussi à remercier les autres membres du jury, Prof. Paul Janssen, Prof. Leopold Simar et Prof. Sébastien Van Bellegem pour avoir lu cette thèse et de m'avoir apporté de nombreux commentaires intéressants.

Je remercie également tous les membres de l'Institut de Statistique, personnel administratif, académique et scientifique, qui ont contribué à une atmosphère de travail très agréable. Un grand merci à Jean-Luc Marion pour son aide précieux en informatique et pour sa disponibilité (même en étant en congé).

Je remercie tous mes amis à commencer par les « poulettes » (et co.) pour leur amitié, leurs bons conseils et les bons moments partagés. Merci à Afroditi et à Christian pour leur soutien moral et pour m'avoir « supporté » pendant cette dernière année, ainsi qu'à Anastasia pour les moments partagés de sa vie. Gracias a Juan Carlos por su amistad, por los momentos compartidos y por la ayuda « informatica » (sin su algoritmo, mis simulationes se habrían tardado aún más). Merci à Abderrahim et à Daniela pour m'avoir pris « en charge » dés mon arrivé en Belgique. Et la liste pourra continuer ...

Multumesc de asemenea prietenilor din Romania (si nu numai) pentru ajutorul moral « la distanta ». Multumesc parintilor pentru ajutorul si dragostea neconditionata oferite in ultimii 30 de ani si pentru ca au « acceptat » plecarea in Belgia pentru a realiza acest doctorat.

En fin, je ne peux pas terminer cette liste (non exhaustive) sans remercier Anouar, mon ami, mon compagnon, pour sa présence dans ma vie, pour ses encouragements, pour son aide, pour tout !

Thanks are due to Sonia Pértega Díaz and Salvador Pita Fernández at the University Hospital of A Coruña for providing the gastric cancer data set.

Contents

Chapter 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. Two data sets are being considered in this context.

The first one can be found in Casariego-Vales et al (2001) and Rabuñal et al (2004). It concerns a retrospective cohort study on patients diagnosed with gastric adenocarcinoma between January 1975 and December 1993 in Xeral Hospital in Lugo and University Hospital of A Coruña (northwest of Spain). 2334 patients were introduced into the study. For each individual we observed the time (in days) from the first symptoms to diagnostic, the time from first symptoms to death or loss of follow up, the death indicator (dead or alive) and some risk factors like the age at diagnosis, the extension of the tumor (TNM classification), the presence of metastasis (yes or no) and the type of surgery. Censoring occurred when patients died from a reason not related to their disease, were lost to follow up or were still alive at last visit. Patients were followed until at least 1st of August 1996. On the other hand, those that were not diagnosed before death, thus they do not appear in the registers of the hospital, were considered as being truncated. The data are shown in Figure 1.1.

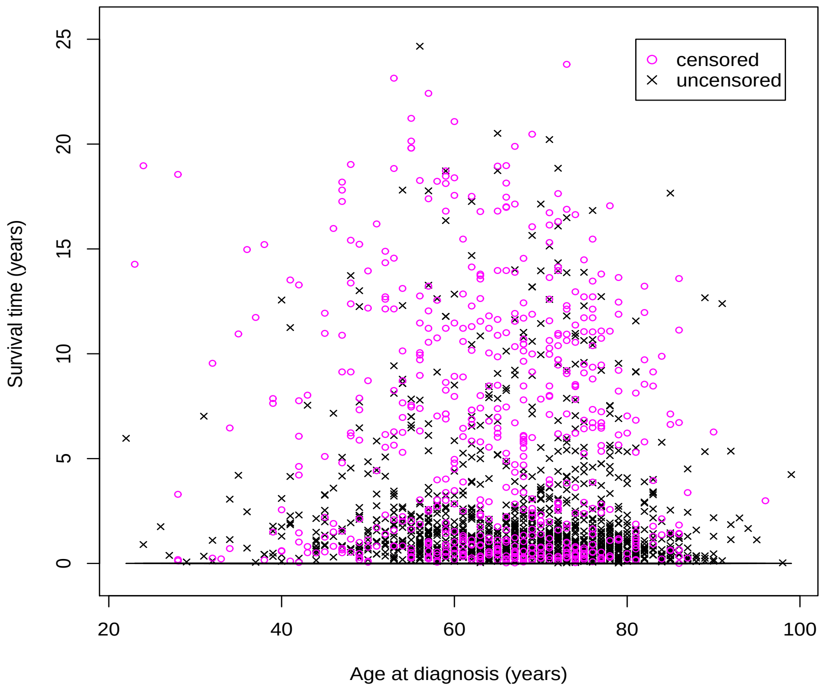

Figure 1.1 : Gastric cancer data.

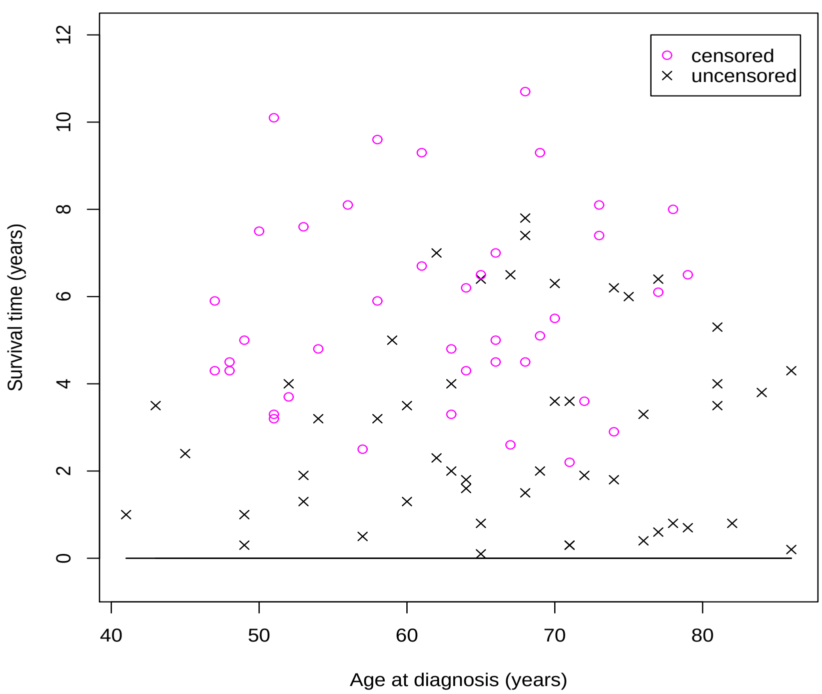

Figure 1.2 : Larynx cancer data.

The second data set was introduced in Kardaun (1983) and Klein and Moeschberger (1997). 90 males diagnosed with cancer of the larynx during the period 1970-1978 at a Dutch hospital were considered. Times recorded are the intervals (in years) between first treatment and either death or the end of the study (January 1, 1983). Also recorded are the patient's age at the time of diagnosis, the year of diagnosis, and the stage of the patient's cancer (TNM classification). 40 patients were still alive at the end of the study, thus censored. The data are presented in Figure 1.2.

In the case of censored and/or truncated responses, lots of models exist in the literature that describe the relationship between the survival function and the covariates (proportional hazards model or Cox model, log-logistic model, accelerated failure time model, additive risks 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 examples, certain covariates (e.g. age at diagnosis, type of surgery, extension of tumor, etc.) can have a relatively high impact on early age survival, but a lower influence at higher age. This motivated a number of authors to extend the Cox model to allow for time-dependent coefficients or consider other type of time-dependent coefficients models like the additive hazards model. In practice it is of great use to have at hand a method to check the validity of the above mentioned models. The case where the coefficients are time-independent was largely studied in the literature. But when the coefficients are time-dependent little work is available. Details on the above models and procedures are given in the rest of this chapter which is organized as follows. Existing conditional models in survival analysis are presented in Section 1.1. The product limit estimator of the survival function for censored and truncated data is introduced in Section 1.2. The likelihood function for binary data is given in Section 1.3. Its application to obtain the maximum likelihood estimator in generalized linear models with time-dependent coefficients is explained in Section 1.4, while least squares estimator in generalized linear models with time-constant coefficients is presented in Section 1.5. Finally goodness-of-fit tests in survival analysis are detailed in Section 1.6.

The objective of this theses is twofold. First, in Chapter 2, we consider a very general model, which includes as special cases the above mentioned models (Cox model, additive model, log-logistic model, linear transformation models, etc.) with time-dependent coefficients and study the parameter estimation by means of a least squares approach. The response is allowed to be subject to right censoring and/or left truncation. The second objective is to propose an omnibus goodness-of-fit test that will test if the general time-dependent model considered above fits the data, see Chapter 3. A bootstrap version, to approximate the critical values of the test is also proposed.

The methodology developed in Chapter 2 and Chapter 3 will be applied to the two data sets described above and the results are presented in Chapter 4.

1.1 Conditional models in survival analysis

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ be a vector of covariates. The dependency between the survival time Y and $\boldsymbol{X}$, can be modeled in many different ways. In the literature lots of models are available that describe this dependency and many of these models rely on the hazard rate and on the cumulative hazard function. The hazard rate $\lambda(y|\boldsymbol{x})$ is the instantaneous risk that an individual of age y with given covariate values $\boldsymbol{x}$ experiences the event of interest in the next instant :

$$\lambda(y|\boldsymbol{x}) = \lim_{\Delta y \to 0} \frac{P[y \leq Y < y + \Delta y | Y \geq y, \boldsymbol{X} = \boldsymbol{x}]}{\Delta y}$$

If Y is continuous, then $\lambda(y|\boldsymbol{x}) = f(y|\boldsymbol{x})/S(y|\boldsymbol{x}) = \frac{-\partial \ln[S(y|\boldsymbol{x})]}{\partial y}$. The cumulative hazard function is defined as follows :

$$\Lambda(y|\boldsymbol{x}) = \int_0^y \lambda(u|\boldsymbol{x}) du = -\ln[S(y|\boldsymbol{x})]$$

An important and well known model is the proportional hazards model of Cox (1972), which stipulates that the covariates act multiplicatively on the hazard function. It takes the form

$$\lambda(y|\boldsymbol{x}) = \lambda_0(y) \exp(\boldsymbol{\beta}' \boldsymbol{x}), \tag{1.1.1}$$

where $\lambda(y|\boldsymbol{x})$ denotes the hazard rate of the response, given the covariate vector $\boldsymbol{X} = \boldsymbol{x}$, $\lambda_0(y)$ is the unknown baseline hazard rate and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ is the unknown parameter vector. This model can also be expressed in terms of the cumulative hazard function

$$\Lambda(y|\boldsymbol{x}) = \Lambda_0(y) \exp(\boldsymbol{\beta}' \boldsymbol{x}) \tag{1.1.2}$$

A model related to the proportional hazards model is the accelerated failure time model introduced by Cox and Oakes (1984) :

$$\lambda(y|\boldsymbol{x}) = \lambda_0(ye^{-\boldsymbol{\beta}'\boldsymbol{x}})\exp(-\boldsymbol{\beta}'\boldsymbol{x}), \tag{1.1.3}$$

that can be equivalently written as :

$$\ln(Y) = \boldsymbol{\beta}'\boldsymbol{X} + \epsilon, \tag{1.1.4}$$

with ϵ independent errors with common distribution G. This model has the advantage of providing a direct characterization of the effects of covariates on the failure time.

An alternative approach to model the conditional hazard rate is to consider the additive risks model

$$\lambda(y|\boldsymbol{x}) = \lambda_0(y) + \boldsymbol{\beta}'\boldsymbol{x}, \tag{1.1.5}$$

that assumes that the covariates act in an additive way on the hazard rate, contrary to the multiplicative way of the Cox model. It is clear that some conditions on $\boldsymbol{\beta}'\boldsymbol{x}$ are needed in order for $\lambda(y|\boldsymbol{x})$ to be a hazard rate, see for example Lin and Ying (1994). This model is a special case of the multiplicative intensity model suggested by Aalen (1980).

The proportional odds model assumes the following relationship between the hazard rate of Y and the vector of covariates $\boldsymbol{X}$

$$\mathrm{logit}(1 - \exp(-\Lambda(\mathrm{y}|\boldsymbol{x}))) = \alpha(\mathrm{y}) + \boldsymbol{\beta}'\boldsymbol{x}, \tag{1.1.6}$$

where $\alpha(y)$ is an increasing function and $\mathrm{logit(u)} = \ln\left(\frac{\mathrm{u}}{1-\mathrm{u}}\right)$. It was first applied to survival analysis by Pettitt (1982) and Bennett (1983).

The model can be rewritten in an equivalent form

$$\ln\left(\frac{F(y|\boldsymbol{x})}{1 - F(y|\boldsymbol{x})}\right) = \alpha(y) + \boldsymbol{\beta}'\boldsymbol{x}. \tag{1.1.7}$$

Another interesting model is the general polynomial regression model

$$T(F(\cdot|x)) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p \tag{1.1.8}$$

where the functional T is given by the equation $T(N) = \int_0^1 N^{-1}(s)J(s)ds$, for any distribution function, N,

$$N^{-1}(s) = \inf\{u : N(u) \geq s\}.$$

and $J(s)$ is a score function satisfying $\int_0^1 J(s)ds = 1$. Different choices of $J(s)$ give different quantities of interest. For example, the choice $J(s) = 1_{\{0\leq s\leq 1\}}$ leads to the conditional mean and $J(s) = (q-p)^{-1}1_{\{p\leq s\leq q\}}$, for some $0 \leq p < q \leq 1$, leads to a trimmed mean.

McCullagh and Nelder (1989) introduced a unifying framework for all these models : the generalized linear models (GLM), that provide a way to model conditional survival functions. The main idea is that the mean response is linked to a linear predictor via a so called link-function ϕ which may take various forms. Here the mean response is the survival function $S(y|\boldsymbol{x}) = 1 - F(y|\boldsymbol{x})$. It is easy to see that :

$$S(y|\boldsymbol{x}) = P(Y > y|\boldsymbol{x}) = P(W = 1|\boldsymbol{x}) = E(W|\boldsymbol{x}),$$

with $W = 1_{\{Y>y\}}$, $W \sim Bi(1,p)$ and $p = S(y|\boldsymbol{x})$. Since the survival time takes only values between 0 and 1, we need a link that will map the interval $(0,1)$ on to the whole real line :

$$\phi(S(y|\boldsymbol{x})) = \alpha(y) + \boldsymbol{\beta}'\boldsymbol{x} \tag{1.1.9}$$

Common link functions for binary responses include the logit $\phi_1(\pi) = \ln\{\pi/(1-\pi)\}$, complementary log-log $\phi_2(\pi) = \ln\{-\ln(1-\pi)\}$, probit $\phi_3(\pi) = \Phi^{-1}(\pi)$ where Φ is the Normal density function and $\phi_4(\pi) = -\ln(\pi)$. These link functions induce the above mentioned models. For example, for the complementary log-log link we get the proportional hazards model, the logit link induces the proportional odds model and the probit link, the probit model (Kalbfleisch and Prentice (2002)). Note that in the classical linear models the link-function is the identity link, since the response, usually, can take any value on the real line.

An alternative approach is to consider these models as special cases of linear transformation models, that are of the following form

$$\varphi(y) = \boldsymbol{\beta}\boldsymbol{x} + \epsilon, \tag{1.1.10}$$

with Y the survival time, φ an unknown monotone increasing function, ϵ a random variable with known distribution function, F_ϵ, independent of the covariates, $\boldsymbol{X}$, and $\boldsymbol{\beta}$ is a vector of constant coefficients or, equivalently,

$$\varsigma\left(S(y|\boldsymbol{x})\right) = \varphi(y) - \boldsymbol{\beta}\boldsymbol{x}, \tag{1.1.11}$$

where ς is a known increasing function and $\varsigma^{-1}(u) = 1 - F_\epsilon(u)$. We get the proportional hazards model and the proportional odds models by taking in (1.1.10) ϵ to follow the extreme-value distribution and the standard logistic distribution, respectively or, by taking in (1.1.11) $\varsigma(u) = \ln(-\ln(u))$ and $\varsigma(u) = \ln(\frac{u}{1-u})$, respectively. For more details on transformation models see, for example, Doksum (1987), Cheng et al. (1995, 1997), Fine et al. (1998), Fine and Gray (1999) or Chen et al. (2002).

For the estimation of the above mentioned models we have mainly two different methodologies : maximum likelihood and least squares.

In the case of censored responses, estimation in Cox models has been considered by many authors. For an overview on maximum likelihood estimators, their large sample properties, as well as their small sample performances and related research we refer to Akritas and LaValley (1997). Maximum likelihood estimation in additive risks models with censored responses has been considered by Breslow and Day (1980, p 55, 58), Pocock et al. (1982), Buckley (1984), Cox and Oakes (1984), Pierce and Preston (1984), Thomas (1986), while the estimator of $\boldsymbol{\beta}$ in (1.1.6) for the proportional odds model has been proposed by Bennett (1983), Pettitt (1984), Dabrowska and Doksum(1988) and Cheng et al. (1995) without estimating the baseline function $\alpha(y)$ directly. Shen (1998) proposed a procedure for estimating both $\alpha(y)$ and $\boldsymbol{\beta}$. For the accelerated failure time model a natural approach is the least squares method. See Kalbfleisch and Prentice (2002) for a review on existing estimators and Jin et al. (2006) for recent developments.

Akritas (1996) introduced a least squares estimator for the coefficients in the general polynomial model (1.1.8) in the case where the responses may be censored or truncated, but not both. Grigoletto and Akritas (1999) presented a similar least squares estimator for the Cox model, the additive risks model and the proportional odds model in the same case of censored or truncated responses. Cao and González-Manteiga (2008) extended the estimators in Akritas (1996) and in Grigoletto and Akritas (1999) to the case where the responses may be censored and truncated. See more details on the estimation procedure in Section 2.2.

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

$$\lambda(y|\boldsymbol{x}) = \lambda_0(y)\exp(\boldsymbol{\beta}'(y)\boldsymbol{x}). \tag{1.1.12}$$

To estimate the coefficient functions $\boldsymbol{\beta}(z)$, Murphy and Sen (1991) proposed a histogram sieve estimation procedure by assuming that the coefficient functions are piecewise constant. Gamerman (1991) described the dynamic linear model approach by assuming that $\lambda_0(y)$ and $\boldsymbol{\beta}(y)$ are piecewise constant functions, constant between distinct failure times, and using the full likelihood for estimation updating terms sequentially in time. The assumption that the baseline hazard function and coefficient functions are piecewise constant may not be appropriate in real applications. To overcome this drawback, Zucker and Karr (1990) and Hastie and Tibshirani (1993) used smoothing spline partial likelihood method, leaving $\lambda_0(y)$ unspecified but modeling the coefficient functions $\boldsymbol{\beta}(y)$ smoothly. Nan and Lin (2003), Lambert and Eilers (2004) and Kauermann (2005) improved this method by using instead a penalized spline fitting smoothing technique. Another approach is based on local techniques, see e.g. Fan et al. (1997) or Cai and Sun (2003). The basic idea is a simple extension of the local fitting technique used in the scatterplot smoothing coupled with the partial likelihood. Martinussen et al. (2002) proposed an iterative estimation procedure for the cumulative coefficient function $\boldsymbol{B}(y) = \int_0^y \boldsymbol{\beta}(s)ds$, by using a modified likelihood approach and the Newton-Raphson iterative algorithm.

Also, other time-dependent survival models have been considered, like the additive hazards model

$$\lambda(y|\boldsymbol{x}) = \beta_0(y) + \boldsymbol{\beta}'(y)\boldsymbol{x}, \tag{1.1.13}$$

see for example Aalen (1980), Cox and Oakes (1984), Huffer and McKeague (1991), Andersen et al. (1992), McKeague and Sasieni (1998) or Lin and Ying (1994) who proposed least squares estimators for the cumulative regression function $\boldsymbol{B}(y) = \int_0^y \boldsymbol{\beta}(y)dt$.

Jung (1996) considered an extension of the general linear model (1.1.9) that allows for time-varying coefficients,

$$\phi(S(y|\boldsymbol{x})) = \beta_0(y) + \beta_1(y)x_1 + \ldots + \beta_p(y)x_p \tag{1.1.14}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ denotes a vector of discrete covariates and ϕ is a monotone differentiable transformation in $[0, 1]$. Particular choices of ϕ give well known models in survival analysis, but extended to time-dependent coefficients. The choice $\phi(u) = \ln(\frac{u}{1-u})$ gives the logistic model

$$S(y|\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{\beta}'(y)\boldsymbol{x})},$$

$\phi(u) = -\ln(u)$ gives the additive hazards model

$$-\ln(S(y|\boldsymbol{x})) = H(y|\boldsymbol{x}) = \beta_0(y) + \beta_1(y)x + \ldots + \beta_p(y)x^p$$

and $\phi(u) = \ln(-\ln(u))$ leads to the proportional hazards model

$$-\ln S(y|\boldsymbol{x}) = H(y|\boldsymbol{x}) = \exp(-\boldsymbol{\beta}^{'}(y)\boldsymbol{x}).$$

Unfortunately, Jung's estimation procedure for $\boldsymbol{\beta}(y)$ is only valid in the case where the observations are censored, the covariates are discrete and the censoring is independent of the covariates (see Section 1.4 for more details).

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 :

$$\phi(S(y|X=x)) = \beta_0(y) + \beta_1(y)x + \ldots + \beta_p(y)x^p, \tag{1.1.15}$$

where X is continuous and p is known. For more details on these models see Section 1.4.

In Chapter 2, we extend these models, (1.1.14) and (1.1.15), by allowing the response to be subject to right censoring and/or left truncation, by allowing to have a combination of discrete covariates and a one dimensional continuous covariate and by using a least squares procedure instead of the maximum likelihood method used in the above models.

1.2 Product limit estimators (PLE)

Let Y denote the survival time, T the truncation time, C the censoring time and X a (one-dimensional) covariates , typically a risk factor, that affects the survival time Y. When data are left-truncated and right-censored we observe (T, Z, δ, X) only if $Z \geq T$, where $Z = \min\{Y, C\}$ and $\delta = 1_{\{Y \leq C\}}$. Let $(T_i, Z_i, \delta_i, X_i)$, $i = 1, \ldots, n$ be an i.i.d. sample from (T, Z, δ, X). A very common assumption that is made in this setup is that Y, T and C are independent given X. Let $F(y|x)$ be the conditional distribution of Y when $X = x$. Let us define the following curves : $M(x) = P(X \leq x)$, $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $L(y|x) = P(T \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H_1(y|x) = P(Z \leq y, \delta = 1|x)$, $L(y) = P(T \leq y)$, $H(y) = P(Z \leq y)$, $H_1(y) = P(Z \leq y, \delta = 1)$, $C(y|x) = P(T \leq y \leq Z|x, T \leq Z)$, and $\alpha(x) = P(T \leq Z|x)$, which is the probability of absence of truncation conditionally on $X = x$. For any distribution function $W(y) = P(\eta \leq t)$, we denote the left and right support endpoints by $a_W = \inf\{t|W(y) > 0\}$ and $b_W = \sup\{t|W(y) < 1\}$, respectively. We define $W^*(y) = P(\eta \leq t|T \leq Z)$. Finally, let m denote the density of X and m^* the density of X conditionally on $T \leq Z$.

In the single continuous covariate context a general product-limit estimator (GPLE) has been introduced by Iglesias-Pérez and González-Manteiga (1999) :

$$\hat{S}_n(y|x) = 1 - \hat{F}_n(y|x) = \prod_{i=1}^{n} \left(1 - \frac{1_{\{Z_i \leq y, \delta_i = 1\}} B_{ni}(x)}{C_n(Z_i|x)}\right), \tag{1.2.1}$$

where $\{B_{ni}(x)\}_{i=1}^{n}$ denotes a sequence of suitable type of weights (Gasser-Müller, Nadaraya-Watson, ...) and $C_n(u|x) = \sum_{j=1}^{n} 1_{\{T_j \leq u \leq Z_j\}} B_{nj}(x)$.

Note that the GPLE reduces to the estimator defined in La Valley and Akritas (2005) when there is no right censoring ($\delta = 1, Z = Y$) and to the estimator given in Beran (1981) and González-Manteiga and Cadarso-Suárez (1994) when there is no left truncation ($T = 0$), (provided that the nonparametric weights are the same in each considered estimator). On the other hand, the GPLE reduces to the PLE estimator defined in Tsai et al (1987) in absence of covariables ($B_{ni}(x) = \frac{1}{n}, \forall i$). Also, when no right censoring is present we obtain the estimator proposed by Lynden-Bell (1971) and when no truncation is present we get the well known estimator proposed by Kaplan-Meier (1958). Iglesias-Pérez and González-Manteiga (1999) also provided an i.i.d. representation for $\hat{F}_n - F$ that will be useful in the next chapters.

Because of the nature of truncated data, a random design has been considered. Also because of the curse of dimensionality only one real covariable has been used. In this context, it is common to assume that the sequence of weights are of Nadaraya-Watson type :

$$B_{ni}(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)}, \qquad i = 1, 2, \ldots, n \tag{1.2.2}$$

where K denotes a kernel function and $h = h_n \geq 0$ is the bandwidth parameter.

The following three results can be found in Iglesias-Pérez and González-Manteiga (1999), as Theorem 2(c), Corollary 3 and Lemma 5, respectively.

Theorem 1.2.1 *Suppose that Conditions (H1), (H2'), (H3)-(H9) of Theorem 2 in Iglesias-Pérez and González-Manteiga (1999) hold together with $h \to 0$, $nh \to \infty$, $\frac{\ln n}{nh} \to 0$ and $nh^5/\ln n = O(1)$, then, if $a < a_{H(\cdot|x)}$, for all $x \in I$, we have :*

$$\hat{F}_n(y|x) - F(y|x) = (1 - F(y|x)) \sum_{i=1}^{n} B_{ni}(x)\xi(Z_i, T_i, \delta_i, y, x) + R_n(y|x),$$

for $x \in I, y \in [a, b]$, where

$$\xi(Z, T, \delta, y, x) := \frac{1_{\{Z \leq y, \delta = 1\}}}{C(Z|x)} - \int_0^y \frac{1_{\{T \leq u \leq Z\}}}{C^2(u|x)} dH_1^*(u|x) \tag{1.2.3}$$

$$\sup_{[a,b]\times I} |R_n(y|x)| = O_p\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \tag{1.2.4}$$

and a, b and I are defined in condition (H2').

An important application of this theorem is the asymptotic normality of the generalized PLE for left-truncated and right-censored (LTRC) data given by (1.2.1), described in the following corollary :

Corollary 1.2.2 *Under the hypotheses of Theorem 1.2.1 for $x \in I$ and for $y \in [a, b]$, it follows that :*

(a) If $nh^5 \to 0$ and $(\ln n)^3/nh \to 0$ then

$$(nh)^{1/2}[\hat{F}_n(y|x) - F(y|x)] \xrightarrow{d} N(0, s^2(y|x))$$

where $s^2(y|x) = (1 - F(y|x))^2(\int K^2(z)dz)(\int_0^y (dH_1^*(u|x)/C^2(u|x)))/m^*(x)$

(b) If $h = Cn^{-1/5}$, for some $C > 0$, then

$$(nh)^{1/2}[\hat{F}_n(y|x) - F(y|x)] \xrightarrow{d} N(b(y|x), s^2(y|x))$$

where $b(y|x) = C^{5/2}(1-F(y|x))(\int z^2K(z)dz)(\beta''(x)m^*(x)+2\beta'(x)m^{*'}(x))/2m^*(x)$ and $\beta(u) = E(\xi(Z, T, \delta, y, x)|X = u, T \leq Z)$.

Also, the uniform strong consistency of the kernel estimator of the distribution (or subdistribution) function when a Nadaraya-Watson sequence of weights is used and when we have right-censoring and left-truncation is being given :

Lemma 1.2.3 *(Uniform strong consistency) Let* η *a continuous and positive r.v., and let* $\underset{\sim}{S}(y|x) = P(\eta > y|X = x)$ *denote the conditional survival function at* $X = x$*, where* X *is a continuous r.v., with density* m*. The kernel estimator of* $\underset{\sim}{S}(y|x)$ *with Nadaraya-Watson weights is given by* $\hat{\underset{\sim}{S}}_n(y|x) = \hat{m}_n^{-1}(x)\hat{A}_n(y,x)$ *where*

$$\hat{A}_n(y,x) = \frac{1}{nh}\sum_{i=1}^{n} 1_{\{\eta_i > y\}} K\left(\frac{x - x_i}{h}\right)$$

$$\hat{m}_n(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)$$

Under the assumptions (h1)-(h5) in Lemma 5 of Iglesias-Pérez and González-Manteiga (1999) we have that :

$$\sup_{[0,\infty)\times I} |\hat{\underset{\sim}{S}}_n(t|x) - \underset{\sim}{S}(t|x)| = O(\ln n/nh)^{1/2} + O(h^2) \quad a.s.$$

Moreover, if $nh^5/\ln n = O(1)$ *then :*

$$\sup_{[0,\infty)\times I} |\hat{\underset{\sim}{S}}_n(t|x) - \underset{\sim}{S}(t|x)| = O(\ln n/nh)^{1/2} \quad a.s.$$

Obviously, the previous lemma holds, when substituting survival functions by distribution functions, and it can also be proved for subdistribution functions of the type $P(\eta \le y, \delta = 1|X = x)$, where δ is an indicator function. For more details see Iglesias-Pérez and González-Manteiga (1999).

A "conditional obvious" bootstrap method for approximating the distribution of $(nh)^{1/2}[\hat{F}_n(y|x) - F(y|x)]$ of LTRC data with covariables has also been proposed and studied by Iglesias-Pérez and González-Manteiga (2003).

Let us first define $\hat{G}_n(y|x)$, the GPLE of $G(y|x)$. Since Y, T and C are conditionally independent given $X = x$ and taking into account that C plays a similar role as Y but with an indicator variable $1 - \delta_i$ instead of δ_i in the LTRC model, $\hat{G}_n(y|x)$ is obtained by substituting in formula (1.2.1) $\delta_i = 1$ by $\delta_i = 0$:

$$\hat{G}_n(y|x) = 1 - \prod_{i=1}^{n}\left(1 - \frac{1_{\{Z_i \le y, \delta_i = 0\}} B_{ni}(x)}{C_n(Z_i|x)}\right). \tag{1.2.5}$$

Also define $\hat{L}_n(y|x)$, the estimator of $L(y|x)$ in an analogous way as $\hat{F}_n(y|x)$:

$$\hat{L}_n(y|x) = \prod_{i=1}^{n} \left(1 - \frac{1_{\{T_i > y, \delta_i = 0\}} B_{ni}(x)}{C_n(T_i|x)} \right), \tag{1.2.6}$$

with $B_{ni}(x)$ the same sequence of nonparametric weights as for $\hat{F}_n(y|x)$.

Note that the GPLE given by (1.2.6) equals the PLE of the distribution function of T studied by Woodroofe (1985) in the absence of a covariable ($B_{ni} = \frac{1}{n}, \forall i$). Consistency of the PLE was proved in Woodroofe's paper.

On the other hand, the consistency of $\hat{G}_n(y|x)$ is a consequence of the consistency of $\hat{F}_n(y|x)$ and the relationship $H_0(y|x) = H(y|x) - H_1(y|x)$.

The conditional obvious bootstrap procedure for LTRC data proposed by Iglesias-Pérez and González-Manteiga (2003) is as follows :

1. Choose a pilot bandwidth, g, (usually larger than h) to estimate $F(z|X_i)$, $G(z|X_i)$ and $L(z|X_i)$ by $\hat{F}_g(z|X_i)$, $\hat{G}_g(z|X_i)$ and $\hat{L}_g(z|X_i)$, respectively ($i = 1, \ldots, n$). $\hat{F}_g(z|X_i)$, $\hat{G}_g(z|X_i)$ and $\hat{L}_g(z|X_i)$ are the estimators in (1.2.1), (1.2.5) and (1.2.6), respectively, but with bandwidth sequence g instead of h.

2. For every $i = 1, \ldots, n$ independently obtain Y_i^* from $\hat{F}_g(\cdot|X_i)$, C_i^* from $\hat{G}_g(\cdot|X_i)$ and T_i^* from $\hat{L}_g(\cdot|X_i)$. 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 $T_i^* > Z_i^*$.

3. Use this resample $\{(T_1^*, Z_1^*, \delta_1^*, X_1), \ldots, (T_n^*, Z_n^*, \delta_n^*, X_n)\}$ to estimate the bootstrap analogue of the GPLE of $F(y|x)$ in (1.2.1), using the bandwidth h :

$$\hat{F}_n^*(y|x) = 1 - \prod_{i=1}^{n} \left(1 - \frac{1_{\{Z_i^* \leq y, \delta_i^* = 1\}} B_{ni}(x)}{\sum_{j=1}^{n} 1_{\{T_j^* \leq Z_i^* \leq Z_j^*\}} B_{nj}(x)} \right).$$

4. Approximate the distribution of $(nh)^{1/2}[\hat{F}_n(y|x) - F(y|x)]$ by the bootstrap distribution of $(nh)^{1/2}[\hat{F}_n^*(y|x) - \hat{F}_g(y|x)]$.

Theorem 1.2.4 *(Uniform weak consistency). Under the assumptions of Theorem 4.2 in Iglesias-Pérez and González-Manteiga (2003), for $x \in I$ and $y \in [a, b]$, we have that*

$$\sup_{t \in R} \left| P^* \left[(nh)^{1/2} (\hat{F}_n(y|x) - F(y|x)) \leq t \right] - P \left[(nh)^{1/2} (\hat{F}_n^*(y|x) - \hat{F}_g(y|x)) \leq t \right] \right|$$

converges to zero in probability, where P^* denotes the conditional probability with respect to the original sample $(T_i, Z_i, \delta_i, X_i)$, $i = 1, \ldots, n$.

1.3 Likelihood function for binary data without censoring

In this section we will present the maximum likelihood method for binary data, when no censoring is present. This methodology will be extended later to the case where censoring is present.

We have $Y_1, Y_2, \ldots, Y_n$ survival times for n patients in a clinical trial. Suppose for the moment, that we have no censoring and no truncation, meaning that for all patients that are in the study, the time of the event of interest has been observed during the period of the study. We are interested in investigating the relationship between the response probability $S(y|\boldsymbol{x})$ and the covariate vector $\boldsymbol{X} = (X_0, X_1, \ldots, X_p)$, $X_0 = 1$. We have already seen that we can model this relationship via a generalized linear model with time-dependent coefficients :

$$\phi(S(y|\boldsymbol{x})) = \beta_0(y) + \beta_1(y)x_1 + \cdots + \beta_p(y)x_p = \boldsymbol{\beta}'(y)\boldsymbol{x}, \tag{1.3.1}$$

with

$$S(y|\boldsymbol{x}) = P(Y \geq y|\boldsymbol{X} = \boldsymbol{x}) = \phi^{-1}(\boldsymbol{\beta}'(y)\boldsymbol{x}) = \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}),$$

ϕ a known monotonic and differentiable link function defined on [0,1] and $\boldsymbol{\beta}(y) = (\beta_0(y), \beta_1(y), \ldots, \beta_p(y))'$ a $(p+1)$ dimensional unknown vector of regression parameters, that depends on y and $\pi = \phi^{-1}$.

We can now apply the likelihood function for binary data in order to estimate the parameters $\boldsymbol{\beta}(y)$. For $Y_1, Y_2, \ldots, Y_n$, we have that $V_i = 1_{\{Y_i \geq y\}} \sim Bi(1, \pi(\boldsymbol{\beta}'(y)x_i))$ with the following probability function

$$P(Y_i \geq y)1_{\{Y_i \geq y\}} + (1 - P(Y_i \geq y))1_{\{Y_i < y\}} = P(Y_i \geq y)^{1_{\{Y_i \geq y\}}}(1 - P(Y_i \geq y))^{1_{\{Y_i < y\}}},$$

which gives the likelihood function :

$$\begin{aligned} L_n(\boldsymbol{\beta}(y)|\boldsymbol{x}, \pi(\boldsymbol{\beta}'(y)\boldsymbol{x})) &= \prod_{i=1}^{n} P(Y_i \geq y|\boldsymbol{X} = \boldsymbol{x}_i)^{1_{\{Y_i \geq y\}}}(1 - P(Y_i \geq y|\boldsymbol{X} = \boldsymbol{x}_i))^{1_{\{Y_i < y\}}} \\ &= \prod_{i=1}^{n} \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)^{1_{\{Y_i \geq y\}}}(1 - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i))^{1_{\{Y_i < y\}}} \\ \ln L_n(\boldsymbol{\beta}(y)|\boldsymbol{x}, \pi(\boldsymbol{\beta}'(y)\boldsymbol{x})) &= \sum_{i=1}^{n} \Big[1_{\{Y_i \geq y\}} \ln \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) + 1_{\{Y_i < y\}} \ln(1 - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i))\Big] \end{aligned}$$

The parameters $\boldsymbol{\beta}(y)$ are the solution to the following equation :

$$\frac{\partial \ln L_n(\boldsymbol{\beta}(y)|\boldsymbol{x}, \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}))}{\partial \boldsymbol{\beta}(y)} = 0.$$

$$\frac{\partial ln L_n(\boldsymbol{\beta}(y)|x, \pi(\boldsymbol{\beta}'(y)x))}{\partial \beta_r(y)} = \sum_{i=1}^{n} \left[I_{\{Y_i \geq y\}} \frac{\partial ln\pi(\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)} + I_{\{Y_i < y\}} \frac{\partial ln(\bar{\pi}(\boldsymbol{\beta}'(y)x_i))}{\partial \beta_r(y)} \right]$$

$$= \sum_{i=1}^{n} \left[I_{\{Y_i \geq y\}} \frac{\partial ln\pi(\boldsymbol{\beta}'(y)x_i)}{\partial \pi(\boldsymbol{\beta}'(y)x_i)} \frac{\partial \pi(\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)} + I_{\{Y_i < y\}} \frac{\partial ln(\bar{\pi}(\boldsymbol{\beta}'(y)x_i))}{\partial \pi(\boldsymbol{\beta}'(y)x_i)} \frac{\partial \pi(\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)} \right]$$

$$= \sum_{i=1}^{n} \left[I_{\{Y_i \geq y\}} \frac{1}{\pi(\boldsymbol{\beta}'(y)x_i)} \frac{\partial \pi(\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)} - I_{\{Y_i < y\}} \frac{1}{\bar{\pi}(\boldsymbol{\beta}'(y)x_i} \frac{\partial \pi(\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)} \right]$$

$$= \sum_{i=1}^{n} \left[I_{\{Y_i \geq y\}} - \pi(\boldsymbol{\beta}'(y)x_i) \right] \frac{1}{\pi(\boldsymbol{\beta}'(y)x_i)\bar{\pi}(\boldsymbol{\beta}'(y)x_i)} \frac{\partial \pi(\boldsymbol{\beta}'(y)x_i)}{\partial (\boldsymbol{\beta}'(y)x_i)} \frac{\partial (\boldsymbol{\beta}'(y)x_i)}{\partial \beta_r(y)}$$

$$= \sum_{i=1}^{n} \left[I_{\{Y_i \geq y\}} - \pi(\boldsymbol{\beta}'(y)x_i) \right] \frac{1}{\pi(\boldsymbol{\beta}'(y)x_i)\bar{\pi}(\boldsymbol{\beta}'(y)x_i)} \pi_\phi(\boldsymbol{\beta}'(y)x_i)x_{ir}$$

Thus

$$\frac{\partial \ln L_n(\boldsymbol{\beta}(y)|\boldsymbol{x}, \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}))}{\partial \boldsymbol{\beta}(y)} = \sum_{i=1}^{n} \left[1_{\{Y_i \geq y\}} - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) \right] \frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{x}_i))} \boldsymbol{x}_i,$$

with $\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) = \dfrac{\partial \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\partial (\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}$ and $\bar{\pi}(u) = 1 - \pi(u)$.

So, we get the equation

$$\tilde{U}_n(\boldsymbol{\beta}(y)) = \sum_{i=1}^{n} \left[1_{\{Y_i \geq y\}} - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) \right] \frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)} \boldsymbol{x}_i = 0. \tag{1.3.2}$$

The maximum-likelihood (ML) estimates of the parameters $\boldsymbol{\beta}(y)$ are then obtained by iterative weighted least squares (LS) (Newton-Raphson method), defined below.

The Fisher information for $\boldsymbol{\beta}(y)$ is

$$-E\left(\frac{\partial^2 \ln L_n(\boldsymbol{\beta}(y)|\boldsymbol{x}, \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}))}{\partial \beta_r(y) \partial \beta_s(y)} \right) = \sum_{i=1}^{n} \frac{\pi_\phi^2(\boldsymbol{\beta}'(y)x_i)}{\pi_i \bar{\pi}_i} x_{ir} x_{is}$$

$$= \{\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X}\}_{rs} \tag{1.3.3}$$

where $\boldsymbol{W}$ is a diagonal matrix of weights given by

$$\boldsymbol{W} = \mathrm{diag}\{\pi_\phi^2(\boldsymbol{\beta}'(y)x_i)/\pi_i \bar{\pi}_i\},$$

with $\pi_i = \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) = S(y|\boldsymbol{x}_i)$.

Following the Newton-Raphson procedure, parameter estimates may be obtained in the following way. Given initial estimates $\hat{\boldsymbol{\beta}}^0(y)$, we may compute the vectors $\hat{\boldsymbol{\pi}}^0$ and $\hat{\boldsymbol{\eta}}^0 = \hat{\boldsymbol{\beta}}^{0'}(y)\boldsymbol{x}$. Using these values, define the adjusted dependent variable, $\boldsymbol{Z}$, with components

$$z_i = \hat{\eta}_i^0 + (w_i - \hat{\pi}_i^0) \cdot [\pi_\phi(\boldsymbol{\beta}^{0'}(y)x_i)]^{-1}$$

where v_i is the observed values of $V_i = 1_{\{Y_i > y\}}$.

Maximum likelihood estimates satisfy the following equation :

$$\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}(y) = \boldsymbol{X}^t\boldsymbol{W}\boldsymbol{Z} \tag{1.3.4}$$

which can be solved interatively using standard LS methods. The revised estimate is

$$\hat{\boldsymbol{\beta}}^1(y) = (\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{Z},$$

where all quantities appearing on the right are computed using the initial estimate. Continue this procedure until convergence of the algorithm.

To a first order of approximation, ML estimates are unbiased with asymptotic variance equal to the inverse Fisher information matrix (1.3.3). Specifically, for large n,

$$E(\hat{\boldsymbol{\beta}}(y) - \boldsymbol{\beta}(y)) = O(n^{-1})$$
$$cov(\hat{\boldsymbol{\beta}}(y)) = (\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X})^{-1}\{1 + O(n^{-1})\}. \tag{1.3.5}$$

McCullagh and Nelder (1989) showed that failure to converge is rarely a problem, unless one or more components of $\hat{\boldsymbol{\beta}}(y)$ are infinite, which usually implies that some of the fitted probabilities are either zero or one. Also, infinite parameter estimates can occur if the data are sparse and $v_i = 0$ or $v_i = 1$ for certain components of the response vector. Some criteria are proposed to detect abnormal convergence due to the problems mentioned before (for details see Chapter 4 of McCullagh and Nelder (1989)). Abnormal convergence means that the loglikelihood is either very flat or, more likely, has an asymptote.

Some results concerning the existence and uniqueness of parameter estimates have been given by Wedderburn (1976) and by Haberman (1977). These results show that if the link function is log concave, as it is for the four functions discussed in the previous sections, and if $0 < v_i < 1$ for each i, then the log likelihood has a unique maximum at $\hat{\boldsymbol{\beta}}(y)$, which is finite.

1.4 Maximum likelihood estimation in GLM with time-dependent coefficients under censoring

Now consider that the survival time Y can be censored, thus Y_i, $i = 1, 2, \ldots, n$ are not all observed. Let $C_i, i = 1, \ldots, n$, be i.i.d. censoring times with common distribution function (DF) G. Let $Z_i = \min\{Y_i, C_i\}$ and $\delta_i = 1_{\{Y_i \leq C_i\}}$. Jung (1996) took $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ to be a vector of discrete covariates and considered the following model

$$\phi(S(y|\boldsymbol{x})) = \boldsymbol{\beta}'(y)\boldsymbol{x} = \beta_0(y) + \beta_1(y)x_1 + \ldots + \beta_p(y)x_p. \tag{1.4.1}$$

He supposed that Y and C are conditionally independent given $\boldsymbol{X}$ and that C is independent of $\boldsymbol{X}$. This implies that :

$$E\{1_{\{Z\geq y\}}|\boldsymbol{x}\} = P(Y \geq y|\boldsymbol{x})(1 - G(y))$$

Using this and a similar reasoning as for the non-censored case, Jung proposed the following estimating equation for survival data with random censoring :

$$U_n(\boldsymbol{\beta}(y)) := n^{-1/2}\sum_{i=1}^{n}\left\{\frac{1_{\{Z_i\geq y\}}}{1-\hat{G}(y)} - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\right\}\frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\bar{\pi}(\boldsymbol{\beta}'(z)\boldsymbol{x}_i)}\boldsymbol{x}_i = 0, \tag{1.4.2}$$

where $\hat{G}(y)$ is the Kaplan-Meier estimator of $G(y)$. Like $\tilde{U}_n(\boldsymbol{\beta}(y))$ in (1.3.2), $U_n(\boldsymbol{\beta}(y))$ is a monotonic differentiable function of $\boldsymbol{\beta}(y)$ and equation (1.4.2) can be solved by the Newton-Raphson method.

Theorem 1.4.1 *Under the regularity assumptions in Appendix 1 and Appendix 2 in Jung (1996) the estimator $\hat{\boldsymbol{\beta}}(y)$ as the solution of (1.4.2) is consistent and $n^{1/2}(\hat{\boldsymbol{\beta}}(y) - \boldsymbol{\beta}(y))$ is asymptotically normal with zero mean and variance-covariance matrix $\boldsymbol{V} = \boldsymbol{A}^{-1}\Gamma\boldsymbol{A}^{-1}$ where $\Gamma = E[\xi_1\xi_1']$, with*

$$\boldsymbol{A} = E[\boldsymbol{X}\boldsymbol{X}'\pi_\phi^2(\boldsymbol{\beta}'(y)\boldsymbol{X})/\{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X})\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X})\}],$$

$$\xi_i = \left\{\frac{1_{\{Z_i\geq y\}}}{1-G(y)} - \pi_i\right\}\frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}\boldsymbol{X}_i + h(\boldsymbol{\beta}(y))\int_0^y \frac{dN_i^c(s) - W_i(s)d\Lambda^c(s)}{y(s)},$$

where $h(\boldsymbol{\beta}(y))$ is the a.s. limit of $h_n(\boldsymbol{\beta}(y))$ as $n \to \infty$,

$$h_n(\boldsymbol{\beta}(y)) = \{1 - G(y)\}^{-1}n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\frac{1_{\{Z_i\geq y\}}\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}$$

and $N_i^c(s) = 1_{\{Z_i\leq s,\delta_i=0\}} = (1-\delta_i)1_{\{Z_i\leq s\}}$, $W_i(s) = 1_{\{Z_i\geq s\}}$, $\Lambda^c(s) = -\ln(1 - G(s))$.

A major assumption for the estimator proposed by Jung (1996) was that the censoring variables C_i, $i = 1, \ldots, n$ are independent of the covariates. However, in some clinical trials, the censoring mechanism can depend on the covariates.

Subramanian (2001) proposes another estimator for the case where the censoring C can depend on the covariate and when the covariate takes values in a finite set. The idea is that when there are discrete covariates and there is no censoring, the basic score function for logistic regression (1.3.2) analyzed by Jung (1996), can be expressed in terms of the conditional empirical survival function of Y given $\boldsymbol{X}$. Then, when the failure times are censored, we can replace the conditional empirical survival function with a conditional Kaplan-Meier (KM) estimator, and this yields a new estimating function of $\boldsymbol{\beta}(z)$.

For uncensored data, rewrite the sample $(Y_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, as $Y_{k,m}$, $m = 1, \ldots, n_k$ for $k = 1, \ldots, K$, where $Y_{k,m}$ corresponds to Y_i with covariate $\boldsymbol{X}_i$ having the value $\boldsymbol{x}_k$. Let $S^{(n)}(y|\boldsymbol{x}_k) = \frac{1}{n_k} \sum_{m=1}^{n_k} 1_{\{Y_{k,m} > y\}}$ denote the empirical survival function based on $Y_{k,m}$, $m = 1, \ldots, n_k$. With this notation, equation (1.3.2) can be rewritten in an alternative form

$$U_n^*(\boldsymbol{\beta}(y)) = n^{-1/2} \sum_{i=1}^{n} \left\{ S^{(n)}(y^-|\boldsymbol{x}_i) - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) \right\} \frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)} \boldsymbol{x}_i. \qquad (1.4.3)$$

When data are censored, rewrite the sample (Z_i, δ_i, X_i), $i = 1, \ldots, n$ as $(Z_{k,m}, \delta_{k,m})$, $m = 1, \ldots, n_k$, for $k = 1, \ldots, K$, where $(Z_{k,m}, \delta_{k,m})$ corresponds to (Z_i, δ_i) with covariate $\boldsymbol{X}_i$ having the value $\boldsymbol{x}_k$. Let $\hat{S}(y|\boldsymbol{x}_k)$ be the local KM estimator based on the pairs $(Z_{k,m}, \delta_{k,m})$, $m = 1, \ldots, n_k$, i.e. $\hat{S}(y|\boldsymbol{x}_k) = \displaystyle\prod_{\substack{m=1 \\ Z_{k,m} \le y}}^{n_k} \left(1 - \frac{1}{n_k - m + 1}\right)^{\delta_{k,m}}$. Now replace the local empirical survival function by the local KM estimator in (1.4.3) to get

$$\hat{U}_n(\boldsymbol{\beta}(y)) = n^{-1/2} \sum_{i=1}^{n} \left\{ \hat{S}(y^-|\boldsymbol{x}_i) - \pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i) \right\} \frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{x}_i)} \boldsymbol{x}_i. \qquad (1.4.4)$$

Notice that for the discrete case, the estimator proposed by Subramanian (2001) is the same as the one proposed by Jung (1996) in the case where the censoring can depend on the covariate.

The following result describes the asymptotic properties of $\hat{\boldsymbol{\beta}}(y)$.

Theorem 1.4.2 *Under the assumptions of Theorem 1 in Subramanian (2001), the estimator* $\hat{\boldsymbol{\beta}}(y)$ *as the solution of (1.4.4) is consistent and* $n^{1/2}(\hat{\boldsymbol{\beta}}(y) - \boldsymbol{\beta}(y))$ *is asymptotically normal with zero mean and variance-covariance matrix* $\boldsymbol{A}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}^{-1}$, *where* $\boldsymbol{\Gamma} = E[\xi_1^* \xi_1^{*'}]$ *with*

$$\boldsymbol{A} = E[\boldsymbol{X}\boldsymbol{X}'\pi_\phi^2(\boldsymbol{\beta}'(y)\boldsymbol{X})/\{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X})\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X})\}],$$

$$\xi_i^* = \xi_i(Z_i, \delta_i, X_i, y-) \cdot \frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X}_i)}\boldsymbol{X}_i$$

where

$$\xi(Z, \delta, \boldsymbol{X}, y) = 1 - F(y|\boldsymbol{X})\left[g(Z \wedge y|\boldsymbol{X}) + \frac{1}{\bar{H}(y|\boldsymbol{X})}1_{\{Z \le y, \delta = 1\}}\right],$$

$g(y|\boldsymbol{x}) = \int_0^y [\bar{H}(s|\boldsymbol{x})]^{-2} d\bar{H}_1(s|\boldsymbol{x})$, $\bar{F}(s|\boldsymbol{x}) = 1 - F(s|\boldsymbol{x})$ *and* $\bar{H}(s|\boldsymbol{x}) = 1 - H(s|\boldsymbol{x})$.

Subramanian (2004) proposed an extension of the estimator in Subramanian (2001) to the case where the censoring depends on a one-dimensional continuous covariate, X. He used the kernel conditional Kaplan-Meier estimator, $\hat{\Lambda}(y|x)$, of the conditional cumulative hazard function, $\Lambda(y|x)$, proposed by Dabrowska (1989) to estimate $S(y|x)$ nonparametrically, and plugged it into equation (1.4.4).

Specifically, $\hat{S}(y|x) = \prod_{s \le y}(1 - d\hat{\Lambda}(s|x))$ and the resulting estimating equation is the following

$$\bar{U}_n(\boldsymbol{\beta}(y)) = \sum_{i=1}^n \left\{\hat{S}(y - |x_i) - \pi(\boldsymbol{\beta}'(y)x_i)\right\} \frac{\pi_\phi(\boldsymbol{\beta}'(y)x_i)}{\pi(\boldsymbol{\beta}'(y)x_i)\bar{\pi}(\boldsymbol{\beta}'(y)x_i)} x_i. \tag{1.4.5}$$

Theorem 1.4.3 *Suppose that the assumptions of Theorem 1 in Subramanian (2004) hold, then the estimator* $\hat{\boldsymbol{\beta}}(y)$ *as solution of (1.4.5) is consistent and asymptotically normal :*

$$n^{1/2}(\hat{\boldsymbol{\beta}}(y) - \boldsymbol{\beta}(y)) \xrightarrow{d} N(0, \boldsymbol{A}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}^{-1}).$$

with

$$\boldsymbol{A} = E[\boldsymbol{X}\boldsymbol{X}'\pi_\phi^2(\boldsymbol{\beta}'(y)\boldsymbol{X})/\{\pi(\boldsymbol{\beta}'(y)\boldsymbol{X})\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X})\}]$$

and

$$\Gamma = E\left(\frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{X})}{\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X})}\left(\frac{\pi_\phi(\boldsymbol{\beta}'(y)\boldsymbol{X})}{\bar{\pi}(\boldsymbol{\beta}'(y)\boldsymbol{X})}\right)' \int_0^\tau \frac{dH_1(s|\boldsymbol{X})}{H^2(s|\boldsymbol{X})}\right)$$

In Chapter 2, our proposed estimator will be compared with the estimators of Jung (1996) and Subramanian (2001, 2004).

1.5 Least squares estimation in GLM with time-constant coefficients under censoring and/ or truncation

An alternative approach to the maximum likelihood estimator is the least squares one. Akritas (1996) defined a least squares estimator $\hat{\boldsymbol{\beta}}$ for the parameters $\boldsymbol{\beta}$ of the general polynomial model defined in (1.1.8) :

$$T(F) = \int_0^1 F^{-1}(s|x)J(s)ds = \beta_0 + \beta_1 x + \ldots + \beta_p x^p,$$

in the case where we have censoring or truncation and Cao and González-Manteiga (2008) extended this estimator to the case where we can have both censoring and truncation.

The idea is very simple. Let $\hat{m}_i = T(\hat{F}_n(\cdot|X_i))$ be an estimator of $m_i = T(F(\cdot|X_i))$ for $i = 1, 2, \ldots, n$, where $\hat{F}_n(\cdot|X_i)$ is an estimator for the conditional distribution function corresponding to the case under study : the conditional estimator of Beran (1981) for the censoring case, the conditional estimator of La Valley and Akritas (2005) for the truncation case or the one of Iglesias-Perez and González-Manteiga (1999) which is valid for both censoring and truncation. To avoid definition problems with $T(\hat{F}_n(\cdot|X_i)) = \int_0^1 \hat{F}_n^{-1}(s|X_i)J(s)ds$ we modify the estimator $\hat{F}_n(\cdot|X_i)$, if necessary, forcing it to attain the value 1 in the largest point with positive probability mass. Then define the least squares estimator as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \hat{\psi}_n(\boldsymbol{\beta}), \tag{1.5.1}$$

where

$$\hat{\psi}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \left(\hat{m}_i - (\beta_0 + \beta_1 X_i + \ldots + \beta_p X_i^p)\right)^2, \tag{1.5.2}$$

or equivalently

$$\hat{\boldsymbol{\beta}} = (\mathcal{X}^t\mathcal{X})^{-1}\mathcal{X}^t\hat{m},$$

where $\hat{m} = (\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_n)^t$ and $\mathcal{X} = \begin{pmatrix} 1 & X_1 & \ldots & X_1^p \\ 1 & X_2 & \ldots & X_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \ldots & X_n^p \end{pmatrix}$ is the design matrix.

Akritas (1996) proved (see the following theorem) the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$ in the case of censoring or truncation.

Theorem 1.5.1 *Assume the covariate X has finite moments of order $2p$, and that $n^{-1}\mathcal{X}'\mathcal{X} \rightarrow C$, $n^{-1}\mathcal{X}'\Sigma_J\mathcal{X} \rightarrow D_J$, almost surely, as $n \rightarrow \infty$ with C a nonsingular matrix. Further, assume that Conditions C1, C2 and C3 in Theorem 2.1 of Akritas (1996) hold. Then, conditionally on $X_1 = x_1, \ldots, X_n = x_n$,*

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

where N is the (p+1)-variate normal distribution, $\Sigma_\beta = C^{-1}D_JC^{-1}$, $\Sigma_J = diag(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_i^2 = var(\int_{-\infty}^{\infty} \xi(W_i, X_i, y)J(F(y|X_i))dy|X_i)$ and $\xi(W, X, y)$ a suitable function which, conditionally on X, has mean zero and finite second moment (see Akritas (1996) for details).

using the same arguments we can also obtain the asymptotic normality of $\hat{\boldsymbol{\beta}}$ in the case of censoring and truncation

Grigoletto and Akritas (1999) defined and studied a similar estimator for the proportional hazards model (1.1.1), the additive risks model (1.1.5) and the proportional odds model (1.1.6) in the case where censoring or truncation were present. Cao and González-Manteiga (2008) extended this estimator to the case where we can have both censoring and truncation. The basic idea for their estimator is that any of the mentioned semiparametric models imply that some transformation Ω_i of $\Lambda(y|X_i)$ follows a linear model in terms of the covariate, i.e.

$$\Omega_i = \beta_0 + \beta_1 X_i + \ldots + \beta_p X_i^p, \quad i = 1, \ldots, n$$

Then we can estimate $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \bar{\psi}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\Omega}_i - (\beta_0 + \beta_1 X_i + \ldots + \beta_p X_i^p) \right)^2 \qquad (1.5.3)$$

where $\hat{\Omega}_i$ is some estimation of Ω_i that will change from one model to other. An explicit expression for the estimator in (1.5.3) can be easily derived :

$$\hat{\boldsymbol{\beta}} = (\mathcal{X}^t\mathcal{X})^{-1}\mathcal{X}^t\hat{\boldsymbol{\Omega}},$$

with $\hat{\boldsymbol{\Omega}} = (\hat{\Omega}_1, \hat{\Omega}_2, \ldots, \hat{\Omega}_n)^t$.

For the proportional hazards model

$$\hat{\Omega}_i = \int_0^\infty \ln \hat{\Lambda}_n(s|X_i)dW(s) \tag{1.5.4}$$

with $W(s)$ a weight function, $W(s) \geq 0$, $s \in [0,\infty)$, $\int_0^\infty dW(s) = 1$ and $\hat{\Lambda}_n(y|x)$ is the conditional cumulative hazard function estimator based on the conditional estimator (1.2.1), i.e.

$$\hat{\Lambda}_n(y|x) = \int_{-\infty}^y \frac{d\hat{H}_{1n}^*(s|x)}{\hat{C}_n(s|x)}. \tag{1.5.5}$$

For the additive risks model

$$\hat{\Omega}_i = \int_0^\infty \ln \hat{\Lambda}_n(s|X_i)d\tilde{W}(s), \tag{1.5.6}$$

with $\tilde{W} = \dfrac{W}{\int_0^\infty sdW(s)}$.

For the proportional odds model

$$\hat{\Omega}_i = \int_0^\infty \mathrm{logit}\left(1 - \exp\left(-\hat{\Lambda}_\mathrm{n}(\mathrm{s}|\mathrm{X_i})\right)\right)\mathrm{dW(s)}. \tag{1.5.7}$$

When censoring or truncation is present, Grigoletto and Akritas (1999) showed the asymptotic normality of the estimator in (1.5.3).

Theorem 1.5.2 *Suppose Condition 1 in Grigoletto and Akritas (1999) is satisfied, the covariate X has finite second-order moments, and let $n^{-1}\mathcal{X}'\boldsymbol{\Sigma}_W\mathcal{X} \xrightarrow{a.s.} \boldsymbol{D}_W$ and $n^{-1}\mathcal{X}'\mathcal{X} \xrightarrow{a.s.} \boldsymbol{M}$ as $n \to \infty$. In addition, assume that the matrix $\boldsymbol{M}$ is not singular. Then, conditionally on $X_1 = x_1, \ldots, X_n = x_n$, the estimator $\hat{\boldsymbol{\beta}}$ satisfies*

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

where N is the $(p+1)$-variate normal distribution, $\boldsymbol{\Sigma}_W(\hat{\boldsymbol{\beta}}) = \boldsymbol{M}^{-1}\boldsymbol{D}_W\boldsymbol{M}^{-1}$, $\boldsymbol{\Sigma}_W = diag(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_i^2 = var(\int_0^\infty \psi(D_i, X_i, y)W(dy)|X_i)$ and $\psi(D, X, y)$ a suitable function that will change from one model to another and which, conditionally on X, has mean zero (see Grigoletto and Akritas (1999) for details).

As for the polynomial model, we can extend the theorem to the case where censoring and truncation are present, by replacing the conditional estimator of $F(y|x)$ by the estimator in Iglesias-Pérez and González-Manteiga (1999) valid for both censoring and truncation.

In Chapter 2, we will extend the estimator of Cao and González-Manteiga (2008) to the case where the coefficients are time-dependent. The asymptotic properties of the proposed estimator are also studied.

1.6 Goodness-of-fit tests in survival analysis

In practice it is of great use to have at hand a method to check the validity of the models mentioned in Section 1.2.

When considering models with time-constant coefficients, lots of methods exist in order to test if the model is suitable for the data. For the Cox model, (1.1.1), several graphical techniques have been proposed by, e.g. Kay (1977), Cox (1979), Kalbfleisch and Prentice (1980), Andersen (1982), Arjas (1988), all based on the plot of the cumulative hazard function. Schoenfeld (1980), Moreau et al (1985), Moreau et al (1986), Andersen et al (1992), Grambsch and Therneau (1994), Parzen and Lipsitz (1999) proposed instead an omnibus goodness-of-fit test for (1.1.1) based on the notion of partitioning the subjects into mutually exclusive regions based on their covariate values. A goodness-of-fit statistic is then calculated as a quadratic form in the observed minus predicted number of failures in these regions. Gill and Schumacher (1987) proposed instead a test based on the comparison of different generalized rank estimators of the relative risk. For the accelerated failure time model Yang (1998) extended the idea in Gill and Schumacher (1987), while Chen (2001) presented a Kolmogorov-Smirnov type test. The additive risks model has been studied, e.g. by Yuen and Burke (1997) with a Kolmogorov-Smirnov and a Cramer-von Mises type statistics and Kim and Lee (1998) who presented two tests : one is based on the martingale residuals and the other on the difference between weighted estimators of the excess risk. All the above tests were studied only in the case of censored responses.

Cao and González-Manteiga (2008) presented a general goodness-of-fit test for the general polynomial regression model (1.1.8), proportional hazard model (1.1.1), additive risks model (1.1.5) and proportional odds model (1.1.6) in the case where censoring and/or truncation are present.

For the general polynomial regression model, the problem under study is to test

$$\begin{aligned} &H_0 : \exists \boldsymbol{\beta} \in \mathbb{R}^{p+1} \text{ such that equation (1.1.8) holds} \\ &\qquad \text{versus the alternative} \\ &H_1 : \text{equation (1.1.8) does not hold for any } \boldsymbol{\beta} \in \mathbb{R}^{p+1}. \end{aligned} \tag{1.6.1}$$

An intuitive way of measuring the discrepancy between the hypothesized model and the data is to consider the stochastic process (1.5.2) at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$:

$$\hat{\psi}_n(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{m}_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + \ldots + \hat{\beta}_p X_i^p) \right)^2 .$$

This is a kind of L_2-distance between the transformed data and the response predicted by the fitted model. Hence, it is reasonable to reject H_0 when $\hat{\psi}(\hat{\boldsymbol{\beta}})$ is large.

In the proportional hazards model, additive risks model and proportional odds model, the hypothesis testing is completely similar to (1.6.1) but replacing equation (1.1.8) by (1.1.1), (1.1.5) or (1.1.6), respectively. The definition of the statistic is also parallel for these models :

$$\bar{\psi}_n(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\Omega}_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + \ldots + \hat{\beta}_p X_i^p) \right)^2 .$$

These test statistics have to be multiplied by a normalizing sequence in order to have a limit distribution. This leads to the following test statistics :

$$\begin{aligned} T_n^{(1)} &= n\sqrt{h}\hat{\psi}_n(\hat{\boldsymbol{\beta}}) \text{ for model (1.1.8)} \\ T_n^{(2)} &= n\sqrt{h}\bar{\psi}_n(\hat{\boldsymbol{\beta}}) \text{ for models (1.1.1), (1.1.5) and (1.1.6).} \end{aligned} \tag{1.6.2}$$

Theorem 1.6.1 *Assume that conditions C1-C14 in Theorem 1 of Cao and González-Manteiga (2008) hold. Then, under H_0,*

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

with $b_{0h} = h^{-1/2} K^{(2)}(0) \int \sigma^2(x)dx$ and $V = 2K^{(4)}(0) \int \sigma^4(x)dx$. Here $K^{(2)}(u) = K * K(u)$ and $K^{(4)}(u) = K * K * K * K(u)$, where $*$ denotes convolution and

$$\sigma^2(x) = Var(\eta(Z, T, \delta, x)|X = x),$$

with

$$\eta(Z,T,\delta,x) = \int (1 - F(y|x))\xi(Z,T,\delta,x,y)J(F(y|x))dy$$

and $\xi(Z,T,\delta,x,y)$ defined in (1.2.3).

Theorem 1.6.2 *Assume conditions $C1'$, C3-C10, $C11'$, $C12'$, $C13'$ and $C14'$ in Theorem 3 of Cao and González-Manteiga (2008). Then , under H_0,*

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

with $b_{0h} = h^{-1/2}K^{(2)}(0)\int \sigma^2(x)dx$ and $V = 2K^{(4)}(0)\int \sigma^4(x)dx$ and

$$\sigma^2(x) = \begin{cases} \sigma_1^2(x) = Var(\eta_1(Z,T,\delta,x)|X=x) & \text{for model (1.1.1)} \\ \sigma_2^2(x) = Var(\eta_2(Z,T,\delta,x)|X=x) & \text{for model (1.1.5)} \\ \sigma_3^2(x) = Var(\eta_3(Z,T,\delta,x)|X=x) & \text{for model (1.1.6)} \end{cases}$$

$\eta_1(Z,T,\delta,x) = \int \xi(Z,T,\delta,x,y)\frac{dW(y)}{\Lambda(y|x)}$, $\eta_2(Z,T,\delta,x) = \int \xi(Z,T,\delta,x,y)d\tilde{W}(y)$, $\eta_3(Z,T,\delta,x) = \int \xi(Z,T,\delta,x,y)\frac{dW(y)}{F(y|x)}$ and $\xi(Z,T,\delta,x,y)$ is defined in (1.2.3).

When the coefficients are time-dependent, little work is available, see e.g. Murphy (1993) and Marzec and Marzec (1997) who proposed a goodness-of-fit test of the Kolmogorov-Smirnov and Cramér-von Mises types for the Cox model with time-dependent coefficients. For the additive hazard model, a graphical method was proposed by Klein and Moeschberger (1997), while McKeague and Utikal (1991) proposed a goodness-of-fit test. All these tests are valid only for censored responses.

In Chapter 3 we extend the methodology of Cao and González-Manteiga (2008) to the case where the coefficients are time-dependent. We study this problem in the framework of the generalized linear models (1.1.15).

Chapter 2

Least Squares Estimator

We will focus in this chapter on a least squares estimator of the time-dependent coefficients in the generalized conditional linear model (1.1.15), when the response variable may be subject to left-truncation and/or right-censoring.

Some notations as well as the model are given in Section 2.1. The least squares estimator is introduced in Section 2.2. In Section 2.3 we give the assumptions needed for the asymptotic properties presented in Section 2.4. In Section 2.5 we propose a bootstrap based method for the selection of the smoothing parameter needed in the nonparametric estimation (1.2.1) of $S(z|X)$, while in Section 2.6 we study the finite sample properties of the estimator, as well as a comparison with the maximum likelihood estimator proposed in Subramanian (2001, 2004). Finally Section 2.7 contains the proofs. In the Appendix of this chapter some technical details used in the proofs are given.

This chapter is mainly based on Teodorescu et al (2008).

2.1 Introduction and notations

As already mentioned in Chapter 1, in many practical situations the parameters in the Cox model, proportional odds model, etc are better considered to be time-dependent in order to better capture the possible changes in the influence of the covariates on the survival function.

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

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \ldots + \beta_p(z)X^p, \tag{2.1.1}$$

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 $\phi(u) = \ln(\frac{u}{1-u})$ gives the logistic model, $\phi(u) = -\ln(u)$ gives the additive risk model and $\phi(u) = \ln(-\ln(u))$ leads to the proportional hazards model.

We also reintroduce 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 $\alpha(x) = P(T \leq Z|x)$, 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 $T \leq Z$.

2.2 Least squares estimator

The estimator $\hat{\boldsymbol{\beta}}(z)$ that we propose for the parameter $\boldsymbol{\beta}(z) = (\beta_0(z), \ldots, \beta_p(z))^t$ in model (2.1.1) 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 (1.2.1) of the conditional distribution, proposed by Iglesias-Pérez and González-Manteiga (1999) :

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

where

$$B_{ni}(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_j K\left(\frac{x-X_j}{h}\right)}$$

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

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

$$\hat{\boldsymbol{\beta}}(z) = \begin{pmatrix} \hat{\beta}_0(z) \\ \hat{\beta}_1(z) \\ \vdots \\ \hat{\beta}_p(z) \end{pmatrix} = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \hat{\boldsymbol{\phi}}(z), \tag{2.2.1}$$

where

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

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 described 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. For $0 < z < Z_{(1)}$ it is clear that $\hat{\boldsymbol{\beta}}(z) = \boldsymbol{\beta}(0)$, where $Z_{(i)}$, $i = 1, \dots, n$, denote the ordered statistics of the observed z.

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. Two examples are given in Chapter 4, where we analyze the larynx cancer data and the gastric cancer data described in the Introduction, that contain one continuous and three binary covariates and one continuous and five binary covariates, respectively.

2.3 Assumptions

The following assumptions are needed in order to obtain the asymptotic properties of the proposed estimator $\hat{\boldsymbol{\beta}}(z)$. Conditions (H1)–(H6) come from the i.i.d. representation of $\hat{F}_n(y|x) - F(y|x)$ in Iglesias-Pérez and González-Manteiga (1999) (See Theorem 1.2.1), that is needed in our proof.

(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 $a, b \in \mathbb{R}$, $a < b$ 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 $\boldsymbol{A} = (a_{ij})_{i,j=0}^{p}$, with $a_{ij} = E(X^{i+j} w(X))$, is nonsingular.

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

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

Remark 2.2.1 Condition (H2) is in Dabrowska (1989) and comes from the fact that we need to stay away from the boundaries of the domain of the covariate while estimating the survival function $S(z|x)$, to avoid boundary effects.

2.4 Asymptotic properties of the least squares estimator

Let $\boldsymbol{\phi}(z) = (\phi(S(z|X_1)), \ldots, \phi(S(z|X_n)))^t$. Model (2.1.1) implies that $\boldsymbol{\phi}(z) = \boldsymbol{X}\boldsymbol{\beta}(z)$, which leads to

$$\boldsymbol{\beta}(z) = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{\phi}(z), \tag{2.4.1}$$

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 2.4.1 *Suppose that conditions (H1) through (H11) hold. Then, for any $a \leq z \leq b$,*

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

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

$$\sigma_{ij}(z) = \int_I x^{i+j} \tilde{w}^2(x) S^2(z|x) \phi'(S(z|x))^2 \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)} m^*(x) dx, \tag{2.4.2}$$

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

Corollary 2.4.2 *Suppose that conditions (H1) through (H11) hold. Then*

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

Remark 2.4.1 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 several 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 where we also have truncation (Tsai et al (1987)).

Remark 2.4.2 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 (2.1.1). Note that this estimator is in general non-monotone. A convenient and satisfactory solution is to keep the estimator constant until it starts decreasing again.

Remark 2.4.3 The choice of the weight matrix $\boldsymbol{W}$ can be done in such a way to accommodate heteroscedasticity in the model.

Remark 2.4.4 The $\hat{\beta}_j(z)$ are step functions, but can be smoothed in z using, e.g., the kernel method.

Remark 2.4.5 Note that the above methodology is not only applicable to polynomial models. Consider for example a general nonlinear model of the form :

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

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

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

When $v(\boldsymbol{\beta}(z), x) \equiv \beta_0(z) + \beta_1(z)x + \ldots + \beta_p(z)x^p$, this estimator coincides with (2.2.1). Theorem 2.4.1 can be extended in a similar way to this more general estimator.

2.5 Bandwidth selection

The estimator $\hat{\boldsymbol{\beta}}(z)$ defined in Section 2.2, is based on the kernel estimator (1.2.1) of the conditional survival function $S(z|X)$. Therefore, a bandwidth parameter h needs to be selected. We propose a bootstrap procedure which selects for a fixed z, the bandwidth for which the estimated mean squared error (MSE) of $\hat{\boldsymbol{\beta}}(z)$ is minimal. It suffices to consider the uncensored observations, since the estimator $\hat{\boldsymbol{\beta}}(z)$ only changes at these points. The procedure is as follows :

1. For fixed z consider values for $h \in \{h_1, \ldots , h_r\}$, a fine grid of bandwidths in the interval $(0, \mu(\text{supp}(X)))$, where μ is the Lebesgue measure.
2. For each h_j $(j = 1, \ldots , r)$:

   a) Choose a pilot bandwidth, g_j, (usually larger than h_j) to estimate $S(z|X_i)$, $G(z|X_i)$ and $L(z|X_i)$ by $\hat{S}_{g_j}(z|X_i)$, $\hat{G}_{g_j}(z|X_i)$ and $\hat{L}_{g_j}(z|X_i)$, respectively $(i = 1, \ldots, n)$, where $\hat{S}_{g_j}(z|X_i)$ is the estimator in (1.2.1),
   $$\hat{G}_{g_j}(z|x) = 1 - \prod_{i=1}^{n}\left(1 - \frac{1_{\{Z_i \le z, \delta_i = 0\}} B_{ni}(x)}{C_n(Z_i|x)}\right) \ , \ \hat{L}_{g_j}(z|x) = \prod_{i=1}^{n}\left(1 - \frac{1_{\{T_i > z\}} B_{ni}(x)}{C_n(T_i|x)}\right),$$
   and the subscript g_j indicates the bandwidth we are working with.

   b) Replace $S(z|X_i)$ by $\hat{S}_{g_j}(z|X_i)$ in (2.1.1) and estimate $\beta_0(z), \ldots , \beta_p(z)$ by the least squares estimator in (2.2.1) to obtain $\hat{\beta}_{0,g_j}(z), \ldots , \hat{\beta}_{p,g_j}(z)$. Plug these estimators into (2.1.1) 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 $Z_i^* = \min\{Y_i^*, C_i^*\}$, $\delta_i^* = 1_{\{Y_i^* \le C_i^*\}}$ and simulate new values Y_i^*, C_i^* and T_i^* if $T_i^* > Z_i^*$.

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

   e) Repeat the steps c)-d) B times and compute the bootstrap estimator of the mean squared error (MSE) :
   $$MSE^*_{h_j}(z) = \sum_{k=0}^{p}\left\{\frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}^*_{k,h_j,b}(z) - \hat{\beta}_{k,g_j}(z))^2\right\} \qquad (2.5.1)$$

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

Remark 2.5.1 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 sizes, this might lead to better approximations than the normal limit established in Theorem 2.4.1.

Remark 2.5.2 The asymptotic validity of a slight variation of the above bootstrap procedure has been established in Theorem 1.2.4 by Iglesias-Pérez and González-Manteiga (2003). In fact, they resampled from $\hat{S}_g(z|X_i)$, $\hat{G}_g(z|X_i)$ and $\hat{L}_g(z|X_i)$ for each X_i $(i = 1, \ldots, n)$ in order to obtain Y^*_j, C^*_j and T^*_j respectively. Bootstrapping from $\tilde{S}$ instead of $\hat{S}$ allows us to actually mimic the model.

Remark 2.5.3 In practice one may want to choose a « global » optimal bandwidth for all z. For this, we can use the following mean integrated squared error (MISE) instead of the MSE in (2.5.1)

$$MISE^*_{h_j} = \int_a^b MSE^*_{h_j}(z)dz, \tag{2.5.2}$$

where a and b are defined in condition (H2) (d). Thus, we choose the h among $\{h_1, \ldots, h_r\}$ that minimizes the $MISE^*_{h_j}$, $j = 1, \ldots, r$.

Remark 2.5.4 The optimal bandwidth selected with this procedure is the optimal among the grid of bandwidths considered in the interval $(0, \mu(\text{supp}(X)))$, thus the procedure is influenced by the choice of the grid.

2.6 Numerical results

In this section, we will first conduct some simulations in order to compare the proposed least squares estimator (LS) in (2.2.1) with the maximum likelihood estimator (ML) in (1.4.2) proposed by Jung (1996) and the ones in (1.4.3) and (1.4.5) proposed by 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 2.5.

Along the simulations, the following model is considered :

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x. \tag{2.6.1}$$

In the discrete case, model 1 considers that X is uniformly distributed in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, $Y|_{X=x} \sim \text{Logistic}(0, \frac{\pi^2}{3(4x)^2})$ (i.e. $F(y|x) = 1/(1 + \exp(4xy))$, $E(Y|x) = 0$ and $\text{Var}(Y|x) = \pi^2/\{3(4x)^2\}$), $\exp(C)|_{X=x} \sim U[0, dx]$, where $d > 0$ will be chosen according to the desired censoring probability, and $\phi(u) = \ln(\frac{u}{1-u})$ (logistic model), which gives us the true model $\phi(S(z|x)) = -4zx$. A similar model has also been considered by Subramanian (2001). The sample size is taken $n = 200$, the number of Monte Carlo simulations is $M = 10000$ and d is taken to be 0.008, 0.022, 0.05 and 0.1 in order to give 10%, 20%, 30% and 40% of censoring, respectively. For estimating the survival function we use the Kaplan-Meier estimator, since there is no truncation and no smoothing is required. For $z = 0.1$ the results are given in Table 2.1. We notice that the results are very similar for the two methods in the case of censoring and in the presence of discrete covariates. Other simulations not reported here lead to similar conclusions : the difference between the two procedures is only very minor, regarding both bias and variance.

Model 2 deals with the continuous case, $X \sim U[0, 1]$, $Y|_{X=x} \sim \text{Exp}(4x)$, $C|_{X=x} \sim \text{Exp}(dx)$ with $d > 0$ that gives different censoring probabilities, and $\phi(u) = \ln(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = -4zx$. Note that $\phi(S(z|X = 0)) = \beta_0(z)$ and $\phi(S(z|X = 1)) = \beta_0(z) + \beta_1(z)$. The sample size is taken $n = 100$, $M = 10000$ Monte Carlo simulations are conducted and d is taken to be 1 and 8/3 in order to give 20% and 40% of censoring, respectively. Since we have a one-dimensional continuous covariate, a bandwidth, h, is needed in order to estimate $S(z|x)$. We worked with $h \in \{0.2, 0.25, 0.3, 0.35, 0.4\}$. The Nadaraya-Watson weights are calculated based on the Epanechnikov kernel $K(u) = 1_{\{-1 \le u \le 1\}} \cdot 3(1 - u^2)/4$ and the weight function $\tilde{w}(x) = 1_{\{0.025 \le x \le 0.975\}}$ has been chosen in order to avoid boundary problems.

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

Table 2.1 : Comparison between the ML and LS methods for model 1, at point $z = 0.1$.

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

Table 2.2 : Comparison between the ML and LS methods for model 2, at point $z = 0.3$ ($x_2 = 1$).

Table 2.2 shows the results for $z = 0.3$. The table shows the bias and MSE of the estimators of $\beta_0(z)$ and $\beta_1(z)$, and also of the estimators of the regression function

$\beta_0(z) + \beta_1(z)x$, evaluated at the endpoints of the support of X, namely at 0 and 1. The values are computed with the LS and the ML methods and they correspond to the minimum in h of the total MSE, i.e. $TMSE(h) = MSE_h(\hat{\beta}_0(z)) + MSE_h(\hat{\beta}_1(z))$. The optimal bandwidth, that minimizes the TMSE(h), is also displayed. The results in Table 2.2 show that the LS method behaves better than the ML procedure. In the simulation studies, we also noticed that the ML algorithm had some serious convergence problems. In the case of 20% censoring we noticed 393 divergences on a total of 10000, while for the 40% censoring case 1323 on a total of 10000.

Table 2.3 gives the results for model 2 computed only with the LS method and with the bandwidth estimated by means of the bootstrap procedure described in Section 2.5. The bandwidth is selected from the grid $\{0.2, 0.25, 0.3, 0.35, 0.4\}$. $B = 100$ bootstrap replications are generated each time in order to compute the bootstrap version of the MSE and $M = 1000$ Monte Carlo simulations are conducted. We notice that the results are very comparable to those of Table 2.2.

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

Table 2.3 : MSE of the LS estimator for model 2 using the bootstrap bandwidth selector, at point $z = 0.3$ ($x_2 = 1$).

Model 2.3 is a variation of model 2, where a truncation variable has been added : $T|_{X=x} \sim \text{Exp}(rx)$, where $r > 0$ controls the probability of truncation and is taken to be 45 and 22 in order to give 10% and 20% of truncation, respectively. The results are given in Tables 2.4 and 2.5. No comparison with other methods is possible here. The tables show similar results to those obtained for model 2, when we had only censoring, thus the MSE increases with the percentage of censoring and truncation. For Table 2.4 we conducted $M = 10000$ Monte Carlo simulations, while for Table 2.5, $M = 1000$ and $B = 100$ for the bootstrap replications.

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

Table 2.4 : MSE of the LS estimator for model 3, at point $z = 0.3$ ($x_2 = 1$).

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

Table 2.5 : MSE of the LS estimator for model 3 using the bootstrap bandwidth selector, at point $z = 0.3$ ($x_2 = 1$).

2.7 Proofs

2.7.1 Proof of Theorem 2.4.1

From (2.2.1) and (2.4.1) we may write

$$\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = (\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}(\hat{\boldsymbol{\phi}}(z) - \boldsymbol{\phi}(z)) = \hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{b}}(z) \tag{2.7.1}$$

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

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

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

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

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

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

and

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

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

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

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

Applying the uniform consistency of $\hat{F}_n(z|x)$, given by Lemma 1.2.3, together with conditions (H7) and (H8) gives that $\hat{b}_i^{(2)}(z) = o_p\left(n^{-1/2}\right)$ uniformly in z.

Let us now concentrate on $\hat{b}_i^{(1)}(z)$. Using the i.i.d. representation for $\hat{S}_n(z|X)$ given in Theorem 1.2.1, we have :

$$\hat{S}_n(z|X_j) - S(z|X_j) = \sum_{l=1}^{n} B_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z) + R_n(z|X_j), \qquad (2.7.2)$$

where $\xi(Z, T, \delta, x, y)$ and $R_n(y|x)$ are defined in (1.2.3) and (1.2.4), respectively.

Observe that $E[\xi(Z, T, \delta, x, y)|X = x] = 0$. We plug (2.7.2) into $\hat{b}_i^{(1)}(z)$ to obtain :

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

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

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

with

$$\hat{b}_i^{(111)}(z) = \frac{1}{n}\sum_{j=1}^n X_j^i w(X_j)\phi'(S(z|X_j))\sum_{\substack{l\neq j\\ l=1}}^n \tilde{B}_{nl}(X_j)S(z|X_j)\xi(Z_l, T_l, \delta_l, X_j, z), \quad (2.7.3)$$

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

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

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

For $\hat{b}_i^{(R)}(z)$ we have from (1.2.4) and using condition (H10) that

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

uniformly in z.

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

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

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

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

Since

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

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

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

As done for the term $\hat{b}_i^{(111)}(z)$ in the proof of Corollary 2.4.2 we can show that under condition (H11) $\hat{b}_i^{(112)}(z)$ is $O_p(n^{-1/2})$ uniformly in z.

So far we have proved that

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

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

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

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

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

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

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

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

where

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

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

$$D_i^{(3)} = \frac{n-1}{n} \sum_{k=1}^{n} h_i^{(3)}(\boldsymbol{V}_k),$$
$$D_i^{(4)} = (n-1)E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)],$$

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 and the fact that $E[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)|\boldsymbol{V}_1] = E[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)|\boldsymbol{V}_2] = 0$ show that

$$\begin{aligned} Var(D_i^{(1)}) &= \frac{4}{n^2} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k>j}}^{n} \sum_{l=1}^{n} \sum_{\substack{r=1 \\ k>r}}^{n} E[h^{(1)}(\boldsymbol{V}_j, \boldsymbol{V}_k) \cdot h^{(1)}(\boldsymbol{V}_l, \boldsymbol{V}_r)] \\ &= \begin{cases} \frac{2(n-1)}{n} E\{[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)]^2\}, & \text{if } j = l \text{ and } k = r \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

From Appendix A.2.1 we have that

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

and that $E[\tilde{h_i}^2(\boldsymbol{V}_1, \boldsymbol{V}_2)] \le E[h_i^2(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^{-1}n^{-2})$ from Appendix A.2.2. These imply that $E[h_i^{(1)}(\boldsymbol{V}_1, \boldsymbol{V}_2)^2] = O(h^{-1}n^{-2})$ and, consequently, $\mathrm{Var}(D_i^{(1)}) = O(h^{-1}n^{-2})$, which gives

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

Now, from see Appendix A.2.3, $D_i^{(4)} = (n-1)E[\tilde{h_i}(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^2) = o_p(n^{-1/2})$.

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

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

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 $\Delta_{ij}^{(1)} = \frac{1}{4}E[h_i(\boldsymbol{V}_3,\boldsymbol{V}_2)h_j(\boldsymbol{V}_1,\boldsymbol{V}_2)]$, $\Delta_{ij}^{(2)} = \frac{1}{4}E[h_i(\boldsymbol{V}_2,\boldsymbol{V}_3)h_j(\boldsymbol{V}_1,\boldsymbol{V}_2)]$, $\Delta_{ij}^{(3)} = \frac{1}{4}E[h_i(\boldsymbol{V}_3,\boldsymbol{V}_2)$
$h_j(\boldsymbol{V}_2,\boldsymbol{V}_1)]$ and $\Delta_{ij}^{(4)} = \frac{1}{4}E[h_i(\boldsymbol{V}_2,\boldsymbol{V}_3)h_j(\boldsymbol{V}_2,\boldsymbol{V}_1)]$. It can be easily seen that

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

from Appendix A.2.4.

$\Delta_{ij}^{(2)} = \Delta_{ij}^{(3)} = O(h^4n^{-2})$ and $\Delta_{ij}^{(4)} = O(h^4n^{-2})$, since $E[h_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1] = O(h^2n^{-1})$. See more details in Appendix A.2.5 and Appendix A.2.6.

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

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

Direct application of the Cramér-Wold device implies that

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

From this and the fact that $\hat{\boldsymbol{\beta}}(z) - \boldsymbol{\beta}(z) = \hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{b}}(z)$ (see (2.7.1)), 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.

2.7.2 Proof of Corollary 2.4.2

From the proof of Theorem 2.4.1 it follows that we only have to prove that $\hat{b}_i^{(111)}(z)$ (defined in (2.7.3)) it is $O_p(n^{-1/2})$ uniformly in z. For this, partition the interval $[a,b]$ into $U_n \sim n^{1/2}$ subintervals $a = Z_0^{'} \leq Z_1^{'} \leq \ldots \leq Z_{U_n}^{'} = b$, of length $O(n^{-1/2})$. Then,

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

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

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

where we used Chebysev's inequality and the variance that appears in the expression above is proved to be $O(n^{-1}h^2)$ by Taylor developments and expectations computations, similar to those of Appendix A.2.4, A.2.5 and A.2.6. Next, consider R_1 :

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

Write

$$A_1 = \max_{1\leq r\leq U_n}\frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l\neq j\\l=1}}^{n} |X_j|^i\cdot w(X_j)\cdot\tilde{B}_{nl}(X_j)\cdot \sup_{Z'_{r-1}\leq z\leq Z'_r}\frac{1_{\{Z_l\leq z,\delta_l=1\}}}{C(Z_l|X_j)}$$

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

Using the fact that $\sup_{Z'_{r-1}\le z\le Z'_r} |\phi'(S(z|X_j))S(z|X_j) - \phi'(S(Z'_r|X_j))S(Z'_r|X_j)| = O(U_n^{-1})$ (see Appendix A.2.7) and condition (H11) it is easy to prove that $A_{11} = O(U_n^{-1}) = O_p(n^{-1/2})$, and in a similar way it can be shown that $A_2 = O(U_n^{-1}) = O_p(n^{-1/2})$.

Similar techniques as for R_2 combined with

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

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

Appendix A.2.1

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

The expectation $E[h_i^{(1)}(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]$ can be decomposed as follows :

$$
E[(a(\boldsymbol{V}_1,\boldsymbol{V}_2) - b(\boldsymbol{V}_2))^2] = E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)^2] + E[b(\boldsymbol{V}_2)^2] - 2E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)b(\boldsymbol{V}_2)] \quad (2.7.4)
$$

with

$$
a(\boldsymbol{V}_1,\boldsymbol{V}_2) = \tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2) - E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1]
$$

and

$$
b(\boldsymbol{V}_2) = E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_2] + E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)].
$$

Since

$$
\begin{aligned}
&E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)b(\boldsymbol{V}_2)] \\
&= E\left[E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2) - E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1]\boldsymbol{V}_2]\cdot(E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_2] - E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)])\right] \\
&= E\left[\{E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_2] - E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)]\}^2\right] \\
&= E[b(\boldsymbol{V}_2)^2],
\end{aligned}
$$

we get that

$$E[(a(\boldsymbol{V}_1,\boldsymbol{V}_2)-b(\boldsymbol{V}_2))^2]=E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]-E[b(\boldsymbol{V}_2)^2]\leq E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]. \tag{2.7.5}$$

On the other hand

$$E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]=E\left[\left(\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)-E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1]\right)^2\right]$$

$$=E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]-2E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1]]+E\left[E^2[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)|\boldsymbol{V}_1]\right]$$

$$=E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]-2E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)|\boldsymbol{V}_1]]+E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\cdot\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)]$$

$$=E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]-2E\left[E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)|\boldsymbol{V}_1,\boldsymbol{V}_2]\right]+E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\cdot\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)]$$

$$=E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]-2E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)]+E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\cdot\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)]$$

$$=E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]-E[\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_2)\tilde{h}_i(\boldsymbol{V}_1,\boldsymbol{V}_3)]$$

Which implies that $E[a(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]\leq E[\tilde{h}_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)]$, thus from (2.7.4) and (2.7.5) we obtain

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

Appendix A.2.2

We will prove that $E[h_i^2(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]=O(h^{-1}n^{-2})$.

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

$$=E\left[X_1^{2i}w(X_1)^2\phi^{'}(S(z|X_1))^2S(z|X_1)^2\tilde{B}_{n2}^2(X_1)\xi^2(Z_2,T_2,\delta_2,X_1,z)\right]$$

$$=E\left[X_1^{2i}w(X_1)^2\phi^{'}(S(z|X_1))^2S(z|X_1)^2E[\tilde{B}_{n2}^2(X_1)\xi^2(Z_2,T_2,\delta_2,X_1,z)|X=X_1,T_1\leq Z_1]\right] \tag{2.7.6}$$

First, we have that

$$E\left[\tilde{B}_{n2}^2(x)\xi^2(Z_2,T_2,\delta_2,x,z)|X=x,T_1\leq Z_1\right]$$

$$=E\left[\frac{\frac{1}{n^2h^2}K^2\left(\frac{x-X_2}{h}\right)}{m^*(x)^2}\xi^2(Z_2,T_2,\delta_2,x,z)|X=x,T_1\leq Z_1\right]$$

$$=\frac{1}{m^*(x)^2}\cdot\frac{1}{n^2h^2}\int K^2\left(\frac{x-v}{h}\right)\underbrace{E\left[\xi^2(Z_2,T_2,\delta_2,x,z)|X_2=v,T_1\leq Z_1\right]}_{\mu(z,v)}\cdot m^*(u)du.$$

By means of change of variable ($v = x - hu$) and by Taylor expansion of $\mu \cdot m^*$ around x we obtain

$$
\begin{aligned}
& E\left[\tilde{B}_{n2}^2(x)\xi^2(Z_2, T_2, \delta_2, x, z)|X = x, T_1 \le Z_1\right] && (2.7.7)\\
&= \frac{1}{m^*(x)^2} \cdot \frac{1}{n^2h^2} \int K^2(u)\mu(z, x - uh)m^*(x - uh)hdu \\
&= \frac{1}{m^*(x)^2} \cdot \frac{1}{n^2h} \int K^2(u)\left[\mu(z,x)m^*(x) - (\mu^{'}(z,x)m^*(x) + \mu(z,x)m^{*'}(x))uh + o(h)\right] du \\
&= \frac{1}{m^*(x)} \cdot \frac{1}{n^2h}\mu(z,x)\left(\int K^2(u)du\right) + o(n^{-2}). && (2.7.8)
\end{aligned}
$$

In the following we will show that

$$\mu(z,x) = \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)}, \tag{2.7.9}$$

which does not depend on n, nor on h.

Thus, (2.7.8) and (2.7.9) give us

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

Now we will prove (2.7.9)

$$
\begin{aligned}
\mu(z,x) &= E\left[\xi^2(Z_2, T_2, \delta_2, x, z)|X_2 = x\right] \\
&= E\left[\left(\frac{I_{\{Z_2 \le z, \delta_2 = 1\}}}{C(Z_2|x)}\right)^2 |X_2 = x, T_2 \le Z_2\right] \\
&\quad + E\left[\left(\int_0^z \frac{I_{\{T_2 \le u \le Z_2\}}}{C^2(u|x)} dH_1^*(u|x)\right)^2 |X_2 = x, T_2 \le Z_2\right] \\
&\quad - 2E\left[\left(\frac{I_{\{Z_2 \le z, \delta_2 = 1\}}}{C(Z_2|x)}\right)\left(\int_0^z \frac{I_{\{T_2 \le u \le Z_2\}}}{C^2(u|x)} dH_1^*(u|x)\right) |X_2 = x, T_2 \le Z_2\right] \\
&= A + B - 2C
\end{aligned}
$$

We will see that $B - 2C = 0$, which gives

$$\mu(z,x) = A = E\left[\frac{I_{\{Z_2 \le z, \delta_2 = 1\}}}{C^2(Z_2|x)}|X_2 = x, T_2 \le Z_2\right] = \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)}.$$

Now we shall prove that $B - 2C = 0$. First,

$$
\begin{aligned}
B =& \int_0^\infty \int_0^\infty \int_{T_i}^{Z_i \wedge z} \frac{1}{C^2(u|x)} dH_1^*(u|x) \\
& \int_{T_i}^{Z_i \wedge z} \frac{1}{C^2(v|x)} dH_1^*(v|x) dP(Z \le Z_i, T \le T_i | X = x, T \le Z) \\
=& \int_0^z \frac{1}{C^2(u|x)} \int_0^z \frac{1}{C^2(v|x)} \\
& \int_0^{u \wedge v} \int_{u \vee v}^\infty dP(Z \le Z_i, T \le T_i | X = x, T \le Z) dH_1^*(v|x) dH_1^*(u|x) \\
=& \frac{2}{\alpha} \int_0^z \frac{1}{C^2(u|x)} \int_0^u \frac{1}{C^2(v|x)} L(v|x) dH_1^*(v|x)(1 - H(u|x)) dH_1^*(u|x)
\end{aligned}
$$

where

$$
\begin{aligned}
& \int_0^{u \wedge v} \int_{u \vee v}^\infty dP(Z \le Z_i, T \le T_i | X = x, T \le Z) \\
&= P(Z \ge u \vee v, T \le u \wedge v | X = x, T \le Z) \\
&= \frac{P(Z \ge u \vee v, T \le u \wedge v | X = x)}{P(T \le Z)} \\
&= \frac{(1 - H(u \vee v|x)) L(u \wedge v|x)}{P(T \le Z)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
C =& \int_0^\infty \int_0^\infty \int_0^z \frac{I_{\{T_i \le u \le Z_i\}}}{C^2(u|x)} dH_1^*(u|x) \\
& \frac{1}{C(Z_i|x)} I_{\{Z_i \le z\}} dP(Z \le Z_i, T \le T_i, \delta_i = 1 | X = x, T \le Z) \\
=& \int_0^z \frac{1}{C^2(u|x)} \int_u^z \int_0^u \frac{1}{C(Z_i|x)} dP(Z \le Z_i, T \le T_i, \delta_i = 1 | X = x, T \le Z) dH_1^*(u|x) \\
=& -\int_0^z \frac{1}{C^2(u|x)} \int_u^z \frac{1}{C(Z_i|x)} dP(Z \ge Z_i, T \le T_i, \delta_i = 1 | X = x, T \le Z) dH_1^*(u|x) \\
=& -\frac{1}{\alpha} \int_0^z \frac{1}{C^2(u|x)} \int_u^z \frac{1}{C(Z_i|x)} dP(Z \ge Z_i, T \le T_i, \delta_i = 1 | X = x) dH_1^*(u|x) \\
=& \frac{1}{\alpha} \int_0^z \frac{1}{C^2(u|x)} \int_u^z \frac{1}{C(v|x)} dH_1(v|x) L(u|x) dH_1^*(u|x)
\end{aligned}
$$

$$= \frac{1}{\alpha} \int_0^z \frac{1}{C^2(u|x)} \int_0^u \frac{1}{C^2(v|x)} L(v|x) dH_1^*(v)(1 - H(u|x)) dH_1^*(u|x)$$
$$= \frac{B}{2}$$

where

$$\begin{aligned} C(v|x)dH_1(v|x) &= \frac{1}{\alpha}(1 - H(v|x))L(v|x)dH_1(v|x) \\ &= \frac{1}{\alpha}(1 - H(v|x))L(v|x)(1 - G(v|x))dF(v|x) \\ &= (1 - H(v|x))\frac{1}{\alpha}P(T \le v \le C|x)dF(v|x) \\ &= (1 - H(v|x))dH_1^*(v|x) \end{aligned}$$

Appendix A.2.3

We will prove that $E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^2 n^{-1})$.

First notice that

$$E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)] = E\left\{\frac{1}{2}[h_i(\boldsymbol{V}_1, \boldsymbol{V}_2) + h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)]\right\} = E[h_i(\boldsymbol{V}_1, \boldsymbol{V}_2)], \qquad (2.7.10)$$

since $E[h_i(\boldsymbol{V}_1, \boldsymbol{V}_2)] = E[h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)]$.

$$\begin{aligned} & E[h_i(\boldsymbol{V}_1, \boldsymbol{V}_2)] \\ &= E\left[X_1^i w(X_1)\phi'(S(z|X_1))S(z|X_1)\tilde{B}_{n2}(X_1)\xi(Z_2, T_2, \delta_2, X_1, z)\right] \\ &= E\left\{E[X_1^i w(X_1)\phi'(S(z|X_1))S(z|X_1)\tilde{B}_{n2}(X_1)\xi(Z_2, T_2, \delta_2, X_1, z)|\boldsymbol{V}_1]\right\} \\ &= E\left\{X_1^i w(X_1)\phi'(S(z|X_1))S(z|X_1)E[\tilde{B}_{n2}(X_1)\xi(Z_2, T_2, \delta_2, X_1, z)|\boldsymbol{V}_1]\right\} \qquad (2.7.11) \end{aligned}$$

with

$$E[\tilde{B}_{n2}(X_1)\xi(Z_2, T_2, \delta_2, X_1, z)|\boldsymbol{V}_1] = O(h^2 n^{-1}), \qquad (2.7.12)$$

which will give $E[\tilde{h}_i(\boldsymbol{V}_1, \boldsymbol{V}_2)] = O(h^2 n^{-1})$.

Lets now prove (2.7.12)

$$
\begin{aligned}
&E[\tilde{B}_{n2}(x)\xi(Z_2,T_2,\delta_2,x,z)|\boldsymbol{V}_1]\\
&= E\left[\frac{\frac{1}{nh}K\left(\frac{x-X_2}{h}\right)}{m^*(x)}\xi(Z_2,T_2,\delta_2,x,z)|X=x,T_1\le Z_1\right]\\
&= \frac{1}{m^*(x)}\cdot\frac{1}{nh}\int K\left(\frac{x-v}{h}\right)E[\xi(Z_2,T_2,\delta_2,x,z)|X_2=v,T_2\le Z_2]m^*(v)dv. \qquad (2.7.13)
\end{aligned}
$$

By means of change of variables ($v = x - uh$) and by Taylor expansion of $\psi \cdot m^*$ around x, (2.7.13) can be written as

$$
\begin{aligned}
&\frac{1}{m^*(x)}\frac{1}{n}\int K(u)\psi(z,x-uh)m^*(x-uh)du\\
&= \frac{1}{m^*(x)}\frac{1}{n}\int K(u)\big[\psi(z,x)m^*(x) - (\psi^{'}(z,x)m^*(x) + \psi(z,x)m^{*'}(x))uh\\
&\qquad + \frac{1}{2}(\psi^{''}(z,x)m^*(x) + \psi(z,x)m^{*''}(x) + 2\psi^{'}(z,x)m^{*'}(x))u^2h^2 + o(h^2)\big]du\\
&= \frac{1}{2}\left(\psi^{''}(z,x) + 2\psi^{'}(z,x)\frac{m^{*'}(x)}{m^*(x)}\right)\frac{h^2}{n}\underbrace{\left(\int K(u)u^2du\right)}_{\mu_2} + \frac{1}{m^*(x)}o(h^2n^{-1}) \qquad (2.7.14)
\end{aligned}
$$

since $\psi(z,x) = 0$.

Now, by using the assumptions from Corollary (1.2.2), ψ has bounded first and second derivatives (it can be written as a function of other function with bounded first, respectively second derivatives) then from (2.7.11), (2.7.13) and (2.7.14)

$$E[\tilde{B}_{n2}(x)\xi(Z_2,T_2,\delta_2,x,z)|X_1=x,\boldsymbol{V}_1] = O(h^2n^{-1})$$

Appendix A.2.4

We will prove that

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

$$\Delta_{ij}^{(1)} = \frac{1}{4}E\left[E[h_i(\boldsymbol{V}_3,\boldsymbol{V}_2)h_j(\boldsymbol{V}_1,\boldsymbol{V}_2)|\boldsymbol{V}_1,\boldsymbol{V}_3]\right]$$

$$= \frac{1}{4}E\Bigg(\frac{1}{(nh)^2}X_3^iX_1^j\tilde{w}(X_3)\tilde{w}(X_1)S(z|X_3)S(z|X_1)\phi^{'}(S(z|X_3))\phi^{'}(S(z|X_1))\frac{1}{m^*(X_3)}\frac{1}{m^*(X_1)}$$

$$\times E\left[K\left(\frac{X_3-X_2}{h}\right)K\left(\frac{X_1-X_2}{h}\right)E[\xi(Z_2,T_2,\delta_2,X_3,z)\xi(Z_2,T_2,\delta_2,X_1,z)|X_2=x_2]\right]\Bigg)$$

By Taylor expansion of $\xi(Z_2,T_2,\delta_2,X_1,z)$ and $\xi(Z_2,T_2,\delta_2,X_3,z)$ around x_2

$$\Delta_{ij}^{(1)}$$

$$= \frac{1}{4(nh)^2}E\Bigg(X_3^iX_1^j\tilde{w}(X_3)\tilde{w}(X_1)S(z|X_3)S(z|X_1)\phi^{'}(S(z|X_3))\phi^{'}(S(z|X_1))\frac{1}{m^*(X_3)}\frac{1}{m^*(X_1)}$$

$$\times E\left[K\left(\frac{X_3-X_2}{h}\right)K\left(\frac{X_1-X_2}{h}\right)E[\xi(Z_2,T_2,\delta_2,X_2,z)^2+O(h)|X_2=x_2]\right]\Bigg)$$

$$= \frac{1}{4n^2}\int_I\int_0^z\frac{dH_1^*(u|x_2)}{C^2(u|x_2)}m^*(x_2)$$

$$\times\left(\frac{1}{h}\int_I x_3^i\tilde{w}(x_3)S(z|x_3)\phi^{'}(S(z|x_3))K\left(\frac{x_3-x_2}{h}\right)dx_3\right)$$

$$\times\left(\frac{1}{h}\int_I x_1^j\tilde{w}(x_1)S(z|x_1)\phi^{'}(S(z|x_1))K\left(\frac{x_1-x_2}{h}\right)dx_1\right)dx_2+O(h^{-1}n^{-2})$$

$$= \frac{1}{4n^2}\int_I x_2^{i+j}\tilde{w}^2(x_2)S^2(z|x_2)\phi^{'}(S(z|x_2))^2\int_0^z\frac{dH_1^*(u|x_2)}{C^2(u|x_2)}m^*(x_2)dx+O(h^{-1}n^{-2})$$

where by means of change of variables ($x_1 = x_2+uh$) and Taylor development of $g_j(z,x)$ around x_2.

$$\frac{1}{h}\int_I x_1^jw(x_1)S(z|x_1)\phi^{'}(S(z|x_1))K\left(\frac{x_2-x_1}{h}\right)dx_1$$

$$= \frac{1}{h}\int_I g_j(z,x_1)K\left(\frac{x_2-x_1}{h}\right)dx_1$$

$$= \frac{1}{h}\int g_j(z,x_2+hu)K(u)h_idu$$

$$= \int[g_j(z,x_2)+hug_j^{'}(z,x_2)+\frac{1}{2}h^2u^2g_j^{'}(z,\xi_{x_2})]K(u)du$$

$$= g_j(z,x_2)+O(h^2) \qquad (2.7.15)$$

with

$$g_j(z,x_1) := x_1^jw(x_1)S(z|x_1)\phi^{'}(S(z|x_1)). \qquad (2.7.16)$$

Appendix A.2.5

Will prove that $4\Delta_{ij}^{(2)} = O(h^4 n^{-2})$.

$$\begin{aligned}
4\Delta_{ij}^{(2)} &= E[h_i(\boldsymbol{V}_2, \boldsymbol{V}_3)h_j(\boldsymbol{V}_1, \boldsymbol{V}_2)] \\
&= E[E[h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)h_j(\boldsymbol{V}_1, \boldsymbol{V}_3)|\boldsymbol{V}_1, \boldsymbol{V}_2]] \\
&= E[h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)E[h_j(\boldsymbol{V}_1, \boldsymbol{V}_3)|\boldsymbol{V}_1]] \\
&= E\left[h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)X_1^j w(X_1)\phi'(S(z|X_1))S(z|X_1)E[\tilde{B}_{n3}(X_1)\xi(Z_3, T_3, \delta_3, X_1, z)|\boldsymbol{V}_1]\right]
\end{aligned}$$

By (2.7.12)

$$\begin{aligned}
4\Delta_{ij}^{(2)} &= E\Bigg[h_i(\boldsymbol{V}_2, \boldsymbol{V}_1)X_1^j w(X_1)\phi'(S(z|X_1))S(z|X_1) \\
&\qquad \times \left\{\frac{1}{2}\left(\psi''(z, X_1) + 2\psi'(z, X_1)\frac{m^{*'}(X_1)}{m^*(X_1)}\right)\frac{h^2}{n}\mu_2 + \frac{1}{m^*(X_1)}o(h^2 n^{-1})\right\}\Bigg] \\
&= \frac{1}{2}\frac{h^2}{n}E\Bigg[X_1^j X_2^i w(X_1)w(X_2)\phi'(S(z|X_1))\phi'(S(z|X_2))S(z|X_1)S(z|X_2)\mu_2 \\
&\qquad \times \left(\psi''(X_1) + 2\psi'(X_1)\frac{m^{*'}(X_1)}{m^*(X_1)}\right)\frac{\frac{1}{nh}K\left(\frac{X_2 - X_1}{h}\right)}{m^*(X_2)}E[\xi(Z_1, T_1, \delta_1, X_2, z)|X_1, X_2]\Bigg].
\end{aligned}$$

Denote

$$\lambda_j(z, X_1) = X_1^j w(X_1)\phi'(S(z|X_1))S(z|X_1)\left(\psi''(X_1) + 2\psi'(X_1)\frac{m^{*'}(X_1)}{m^*(X_1)}\right)$$

and

$$\rho_j(z, X_1, X_2) = g_j(z, X_2)E[\xi(Z_1, T_1, \delta_1, X_2, z)|X_1, X_2]$$

with $g_j(z, x)$ defined in (2.7.16).

Then, by Taylor expansion of $\rho(z, X_1, X_2)$ around (X_1, X_1)

$$4\Delta_{ij}^{(2)} = \frac{1}{2}\frac{h}{n^2}E\left[\lambda(z, X_1)\mu_2 K\left(\frac{X_2 - X_1}{h}\right)\rho_j(z, X_1, X_2)\frac{1}{m^*(X_2)}\right]$$

Taylor expansion of $\rho(z, X_1, X_2)$ around (X_1, X_1)

$$= \frac{1}{2}\frac{h}{n^2}E\Bigg\{\lambda(z, X_1)\mu_2 K\left(\frac{X_2 - X_1}{h}\right)\frac{1}{m^*(X_2)}\Bigg[\rho_j(z, X_1, X_1) + (X_2 - X_1)\rho_j^{'}(z, X_1, X_1)$$
$$+ \frac{1}{2}(X_2 - X_1)^2\rho_j^{''}(z, X_1, X_1) + \frac{1}{6}(X_2 - X_1)^3\rho_j^{'''}(z, X_1, X_1)$$
$$- \frac{1}{24}(X_2 - X_1)^4\rho_j^{''''}(z, X_1, \xi_{X_1})\Bigg]\Bigg\}$$
$$= \frac{h}{n^2}\frac{\mu_2}{2}\int\int\lambda(z, x_1)K\left(\frac{x_2 - x_1}{h}\right)\frac{1}{m^*(x_2)}\Bigg[(x_2 - x_1)\rho_j^{'}(z, x_1, x_1)$$
$$+ \frac{1}{2}(x_2 - x_1)^2\rho_j^{''}(z, x_1, x_1) + \frac{1}{6}(x_2 - x_1)^3\rho_j^{'''}(z, x_1, x_1)$$
$$- \frac{1}{24}(X_2 - X_1)^4\rho_j^{''''}(z, X_1, \xi_{X_1})\Bigg] \times m^*(x_1)m^*(x_2)dx_1dx_2$$

By a change of variable $(x_2 - x_1 = uh)$ we get

$$4\Delta_{ij}^{(2)} = \frac{h}{n^2}\frac{\mu_2}{2}\int\int_I\lambda(z, x_1)K(u)\Bigg[hu\rho_j^{'}(z, x_1, x_1) + \frac{1}{2}(hu)^2\rho_j^{''}(z, x_1, x_1) + \frac{1}{6}(hu)^3\rho_j^{'''}(x_1, x_1)$$
$$- \frac{1}{24}(X_2 - X_1)^4\rho_j^{''''}(z, X_1, \xi_u)\Bigg]m^*(x_1)dx_1(hdu)$$
$$= \frac{h^2}{n^2}\frac{\mu_2}{2}\int_I\lambda(z, x_1)\Bigg[\rho_j^{'}(z, x_1, x_1)\int K(u)hudu + \frac{1}{2}\rho_j^{''}(z, x_1, x_1)\int K(u)(hu)^2du$$
$$+ \frac{1}{6}\int K(u)(hu)^3\rho_j^{'''}(z, x_1, x_1)du - \frac{1}{24}\int K(u)(hu)^4\rho_j^{''''}(z, X_1, \xi_u)du\Bigg]m^*(x_1)dx_1$$
$$= \frac{h^4}{n^2}\frac{\mu_2}{2}\int_I\lambda(z, x_1)\Bigg[\frac{1}{2}\rho_j^{''}(z, x_1, x_1)\mu_2 - \frac{1}{24}\int h^2u^4K(u)\rho_j^{''''}(z, x_1, \xi_u)du\Bigg]m^*(x_1)dx_1$$

Thus

$$\Delta_{ij}^{(2)} = O(h^4n^{-2}).$$

Appendix A.2.6

We will prove that $\Delta_{ij}^{(4)} = O(h^4 n^{-2})$.

$$
\begin{aligned}
4\Delta_{ij}^{(4)} &= E[h_i(\boldsymbol{V}_2, \boldsymbol{V}_3) h_j(\boldsymbol{V}_2, \boldsymbol{V}_1)] \\
&= E\Big[X_2^{i+j} w^2(X_2) \phi^{'}(S(z|X_2))^2 S(z|X_2)^2 \\
&\qquad \times E[\tilde{B}_{n3}(x_2)\tilde{B}_{n1}(x_2)\xi(Z_3, T_3, \delta_3, x_2, z)\xi(Z_1, T_1, \delta_1, x_2, z)|X_2 = x_2]\Big] \\
&= E\Big[X_2^{i+j} w^2(X_2) \phi^{'}(S(z|X_2))^2 S(z|X_2)^2 \\
&\qquad \times E[\tilde{B}_{n3}(x_2)\xi(Z_3, T_3, \delta_3, x_2, z)|X_2 = x_2] E[\tilde{B}_{n1}(x_2)\xi(Z_1, T_1, \delta_1, x_2, z)|X_2 = x_2]\Big] \\
&= O(h^4 n^{-2}) E[X_2^{i+j} w^2(X_2) \phi^{'}(S(z|X_2))^2 S(z|X_2)^2], \text{ by (2.7.14).}
\end{aligned}
$$

Since $E[X_2^{i+j} w^2(X_2) \phi^{'}(S(z|X_2))^2 S(z|X_2)^2] < \infty$, we get $\Delta_{ij}^{(4)} = O(h^4 n^{-2})$.

Appendix A.2.7

We will show that

$$
\sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j))S(z|X_j) - \phi^{'}(S(Z'_r|X_j))S(Z'_r|X_j)| = O(U_n^{-1}).
$$

$$
\begin{aligned}
&\sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j))S(z|X_j) - \phi^{'}(S(Z'_r|X_j))S(Z'_r|X_j)| \\
&\le \sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j))S(z|X_j) - \phi^{'}(S(z|X_j))S(Z'_r|X_j)| \\
&\qquad + \sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j))S(Z'_r|X_j) - \phi^{'}(S(Z'_r|X_j))S(Z'_r|X_j)| \\
&\le \sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j))| \cdot |S(z|X_j) - S(Z'_r|X_j)| \\
&\qquad + \sup_{Z'_{r-1} \le z \le Z'_r} |\phi^{'}(S(z|X_j)) - \phi^{'}(S(Z'_r|X_j))| \cdot |S(Z'_r|X_j)| \\
&= O(U_n^{-1}),
\end{aligned}
$$

with

$$
\begin{aligned}
&\sup_{Z'_{r-1}\le z\le Z'_r} |S(z|X_j) - S(Z'_r|X_j)| \\
&= \sup_{Z'_{r-1}\le z\le Z'_r} |S(Z'_r|X_j) + (z - Z_r)S^{'}(\xi_i|X_j) - S(Z_r|X_j)| \\
&= \sup_{Z'_{r-1}\le z\le Z'_r} |(z - Z_r)(z - Z_r S^{'}(\xi_r|X_j)| = O(U_n^{-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&\sup_{Z'_{r-1}\le z\le Z'_r} |\phi^{'} S(z|X_j) - \phi^{'} S(Z'_r|X_j)| \\
&= \sup_{Z'_{r-1}\le z\le Z'_r} |\phi^{''} S(\xi_r|X_j) \left(S(z|X_j) - S(Z_r|X_j)\right)| \\
&= \sup_{Z'_{r-1}\le z\le Z'_r} |S(z|X_j) - S(Z'_r|X_j)| = O(U_n^{-1}).
\end{aligned}
$$

Chapter 3

Goodness-of-fit Tests

In Chapter 2 we introduced a least squares estimator for the time-dependent parameters $\boldsymbol{\beta}(z)$ of model (2.1.1) in the case were the response may be subject to left-truncation and/or right-censoring. In practice one would like to check first the validity of the considered model for the data. We introduce in this chapter a goodness-of-fit test for model (2.1.1) when the response variable is subject to left-truncation and/or right-censoring. The test is based on the comparison of the parametric modeling of regression data with a nonparametric estimator of the response, $\phi(S(z|\boldsymbol{X}))$.

Section 3.1 introduces the test statistic and its asymptotic distribution, as well as the assumption under which the latter is valid. The convergence to the asymptotic distribution may be slow, so a bootstrap method is proposed in Section 3.2 in order to estimate the critical values. In Section 3.3 we present some numerical results about the finite sample properties of the test statistic. Finally, Section 3.4 contains the proof of the main result, Section 3.5 contains the lemmas needed in Section 3.4 as well as their proofs, while the Appendix gathers some technical details needed in the proofs of the previous sections.

This chapter is mainly based on Teodorescu and Van Keilegom (2008).

3.1 The test statistic

Same notations and assumptions as in Section 2.2 are being used.

If the following model holds

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \ldots + \beta_p(z)X^p, \tag{3.1.1}$$

where ϕ is a known monotone transformation, $\phi : [0,1] \to \mathbb{R}$, X is a one dimensional continuous covariate and p is known, then the coefficients $\boldsymbol{\beta}(z)$ can be estimated via the weighted least squares procedure presented in Chapter 2. Note that model (3.1.1) is exactly the same model as the one in (2.1.1), but we write it again for easier reference. More precisely,

$$\hat{\boldsymbol{\beta}}(z) = (\hat{\beta}_0(z), \hat{\beta}_1(z), \ldots, \hat{\beta}_p(z))^t = (\boldsymbol{X}^t\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{W}\hat{\boldsymbol{\phi}}(z), \tag{3.1.2}$$

Now we want to check the appropriateness of the semi-parametric model (3.1.1). For a given ϕ and p we would like to test

$$H_0 : \text{for all } z \text{ there exists a vector } \boldsymbol{\beta}(z) \in \mathbb{R}^{p+1} \text{ such that (3.1.1) holds,}$$

$$\text{versus } H_a : \text{there exists a } z \text{ such that (3.1.1) does not hold for any } \boldsymbol{\beta}(z) \in \mathbb{R}^{p+1}.$$

A natural way to proceed is to measure the distance between $\hat{\boldsymbol{\phi}}(z)$ and the hypothesized model (3.1.1) and to use this distance as test statistic. Here we study a kind of L_2 - distance between these two :

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

Hence, it is reasonable to reject H_0 when $\hat{\Phi}_n(\hat{\boldsymbol{\beta}}(z))$ 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 Section 2.2 and define :

$$T_n = \int_a^b \hat{\Phi}_n(\hat{\boldsymbol{\beta}}(z))dz \tag{3.1.3}$$

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, the same conditions (H1)-(H11) as for Theorem 2.4.1 in Section 2.2 are to be assumed, plus

(H12) $g(z_1, z_2, x)$ defined in (3.1.5) has bounded second derivative with respect to x.

Theorem 3.1.1 *Suppose that conditions (H1) through (H12) hold. Then, under H_0,*

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

where $b_{0h} = h^{-1/2}K^{(2)}(0)\int_a^b \int_I g(z,z,x)dxdz,\ K^{(2)}(x) = \int_{-\infty}^{+\infty} K(x-t)K(t)dt,$

$$V = 2\int K * K(v)^2 dv \int_a^b \int_a^b \int_I g^2(z_1,z_2,x)dxdz_1dz_2, \tag{3.1.4}$$

and

$$g(z_1,z_2,x) = \phi^{'}(S(z_1|x))\phi^{'}(S(z_2|x))S(z_1|x)S(z_2|x)\int_a^{z_1\wedge z_2} \frac{dH_1^*(u|x)}{C^2(u|x)}. \tag{3.1.5}$$

Remark 3.1.1 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 a few 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 (1981)).

Remark 3.1.2 We could also consider $T_n^1 = \int_0^1 \hat{\Psi}_n(\hat{\boldsymbol{\beta}}(z))dz$ as an alternative to T_n in the test statistic, where

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

The same arguments as in the proof of Theorem 3.1.1 should work in order to obtain the limiting distribution of $nh^{1/2}T_n^1$.

3.2 Bootstrap version

The quadratic form structure of the dominant terms in the test statistic suggests that the convergence of the distribution of the test statistic $nh^{1/2}T_n$ to a normal distribution will be 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 $z \in \{Z_1, \ldots, Z_n\}$ and $x \in \{X_1, \ldots, X_n\}$:

   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

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

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

   b) Replace $S(z|x)$ by $\hat{S}_g(z|x)$ in (3.1.1) and estimate $\beta_0(z), \ldots, \beta_p(z)$ by the least squares estimator in (3.1.2) to obtain $\hat{\beta}_{0,g}(z), \ldots, \hat{\beta}_{p,g}(z)$. Plug-in these estimators into (3.1.1) and re-estimate $S(z|x)$ by

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

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

   a) For every $i = 1, \ldots, n$, 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 $\{(T_1^*, Z_1^*, \delta_1^*, X_1), \ldots, (T_n^*, Z_n^*, \delta_n^*, X_n)\}$ to estimate a bootstrap version of the conditional survival function, $\hat{S}_h^*(z|X_i)$ $(i = 1, \ldots, n)$ using the bandwidth h. This bootstrap version is used to obtain the bootstrap vector of coefficients $\hat{\boldsymbol{\beta}}_h^{*(b)} = \left(\hat{\beta}_{0,h}^{*(b)}(z), \ldots, \hat{\beta}_{p,h}^{*(b)}(z)\right)$ by means of the least squares estimator, and to obtain the bootstrap version $\hat{\Phi}_n^*\left(\hat{\boldsymbol{\beta}}_h^{*(b)}(z)\right)$ of $\hat{\Phi}_n\left(\hat{\boldsymbol{\beta}}(z)\right)$.

   c) Compute the bootstrap version of the test statistic $nh^{1/2}T_n$, which is given by :

$$nh^{1/2}T_{n,b}^* = nh^{1/2}\int_a^b \hat{\Phi}_n^*\left(\hat{\boldsymbol{\beta}}_h^{*(b)}(z)\right)dz \simeq nh^{1/2}\sum_{i=1}^{n-1}\hat{\Phi}_n^*\left(\hat{\boldsymbol{\beta}}_h^{*(b)}(Z_i^*)\right)(Z_{i+1}^* - Z_i^*)$$

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 3.2.1 A way to choose the smoothing parameter h under H_0 is to apply the bootstrap procedure for the choice of h described in Section 2.5.

Remark 3.2.2 The choice of the pilot bandwidth g used in the bootstrap procedure is not of crucial importance (see Cao and González-Manteiga (2008) for details).

3.3 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 :

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x + \beta_2(z)\sin\left(\frac{\pi x}{2}\right) \tag{3.3.1}$$

$$\phi(S(z|x)) = \beta_0(z) + \beta_1(z)x + \beta_2(z)\exp\left(\frac{2x}{3}\right). \tag{3.3.2}$$

For model (3.3.1), $X \sim U[4, 10]$, $Y|_{X=x} \sim \text{Exp}\left(4x + a_1 \sin\left(\frac{\pi x}{2}\right)\right)$, $C|_{X=x} \sim \text{Exp}(d_1 x)$, 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 $\phi(S(z|x)) = 4zx + a_1 z \sin\left(\frac{\pi x}{2}\right)$. So, $\beta_0(z) = 0$, $\beta_1(z) = 4z$ and $\beta_2(z) = a_1 z$. 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 $K(u) = 1_{\{-1\le u\le 1\}} \cdot 3(1-u^2)/4$ and the weight function $\tilde{w}(x) = 1_{\{4.35\le x\le 9.65\}}$ 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 (3.3.1) for $a_1 = 0$, while as alternatives to H_0 we have considered model (3.3.1) for $a_1 \in \{4, 8, 12, 16, 20\}$, see Figure 3.1, for a graphical representation of H_0 and H_1 for a fixed z. As a_1 becomes larger, the departure from H_0 also increases.

Table 3.1 displays the results for model (3.3.1) 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.

Cens		H_0	H_1				
perc	h	$a_1 = 0$	$a_1 = 4$	$a_1 = 8$	$a_1 = 12$	$a_1 = 16$	$a_1 = 20$
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 3.1 : Power of the test for model (3.3.1) under censoring, with B=500 and M=1000.

Table 3.2 contains the results for model (3.3.1) under both censoring and truncation. The results are similar to those of Table 3.1. We notice that the results for 20% censoring are better than those for 40% of censoring and also that for 10% of truncation displays better results than 20% truncation.

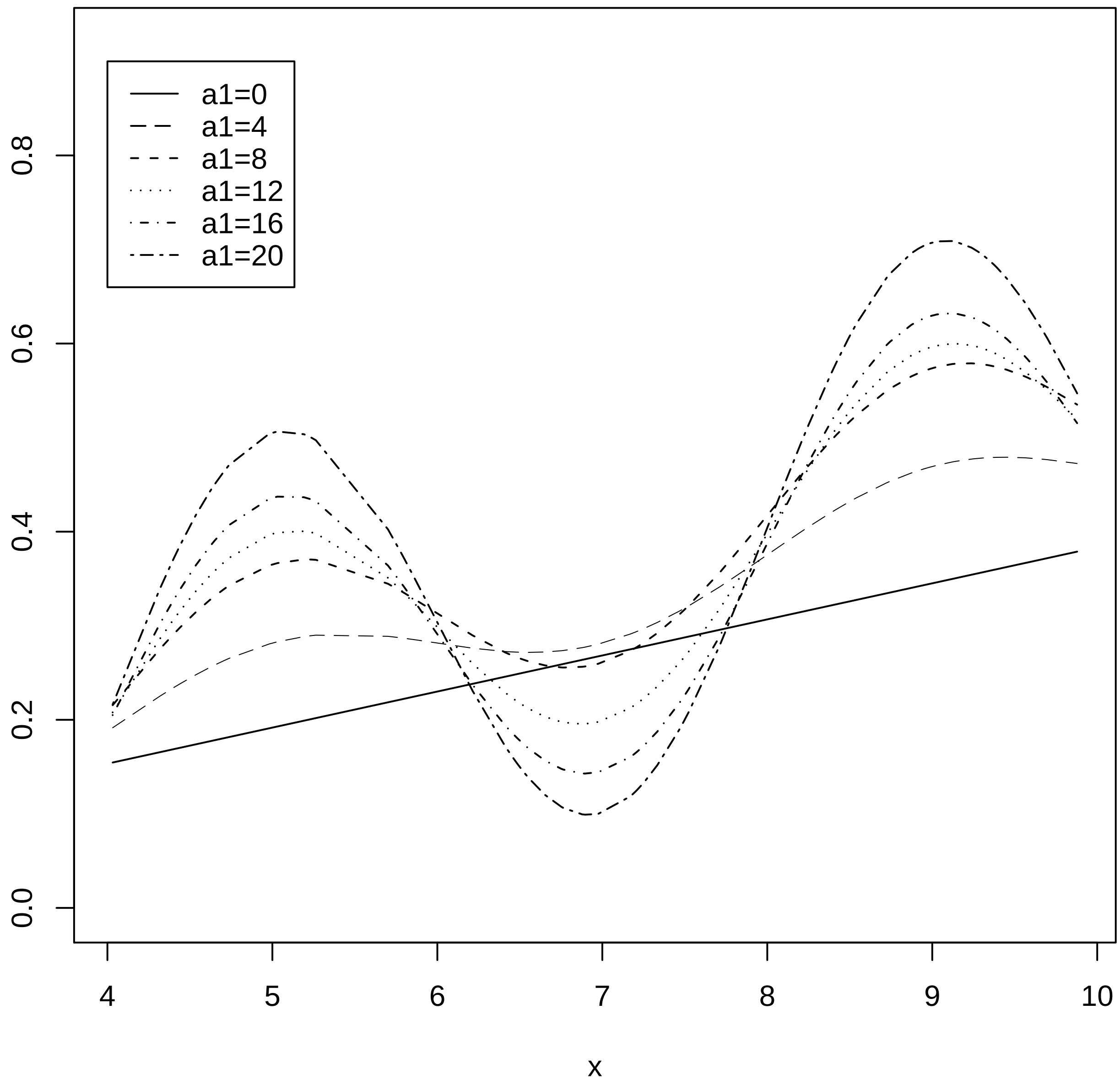

Figure 3.1 : Model (3.3.1) for different values of a_1, at time point $z = 0.01423$.

Cens perc	Trunc perc	h	H_0 $a_1 = 0$	H_1 $a_1 = 4$	$a_1 = 8$	$a_1 = 12$	$a_1 = 16$	$a_1 = 20$
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 3.2 : Power of the test for model (3.3.1) under censoring and truncation, with B=500 and M=1000.

For model (3.3.2), $X \sim U[4,10]$, $Y|_{X=x} \sim \text{Exp}\left(x + a_2 \exp\left(\frac{2x}{3}\right)\right)$, $C|_{X=x} \sim \text{Exp}(d_2 x)$, where $d_2 > 0$ determines the censoring probability, $T|_{X=x} \sim \text{Exp}(r_2 x)$, where $r_2 > 0$ controls the probability of truncation and $\phi(u) = -\log(u)$ (additive hazards model), which gives the true model $\phi(S(z|x)) = zx + a_2 z \exp\left(\frac{2x}{3}\right)$. The sample size is taken $n = 100$, $M = 1000$ Monte Carlo simulations are conducted, $B = 500$ bootstrap re-samples 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 (3.3.1). The level of the test was taken to be $\alpha = 0.05$, H_0 is model (3.3.2) for $a_2 = 0$, while the alternative, H_1, is model (3.3.2) for values of a_2 in $\{0.01, 0.1, 1, 10, 100\}$, see Figure 3.2 for a graphical representation of H_0 and H_1 for a fixed z. As a_2 increases, we go further away from H_0.

Cens	h	H_0	H_1				
perc		$a_2 = 0$	$a_2 = 0.01$	$a_2 = 0.1$	$a_2 = 1$	$a_2 = 10$	$a_2 = 100$
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 3.3 : Power of the test for model (3.3.2) under censoring, with B=500 and M=1000.

Table 3.3 shows the results for model (3.3.2) under censoring. We remark more or less the same features as for model (3.3.1) under censoring.

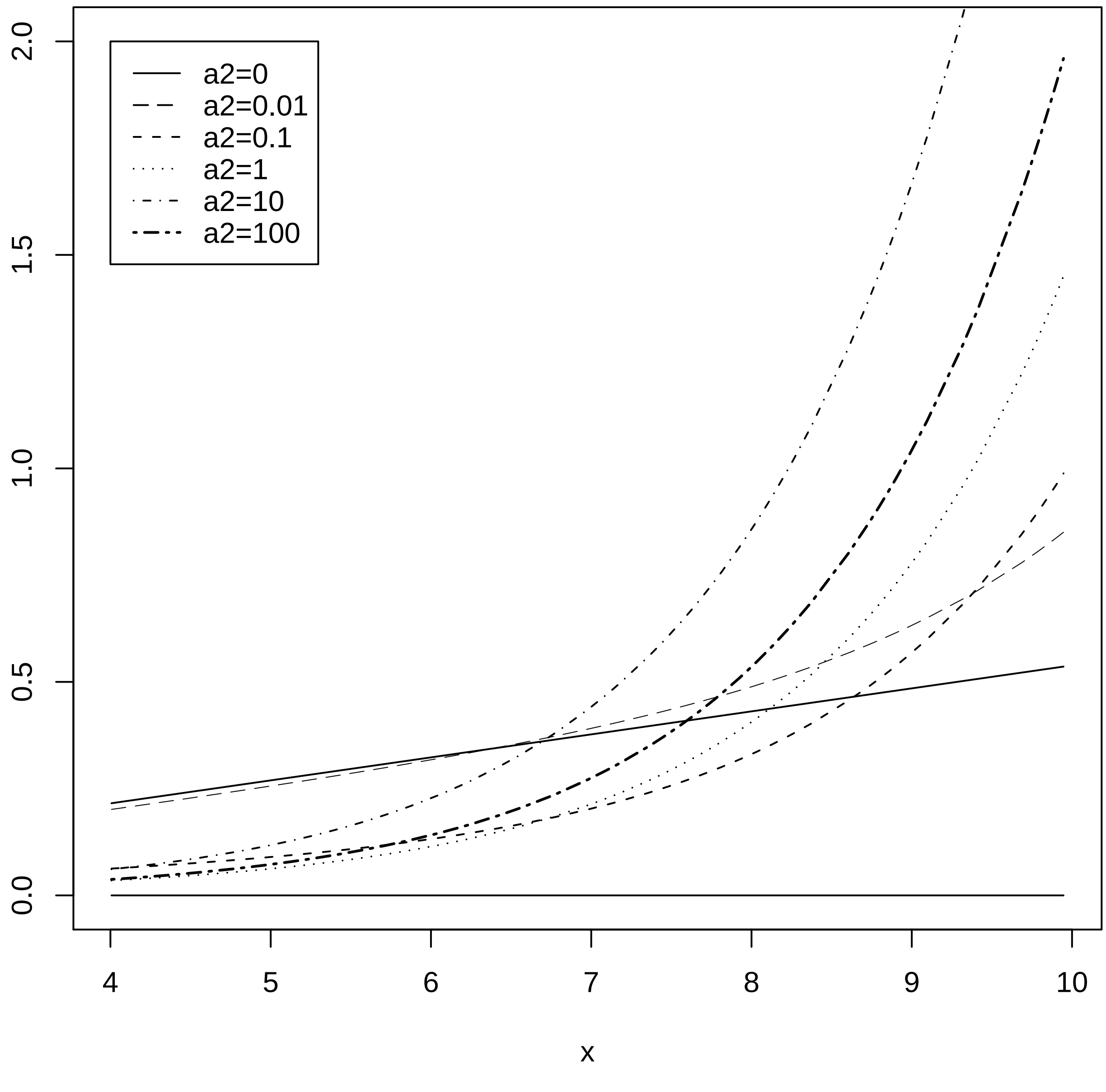

Figure 3.2 : Model (3.3.2) for different values of a_2, at time point $z = 0.000114$.

Cens perc	Trunc perc	h	H_0 $a_2 = 0$	H_1 $a_2 = 0.01$	$a_2 = 0.1$	$a_2 = 1$	$a_2 = 10$	$a_2 = 100$
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 3.4 : Power of the test for model (3.3.2) under censoring and truncation, with B=500 and M=1000.

Table 3.4 displays the results for model (3.3.2) under both censoring and truncation. They look similar to those of Table 3.2, for model (3.3.1) 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.

Remark 3.3.1 The numerical studies give satisfactory results, that can be slightly improved by increasing the number of bootstrap resamples. The results for model (3.3.1) under censoring, with $B = 1000$ bootstrap replications, instead of $B = 500$ as done above and $M = 500$ Monte Carlo simulations are given in Table 3.5.

Cens	h	H_0	H_1				
perc		$a_1 = 0$	$a_1 = 4$	$a_1 = 8$	$a_1 = 12$	$a_1 = 16$	$a_1 = 20$
20	1.2	0.024	0.068	0.212	0.464	0.702	0.910
	1.5	0.046	0.096	0.272	0.562	0.828	0.932
	1.8	0.039	0.134	0.283	0.567	0.837	0.961
	2.1	0.066	0.132	0.311	0.550	0.825	0.950
	2.4	0.070	0.122	0.252	0.484	0.630	0.864
40	1.2	0.022	0.058	0.098	0.212	0.428	0.650
	1.5	0.035	0.098	0.136	0.322	0.496	0.715
	1.8	0.028	0.074	0.168	0.328	0.508	0.786
	2.1	0.048	0.078	0.170	0.291	0.501	0.705
	2.4	0.066	0.129	0.140	0.252	0.397	0.560

Table 3.5 : Power of the test for model (3.3.1) under censoring, with B=1000 and M=500.

3.4 Proof of Theorem 3.1.1

Using (3.1.3), we can write

$$T_n = T_{n,1} + R_{n,1} + R_{n,2}, \tag{3.4.1}$$

with

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

$$R_{n,1} = \int_a^b \frac{1}{n}\sum_{r=1}^{n}\Big[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \ldots + \hat{\beta}_p(z)X_r^p)\Big]^2 dz \tag{3.4.2}$$

$$R_{n,2} = \int_a^b \frac{2}{n}\sum_{r=1}^{n}\Big[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r))\Big]\Big[\phi(S(z|X_r)) - (\hat{\beta}_0(z) + \ldots + \hat{\beta}_p(z)X_r^p)\Big] dz \tag{3.4.3}$$

where $R_{n,1}$ and $R_{n,2}$ are $o_p(n^{-1}h^{-1/2})$ by Lemma 3.5.6 below. The leading term T_{n1} of equation (3.5.15) can be easily handled using a Taylor expansion of ϕ around $S(z|X_r)$:

$$\begin{aligned} T_{n,1} &= \int_a^b \frac{1}{n}\sum_{r=1}^{n}\Big[\phi^{'}(S(z|X_r))(\hat{S}(z|X_r) - S(z|X_r)) + \frac{1}{2}\phi^{''}(\Delta_r(z))(\hat{S}(z|X_r) - S(z|X_r))^2\Big]^2 dz \\ &= T_{n,2} + R_{n,3} + R_{n,4} \end{aligned} \tag{3.4.4}$$

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

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

$$R_{n,3} = \int_a^b \frac{1}{n}\sum_{r=1}^{n}\frac{1}{4}\phi^{''}(\Delta_r(z))^2(\hat{S}(z|X_r) - S(z|X_r))^4 dz = o_p(n^{-1}h^{-1/2}) \tag{3.4.5}$$

$$R_{n,4} = \int_a^b \frac{1}{n}\sum_{r=1}^{n}\phi^{'}(S(z|X_r))\phi^{''}(\Delta_r(z))(\hat{S}(z|X_r) - S(z|X_r))^3 dz = o_p(n^{-1}h^{-1/2}), \tag{3.4.6}$$

where the order of $R_{n,3}$ and $R_{n,4}$ follows from the fact that $\sup_{x\in I}\sup_{a\le z\le b}|\hat{S}(z|x) - S(z|x)| = O_p\left((\ln n)^{1/2}(nh)^{-1/2}\right)$ (see Lemma 1.2.3 in Section 1.1).

By the i.i.d. representation for $\hat{S}(z|X_r)$ given in Theorem 1.2.1,

$$T_{n,2} = \int_a^b \frac{1}{n}\sum_{r=1}^n \phi^{'}(S(z|X_r))^2 \left[S(z|X_r)\sum_{i=1}^n B_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z) + R_n^{'}(z|X_r)\right]^2 dz$$
$$= T_{n,3} + R_{n,5} + R_{n,6}, \tag{3.4.7}$$

where

$$\xi(Z,T,\delta,x,z) = \frac{1_{\{Z\le z,\delta=1\}}}{C(Z|x)} - \int_0^z \frac{1_{\{T\le u\le Z\}}}{C^2(u|x)} dH_1^*(u|x) \tag{3.4.8}$$

and

$$\sup_{\substack{z\in[a,b]\\ x\in I}} |R_n^{'}(z|x)| = O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right). \tag{3.4.9}$$

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

$$R_{n,5} = \int_a^b \frac{1}{n}\sum_{r=1}^n \phi^{'}(S(z|X_r))^2 R_n^{'}(z|X_r)^2 dz = o_p(n^{-1}h^{-1/2})$$

$$\begin{aligned} R_{n,6} &= \int_a^b \frac{2}{n}\sum_{r=1}^n \phi^{'}(S(z|X_r))^2 S(z|X_r) R_n^{'}(z|X_r)\sum_{i=1}^n B_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z)dz \\ &\le \int_a^b \frac{2}{n}\sum_{r=1}^n \phi^{'}(S(z|X_r))^2 S(z|X_r)\sup_z|R_n^{'}(z|X_r)| \left|\sum_{i=1}^n B_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z)\right| dz \\ &= O_p\left((T_{n,3})^{1/2}\cdot\left(\frac{\ln n}{nh}\right)^{3/4}\right). \end{aligned} \tag{3.4.10}$$

By Lemma 3.5.5 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 :

$$\begin{aligned} T_{n,3} &= \int_a^b \frac{1}{n}\sum_{r=1}^n \Big[\sum_{i=1}^n B_{ni}(X_r)\epsilon_i(z) + \Delta_r^{(1)}(z)\Big]^2 dz \\ &= \int_a^b \frac{1}{n}\sum_{r=1}^n \left[\sum_{i=1}^n B_{ni}(X_r)\epsilon_i(z)\right]^2 dz + \int_a^b \frac{1}{n}\sum_{r=1}^n \left[\Delta_r^{(1)}(z)\right]^2 dz \\ &\quad + \int_a^b \frac{1}{n}\sum_{r=1}^n\sum_{i=1}^n B_{ni}(X_r)\epsilon_i(z)\Delta_r^{(1)}(z)dz = T_{n,4} + R_{n,1} + R_{n,8}, \end{aligned} \tag{3.4.11}$$

with

$$\epsilon_i(z) = \phi'(S(z|X_i))S(z|X_i)\xi(Z_i, T_i, \delta_i, X_i, z) \tag{3.4.12}$$

$$\Delta_r^{(1)}(z) = \sum_{j=1}^{n} B_{nj}(X_r)\Big(\phi'(S(z|X_r))S(z|X_r)\xi(Z_j, T_j, \delta_j, X_r, z) - \\ - \phi'(S(z|X_j))S(z|X_j)\xi(Z_j, T_j, \delta_j, X_j, z)\Big).$$

The terms $R_{n,1}$ and $R_{n,8}$ are $o_p(n^{-1}h^{-1/2})$ by Lemma 3.5.4 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 :

$$T_{n,4} = \int_a^b \frac{1}{n}\sum_{r=1}^{n}\left[\sum_{i=1}^{n} B_{ni}(X_r)\epsilon_i(z)\right]^2 dz = T_{n,5} + R_{n,9} \tag{3.4.13}$$

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

$$R_{n,9} = \int_a^b \sum_{i=1}^{n} b_{nii}\epsilon_i^2(z)dz. \tag{3.4.14}$$

In Lemma 3.5.3 it is shown that

$$R_{n,9} = \frac{1}{nh}\left(\int_a^b\int_I g(z,z,x)dxdz\right)K^{(2)}(0) + o_p(n^{-1}h^{-1/2}). \tag{3.4.15}$$

To analyze $T_{n,5}$, 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 $m(x)$:

$$\hat{m}^*(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{3.4.16}$$

which implies that

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

where $\tilde{b}_{nij} = \frac{1}{n}\sum_{r=1}^{n} \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)$.

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

$$T_{n,6} = \int_a^b 2\sum_{i<j} \tilde{b}_{nij}\epsilon_i(z)\epsilon_j(z)dz \tag{3.4.17}$$

$$R_{n,10} = \int_a^b \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} \tilde{B}_{ni}(X_r)B_{nj}(X_r)\frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)}\epsilon_i(z)\epsilon_j(z)dz \tag{3.4.18}$$

$$R_{n,11} = \int_a^b \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} B_{ni}(X_r)\tilde{B}_{nj}(X_r)\frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)}\epsilon_i(z)\epsilon_j(z)dz \tag{3.4.19}$$

$$R_{n,12} = \int_a^b \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} B_{ni}(X_r)B_{nj}(X_r)\frac{[m^*(X_r) - \hat{m}^*(X_r)]^2}{m^*(X_r)^2}\epsilon_i(z)\epsilon_j(z)dz. \tag{3.4.20}$$

From Lemma 3.5.2 we have that $R_{n,10}$, $R_{n,11}$ and $R_{n,12}$ are $o_p(n^{-1}h^{-1/2})$, while from Lemma 3.5.1 we get that

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

The result now follows, since we have shown that

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

where all $R_{n,k}$ $(k = 1, \ldots, 12)$ are $o_p(n^{-1}h^{-1/2})$, except R_{n9}, whose asymptotic expression is given in (3.4.15).

3.5 Lemmas and proofs

3.5.1 Lemma 3.5.1

Lemma 3.5.1

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

with $T_{n,6}$ and V defined in (3.4.17) and (3.1.4), respectively.

Proof : Write

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

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

$$\begin{aligned}
&\mathrm{Cov}\left(\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz, \tilde{b}_{nkl}\int_a^b \epsilon_k(z)\epsilon_l(z)dz\right)\\
&= E\left(\mathrm{Cov}\left(\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz, \tilde{b}_{nkl}\int_a^b \epsilon_k(z)\epsilon_l(z)dz|\tilde{\boldsymbol{X}}\right)\right)+\\
&\qquad +\mathrm{Cov}\left(E\left(\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz|\tilde{\boldsymbol{X}}\right), E\left(\tilde{b}_{nkl}\int_a^b \epsilon_k(z)\epsilon_l(z)dz|\tilde{\boldsymbol{X}}\right)\right)\\
&= E\left(\tilde{b}_{nij}\tilde{b}_{nkl}\mathrm{Cov}\left(\int_a^b \epsilon_i(z)\epsilon_j(z)dz, \int_a^b \epsilon_k(z)\epsilon_l(z)dz|\tilde{\boldsymbol{X}}\right)\right)
\end{aligned}$$

$$= \begin{cases} E\left(\tilde{b}_{nij}^2\mathrm{Var}\left(\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i,X_j\right)\right) = E\left(\tilde{b}_{n12}^2 g_2(X_1,X_2)\right), & \text{if } i=k<j=l\\ 0, & \text{otherwise}\end{cases}$$

where

$$g_2(X_1,X_2) = \int_a^b\int_a^b g(z_1,z_2,X_1)g(z_1,z_2,X_2)dz_1dz_2, \tag{3.5.1}$$

with $g(z_1,z_2,x)$ defined in (3.1.5). Hence,

$$\mathrm{Var}(T_{n,6}) = 4\cdot\frac{n(n-1)}{2}E\left(\tilde{b}_{n12}^2 g_2(X_1,X_2)\right) \tag{3.5.2}$$

$$= \frac{2(n-1)}{n(nh)^4}\Bigg[(n-2)(n-3)E\left(\frac{K\left(\frac{X_3-X_1}{h}\right)K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)^2}\cdot\frac{K\left(\frac{X_4-X_1}{h}\right)K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)^2}g_2(X_1,X_2)\right)$$
$$+(n-2)E\left(\frac{K\left(\frac{X_3-X_1}{h}\right)^2K\left(\frac{X_3-X_2}{h}\right)^2}{m^*(X_3)^4}g_2(X_1,X_2)\right)$$
$$+4(n-2)E\left(\frac{K(0)K\left(\frac{X_1-X_2}{h}\right)}{m^*(X_1)^2}\cdot\frac{K\left(\frac{X_4-X_1}{h}\right)K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)^2}g_2(X_1,X_2)\right)$$
$$+2E\left(\frac{K(0)^2K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^2m^*(X_2)^2}g_2(X_1,X_2)\right)$$
$$+2E\left(\frac{K(0)^2K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^4}g_2(X_1,X_2)\right)\Bigg]$$

From Appendix A.3.1

$$E\left(\frac{K\left(\frac{X_3-X_1}{h}\right)K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)^2}\cdot\frac{K\left(\frac{X_4-X_1}{h}\right)K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)^2}g_2(X_1,X_2)\right) \tag{3.5.3}$$
$$= h^3\int(K*K(v))^2dv\int_a^b\int_a^b\int_I g(z_1,z_2,x)^2dxdz_1dz_2+O(h^4)$$

$$E\left(\frac{K\left(\frac{X_3-X_1}{h}\right)K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)^4}g_2(X_1,X_2)\right)=O(h^2) \tag{3.5.4}$$

$$E\left(\frac{K(0)K\left(\frac{X_1-X_2}{h}\right)}{m^*(X_1)^2}\frac{K\left(\frac{X_4-X_1}{h}\right)K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)^2}g_2(X_1,X_2)\right)=O(h^2) \tag{3.5.5}$$

$$E\left(\frac{K(0)^2K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^2m^*(X_2)^2}g_2(X_1,X_2)\right)=O(h) \tag{3.5.6}$$

$$E\left(\frac{K(0)^2K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^4}g_2(X_1,X_2)\right)=O(h) \tag{3.5.7}$$

So, by using (H10), we get

$$\mathrm{Var}(T_{n,6}) = \frac{2(n-1)}{n(nh)^4}\Bigg[(n-2)(n-3)\Big[h^3\int K*K(v)^2dv\int_a^b\int_a^b\int_I g^2(z_1,z_2,x)dxdz_1dz_2$$

$$+O(h^4)\Big] + O(nh^2) + O(h)\Bigg]$$

$$= \frac{2(n-1)(n-2)(n-3)}{n^5 h}\left(K*K*K*K(0)\int_a^b\int_a^b\int_I g^2(z_1,z_2,x)dxdz_1dz_2\right)$$

$$+O(n^{-2}) + O(n^{-3}h^{-2}) + O(n^{-4}h^{-3}) + O(n^{-4}h^{-4})$$

$$= \frac{2}{n^2h}K*K*K*K(0)\int_a^b\int_a^b\int_I g^2(z_1,z_2,x)dxdz_1dz_2 + o(n^{-2}h^{-1})$$

In order to prove the asymptotic convergence, we will use the central limit theorem for double sums given by de Jong (1981). Using the notations of that theorem, let $W(n) = T_{n,6}$, with $\boldsymbol{V}_i = (X_i, Z_i, T_i, \delta_i)$ and

$$W_{nij}(\boldsymbol{V}_i, \boldsymbol{V}_j) = 2\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz. \tag{3.5.8}$$

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

$$\sigma_{ij}^2 = E(W_{nij}^2) = 4E\left\{\tilde{b}_{nij}^2 \mathrm{Var}\Big[\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i, X_j\Big]\right\}$$

$$= \frac{4}{n^4h}\left(\int K*K(v)^2dv\int_a^b\int_a^b\int_I g^2(z_1,z_2,x)dxdz_1dz_2\right) + o(n^{-4}h^{-1})$$

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

Hence, the first condition in de Jong (1981) is satisfied.

In order to check the second condition, we will use the same terminology as in de Jong (1981), page 266 :

$$E(W(n)^4) = E\left((T_{n,6})^4\right) = G_I + 6G_{II} + 12G_{III} + 24G_{IV} + 6G_V \tag{3.5.9}$$

with

$$G_I = \sum_{i<j} E(W_{ij}^4) \tag{3.5.10}$$

$$G_{II} = \sum_{i<j<k}\left(E(W_{ij}^2W_{ik}^2) + E(W_{ji}^2W_{jk}^2) + E(W_{ki}^2W_{kj}^2)\right) \tag{3.5.11}$$

$$G_{III} = \sum_{i<j<k} \left(E(W_{ij}^2 W_{ki} W_{kj}) + E(W_{ik}^2 W_{ji} W_{jk}) + E(W_{kj}^2 W_{ij} W_{ik}) \right) \tag{3.5.12}$$

$$G_{IV} = \sum_{i<j<k<l} (E(W_{ij} W_{ik} W_{lj} W_{lk}) + E(W_{ij} W_{il} W_{kj} W_{kl}) + E(W_{ik} W_{il} W_{jk} W_{jl})) \tag{3.5.13}$$

$$G_V = \sum_{i<j<k<l} \left(E(W_{ij}^2 W_{kl}^2) + E(W_{ik}^2 W_{jl}^2) + E(W_{il}^2 W_{jk}^2) \right). \tag{3.5.14}$$

From Appendix A.3.2 we have 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,

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

So we get the desired result.

3.5.2 Lemma 3.5.2

Lemma 3.5.2 *$R_{n,10}$, $R_{n,11}$ and $R_{n,12}$ defined in (3.4.18), (3.4.19) and (3.4.20), respectively, are $o_p(n^{-1}h^{-1/2})$.*

Proof :

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

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

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

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

Note that $B_1 \leq 1$ and $A_2 \leq 1$. Moreover, $E(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon) = 0$ and $\mathrm{Var}(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon) = o(n^{-2}h^{-1})$, provided that $\sup_x |\hat{m}^*(x) - m^*(x)| = o_p(1)$. see details in Appendix A.3.3. This proves that $A_1 = o(1)$.

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

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

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

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

3.5.3 Lemma 3.5.3

Lemma 3.5.3

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

where $R_{n,9}$ and $g(z,z,x)$ are defined in (3.4.14) and (3.1.5), 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,

$$\begin{aligned} R_{n,9} &= \int_a^b \sum_{i=1}^n b_{nii}\epsilon_i^2(z)dz \\ &= \int_a^b \frac{1}{n^3h^2}\sum_{r=1}^n \frac{1}{\hat{m}^*(X_r)^2}\sum_{i=1}^n K\left(\frac{X_r - X_i}{h}\right)^2 \epsilon_i^2(z)dz \\ &= R_{n,9}^{(1)} + R_{n,9}^{(2)}, \end{aligned} \tag{3.5.15}$$

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

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

The mean and the variance of $R_{n,9}^{(1)}$ can be studied by using Taylor expansions (see more details in Appendix A.3.4) :

$$E(R_{n,9}^{(1)}) = \frac{1}{nh}\left(\int_a^b \int_I g(z,z,x)dxdz\right)\left(\int K^2(t)dt\right) + O(n^{-1}h),$$
$$\mathrm{Var}(R_{n,9}^{(1)}) = O(n^{-4}h^{-3}),$$

which implies that

$$R_{n,9}^{(1)} = \frac{1}{nh}\left(\int_a^b \int_I g(z,z,x)dxdz\right)\left(\int K^2(t)dt\right) + O(n^{-1}h + n^{-2}h^{-3/2}) \quad (3.5.16)$$

Standard arguments, condition (H10) and Theorem B in Silverman (1918) imply that

$$R_{n,9}^{(2)} \leq \int_a^b |R_{n,9}^{(3)}(z)|dz \leq \frac{\Gamma(1+D)D}{\gamma^2} R_{n,9}, \quad (3.5.17)$$

where

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

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

We now use (3.5.15), (3.5.17) and (3.5.16) to conclude.

3.5.4 Lemma 3.5.4

Lemma 3.5.4 *$R_{n,1}$ and $R_{n,8}$ defined in (3.4.11) are $o_p(n^{-1}h^{-1/2})$.*

Proof : The term $R_{n,1}$ can be handled in a similar way as $T_{n,5}$ in the proof of Theorem 3.1.1 in order to get rid of the random denominator in $B_{ni}(X_r)$. Thus, we get the following :

$$R_{n,1} = \int_a^b \frac{1}{n} \sum_{r=1}^n \left(\Delta_r^{(1)}(z)\right)^2 dz = \int_a^b (1 + o_p(1)) \frac{1}{n} \sum_{r=1}^n \sum_{i,j=1}^n \tilde{\Delta}_{rij}^{(1)}(z)dz \quad (3.5.18)$$

with $\tilde{\Delta}^{(1)}_{rij}(z) = \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)\eta_{ri}(z)\eta_{rj}(z)$ 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 $\tilde{\Delta}^{(1)}_{rij} = 0$ when $r = i$ or $r = j$. On the other hand, using (3.4.8), 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

$$E(\xi(Z, T, \delta, y, z)^2|X = x) = \int_0^z \frac{dH_1^*(u|y)}{C(u|x)^2}. \tag{3.5.19}$$

see Appendix A.2.2.

It is straightforward, but long and tedious to compute the order of $E(\tilde{\Delta}^{(1)}_{rij})$. Standard arguments such as changes of variables and Taylor expansions lead to

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

uniformly in z (see details in Appendix A.3.5). These results imply that

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

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

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

Now (3.5.18) implies that

$$R_{n,1} = \int_a^b \frac{1}{n}\sum_{r=1}^{n} \left(\Delta^{(1)}_r(z)\right)^2 dz = O_p(h^4 + n^{-1}h). \tag{3.5.20}$$

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

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

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

3.5.5 Lemma 3.5.5

Lemma 3.5.5 *$R_{n,6}$ defined in (3.4.10) is $o_p(n^{-1}h^{-1/2})$.*

Proof : Note that

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

From Lemma 3.5.1, 3.5.2, 3.5.3, condition (H10) and the fact that $T_{n,4} = R_{n,9} + T_{n,5}$ we get

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

This limit distribution together with Lemma 3.5.4 gives that

$$\sqrt{n^2h}T_{n,3} - h^{-1/2}K^{(2)}(0)\int_a^b\int_I g(z,z,x)dxdz \xrightarrow{d} N(0,V). \tag{3.5.21}$$

As a consequence, $T_{n,3} = O_p(n^{-1}h^{-1})$, thus $R_{n,6} = O_p(n^{-5/4}h^{-5/4}(\ln n)^{3/4}) = o_p(n^{-1}h^{-1/2})$.

3.5.6 Lemma 3.5.6

Lemma 3.5.6 *$R_{n,1}$ and $R_{n,2}$ defined in (3.4.2) and (3.4.3), respectively, are $o_p(n^{-1}h^{-1/2})$.*

Proof : Write

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

and by Cauchy-Schwarz inequality

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

where $\sup_{z\in[a,b]} |\boldsymbol{\beta}(z) - \hat{\boldsymbol{\beta}}(z)| = O_p(n^{-1/2})$ by Corollary 2.4.2 in Section 2.4, and where $\frac{1}{n}\sum_{r=1}^n \left[\phi(\hat{S}(z|X_r)) - \phi(S(z|X_r))\right]^2 = O_p(n^{-1}h^{-1/2})$ uniformly in z.

Appendix A.3.1

We will prove that the terms (3.5.4), (3.5.5), (3.5.6) and (3.5.7) are $O(h^2)$, $O(h^2)$, $O(h)$ and $O(h)$, respectively and that (3.5.3) is

$$h^3 \int (K * K(v))^2 dv \int_a^b \int_a^b \int_I g(z_1, z_2, x)^2 dx dz_1 dz_2 + O(h^4).$$

The term (3.5.3) is equal to

$$\int_I\int_I\int_I\int_I \frac{1}{m^*(x_3)^2}\frac{1}{m^*(x_4)^2} K\left(\frac{x_3-x_1}{h}\right)K\left(\frac{x_3-x_2}{h}\right)K\left(\frac{x_4-x_1}{h}\right)K\left(\frac{x_4-x_2}{h}\right)$$
$$\times\, g_2(x_1,x_2)m^*(x_1)m^*(x_2)m^*(x_3)m^*(x_4)dx_1dx_2dx_3dx_4$$

$$= \int_a^b\int_a^b\int_I\int_I \frac{1}{m^*(x_3)}\frac{1}{m^*(x_4)}$$
$$\times \int_I K\left(\frac{x_1-x_3}{h}\right)K\left(\frac{x_1-x_4}{h}\right)\cdot g(z_1,z_2,x_1)m^*(x_1)dx_1$$
$$\times \int_I K\left(\frac{x_2-x_3}{h}\right)K\left(\frac{x_2-x_4}{h}\right)g(z_1,z_2,x_2)m^*(x_2)dx_2\cdot dx_3dx_4dz_1dz_2$$
$$= \int_a^b\int_a^b\int_I\int_I \frac{1}{m^*(x_3)}\frac{1}{m^*(x_4)}\Psi(z_1,z_2,x_3,x_4)^2dx_4dx_3dz_1dz_2,$$

where

$$\Psi(z_1,z_2,y_1,y_2) = \int_I K\left(\frac{x-y_1}{h}\right)K\left(\frac{x-y_2}{h}\right)g(z_1,z_2,x)m^*(x)dx.$$

A change of variable ($x = uh + x_3$) and a Taylor expansion of $g\cdot m^*$ around x_3 give

$$\Psi(z_1,z_2,x_3,x_4) = \int_I K\left(\frac{x-x_3}{h}\right)K\left(\frac{x-x_4}{h}\right)g(z_1,z_2,x)m^*(x)dx$$
$$= h\int K(u)K\left(\frac{x_4-x_3}{h}-u\right)\Big[g(z_1,z_2,x_3)m^*(x_3)+o(h)\Big]du$$
$$= hg(z_1,z_2,x_3)m^*(x_3)\int K(u)K\left(\frac{x_4-x_3}{h}-u\right)du + o(h)$$
$$= hg(z_1,z_2,x_3)m^*(x_3)K*K\left(\frac{x_4-x_3}{h}\right)+O(h^2)$$

Thus,

$$\Psi(z_1,z_2,x_3,x_4)^2 = h^2g(z_1,z_2,x_3)^2m^*(x_3)^2\left(K*K\left(\frac{x_4-x_3}{h}\right)\right)^2 +$$
$$+\,O(h^2)hg(z_1,z_2,x_3)m^*(x_3)K*K\left(\frac{x_4-x_3}{h}\right)+O(h^4)$$
$$= h^2g(z_1,z_2,x_3)^2m^*(x_3)^2\left(K*K\left(\frac{x_4-x_3}{h}\right)\right)^2+O(h^3) \qquad (3.5.22)$$

which implies that (3.5.3) is

$$\int_a^b \int_a^b \int_I \int_I \frac{m^*(x_3)}{m^*(x_4)} h^2 g(z_1, z_2, x_3)^2 \left(K * K \left(\frac{x_4 - x_3}{h} \right) \right)^2 dx_3 dx_4 dz_1 dz_2 + O(h^3)$$

Again, a change of variable ($x_4 = hv + x_3$) gives

$$\int_a^b \int_a^b \int_I \frac{1}{m^*(x_4)} h \int h^2 g(z_1, z_2, x_4 - hv)^2 (K * K(v))^2 m^*(x_4 - hv) dv dx_4 dz_1 dz_2 + O(h^3)$$

and a Taylor expansion of $g_1^2 \cdot m^*$ around x_4 gives us the final expression of (3.5.3)

$$h^3 \int (K * K(v))^2 dv \int_a^b \int_a^b \int_I g(z_1, z_2, x)^2 dx dz_1 dz_2 + O(h^4)$$

The term (3.5.4) can be written as follows

$$\begin{aligned} &\int_I \int_I \int_I \frac{1}{m^*(x_3)^4} K \left(\frac{x_3 - x_1}{h} \right)^2 K \left(\frac{x_3 - x_2}{h} \right)^2 \\ &\qquad \cdot \int_a^b \int_a^b g(z_1, z_2, x_1) g(z_1, z_2, x_2) dz_1 dz_2 m^*(x_3) m^*(x_2) m^*(x_1) dx_3 dx_2 dx_1 \\ &= \int_a^b \int_a^b \int_I \frac{1}{m^*(x_3)^3} \Psi(z_1, z_2, x_3, x_3)^2 dx_3 dz_1 dz_2 \end{aligned}$$

which by (3.5.22) is

$$\int_a^b \int_a^b \int_I \frac{1}{m^*(x_3)^3} \left[h^2 g(z_1, z_2, x_3)^2 m^*(x_3)^2 \left(K * K \left(\frac{x_3 - x_3}{h} \right) \right)^2 + O(h^3) \right]^2 dx_3 dz_1 dz_2$$

$= O(h^2)$, since all the quantities involved in the integrals are bounded.

(3.5.5) can be written in a similar way as (3.5.4)

$$\begin{aligned} &\int_I \int_I \int_I \frac{1}{m^*(x_1)^2} \frac{1}{m^*(x_4)^2} K(0) K \left(\frac{x_1 - x_2}{h} \right) K \left(\frac{x_4 - x_1}{h} \right) K \left(\frac{x_4 - x_2}{h} \right) \\ &\qquad \times g_2(x_1, x_2) m^*(x_1) m^*(x_2) m^*(x_4) dx_1 dx_2 dx_4 \\ &= \int_I \int_I \int_I \frac{1}{m^*(x_1)} \frac{m^*(x_2)}{m^*(x_4)} K(0) K \left(\frac{x_1 - x_2}{h} \right) K \left(\frac{x_4 - x_1}{h} \right) K \left(\frac{x_4 - x_2}{h} \right) \\ &\qquad \times \int_a^b \int_a^b g(z_1, z_2, x_1) g(z_1, z_2, x_2) dz_1 dz_2 dx_1 dx_2 dx_4 \\ &= \int_a^b \int_a^b \int_I \int_I \frac{1}{m^*(x_1)} \frac{1}{m^*(x_4)} K(0) K \left(\frac{x_4 - x_1}{h} \right) g(z_1, z_2, x_1) \end{aligned}$$

$$\times \underbrace{\int_I m^*(x_2) K\left(\frac{x_1 - x_2}{h}\right) K\left(\frac{x_4 - x_2}{h}\right) g(z_1, z_2, x_2) dx_2}_{\Psi(z_1,z_2,x_1,x_4)} dx_4 dx_1 dz_1 dz_2$$

$$= \int_a^b \int_a^b \int_I \int_I \frac{1}{m^*(x_1)} \frac{1}{m^*(x_4)} hK(0) K\left(\frac{x_4 - x_1}{h}\right) K * K\left(\frac{x_4 - x_1}{h}\right)$$
$$\times g^2(z_1, z_2, x_1) m^*(x_1) dx_4 dx_1 dz_1 dz_2 + O(h^2)$$

A change of variable ($x_4 = x_1 + hv$) and Taylor expansion of g^2 around x_4 give us

$$= \int_a^b \int_a^b \int_I \int \frac{1}{m^*(x_4)} h^2 K(0) K(v) K * K(v) g^2(z_1, z_2, x_4 - hv) dx_4 dv dz_1 dz_2 + O(h^2)$$

$$= \int_a^b \int_a^b \int_I \int \frac{1}{m^*(x_4)} h^2 K(0) K(v) K*K(v) g^2(z_1, z_2, x_4) dx_4 dv dz_1 dz_2 + O(h^3) + O(h^2)$$
$$= h^2 \underbrace{\int K(0) K(v) K * K(v) dv}_{\text{bounded}} \cdot \underbrace{\int_a^b \int_a^b \int_I \frac{1}{m^*(x_4)} g^2(z_1, z_2, x_4) dx_4 dz_1 dz_2}_{\text{bounded}} + O(h^2)$$
$$= O(h^2)$$

(3.5.6) can be treated in the following way

$$\int_I \int_I K(0)^2 K\left(\frac{x_1 - x_2}{h}\right)^2 \frac{1}{m^*(x_1)^2} \frac{1}{m^*(x_2)^2} g_2(x_1, x_2) m^*(x_1) m^*(x_2) dx_2 dx_1$$
$$= \int_I \int_I K(0)^2 K\left(\frac{x_1 - x_2}{h}\right)^2 \frac{1}{m^*(x_1)} \frac{1}{m^*(x_2)} \int_a^b \int_a^b g(z_1, z_2, x_1) g(z_1, z_2, x_2) dz_1 dz_2 dx_2 dx_1$$

A change of variable ($x_2 = x_1 - hv$)

$$= \int_I \int hK(0)^2 K(v)^2 \frac{1}{m^*(x_2 + hv)} \frac{1}{m^*(x_2)} \int_a^b \int_a^b g(z_1, z_2, x_2 + hv) g(z_1, z_2, x_2) dz_1 dz_2 dv dx_2$$
$$= \int_I hK(0)^2 \frac{1}{m^*(x_2)} \int_a^b \int_a^b g(z_1, z_2, x_2) \int K(v)^2 \frac{g(z_1, z_2, x_2 + hv)}{m^*(x_2 + hv)} dv dz_1 dz_2 dx_2$$

A Taylor expansion of g/m^* around x_2 gives : $\dfrac{g(z_1, z_2, x_2 + hv)}{m^*(x_2 + hv)} = \dfrac{g(z_1, z_2, x_2)}{m^*(x_2)} + O(h)$

$$= \int K(v)^2 dv \int_I hK(0)^2 \frac{1}{m^*(x_2)^2} \int_a^b \int_a^b g(z_1, z_2, x_2)^2 dz_1 dz_2 dx_2 + O(h)$$
$$= O(h)$$

Same computations as for (3.5.6) give that (3.5.7) is $O(h)$.

Appendix A.3.2

We will prove that (3.5.10), (3.5.11), (3.5.12) and (3.5.13) are $O(n^{-6}h^{-3})$, $O(n^{-5}h^{-2})$, $O(n^{-5}h^{-2})$ and $O(n^{-4}h^{-1})$, respectively and that (3.5.14) is equal to $C_V n^{-4}h^{-2} + o(n^{-4}h^{-2})$, for some $C_V > 0$.

First notice that since Z and T are independent, given X, we have that

$$
\begin{aligned}
C(z|x) &= P(T \leq z \leq Z|x) = P(T \leq z|x)P(Z \geq z|x) \\
&= L(z|x)(1 - H(z|x)) > 0, \qquad (3.5.23)
\end{aligned}
$$

uniformly in $x \in I$, $z \leq \tau$

Also, by condition (H2)

$$
\xi(Z,T,\delta,x,y) = \frac{I_{\{Z\leq y,\delta=1\}}}{C(Z|x)} - \underbrace{\int_0^y \frac{I_{\{T\leq u\leq Z\}}}{C^2(u|x)}dH_1^*(u|x)}_{>0}
$$

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

Then, for G_I we get that

$$
\begin{aligned}
G_I &= \sum_{i<j} E(W_{ij}^4) = \sum_{i<j} E\left(\tau^4 \left(\frac{2}{n}\tilde{b}_{nij}\frac{1}{\tau}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right)^4\right) \\
&\leq \sum_{i<j} E\left(\frac{2\tau^3}{n^4}\tilde{b}_{nij}^4 \int_a^b \epsilon_i^4(z)\epsilon_j^4(z)dz\right)
\end{aligned}
$$

From (3.4.12) we have that

$$
\epsilon_i(z) = \phi'(S(z|X_i))S(z|X_i)\xi(Z_i,T_i,\delta_i,X_i,z).
$$

Since $\xi(Z,T,\delta,x,y)$ is bounded uniformly in z, then $\epsilon_i(z)$ is also bounded uniformly in z, thus $\int_a^b \epsilon_i^4(z)\epsilon_j^4(z)dz$ is also bounded by a constant S_1, which gives

$$
G_I \leq S_1 \sum_{i<j} E\left(\frac{1}{n^4}\tilde{b}_{nij}^4\right) = n\cdot nh \cdot \frac{1}{n^4}\cdot O\left(\frac{1}{n^4h^4}\right) = O(n^{-6}h^{-3})
$$

where

$$
\tilde{b}_{nij} = \sum_r \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r) = \begin{cases} 0, & |X_i - X_j| > 2h \\ O(n^{-1}h^{-1}), & |X_i - X_j| \leq 2h \end{cases}
$$

only for X_r in the neighborhood of X_i and X_j and j only such that X_j in neighborhood of X_i.

For G_{II} we have that

$$E(W_{ij}^2 W_{ik}^2) = E\left[\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right)^2 \left(\frac{1}{n}\tilde{b}_{nik}\int_a^b \epsilon_i(z)\epsilon_k(z)dz\right)^2\right]$$
$$\leq E\left[\left(\frac{\tau}{n^2}\tilde{b}_{nij}^2\int_a^b \epsilon_i^2(z)\epsilon_j^2(z)dz\right)\left(\frac{\tau}{n^2}\tilde{b}_{nik}^2\int_a^b \epsilon_i^2(z)\epsilon_k^2(z)dz\right)\right] = O(n^{-8}h^{-4}),$$

since $\int_a^b \epsilon_i^2(z)\epsilon_j^2(z)dz$ is bounded and if i, j, k are "close" we get

$$G_{II} \leq \sum_{i<j<k} O(n^{-8}h^{-4}) = O(n^{-5}h^{-2})$$

G_{III} and G_{IV} can be treated in a similar way as G_{II} and we get $G_{III} = O(n^{-5}h^{-2})$ and $G_{IV} = O(n^{-4}h^{-1})$.

Now,

$$\begin{aligned}
G_V &= \sum_{i<j<k<l}\left(E(W_{ij}^2W_{kl}^2) + E(W_{ik}^2W_{jl}^2) + E(W_{il}^2W_{jk}^2)\right)\\
&= \sum_{i<j<k<l}\left(E(W_{ij}^2)E(W_{kl}^2) + E(W_{ik}^2)E(W_{jl}^2) + E(W_{il}^2)E(W_{jk}^2)\right)\\
&= C_V n^{-4}h^{-2} + O(n^{-4}h^{-2}),\ \text{for some } C_V > O.
\end{aligned}$$

with

$$\sigma_{ij}^2 = E(W_{ij}^2) = E\left(\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right)^2\right) =$$
$$= Var\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right) + E\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right)^2 =$$
$$= Var\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right) + E\left[E\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i,X_j\right)\right]^2 =$$
$$= Var\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right) + E\left[\frac{1}{n}\tilde{b}_{nij}\int_a^b \underbrace{E(\epsilon_i(z)|X_i)}_{=0}\underbrace{E(\epsilon_j(z)dz|X_j)}_{=0}\right]^2 =$$
$$= Var\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz\right) =$$
$$= Var\left(E\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i,X_j\right)\right) + E\left(Var\left(\frac{1}{n}\tilde{b}_{nij}\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i,X_j\right)\right)$$

$$= \frac{1}{n^2} E\left(\tilde{b}_{nij}\tilde{b}_{nkl} Var\left(\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_i, X_j\right)\right)$$
$$= \frac{(n-2)(n-3)h^3}{n^2 n^4 h^4} C_1 + O(n^{-6}h^{-3}),$$

see $Var(T_{n,6})$ for details.

This gives

$$G_V = n^{-4}h^{-2}C_1^2 + o(n^{-4}h^{-2})$$

Appendix A.3.3

We will prove that

$$E(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon) = 0$$

and

$$Var(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon) = O(n^{-2}h^{-1}).$$

First,

$$E(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon)$$
$$= E\left(2\sum_{i<j} \tilde{b}_{nij}^{(1)} \int_a^b \epsilon_i(z)\epsilon_j(z)dz \,\middle|\, \inf_x \hat{m}^*(x) > \epsilon\right)$$
$$= 2\sum_{i<j} E\left[\tilde{b}_{nij}^{(1)} E\left(\int_a^b \epsilon_i(z)\epsilon_j(z)dz|X_1, \ldots, X_n, \inf_x \hat{m}^*(x) > \epsilon\right)|\inf_x \hat{m}^*(x) > \epsilon\right]$$
$$= 2\sum_{i<j} E\left[\tilde{b}_{nij}^{(1)} \int_a^b \underbrace{E(\epsilon_i(z)|X_i)}_{=0}\underbrace{E(\epsilon_j(z)|X_j)}_{=0} dz|\inf_x \hat{m}^*(x) > \epsilon\right] = 0$$

Just as for $T_{n,6}$ in the proof of Lemma 3.5.1 we can show that

$$\Rightarrow Var(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon) = 4 \cdot \frac{n(n-1)}{2} E\left((\tilde{b}_{n12}^{(1)})^2 g_2(X_1, X_2)|\inf_x \hat{m}^*(x) > \epsilon\right)$$

with $g_2(X_1, X_2)$ defined in (3.5.11). And consequently

$$Var(R_{n,10}|\inf_x \hat{m}^*(x) > \epsilon)$$

$$= \frac{4}{n^2} \cdot \frac{n(n-1)}{2} E\Bigg(\sum_{r,s=1}^{n} \tilde{B}_{n1}(X_r)B_{n2}(X_r)\tilde{B}_{n1}(X_s)B_{n2}(X_s)$$
$$\times \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \frac{m^*(X_s) - \hat{m}^*(X_s)}{m^*(X_s)} g_2(X_1, X_2)|\inf_x \hat{m}^*(x) > \epsilon\Bigg)$$

as for $Var(T_{n,6})$ in (3.5.2).

$$= \frac{2(n-1)}{n(nh)^4} E\Big(\sum_{r,s=1}^{n} \frac{K\left(\frac{X_r-X_1}{h}\right) K\left(\frac{X_r-X_2}{h}\right)}{m^*(X_r)\hat{m}^*(X_r)} \cdot \frac{K\left(\frac{X_s-X_1}{h}\right) K\left(\frac{X_s-X_2}{h}\right)}{m^*(X_s)\hat{m}^*(X_s)}$$

$$\times \frac{m^*(X_r)-\hat{m}^*(X_r)}{m^*(X_r)} \frac{m^*(X_s)-\hat{m}^*(X_s)}{m^*(X_s)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \Big)$$

$$= \frac{2(n-1)}{n(nh)^4} \Bigg[(n-2)(n-3) E\Bigg(\frac{K\left(\frac{X_3-X_1}{h}\right) K\left(\frac{X_3-X_2}{h}\right)}{m^*(X_3)\hat{m}^*(X_3)} \cdot \frac{K\left(\frac{X_4-X_1}{h}\right) K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)\hat{m}^*(X_4)}$$

$$\times \frac{m^*(X_3)-\hat{m}^*(X_3)}{m^*(X_3)} \frac{m^*(X_4)-\hat{m}^*(X_4)}{m^*(X_4)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \Bigg) +$$

$$+(n-2) E\Bigg(\frac{K\left(\frac{X_3-X_1}{h}\right)^2 K\left(\frac{X_3-X_2}{h}\right)^2}{m^*(X_3)^2 \hat{m}^*(X_3)^2}$$

$$\times \frac{m^*(X_3)-\hat{m}^*(X_3)}{m^*(X_3)} \frac{m^*(X_3)-\hat{m}^*(X_3)}{m^*(X_3)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \Bigg) +$$

$$+4(n-2) E\Bigg(\frac{K(0)\, K\left(\frac{X_1-X_2}{h}\right)}{m^*(X_1)\hat{m}^*(X_1)} \cdot \frac{K\left(\frac{X_4-X_1}{h}\right) K\left(\frac{X_4-X_2}{h}\right)}{m^*(X_4)\hat{m}^*(X_4)}$$

$$\times \frac{m^*(X_1)-\hat{m}^*(X_1)}{m^*(X_1)} \frac{m^*(X_4)-\hat{m}^*(X_4)}{m^*(X_4)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \Bigg) +$$

$$+2E\Bigg(\frac{K(0)^2\, K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)\hat{m}^*(X_1) m^*(X_2)\hat{m}^*(X_2)}$$

$$\times \frac{m^*(X_1)-\hat{m}^*(X_1)}{m^*(X_1)} \frac{m^*(X_2)-\hat{m}^*(X_2)}{m^*(X_2)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \Bigg) +$$

$$+E\left(\frac{K(0)^2\, K\left(\frac{X_1-X_2}{h}\right)^2}{m^*(X_1)^2 \hat{m}^*(X_1)^2} \frac{m^*(X_1)-\hat{m}^*(X_1)}{m^*(X_1)} \frac{m^*(X_1)-\hat{m}^*(X_1)}{m^*(X_1)} g_2(X_1,X_2) \Big| \inf_x \hat{m}^*(x) > \epsilon \right) \Bigg]$$

The computations for the above terms are similar to the ones for (3.5.3), (3.5.4), (3.5.5), (3.5.6) and (3.5.7), respectively. Thus we get that

$$Var(R_{n,10} | \inf_x \hat{m}^*(x) > \epsilon) = O(n^{-2}h^{-1}).$$

Appendix A.3.4

We will prove that

$$E(R_{n,9}^{(1)}) = \frac{1}{nh}\left(\int_a^b \int_I g(z,z,x)dxdz\right)\left(\int K^2(t)dt\right) + O(n^{-1}h)$$

and

$$\mathrm{Var}(R_{n,9}^{(1)}) = O(n^{-4}h^{-3}).$$

$$E(R_{n,9}^{(1)}) = E\left(\int_a^b \frac{1}{n^3h^2}\sum_{r=1}^n \frac{1}{m^*(X_r)^2}\sum_{i=1}^n K\left(\frac{X_r - X_i}{h}\right)^2 \epsilon_i^2(z)dz\right)$$
$$= \frac{1}{n^3h^2}\sum_{r=1}^n\sum_{i=1}^n E\left[\frac{1}{m^*(X_r)^2}\underbrace{E\left(K\left(\frac{X_r - X_i}{h}\right)^2\int_a^b \epsilon_i^2(z)dz|X_r\right)}_{E_1}\right]$$

with

$$E_1 = h\int_a^b \left(\int K^2(t)dt\right) m^*(X_r)\beta(X_r,z)dz + O(h^3) \tag{3.5.24}$$

Thus,

$$E(R_{n,9}^{(1)}) = \frac{1}{n^3h^2}\sum_{r=1}^n\sum_{i=1}^n E\left[\frac{1}{m^*(X_r)^2}\left(h\int_a^b m^*(X_r)g(X_r,z)dz\int K^2(t)dt + O(h^3)\right)\right]$$
$$= \frac{1}{n^3h^2}\sum_{r=1}^n\sum_{i=1}^n \int_I \frac{1}{m^*(y)^2}hm^*(y)\int_a^b m^*(y)g(y,z)dzdy\int K^2(t)dt$$
$$+\frac{1}{n^3h^2}\sum_{r=1}^n\sum_{i=1}^n \int_I \frac{1}{m^*(y)^2}m^*(y)dyO(h^3)$$

$$E(R_{n,9}^{(1)}) = \frac{1}{nh}\left(\int_a^b\int_I g(y,z)dydz\right)\left(\int K^2(t)dt\right) + \int_I \frac{1}{m^*(y)}dyO(n^{-1}h). \tag{3.5.25}$$

Lets now show (3.5.24)

$$E_1 = E\left(K\left(\frac{X_r - X_i}{h}\right)^2\int_a^b \epsilon_i^2(z)dz|X_r\right)$$
$$= \int_a^b E\left(K\left(\frac{X_r - X_i}{h}\right)^2 E\left(\epsilon_i^2(z)|X_i\right)\big|X_r\right)dz =$$
$$= \int_a^b E\left(K\left(\frac{X_r - X_i}{h}\right)^2 \phi'(S(z|X_i))^2 S(z|X_i)^2 E\left(\xi^2(Z_i,\delta_i,T_i,X_i,z)\big|X_i\right)\big|X_r\right)dz =$$
$$= \int_a^b\int_I K\left(\frac{X_r - x}{h}\right)^2 \phi'(S(z|x))^2 S(z|x)^2 \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)}m^*(x)dxdz,$$

by using (3.5.19).

Denote $\beta(x,z) = \phi^{'}(S(z|x))^2 S(z|x)^2 \int_0^z \frac{dH_1^*(u|x)}{C^2(u|x)}$. By a change of variables ($x = X_r + ht$) and Taylor expansion of m^*g around X_r we get

$$E_1 = \int_a^b \int_I K\left(\frac{X_r - x}{h}\right)^2 m^*(x)\beta(x,z)dxdz =$$

$$= h\int_a^b \int K(t)^2 m^*(X_r+ht)\beta(X_r+ht,z)dtdz =$$
$$= h\int_a^b \int K(t)^2[m^*(X_r)\beta(y,z) + (m^{*'}(X_r)\beta(X_r,z) + m^*(X_r)\beta^{'}(X_r,z))th +$$
$$+\frac{1}{2}(m^{*"}(\xi_{X_r})\beta(\xi,z) + m^*(\xi)\beta^{"}(\xi_{X_r},z) + 2m^{*'}(\xi_{X_r})\beta^{'}(\xi_{X_r},z))t^2h^2]dtdz$$

Thus,

$$E_1 = h\int_a^b \left(\int K^2(t)dt\right) m^*(X_r)\beta(X_r,z)dz + \underbrace{h\int_a^b \int K^2(t)t^2 O(h^2)dtdz}_{O(h^3)}$$

with ξ_{X_r} between y and $y+ht$.

Now, for the variance of $R_{n,9}$ we have

$$Var(R_{n,9}^{(1)}) = Var\left(\frac{1}{n^3h^2}\sum_{r=1}^n\sum_{i=1}^n \frac{1}{m^*(X_r)^2}K\left(\frac{X_r - X_i}{h}\right)^2 \int_a^b \epsilon_i^2(z)dz\right)$$
$$= \frac{1}{n^6h^4}Cov\left(\sum_{r=1}^n\sum_{i=1}^n \frac{1}{m^*(X_r)^2}K\left(\frac{X_r - X_i}{h}\right)^2 \int_a^b \epsilon_i^2(z)dz,\right.$$
$$\left.\sum_{l=1}^n\sum_{j=1}^n \frac{1}{m^*(X_l)^2}K\left(\frac{X_l - X_j}{h}\right)^2 \int_a^b \epsilon_j^2(z)dz\right)$$
$$= \frac{1}{n^6h^4}\sum_{r=1}^n\sum_{l=1}^n\sum_{i=1}^n\sum_{j=1}^n Cov\left(\frac{1}{m^*(X_r)^2}K\left(\frac{X_r - X_i}{h}\right)^2 \int_a^b \epsilon_i^2(z)dz,\right.$$
$$\left.\frac{1}{m^*(X_l)^2}K\left(\frac{X_l - X_j}{h}\right)^2 \int_a^b \epsilon_j^2(z)dz\right)$$
$$= \frac{1}{n^6h^4}\sum_{r=1}^n\sum_{l=1}^n\sum_{i=1}^n\sum_{j=1}^n E\left(\frac{1}{m^*(X_r)^2}\frac{1}{m^*(X_l)^2}K\left(\frac{X_r - X_i}{h}\right)^2 K\left(\frac{X_l - X_j}{h}\right)^2\right.$$
$$\left.\times \int_a^b \epsilon_i^2(z)dz \int_a^b \epsilon_j^2(z)dz\right)$$

$$-\frac{1}{n^6h^4}\sum_{r=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\frac{1}{m^*(X_r)^2}K\left(\frac{X_r-X_i}{h}\right)^2\int_a^b\epsilon_i^2(z)dz\right)$$
$$\times E\left(\frac{1}{m^*(X_l)^2}K\left(\frac{X_l-X_j}{h}\right)^2\int_a^b\epsilon_j^2(z)dz\right)$$
$$= I - II.$$

First we study II :
$II = \frac{1}{n^6h^4}\cdot n\cdot n\cdot n\cdot n\cdot A_1\cdot A_2 = \frac{1}{n^2h^4}R^{(1)} + R^{(2)}$
with

$$R^{(1)} = E\left(\frac{1}{m^*(X_r)^2}K\left(\frac{X_r-X_i}{h}\right)^2\int_a^b\epsilon_i^2(z)dz\right)$$

and

$$R^{(2)} = E\left(\frac{1}{m^*(X_l)^2}K\left(\frac{X_l-X_j}{h}\right)^2\int_a^b\epsilon_j^2(z)dz\right)$$

$$R^{(1)} = E\left[\frac{1}{m^*(X_r)^2}E\left[K\left(\frac{X_r-X_i}{h}\right)^2\int_a^b\epsilon_i^2(z)dz|X_r\right]\right]$$
$$= \int_I\frac{1}{m^*(y)}h\int_a^b\int K^2(t)[g(y,z)+O(h^2)]dtdz\cdot m^*(y)dy$$
$$= h\left(\int K^2(t)dt\right)\left(\int_I\int_a^b g(y,z)dzdy\right)+O(h^3)$$

by using (3.5.24). Thus

$$II = \frac{1}{n^2h^4}[h^2\left(\int K^2(t)dt\right)^2\left(\int_I\int_a^b g(y,z)dzdy\right)^2$$
$$+2h\left(\int K^2(t)dt\right)\left(\int_I\int_a^b g(y,z)dzdy\right)\cdot O(h^3)+O(h^6)]$$
$$= \frac{1}{n^2h^2}\left(\int K^2(t)dt\right)^2\left(\int_I\int_a^b g(y,z)dzdy\right)^2$$
$$+\frac{2}{n^2h^3}\underbrace{\left(\int K^2(t)dt\right)}_{\text{bounded}}\underbrace{\left(\int_I\int_a^b g(y,z)dzdy\right)}_{\text{bounded}}\cdot O(h^3)+O(h^6)$$
$$= \frac{1}{n^2h^2}\left(\int K^2(t)dt\right)^2\left(\int_I\int_a^b g(y,z)dzdy\right)^2+O(n^{-2})+O(h^6).$$

Now, for I we have that.

$$I = \frac{1}{n^6h^4}\sum_{r=1}^{n}\sum_{i=1}^{n}E\left(\frac{1}{m^*(X_r)^4}K\left(\frac{X_r-X_i}{h}\right)^4\left(\int_a^b\epsilon_i^2(z)dz\right)^2\right)$$

$$+\frac{1}{n^6h^4}\sum_{r=1}^{n}\sum_{l\neq r}\sum_{i=1}^{n}\sum_{j\neq i}E\Big(\frac{1}{m^*(X_r)^2}\frac{1}{m^*(X_l)^2}K\left(\frac{X_r-X_i}{h}\right)^2K\left(\frac{X_l-X_j}{h}\right)^2$$
$$\times\int_a^b\epsilon_i^2(z)dz\int_a^b\epsilon_j^2(z)dz\Big)$$
$$=\frac{1}{n^6h^4}\cdot n^2E\left[\frac{1}{m^*(X_r)^4}E\left[K\left(\frac{X_r-X_i}{h}\right)^4\left(\int_a^b\epsilon_i^2(z)dz\right)^2|X_r\right]\right]$$
$$+\frac{1}{n^6h^4}\cdot n^2(n-1)^2E\left[\frac{1}{m^*(X_r)^2}E\left[K\left(\frac{X_r-X_i}{h}\right)^2\int_a^b\epsilon_i^2(z)dz|X_r\right]\right]$$
$$\times E\left[\frac{1}{m^*(X_l)^2}E\left[K\left(\frac{X_l-X_j}{h}\right)^2\int_a^b\epsilon_j^2(z)dz|X_l\right]\right]$$
$$=\frac{1}{n^4h^4}E\left[\frac{1}{m^*(X_r)^4}E\left[K\left(\frac{X_r-X_i}{h}\right)^4\left(\int_a^b\epsilon_i^2(z)dz\right)^2|X_r\right]\right]+\frac{1}{n^2h^4}R^{(1)}R^{(2)}$$
$$=II+\frac{1}{n^4h^4}E\left[\frac{1}{m^*(X_r)^4}E_2\right]$$

with

$$E_2=E\left[K\left(\frac{X_r-X_i}{h}\right)^4\left(\int_a^b\epsilon_i^2(z)dz\right)^2|X_r\right]$$
$$=E\left[K\left(\tfrac{x-X_i}{h}\right)^4\left(\textstyle\int_a^b\epsilon_i^2(z)dz\right)^2\right]$$
$$=E\left[K\left(\tfrac{x-X_i}{h}\right)^4E\left[\left(\textstyle\int_a^b\epsilon_i^2(z)dz\right)^2|X_i\right]\right]$$
$$=\int K\left(\frac{x-u}{h}\right)^4\eta(u)m^*(u)du$$

where $\eta(u)=E\left[\left(\int_a^b\epsilon_i^2(z)dz\right)^2|X_i=u\right]$ and $X_r=x$.

A change of variable ($u=x-vh$) and a Taylor expansion of $\eta\cdot m^*$ around x will give

$$E_2=h\int K(v)^4\eta(x-vh)m^*(x-vh)dv$$
$$=h\int K(v)^4\left[\eta(x)m^*(x)-(\eta'(x)m^*(x)+\eta(x)m^{*'}(x))vh+o(h)\right]dv$$
$$=h\cdot\eta(x)m^*(x)\int K(v)^4dv$$
$$-h^2(\eta'(x)m^*(x)+\eta(x)m^{*'}(x))\underbrace{\int vK(v)^4dv}_{0}+h\cdot o(h)\cdot\int K(v)^4dv$$

Finally we get

$$E_2 = h \cdot \eta(x) m^*(x) \int K(v)^4 dv + o(h^2) \int K(v)^4 dv$$

Now

$$I = II + \frac{1}{n^4h^4} E\left[\frac{1}{m^*(X_r)^4}\left(h\eta(X_r)m^*(X_r)\int K(v)^4dv + o(h^2)\int K(v)^4 dv\right)\right]$$
$$= II + \frac{1}{n^4h^4}\int_I \frac{1}{m^*(x)^4} h \cdot \eta(x) m^*(x) \int K(v)^4 dv m^*(x) dx$$
$$+\frac{1}{n^4h^4}\int_I \frac{1}{m^*(x)^4} m^*(x) dx \cdot o(h^2)\int K(v)^4 dv$$
$$= II + \frac{1}{n^4h^3}\int_I \frac{\eta(x)}{m^*(x)^2} dx \int K(v)^4 dv + o(n^{-4}h^{-2})\int_I \frac{1}{m^*(x)^3} dx \int K(v)^4 dv$$

Thus,

$$Var(R^{(1)}_{n,9}) = II + \frac{1}{n^4h^3}\underbrace{\int_I \frac{\eta(x)}{m^*(x)^2}dx \int K(v)^4 dv}_{bounded} + o(n^{-4}h^{-2})\underbrace{\int_I \frac{1}{m^*(x)^3}dx}_{bounded}\int K(v)^4 dv - II$$

$$Var(R^{(1)}_{n,9}) = O(n^{-4}h^{-3}). \tag{3.5.26}$$

From (3.5.25), (3.5.26) and Chebysev we get that

$$R^{(1)}_{n,9} = E(R^{(1)}_{n,9}) + O_p\left(\sqrt{Var(R^{(1)}_{n,9})}\right) =$$
$$= \frac{1}{nh}\left(\int_a^b \int_I g(y,z) dy dz\right)\left(\int K^2(t) dt\right) + O(n^{-1}h + n^{-2}h^{-1.5})$$

Appendix A.3.5

If $i = j \neq r$, $\tilde{\Delta}^{(1)}_{rij}(z) = \tilde{B}^2_{ni}(X_r)\eta_{ri}(z)^2$

$$= \tilde{B}^2_{ni}(X_r)\Big(\phi'(S(z|X_r))S(z|X_r)\xi(Z_i,T_i,\delta_i,X_r,z)$$
$$-\phi'(S(z|X_i))S(z|X_i)\xi(Z_i,T_i,\delta_i,X_i,z)\Big)^2$$
$$= (1) + (2) + (3)$$

where

$$(1) = \tilde{B}^2_{ni}(X_r)\phi'(S(z|X_r))^2 S(z|X_r)^2 \xi(Z_i,T_i,\delta_i,X_r,z)^2$$
$$(2) = \tilde{B}^2_{ni}(X_r)\phi'(S(z|X_i))^2 S(z|X_i)^2 \xi(Z_i,T_i,\delta_i,X_i,z)^2$$
$$(3) = -2\tilde{B}^2_{ni}(X_r)\phi'(S(z|X_r))\phi'(S(z|X_i))S(z|X_r)S(z|X_i)\xi(Z_i,T_i,\delta_i,X_r,z)\xi(Z_i,T_i,\delta_i,X_i,z)$$

In the following we will prove that $E[(1)] = E[(2)] = -2E[(3)]$.

First we will take care of (1) :

$$E[(1)] = E\Big[\tilde{B}^2_{ni}(X_r)\phi'(S(z|X_r))^2S(z|X_r)^2\xi(Z_i,T_i,\delta_i,X_r,z)^2\Big]$$
$$= E\Big[\phi'(S(z|X_r))^2S(z|X_r)^2E\Big(\tilde{B}^2_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z)^2|X_i,X_r\Big)\Big]$$
$$= \frac{1}{n^2h^2}E\Big[\frac{1}{m^*(X_r)^2}\phi'(S(z|X_r))^2S(z|X_r)^2K\left(\frac{X_r-X_i}{h}\right)^2\underbrace{E\Big[\xi(Z_i,T_i,\delta_i,X_r,z)^2|X_i,X_r\Big]}_{\mu(z,X_r)\text{ defined in (2.7.9)}}\Big]$$
$$= \frac{1}{n^2h^2}E\Big[\frac{1}{m^*(X_r)^2}\phi'(S(z|X_r))^2S(z|X_r)^2\mu(z,X_r)E\Big[K\left(\frac{X_r-X_i}{h}\right)^2|X_r\Big]\Big]$$
$$= \frac{1}{n^2h^2}E\left[\frac{1}{m^*(X_r)}\phi'(S(z|X_r))^2S(z|X_r)^2\mu(z,X_r)\left(\int K(u)^2du\right)h\right] + O(n^{-2}h)$$
$$= \frac{1}{n^2h}\int_I \phi'(S(z|x))^2S(z|x)^2\mu(z,x)dx\int K(u)^2du + O(n^{-2}h)$$

Where, by a change of variable ($x = y - uh$) and by Taylor expansion of m^* around y we had

$$E\Big[K\left(\frac{X_r-X_i}{h}\right)^2|X_r\Big] = \int_I K\left(\frac{y-x}{h}\right)^2 m^*(x)dx = \int K(u)^2m^*(y-uh)hdu$$
$$= \int K(u)^2[m^*(y) - uhm^{*'}(y) + \frac{1}{2}(uh)^2m^{*''}(\xi)]hdu$$
$$= \left(\int K(u)^2du\right)m^*(y)h + O(h^3),\text{ since }\textstyle\int K^2(u)udu = 0\text{ if }K\text{ symmetric.}$$

For (2), $E[(2)] = E\Big[\tilde{B}^2_{ni}(X_r)\phi'(S(z|X_i))^2S(z|X_i)^2\xi(Z_i,T_i,\delta_i,X_i,z)^2\Big]$

$$= E\Big[\frac{1}{n^2h^2}\frac{1}{m^*(X_r)^2}\phi'(S(z|X_i))^2S(z|X_i)^2E\Big(K\left(\frac{X_r-X_i}{h}\right)^2\xi(Z_i,T_i,\delta_i,X_i,z)^2|X_i,X_r\Big)\Big]$$
$$= E\Big[\frac{1}{n^2h^2}\frac{1}{m^*(X_r)^2}\phi'(S(z|X_i))^2S(z|X_i)^2K\left(\frac{X_r-X_i}{h}\right)^2\underbrace{E\Big(\xi(Z_i,T_i,\delta_i,X_i,z)^2|X_i\Big)}_{\mu(z,X_i)}\Big]$$
$$= \frac{1}{n^2h^2}E\Big[\phi'(S(z|X_i))^2S(z|X_i)^2\mu(z,X_i)E\Big(\frac{1}{m^*(X_r)^2}K\left(\frac{X_r-X_i}{h}\right)^2|X_i\Big)\Big]$$
$$= \frac{1}{n^2h}E\Big[\phi'(S(z|X_i))^2S(z|X_i)^2\mu(z,X_i)\frac{1}{m^*(X_i)}\int K(u)^2du + o(h^2)\Big]$$
$$= \frac{1}{n^2h}\int_I \phi'(S(z|x))^2S(z|x)^2\mu(z,x)dx\int K(u)^2du + O(n^{-2}h)$$

Where by a change of variable ($x = y + hu$) and by Taylor expansion of $\dfrac{1}{m^*(y+hu)}$ around y we had

$$E\left(\frac{1}{m^*(X_r)^2}K\left(\frac{X_r - X_i}{h}\right)^2 |X_i\right) = \int_I \frac{1}{m^*(x)^2}K\left(\frac{x-y}{h}\right)^2 m^*(x)dx$$
$$= \int_I \frac{1}{m^*(x)}K\left(\frac{x-y}{h}\right)^2 dx = \int \frac{1}{m^*(y+hu)}K(u)^2 hdu$$
$$= \int \left[\frac{1}{m^*(y)} + o(h)\right]K(u)^2 hdu = \frac{1}{m^*(y)}h\int K(u)^2 du + o(h^3)$$

$$E[(3)] = -2E\Bigg[\tilde{B}^2_{ni}(X_r)\phi^{'}(S(z|X_r))\phi^{'}(S(z|X_i))S(z|X_r)S(z|X_i)\cdot$$
$$\cdot E\Big(\xi(Z_i,T_i,\delta_i,X_r,z)\xi(Z_i,T_i,\delta_i,X_i,z)|X_i,X_r\Big)\Bigg]$$
$$= -2E\Bigg[\tilde{B}^2_{ni}(X_r)\phi^{'}(S(z|X_r))\phi^{'}(S(z|X_i))S(z|X_r)S(z|X_i)\Big(\underbrace{\xi(Z_i,T_i,\delta_i,X_i,z)}_{\mu(z,X_i)}+O(h)\Big)\Bigg]$$
$$= -2E\Bigg[\phi^{'}(S(z|X_r))S(z|X_r)\frac{1}{n^2h^2}\frac{1}{m^*(X_r)^2}\cdot$$
$$\cdot E\left(\phi^{'}(S(z|X_i))S(z|X_i)\mu(z,X_i)K\left(\frac{X_r-X_i}{h}\right)^2|X_r\right)\Bigg]$$
$$= -2E\Bigg[\phi^{'}(S(z|X_r))S(z|X_r)\frac{1}{n^2h^2}\frac{1}{m^*(X_r)^2}\cdot$$
$$\cdot\left(h\phi^{'}(S(z|X_r))S(z|X_r)\mu(z,X_r)m^*(X_r)\int K(u)^2du + o(h^2)\right)\Bigg]$$
$$= -2\frac{1}{n^2h}\int_I \phi^{'}(S(z|x))^2 S(z|x)^2\mu(z,x)dx\int K(u)^2du + O(n^{-2}h)$$

With

$$E\left(\phi^{'}(S(z|X_i))S(z|X_i)\mu(z,X_i)K\left(\frac{X_r-X_i}{h}\right)^2|X_r\right)$$
$$= \int_I \phi^{'}(S(z|x))S(z|x)\mu(z,x)K\left(\frac{y-x}{h}\right)^2 m^*(x)dx$$
$$= h\int K(u)^2\phi^{'}(S(z|y-ux))S(z|y-uh)\mu(z,y-uh)m^*(y-uh)du$$
$$= h\int K(u)^2\left[\phi^{'}(S(z|y))S(z|y)\mu(z,y)m^*(y) + o(h)\right]du$$
$$= h\phi^{'}(S(z|y))S(z|y)\mu(z,y)m^*(y)\int K(u)^2du + o(h^3),$$

by a change of variable ($x = y - uh$) and Taylor expansion of $\phi^{'}\cdot S\cdot g\cdot m^*$ around y.

Thus, we get that for $i = j \neq r$

$$E\left(\tilde{\Delta}_{rij}^{(1)}(z)\right) = O(n^{-2}h),$$

since dominant terms cancel out.

Now, if $i \neq j \neq r$

$$\begin{aligned}
&\tilde{\Delta}_{rij}^{(1)}(z) = \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)\cdot \\
&\quad\cdot\Bigg(\phi^{'}(S(z|X_r))S(z|X_r)\xi(Z_i,T_i,\delta_i,X_r,z) - \phi^{'}(S(z|X_i))S(z|X_i)\xi(Z_i,T_i,\delta_i,X_i,z)\Bigg)\cdot \\
&\quad\cdot\Bigg(\phi^{'}(S(z|X_r))S(z|X_r)\xi(Z_j,T_j,\delta_j,X_r,z) - \phi^{'}(S(z|X_j))S(z|X_j)\xi(Z_j,T_j,\delta_j,X_j,z)\Bigg) \\
&= A_1 + A_2 + A_3 + A_4
\end{aligned}$$

$$\begin{aligned}
E(A_1) = E\Bigg[&\phi^{'}(S(z|X_r))^2 S(z|X_r)^2 \underbrace{E\Bigg(\tilde{B}_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z)|X_r\Bigg)}_{=O(n^{-1}h^2)} \\
&\cdot \underbrace{E\Bigg(\tilde{B}_{nj}(X_r)\xi(Z_j,T_j,\delta_j,X_r,z)|X_r\Bigg)}_{=O(n^{-1}h^2)}\Bigg] = O(n^{-2}h^4), \text{ from (2.7.12).}
\end{aligned}$$

$$\begin{aligned}
E(A_2) = E\Bigg[&\tilde{B}_{nj}(X_r)\phi^{'}(S(z|X_r))S(z|X_r)\phi^{'}(S(z|X_j))S(z|X_j) \\
&\cdot \underbrace{E\Bigg(\tilde{B}_{ni}(X_r)\xi(Z_i,T_i,\delta_i,X_r,z)|X_r\Bigg)}_{=O(n^{-1}h^2)} \underbrace{E\Bigg(\xi(Z_j,T_j,\delta_j,X_j,z)|X_j,X_r\Bigg)}_{=0}\Bigg] = 0
\end{aligned}$$

Same reasoning gives $E(A_3)$.

$$\begin{aligned}
E(A_4) = E\Bigg[&\tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)\phi^{'}(S(z|X_i))S(z|X_i)\phi^{'}(S(z|X_j))S(z|X_j) \\
&\cdot \underbrace{E\Bigg(\xi(Z_i,T_i,\delta_i,X_i,z)|X_i,X_r\Bigg)}_{=0} \underbrace{E\Bigg(\xi(Z_j,T_j,\delta_j,X_j,z)|X_j,X_r\Bigg)}_{=0}\Bigg] = 0
\end{aligned}$$

$$\Rightarrow E[\tilde{\Delta}_{rij}^{(1)}(z)] = O(n^{-2}h^4)$$

Chapter 4

Real Data Analysis

In the previous chapters we conducted simulations studies to quantify the performance of our proposed methods and we have seen that they have given very good results. Now we will apply these methods to two real data sets, one concerning patients suffering from larynx cancer and the other one on patients suffering from gastric adenocarcinoma. We give a short description of the data sets, followed by a detailed analysis.

4.1 Larynx Cancer Data

Larynx cancer is fairly often encountered. Kardaun (1983) reports data on 90 males diagnosed with cancer of the larynx during the period 1970-1978 at a Dutch hospital. Times recorded are the intervals (in years) between first treatment and either death, or the end of the study (January 1, 1983). Also recorded are the patient's age at the time of diagnosis, the year of diagnosis, and the stage of the patient's cancer. The four stages of disease in the study were based on the TNM (primary tumor (T), nodal involvement (N) and distant metastasis (M) grading) classification used by the American Joint Committee for Cancer Staging (1972). The four groups are Stage 1 : T1N0M0 with 33 patients ; Stage 2 : T2N0M0 with 17 patients ; Stage 3 : T3N0M0 and TxN1M0, with 27 patients, where x $= 1, 2$, or 3 ; and Stage 4 : all other TNM combinations with 13 patients. The stages are ordered from least serious to most serious. This data set has also been studied by Klein and Moeschberger (1997) and we will refer later to their work.

For each individual i ($i = 1, \ldots, 90$) let Z_i be the time-to-death or on-study, δ_i be the death indicator (0=alive, 1=dead), $X_{j,i}$ be the indicator of being at stage $j+1$ ($j = 1, 2, 3$) and $X_{4,i}$ be the age at diagnosis minus its mean (64.11 years). Age is centered at its mean for easier interpretation of the results. $\tau = 4.4$ is the largest observed value of Z_i, for which at least one patient is still at risk in each of the four disease stages. No truncation is present, thus $T_i = 0$ with probability one and 40 patients were still alive at the end of the study, thus they were considered as censored. For these data the independence between Y and C given $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$ seems a reasonable assumption. The following table, as well as the box-plots in Figure 4.1, summarize the data :

Characteristics	Number	Age			
TNM classification	obs	Min	Max	Mean	Median
Stage 1	33	43	86	64.18	64
Stage 2	17	47	86	64.82	64
Stage 3	27	49	82	63.81	65
Stage 4	13	41	84	67.07	69

Table 4.1 : Summary statistics for larynx cancer data.

In Klein and Moeschberger (1997) an additive hazards regression model has been fitted to the data, in order to explain how the covariables $\boldsymbol{X}$ influence the survival time of the patients. The assumptions for the additive hazards model have also been verified via two 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. They also adjusted a Cox proportional hazards model with constant coefficients, without checking if the assumptions for this model are satisfied.

Here we would like to apply the goodness-of-fit procedure presented in Chapter 3 to see if the additive hazards model is accepted for this data, as well as the Cox model with time-dependent coefficients.

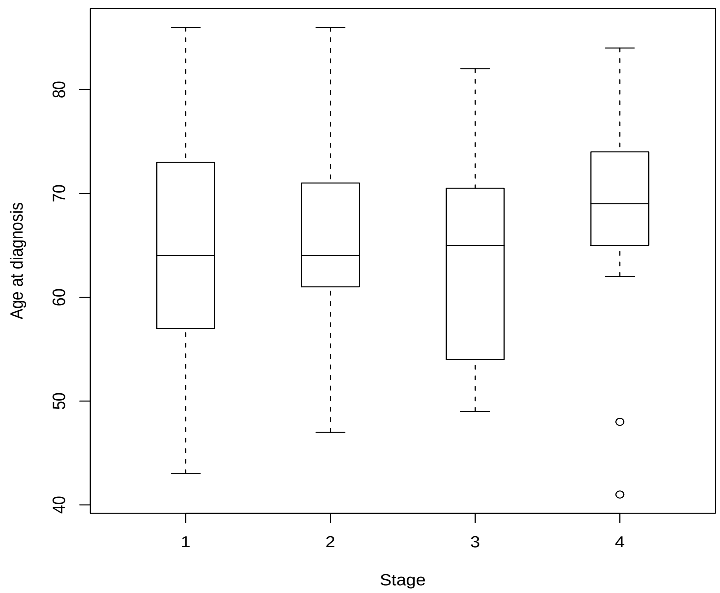

Figure 4.1 : Box-plots of the age at diagnosis in each Stage, for larynx cancer data.

For the accepted model(s) we will estimate the parameters via the least squares procedure in Section 2.2. More precisely, in the sequel, the following model will be considered

$$\phi_k(S(z|\boldsymbol{X})) = \beta_0(z) + \beta_1(z)X_1 + \beta_2(z)X_2 + \beta_3(z)X_3 + \beta_4(z)X_4, \tag{4.1.1}$$

where $\phi_k(u)$, $k = 1, 2$, is either $\phi_1(u) = -\log(u)$ or $\phi_2(u) = \log(-\log(u))$, that will give the additive hazards model or the Cox model with time-dependent coefficients.

For each $\phi_k(u)$, $k = 1, 2$ we will test

$$H_0 : \text{for all } z \text{ there exists a vector } \boldsymbol{\beta}(z) \in \mathbb{R}^5 \text{ such that (4.1.1) holds,}$$

$$\text{versus } H_a : \text{there exists a } z \text{ such that (4.1.1) does not hold for any } \boldsymbol{\beta}(z) \in \mathbb{R}^5.$$

We will first consider the hypothesis testing for $\phi_1(u)$, thus the additive hazards model. In order to approximate the critical values of the test we will apply the bootstrap procedure in Section 3.2. Since X_4 is a continuous variable, a smoothing parameter is needed in order to estimate the survival function $S(z|\boldsymbol{X})$ and the distribution function $G(z|\boldsymbol{X})$ involved in the procedure.

Under H_0, a way to choose the smoothing parameter is to apply the bootstrap procedure described in Section 2.5. Thus, we will choose the optimal bandwidth among a grid $\{h_1, h_2, \ldots, h_m\}$ in $(0, 45)$, the expected range of X_4, as the one that minimizes the $MISE^*_{h_j}$, $j = 1, \ldots, m$, of the estimated coefficients, defined in (2.5.2). We will take the following grid of points $\{20, 25, 30, 35, 40, 45\}$. The pilot bandwidth in the bootstrap procedure was $g_j = h_j$, $j = 1, \ldots, 6$, and we used $B = 10000$ bootstrap replications. The coefficients are estimated for time-points $z \in [0, \tau]$ and the weight function $\tilde{w}(x)$ is chosen in such a way to avoid boundary problems : $\tilde{w}(x) = 1_{\{-18.66 \leq x \leq 19.44\}}$, where -18.66 and 19.44 are the 2.5% and 97.5% percentiles of X_4, respectively. The selected optimal bandwidth was $h = 25$.

Now, we can apply the bootstrap procedure of Section 3.2 with $h = 25$ in order to test H_0 versus H_a. The other needed parameters were taken as follows $g = 1.5h$, $\alpha = 0.05$ and we used $B = 10000$ bootstrap replications. 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).

Now, we can estimate the coefficients of this model, by the estimator in (2.2.1) and compare them to those obtained in Klein and Moeschberger (1997). Note that the procedure used by Klein and Moeschberger (1997), called « classical method » from now on, is also based on least squares, but they didn't need a smoothing parameter in their estimation, since the cumulative hazard function $\Lambda(z|\boldsymbol{X}) = \phi_1(S(z|\boldsymbol{X}))$ was estimated via counting process techniques (see Chapter 10 in their book for more details).

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 :

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

with the variance computed using the formulas presented in Chapter 10 of Klein and Moeschberger (1997), while for the new method they were estimated using the percentile method, via the bootstrap procedure presented in Section 3.2.

As it can be seen in Figure 4.2 the estimator of the cumulative baseline hazard rate (the cumulative hazard rate of a Stage 1 patient aged 64.11 years), $\beta_0(z)$, is almost the same with both methods, as well as its confidence intervals. Similar things happen for the estimators of $\beta_1(z)$, $\beta_2(z)$ and $\beta_3(z)$, in Figure 4.3, 4.4 and 4.5, respectively.

The slope of the estimator $\hat{\beta}_1(z)$ in Figure 4.3 indicates that there is little excess risk due to being a Stage 2 patient, whereas $\hat{\beta}_2(z)$ and $\hat{\beta}_3(z)$ in Figure 4.4 and Figure 4.5, respectively, show that Stage 3 and Stage 4 patients have an elevated risk in the first two years following diagnosis, where the slopes of the two cumulative hazards are non zero. As for the coefficient corresponding to the continuous covariate, $\beta_4(z)$, we notice in Figure 4.6 that the two curves are slightly different from each other but both are very close to zero, which suggests that there is no excess risk due to age at diagnosis.

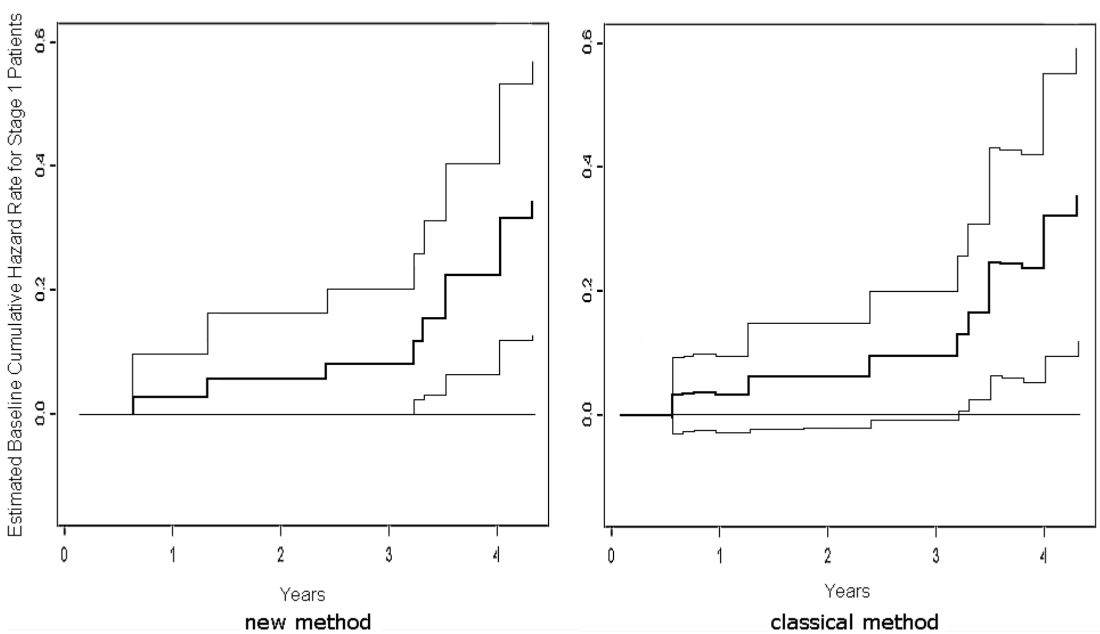

Figure 4.2 : Estimate of the cumulative hazard rate of a Stage 1 patient aged 64.11 years ($\beta_0(z)$) and 95% pointwise confidence intervals.

We also plotted the model-based estimators of the survival functions associated to the four stages of the disease : $\tilde{S}(z|\boldsymbol{X}) = \phi_1\big(\beta_0(z)+\beta_1(z)X_1+\beta_2(z)X_2+\beta_3(z)X_3+\beta_4(z)X_4\big)$. Figure 4.7 illustrates the four survival rates for patients of mean age 64.11. We notice that there is almost no difference between the survival rates of Stage 1 and Stage 2, which explains why there is no excess risk due to being a Stage 2 patient as compared to Stage 1 patient. And that Stage 4 patients have a worse survival rate than Stage 3 patients, who on their turn have a worse survival rate than Stage 1 and Stage 2 patients.

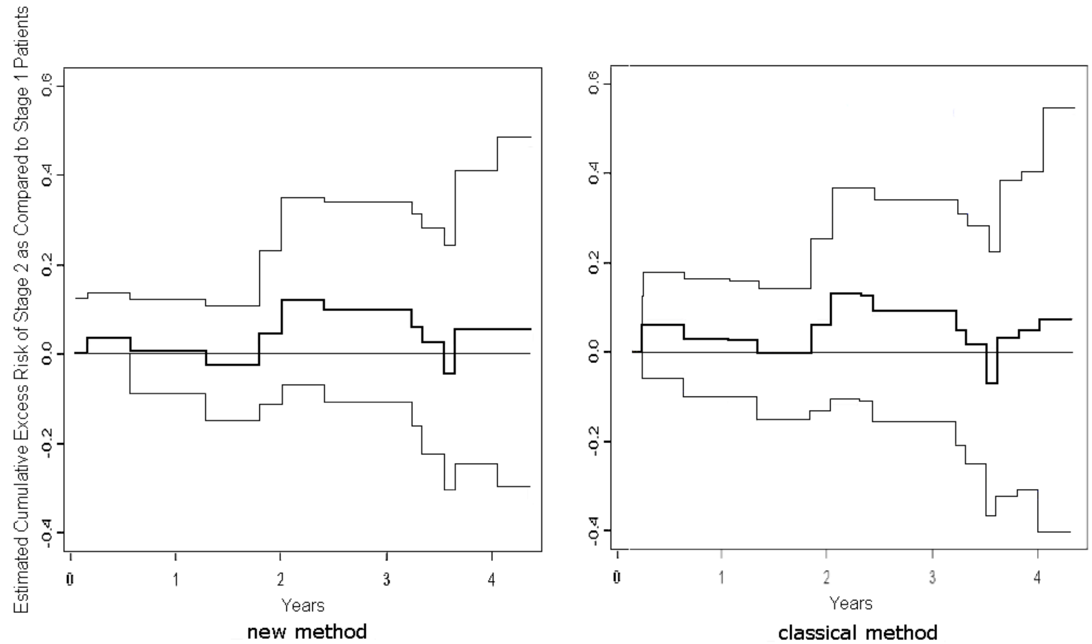

Figure 4.3 : Estimate of the cumulative excess risk of Stage 2 cancer as compared to Stage 1 cancer ($\beta_1(z)$) and 95% pointwise confidence intervals.

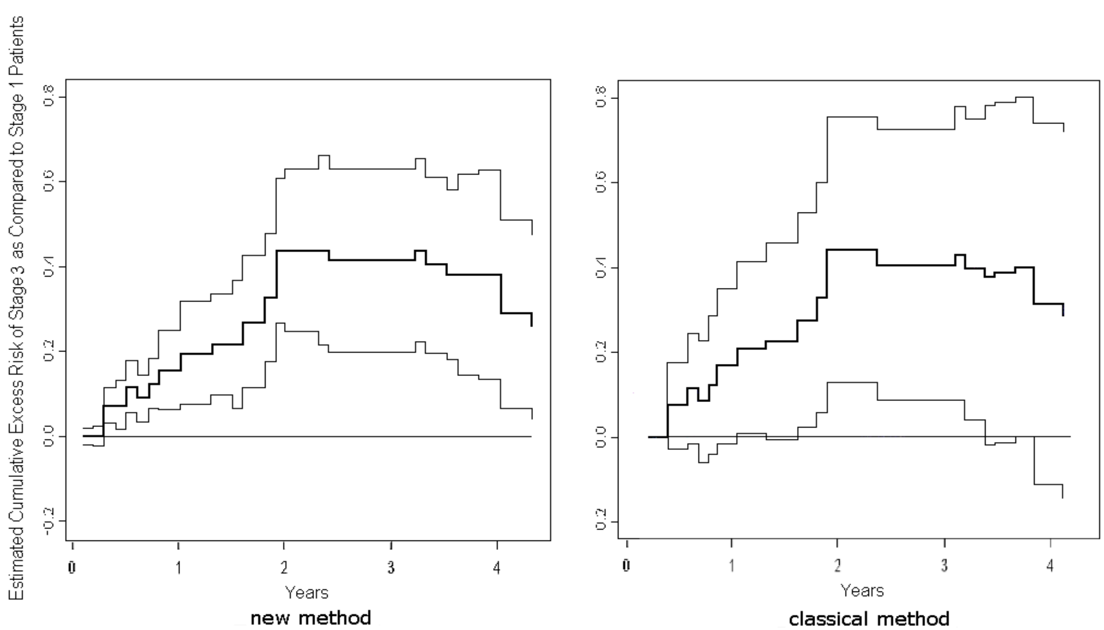

Figure 4.4 : Estimate of the cumulative excess risk of Stage 3 cancer as compared to Stage 1 cancer ($\beta_2(z)$) and 95% pointwise confidence intervals.

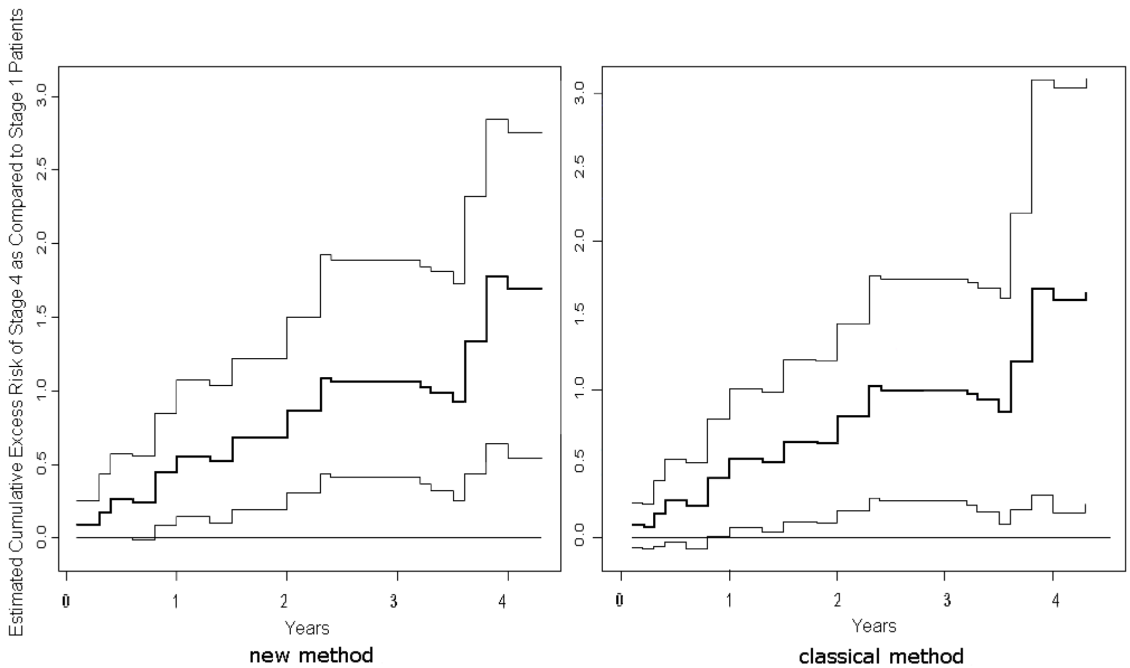

Figure 4.5 : Estimate of the cumulative excess risk of Stage 4 cancer as compared to Stage 1 cancer ($\beta_3(z)$) and 95% pointwise confidence intervals.

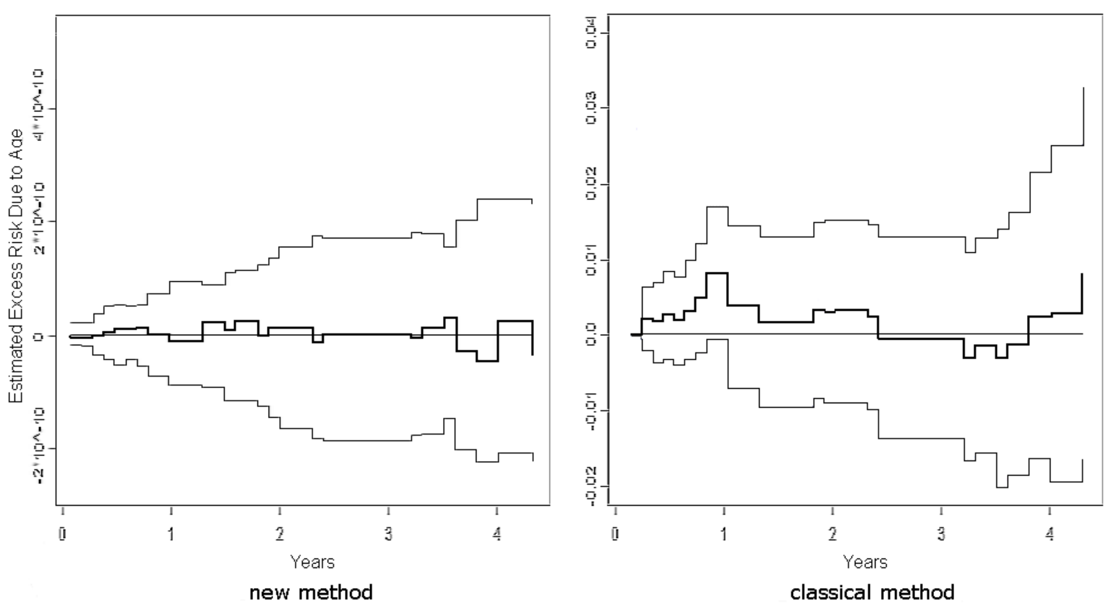

Figure 4.6 : Estimate of the cumulative effect of age ($\beta_4(z)$) and 95% pointwise confidence intervals.

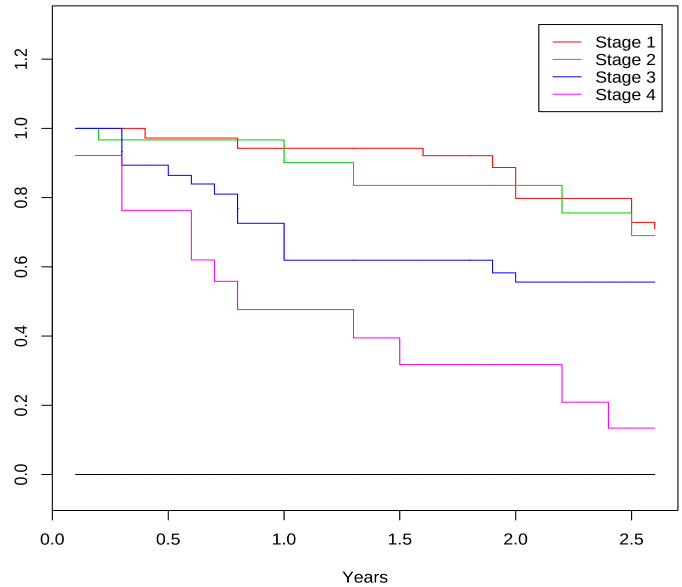

Figure 4.7 : Estimates of the survival function based on the additive hazards model for larynx cancer patients of mean age 64.11.

We also plotted the survival rates for Stage 1 patients aged 50, 64.11 and 80 years (not shown here) and we noticed absolutely no difference between their survival rates, which confirms our conclusion, that age had no effect on the cumulative hazard rate, thus no effect on the survival rate. Same conclusions are observed for the other stages of the disease.

Now, we consider the hypothesis testing for $\phi_2(u)$, thus for the Cox model with time-dependent coefficients. Same procedures as above are being applied. The optimal bandwidth under H_0 is $h = 30$, and the associated p-value is found to be 0.0415, thus we reject the null hypothesis.

4.2 Gastric Cancer Data

A retrospective cohort study has been conducted on patients diagnosed with gastric adenocarcinoma in Xeral Hospital in Lugo and Hospital Universitario de A Coruña (northwest of Spain), between January 1975 and December 1993. The main interest was to analyze the survival of these patients and the factors that can modify their prognosis. This data set has been previously studied by Casariego-Vales et al (2001), Rabuñal et al (2004).

Here we only consider the 1957 patients for which we have the date of first symptoms related to the disease and as the factors that influence the prognoses of the patients, only those that turned out to be significant in previous studies. For each individual i ($i = 1, \dots, 1957$) we observed the time (in days) from the first symptoms to diagnosis T_i, the time (in days) from first symptoms to death or loss to follow up Z_i, the death indicator δ_i (0=alive, 1=dead), the age (in years) at diagnosis (X_i), the extension of the tumor which was established by using the fourth TNM classification published by Hermanek and Soben (1987) and recoded as follows : 1=initial (Stage Ia, Stage Ib, Stage II), 2=advanced (Stage IIIa, Stage IIIb, Stage IV), 3=unknown (the stages are ordered from least serious to most serious), the presence of metastasis (1=no, 2=yes) and the types of surgery (1=curative, 2=palliative, 3=no or minimal surgery). It was considered that some type of surgery had been performed if the intervention was different from a mere inspection or biopsy and was capable of altering the course of the disease in some way. An operation was defined as having curative intent when no appreciable residual disease existed after its performance. In contrast, a palliative resection implied the presence of residual disease. Censoring occurred when patients died from a reason not related to their disease, were lost to follow up or were still alive at last visit. Patients were followed until at least August 1st 1996. On the other hand, those that were not diagnosed before death ($Z_i < T_i$) were considered as being truncated. The date of diagnosis was considered to be the day on which gastric adenocarcinoma was histologically diagnosed and the date of first symptoms, the day when symptoms related to the disease appeared for the first time.

In the following, TNM will denote the extension of tumor (1, 2, 3), RES the type of surgery (1, 2, 3) and META the presence of metastasis (1, 2), with coding as described above. For these patients, only the following combinations between these factors were observed : (TNM=1, RES=1, META=1), (TNM=1, RES=2, META=1), (TNM=1, RES=3, META=1), (TNM=2, RES=1, META=1), (TNM=2, RES=1, META=2), (TNM=2, RES=2, META=1), (TNM=2, RES=2, META=2), (TNM=2, RES=3, META=1), (TNM=2, RES=3, META=2), (TNM=3, RES=1, META=1) and (TNM=3, RES=3, META=1), denoted in the sequel by Group 111, Group 121, Group 131, Group 211, Group 212, Group 221, Group 222, Group 231, Group 232, Group 311 and Group 331, respectively. $b = 2487$ days (6.81 years) is the largest Z_i, for which at least one patient is still at risk in each of the 11 groups and $a = 21$ days is the smallest T_i for which there is at least one observed T in each group.

The following table contains some descriptive statistics of the data. We also provide in Figure 4.8 box plots of the ages at diagnosis in the 11 different observed groups.

Group	Number obs	Cens perc	age			
			min	max	mean	median
Group 111	408	45.09	22	92	65.52	67
Group 121	25	16	46	83	65.44	68
Group 131	45	2.22	34	85.23	67.67	71
Group 211	362	44.47	23	99	64.89	66
Group 212	38	21.05	39	81	65.87	69
Group 221	122	18.85	43	96	67.20	69
Group 222	227	16.29	24	89	66.62	68
Group 231	46	10.86	44	84	69.13	70
Group 232	249	28.91	27	90	65.16	66
Group 311	120	42.5	24	86	63.37	66
Group 331	315	18.41	29	98	72.16	74

Table 4.2 : Summary statistics for gastric cancer data.

The main purpose is to find a regression model that explains how the survival time of these patients is influenced by the factors described above. We will consider two types of models : one that has the factors acting in additive way on the cumulative hazard (additive hazards model) and the other one in a multiplicative way (Cox model with time-dependent coefficients). For the rest of this section, the following model is to be considered :

$$\phi_k(S(z|\boldsymbol{X})) = \beta_0(z) + \beta_1(z)X_1 + \beta_2(z)X_2 + \beta_3(z)X_3 + \beta_4(z)X_4 + \beta_5(z)X_5 + \beta_6(z)X_6, \tag{4.2.1}$$

where $\phi_k(u)$, $k = 1, 2$ is either $\phi_1(u) = -\log(u)$ or $\phi_2(u) = \log(-\log(u))$, giving the additive hazards model or the proportional hazards model with time-dependent coefficients, respectively.

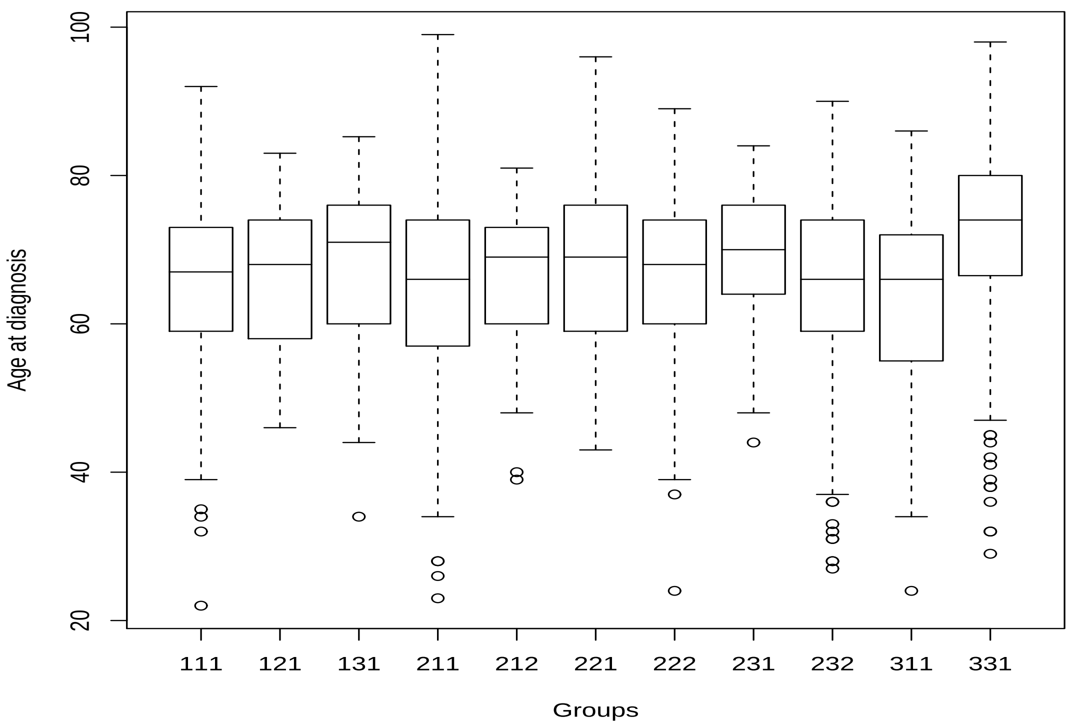

Figure 4.8 : Box-plots of the age at diagnosis in each Group, for gastric cancer data.

X_1 and X_2 are the indicators of having TNM classification 2 and 3, respectively, X_3 and X_4 are the indicators for the type of surgery 2 and 3, respectively, X_5 is the indicator for the presence of metastasis (0=no, 1=yes) and X_6 is the age at diagnosis (X) minus its mean (67.71 years). The age is centered at its mean for easier interpretation of the parameters. The independence between Y and (C, T) given $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ seems a reasonable assumption.

One of the main differences with the previous studies is the target population. Here, we are interested in all the people in the area of Lugo and La Coruña that suffered from gastric adenocarcinoma (diagnosed or not), while previously, they were only interested in those that were diagnosed in the same area. This difference comes from the fact that here the time of interest is the time from first symptoms to death, while before the interest variable was the time from diagnosis to death. Another difference is the fact that here we consider time-dependent coefficients, instead of constant coefficients, that will allow us to detect fluctuations in the influence of the factors on the survival function.

We would like to test if the additive hazards model and/or the Cox model with time-dependent coefficients fit to the data. First, we will test (4.2.1) for $\phi_1(u)$, the additive hazards model :

$$H_0 : \text{for all } z \text{ there exist a vector } \boldsymbol{\beta}(z) \in \mathbb{R}^7 \text{ such that (4.2.1) holds,}$$

$$\text{versus } H_a : \text{there exist a } z \text{ such that (4.2.1) does not hold for any } \boldsymbol{\beta}(z) \in \mathbb{R}^7.$$

As for the larynx cancer data in Section 4.1 we will first need to choose the optimal smoothing parameter h under H_0, since we have one continuous covariate, the age at diagnosis. We will do so by the bootstrap procedure in Section 2.5. The parameters needed in this procedure were taken as follows : $B = 500$ bootstrap replications, $g = h$ and h was to be chosen among $\{25, 30, 35, 40, 45, 50, 55, 60\}$, where $(0, 77)$ is the expected range for X_6. The coefficients were estimated for time-points $z \in [a, b]$ ($a = 21$ days, $b = 2487$ days) and $\tilde{w}(x) = 1_{\{-43.785 \leq x \leq 29.365\}}$, where -43.785 and 29.365 are the 0.025 and 0.975 quantile of X_6, respectively, in order to avoid boundary problems.

The optimal bandwidth, the one that minimizes the $MISE^*_{h_j}$, $j = 1, \ldots, 8$, defined in (2.5.2) was found to be $h = 40$. By using this bandwidth, we now conduct the goodness-of-fit test and we obtain its critical values by the bootstrap procedure in Section 3.2. 500 bootstrap replications were drawn, $\alpha = 0.05$, $g = 1.5h$ and the p-value was found to be 0.984. Thus the additive hazards model fits well to the data.

The estimators of $\beta_k(z)$, $k = 0, \ldots, 6$, in model (4.2.1) as well as their 95% confidence intervals obtained by the percentile method via the bootstrap procedure in Section 2.5, are given in Figure 4.9 and Figure 4.10.

Figure 4.9 (a) shows the estimated cumulative hazard rate of a Group 111 patient aged 67.71 years. Figure 4.9 (b) and (c) show that there is little excess risk due to being a Group 211 and Group 311 patient as compared with Group 111 patients. On the contrary, Figure 4.9 (d) and Figure 4.10 (e) show that Group 121 and Group 131 patients as compared to Group 111 (RES=1) patients have an elevated excess risk during the first 3 years following the first symptoms, where the slopes of the two cumulative hazards are non-zero. This means that whenever curative surgery was practiced, the extension of the tumor did not matter, but for patients having the same extension of tumor, palliative surgery and minimal or no surgery showed excess risk as compared to patients that have undergone curative surgery.

$\hat{\beta}_5(z)$ represents the estimated excess risk of patients having (TNM=1,RES=1, META=2) as compared to Group 111 patients. Patients with this combination have not been observed (no metastasis present when being in an initial stage of tumor), but $\hat{\beta}_5(z)$ is used in order to compare other observed groups with the reference group, Group 111, as for example Group 212. For this we should look at $\hat{\beta}_1(z) + \hat{\beta}_5(z)$, see Figure 4.10 (h). We see that there is excess risk due to being in Group 212, as compared to Group 111 during the first year and a half after the first symptoms.

As for the excess risk due to age there seems to be one during the first two years and a half, where the slope of $\hat{\beta}_6(z)$ in Figure 4.10 (g) is non zero.

The plot of the model-based estimators of the survival functions for the different groups under study in Figure 4.11 indicates that the baseline group Group 111 has the best survival rate, along with those groups that have undergone curative surgery (RES=1). The other groups have more or less the same survival rate, with the groups that have undergone palliative surgery and had no presence of metastasis showing a slightly better survival rate during the first 3 years following the symptoms.

We also notice that Group 311 (patients with unclassified type of tumor that showed no presence of metastasis and had undergone curative surgery) shows a worse survival rate during the first two and a half years following first symptoms as compared to Group 211. The tendency reverses after this period. This may be because those patients for which the type of the tumor could not be classified, have undergone not the best suited treatment for their disease, but that those patients weren't really in the advanced stage of the disease, but more likely in the initial stage or in between, which explains the resulting survival rates. Same patterns as described above are observed for the survival rates computed at other ages (50 and 80).

When comparing the survival rates of patients belonging to the same group, see for example Figure 4.12 for Group 131, but of different ages (50, 67.71 and 80), we notice that as the age at diagnosis increases, the survival rate becomes worse. This is valid for all the observed groups.

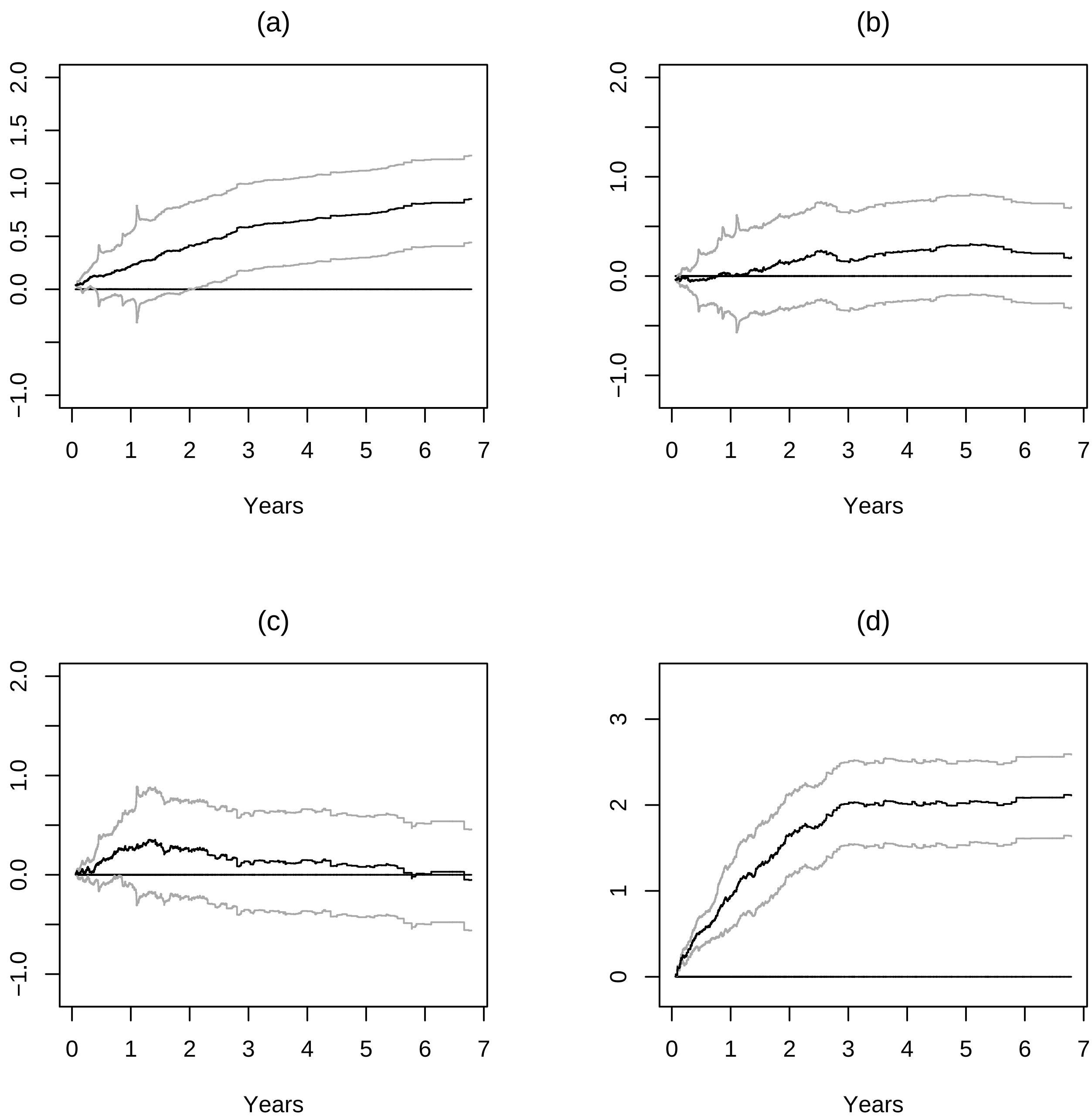

Figure 4.9 : Estimates of the coefficients of model (4.2.1) with $\phi_1(u)$ and 95% pointwise confidence intervals : (a) $\beta_0(z)$; (b) $\beta_1(z)$; (c) $\beta_2(z)$; (d) $\beta_3(z)$.

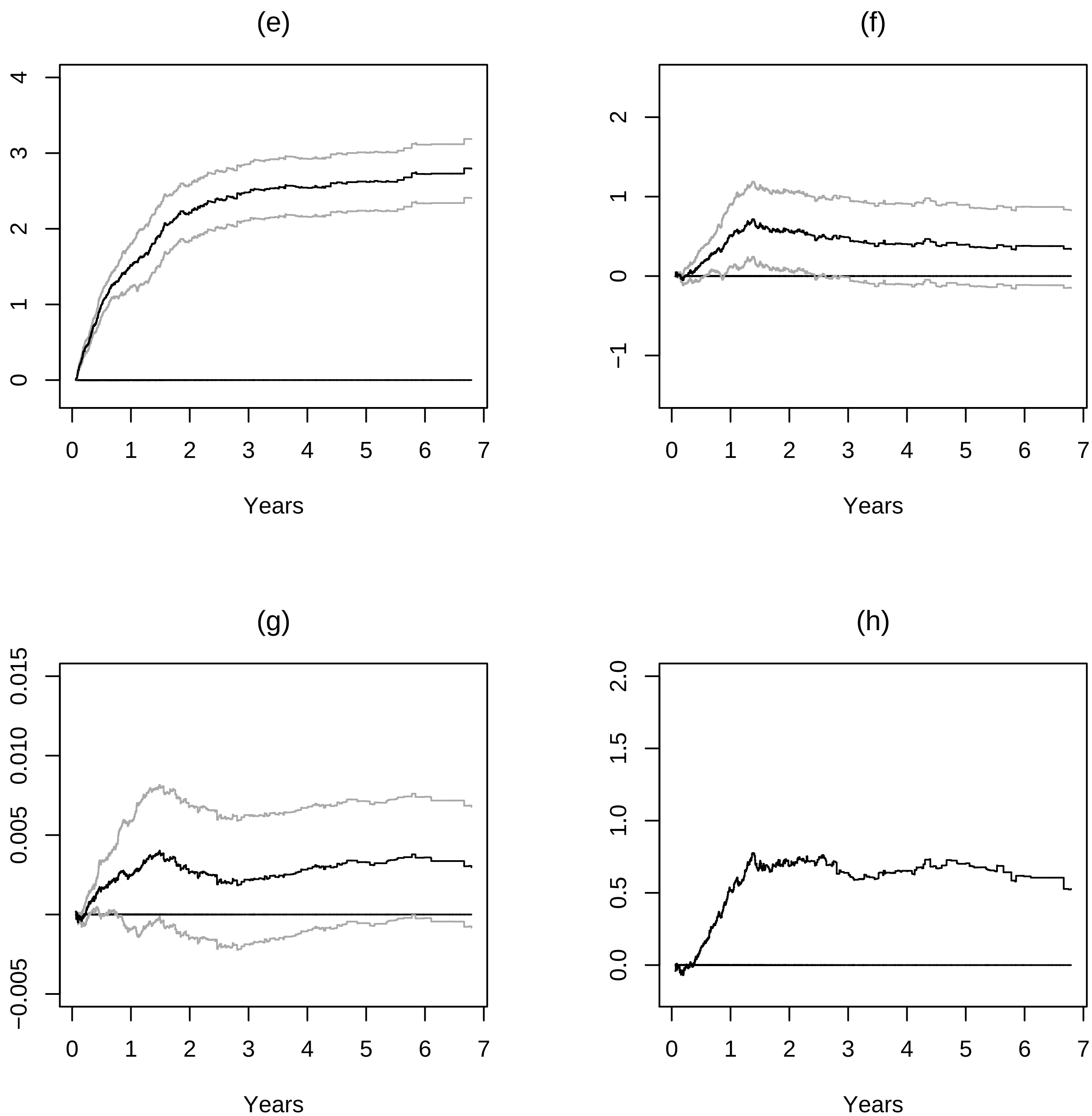

Figure 4.10 : Estimates of the coefficients of model (4.2.1) with $\phi_1(u)$ and 95% pointwise confidence intervals : (e) $\beta_4(z)$; (f)$\beta_5(z)$; (g) $\beta_6(z)$ and the cumulative hazard rate of Group 212 as compared to Group 111 : (h) $\beta_1(z) + \beta_5(z)$.

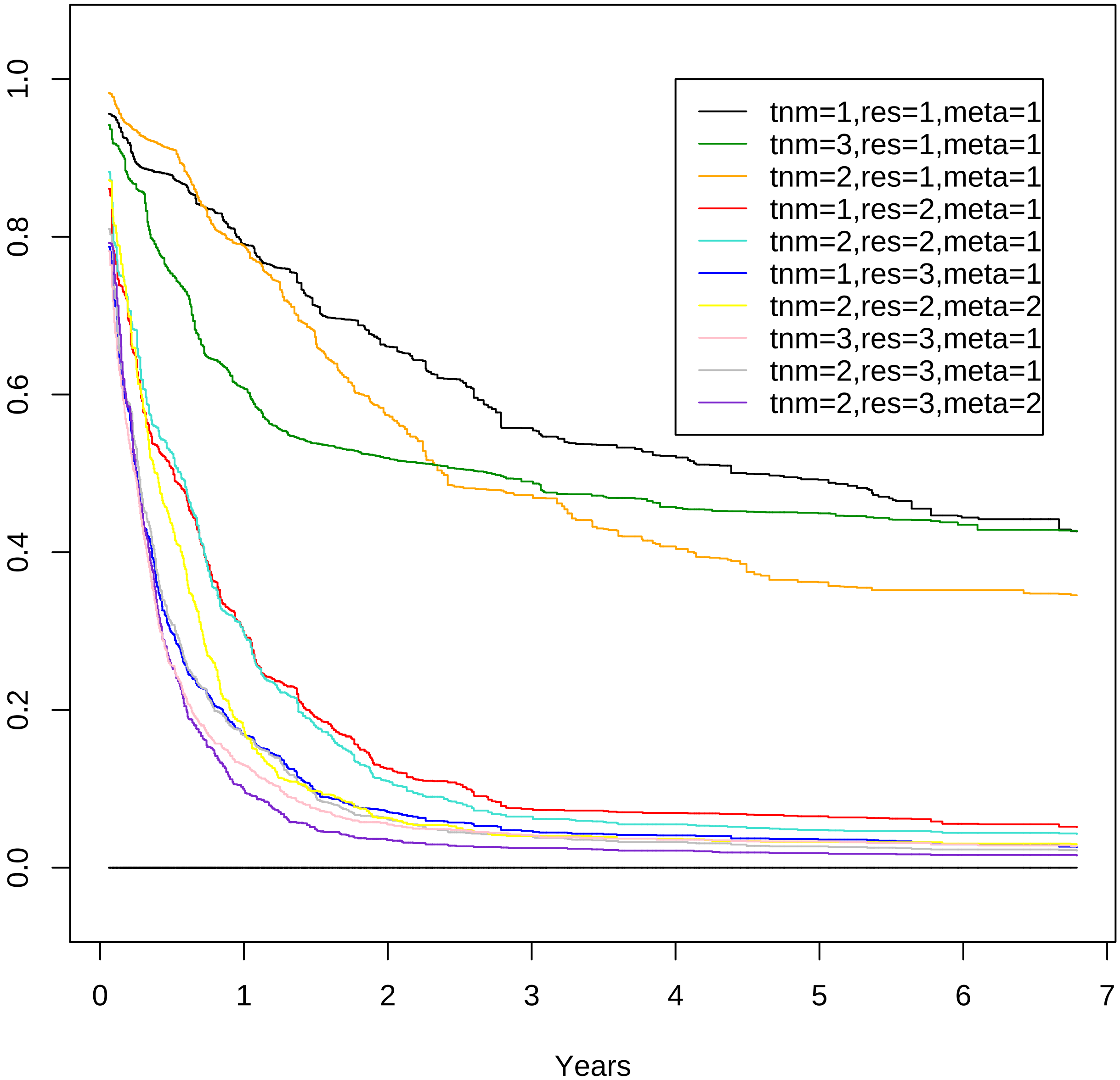

Figure 4.11 : Estimates of the survival functions based on the additive hazards model for gastric cancer patients of mean age 67.71.

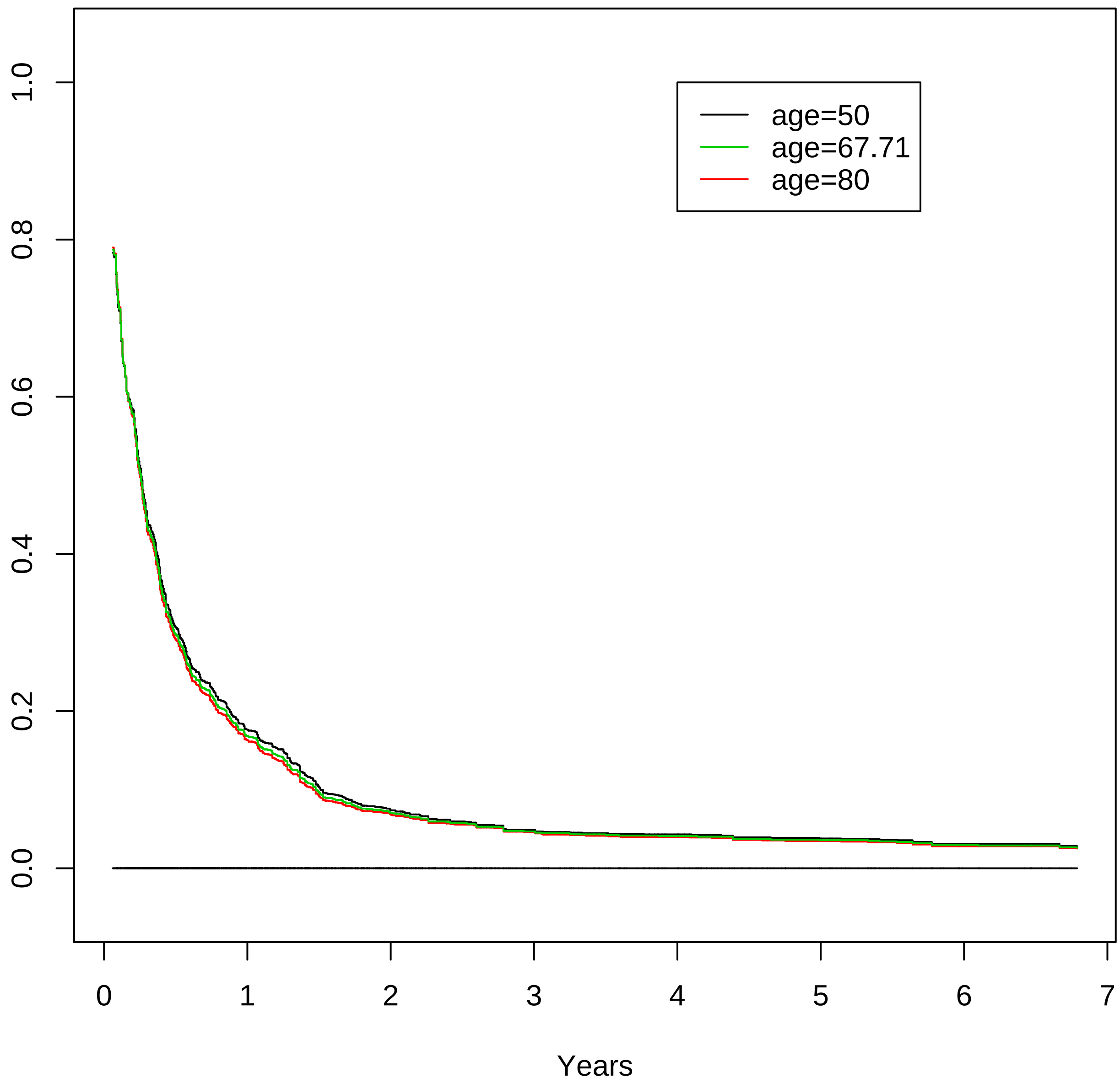

Figure 4.12 : Estimates of the survival functions based on the additive hazards model for gastric cancer patients in Group 131 of age $50, 67.71$ and 80.

Now we consider the hypothesis testing for $\phi_2(u)$, thus for the Cox model with time-dependent coefficients. Same procedures as for $\phi_1(u)$ are being applied. The optimal bandwidth under H_0 is $h = 25$ and the associated p-value was found to be 0.991. Hence, we do not reject H_0, i.e. we conclude that the Cox model with time-dependent coefficients also fits these data. So we can explain the influence of the covariates on the cumulative hazard function both in an additive and in a multiplicative way. We will see that the two models, in this case, complete each other.

The estimators of $\beta_k(z)$, $k = 0, \dots, 6$, as well as their 95 % confidence intervals computed via the percentile method are given in Figure 4.13 and Figure 4.14. As we can see in Figure 4.14 (g) the age at diagnosis seems to have an influence on the survival rate during the first two years and a half (same conclusions as when fitting the additive hazards model).

When fitting a Cox model with time-dependent coefficients, an easier interpretation of the results arises when computing the $\exp(\beta_k(z))$, $k = 1, \dots, 5$, see Figure 4.15 and Figure 4.16, since it represents different ratios of cumulative hazard functions of different groups. For example, $\exp(\beta_1(z))$ in Figure 4.15 (a) represents the ratio between the cumulative hazard rate of Group 211 and the cumulative hazard rate of Group 111 and it indicates that for the first 8 months after the first symptoms the cumulative hazard rate of Group 211 is smaller than that of Group 111, after which it becomes larger, until getting to be 1.5 larger after the first 2 years from the first symptoms. From $\exp(\beta_6(z))$ in Figure 4.16 (h) we see that when comparing patients belonging to the same group, but of different ages, then the risk for patients having age $a + 1$ as compared to those having age a is higher, but decreasing during the first 2 years and a half, to become constant bigger than 1, after this period.

We also plotted in Figure 4.17 the model-based estimators of the survival functions of the 11 groups. We notice more or less the same things as when the additive hazards model was fitted, only that here, the difference between Group 311 and Group 211 is almost inexistent during the first two years and then Group 311 seems to have a better survival rate. For the other groups we clearly see that Group 121 has a better survival rate than the remaining ones, followed by Group 131. But globally we make the same remark as when the additive hazards model was fitted, meaning that almost all patients in the groups with worst survival rate were dead by the 3rd year after the first symptoms have appeared, since their survival rate reaches almost zero.

From the additive hazards model we could see that we had excess risk of patients of different groups as compared to patients of Group 111 varying with time. The coefficients of this model quantified the difference between these cumulative survival functions.

Cox model with time-dependent coefficients allows as to quantify, instead, the ratio between the cumulative hazard functions of different groups. For example for Group 121 patients, we could see that there was excess risk for the first 3 years (non-zero slope in Figure 4.9 (b)). On the other hand, with Cox model, from Figure 4.16 (d), we can say that during the first 8 months after the first symptoms, the cumulative hazard rate of Group 121 was 5 times larger than that of Group 111, after which it decreased to being now only 4 times larger and after a period of 3 years it only remains 3.5 larger.

In conclusion we see that allowing the coefficients to vary with time revealed significant changes of the influence of the different factors on the survival time of the patients. Compared to previous studies on this data, when only Cox model with constant coefficients was fitted, we managed not only to show that the considered factors have an influence on the survival of the patients, but we were also able to show that their influence is not constant as before stated. We could clearly see from the graphic representations that the influence changes during the first 3 years following the first symptoms, and that only after this period the influence becomes constant.

Remark 4.2.1 : We can notice that for the Cox model the estimator of the coefficients has big jumps in the first days, in comparison to the coefficients of the additive hazards model. This is due to the $\log(-\log(u))$ transformation of the survival function $S(z|\boldsymbol{X})$, which is sensitive to small variations when S is close to 1. This is not the case when we only take $-\log(S(z|\boldsymbol{X}))$ as for the additive hazards model.

Remark 4.2.2 : In the bootstrap method, the estimator of the survival function was not necessarily strictly decreasing, thus we kept it constant until it started decreasing again. The plots of the survival function presented in this chapter were monotized via the procedure proposed by Dette et al. (2006).

Remark 4.2.3 : It will be interesting to apply the procedure in Cao and González-Manteiga (2008) in order to test if Cox model or additive hazards model with constant coefficients would have fitted the two data sets presented in this chapter.

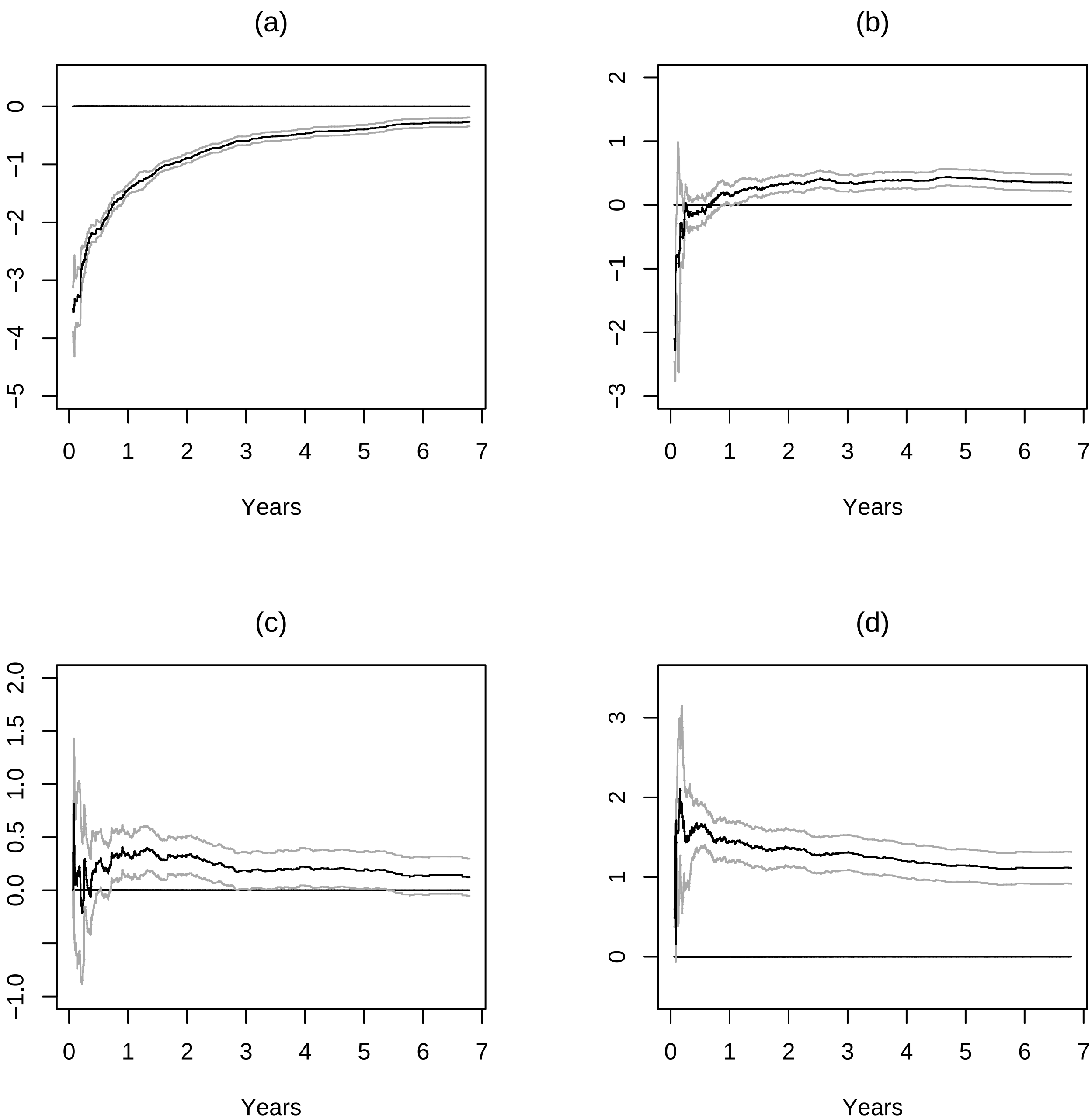

Figure 4.13 : Estimates of the coefficients of model (4.2.1) with $\phi_2(u)$ and 95% pointwise confidence intervals : (a) $\beta_0(z)$; (b) $\beta_1(z)$; (c) $\beta_2(z)$; (d) $\beta_3(z)$.

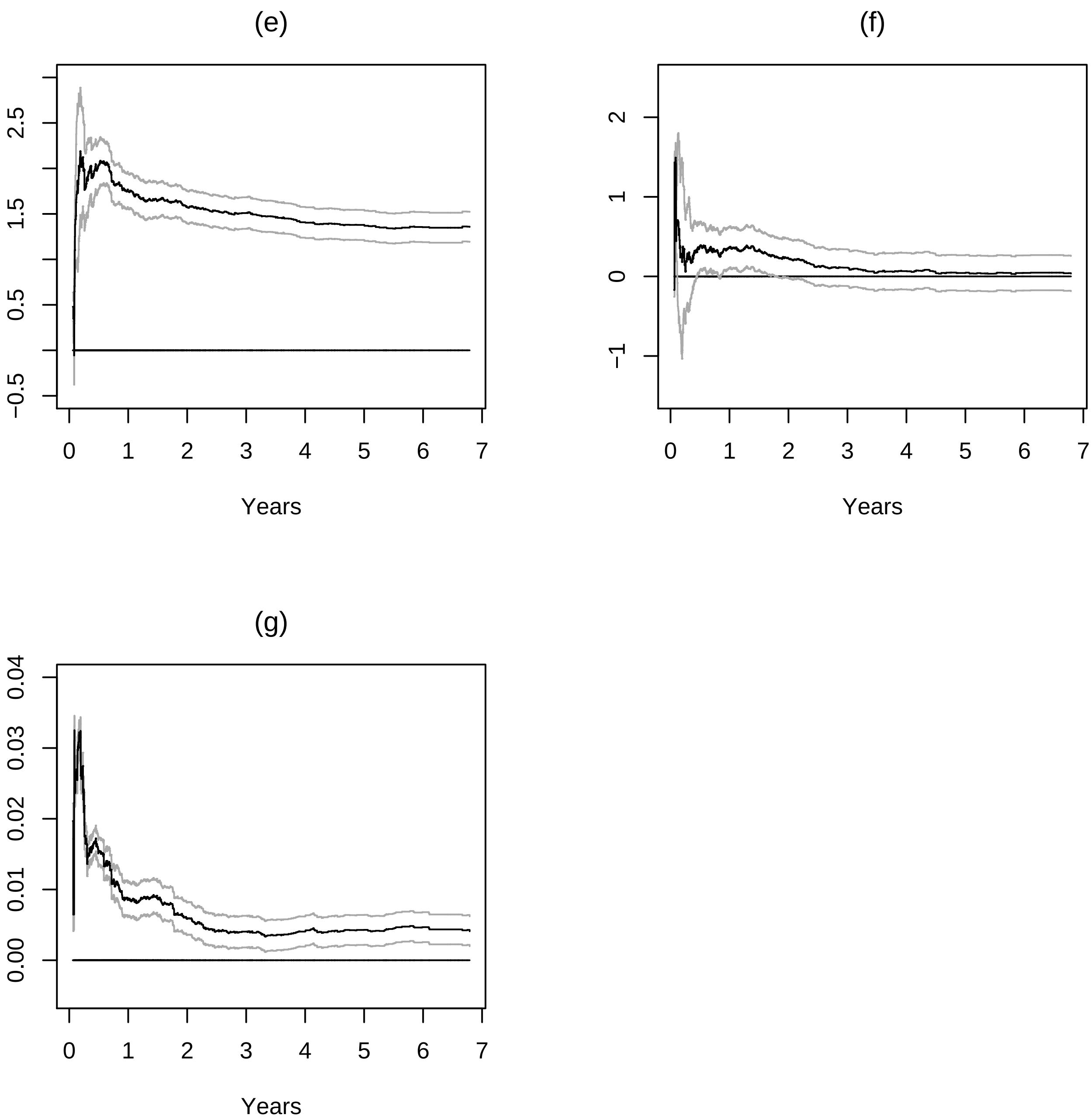

Figure 4.14 : Estimates of the coefficients of model (4.2.1) with $\phi_2(u)$ and 95% pointwise confidence intervals : (e) $\beta_4(z)$; (f) $\beta_5(z)$; (g) $\beta_6(z)$.

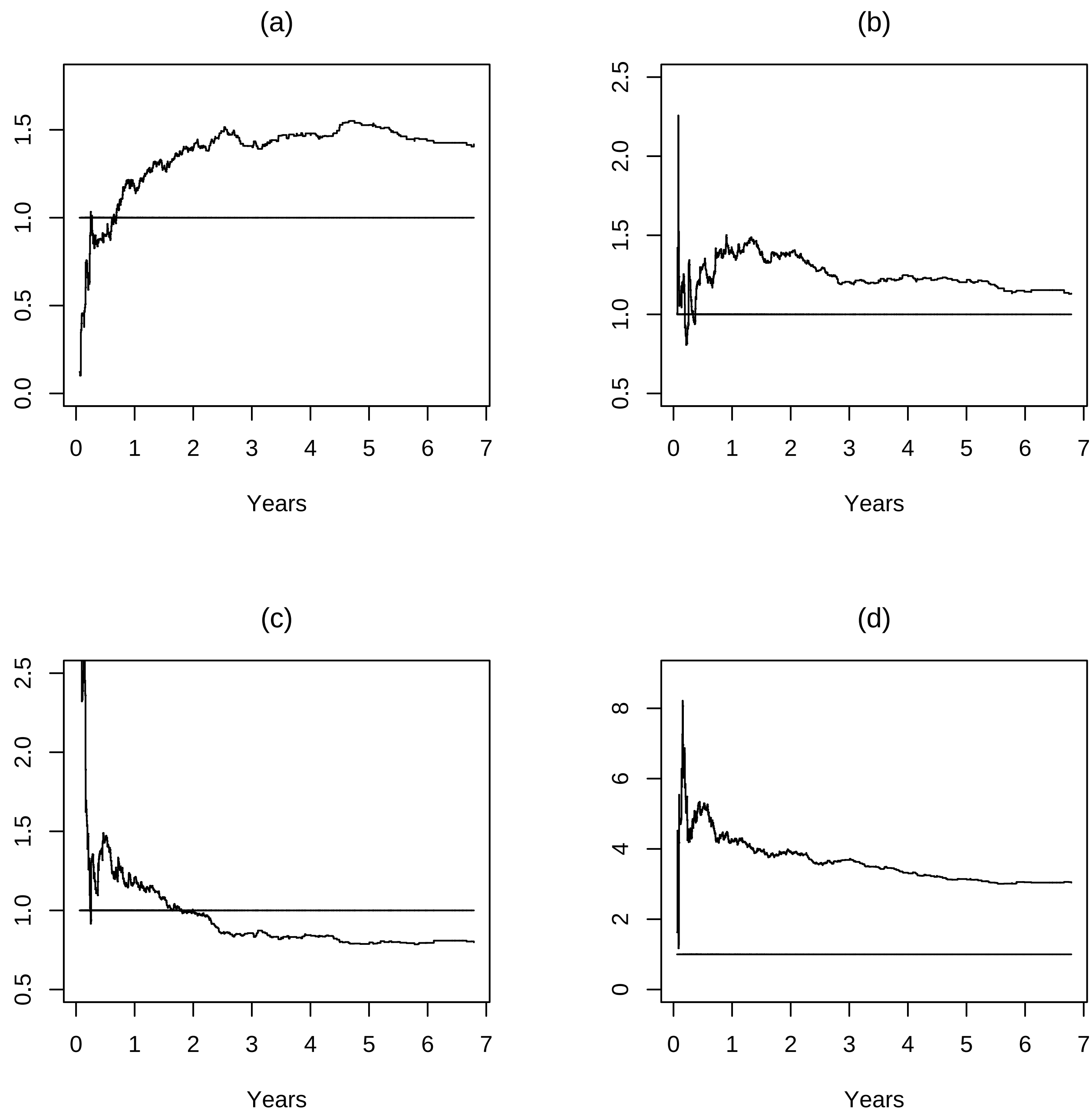

Figure 4.15 : Estimates of (a) $\exp(\beta_1(z))$, (b) $\exp(\beta_2(z))$, (c) $\exp(\beta_2(z) - \beta_1(z))$, (d) $\exp(\beta_3(z))$.

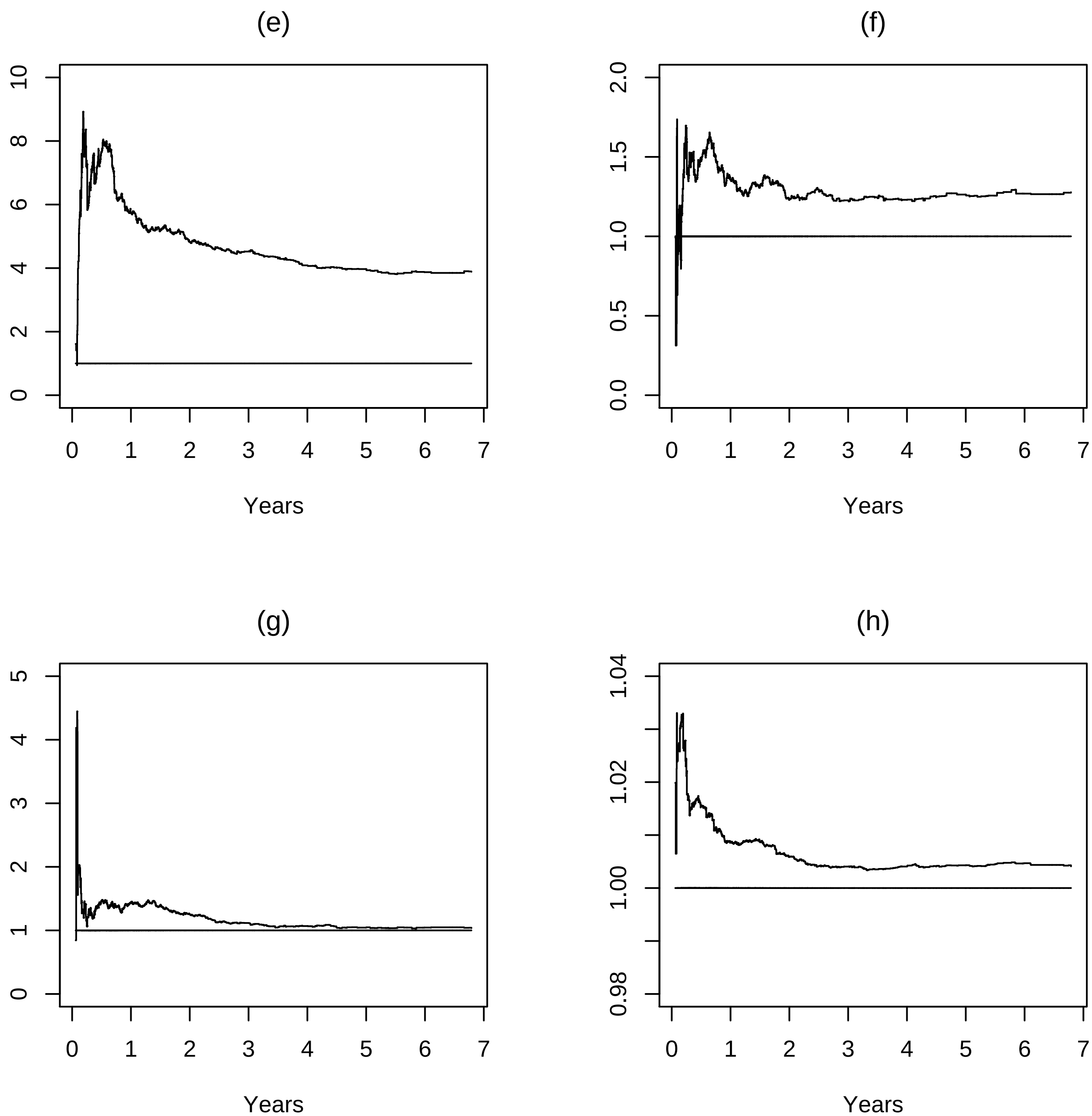

Figure 4.16 : Estimates of (e) $\exp(\beta_4(z))$, (f) $\exp(\beta_4(z) - \beta_3(z))$, (g) $\exp(\beta_5(z))$, (h) $\exp(\beta_6(z))$.

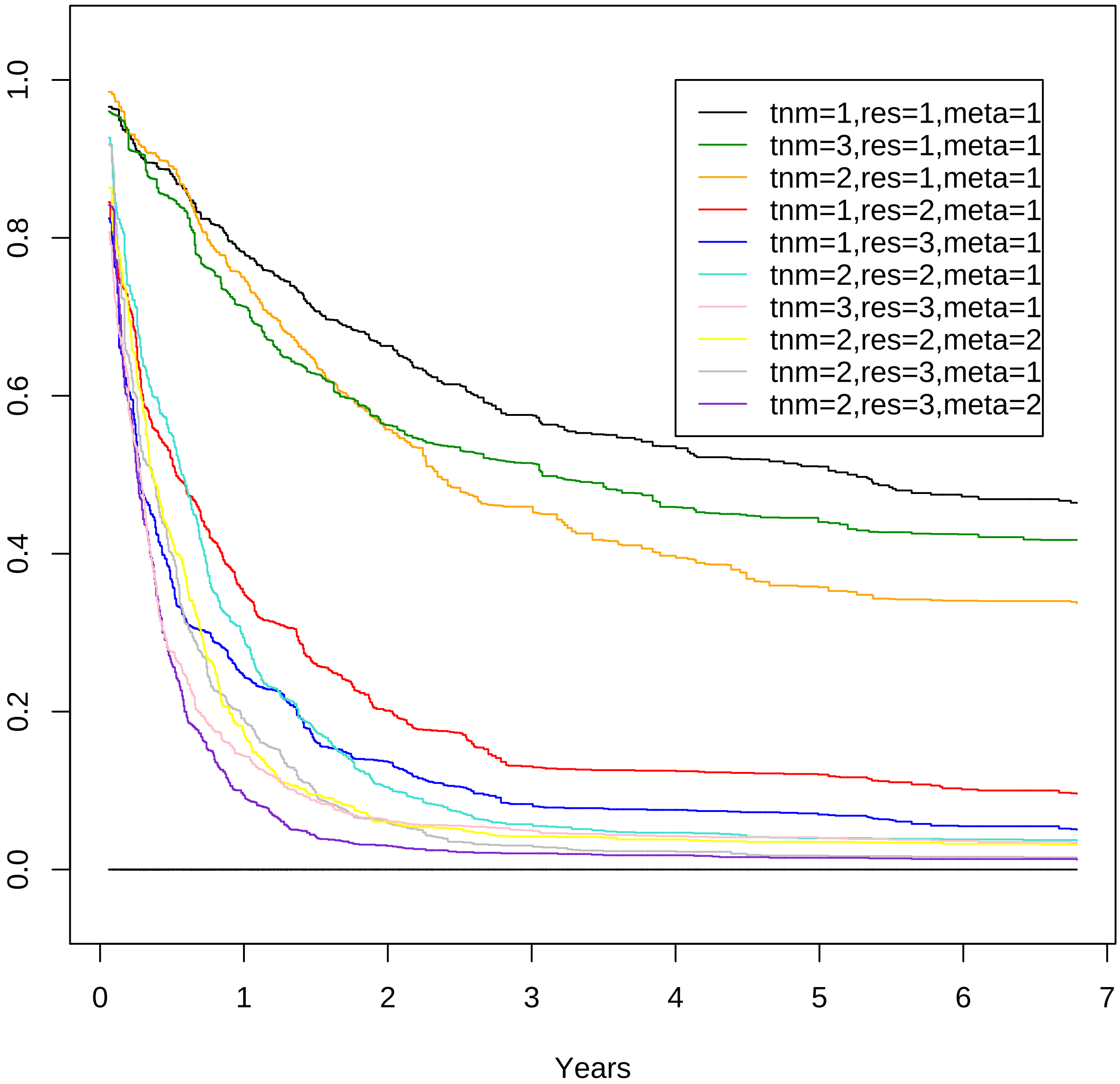

Figure 4.17 : Estimates of the survival functions based on the Cox model with time-dependent coefficients for gastric cancer patients of mean age 67.71.

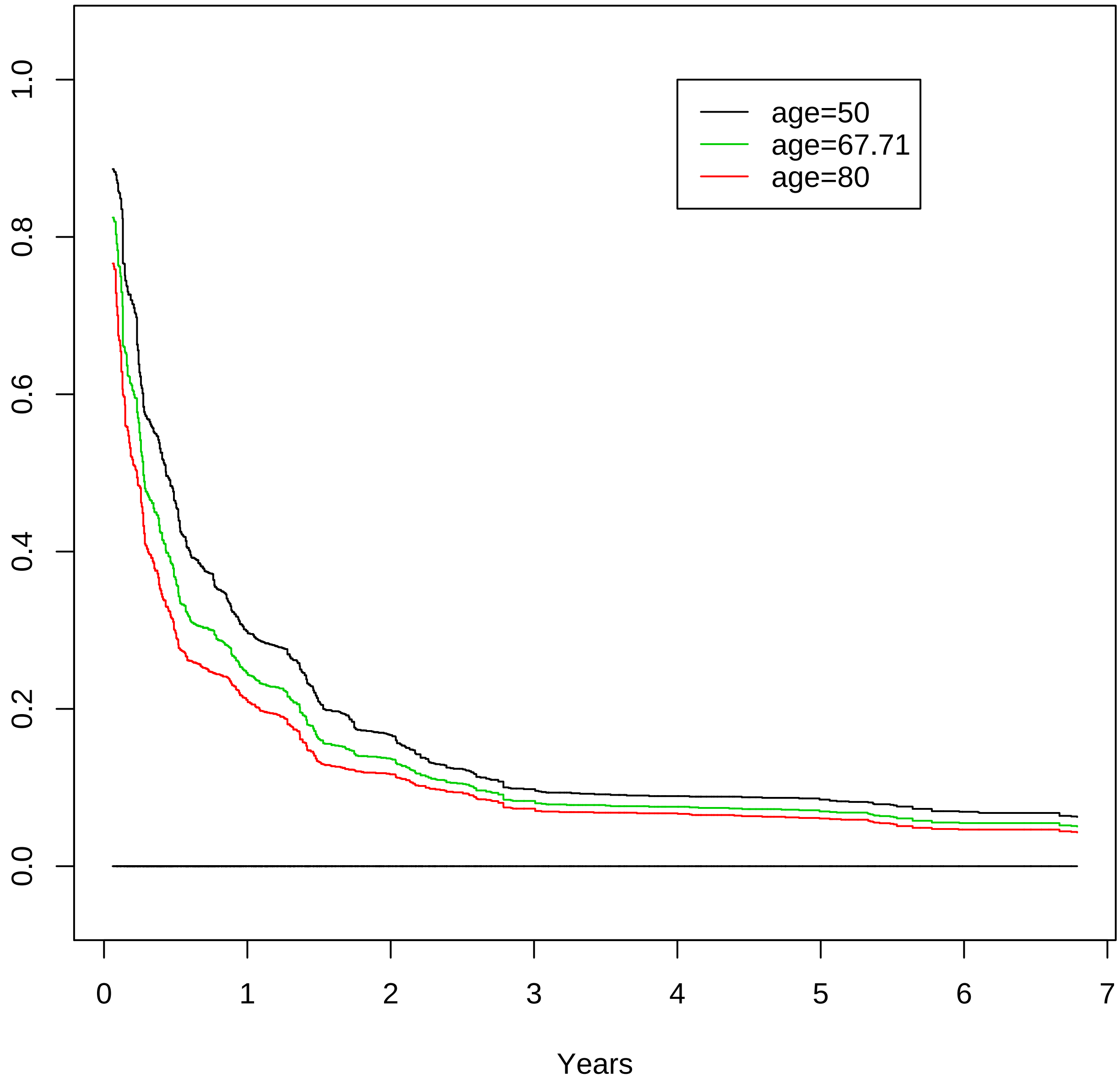

Figure 4.18 : Estimates of the survival functions based on the Cox model with time-dependent coefficients for gastric cancer patients in Group 131 of age $50, 67.71$ and 80.

Chapter 5

Conclusions and Further research

In this thesis we have studied a general class of flexible regression models : GLM with time-dependent coefficients, when the variable of interest can be right-censored and/or left-truncated. In Chapter 2, we proposed a least-squares estimator for model

$$\phi(S(z|\boldsymbol{X}) = \boldsymbol{\beta}(z)\boldsymbol{X}. \tag{5.0.1}$$

We studied its asymptotic properties as well as its finite sample properties. For the choice of the optimal bandwidth involved in the estimation procedure, a bootstrap based method has been proposed. This selection method gives good results, but it should be interesting to develop (practically and theoretically) other bandwidth selection techniques like plug-in principle or cross-validation. In Chapter 3, an omnibus goodness-of-fit test has been proposed to check the validity of model (5.0.1). The asymptotic distribution of the test statistic has been studied and a bootstrap method to approximate the critical values of the test has been proposed and tested via simulations. The bootstrap approximation gives promising results, but needs further investigations including the proof of consistency and application to other data generating scenarios. In Chapter 4, the above developed methodologies were applied to the two data sets described in Section 1.1.

In this work we have focused our attention to model (5.0.1), where all the coefficients are considered time-varying. In practice, it might be the case that some coefficients are time-independent, so one might consider a model with mixed coefficients : some constants and some time-varying. Let's write $\boldsymbol{\beta}(z)$ in model (5.0.1) as $\boldsymbol{\beta}(z) = (\boldsymbol{\beta}_1(z), \boldsymbol{\beta}_2(z)) = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2(z))$, with $\boldsymbol{\beta}_1$ a r-vector of time-independent regression coefficients, $\boldsymbol{\beta}_2(z)$ a s-vector of time-dependent regression coefficients and where $\boldsymbol{X}$ in model (5.0.1) is a $(r+s)$-dimensional vector of covariates. Now let $\hat{\boldsymbol{\beta}}_1(z)$ be the time-varying estimate of $\boldsymbol{\beta}_1(z)$, defined by (2.2.1) in Chapter 2. If $\boldsymbol{\beta}_1$ is time-independent, a natural way to define a time-independent estimator is to consider :

$$\hat{\boldsymbol{\beta}}_1 = \frac{1}{b-a}\int_a^b \hat{\boldsymbol{\beta}}_1(z)dz, \tag{5.0.2}$$

where a, b are defined in condition (H2), in Section 2.3. It can be easily checked that this quantity converges to $\boldsymbol{\beta}_1 = \frac{1}{b-a}\int_a^b \boldsymbol{\beta}_1(z)dz$. Moreover, following the proof of Theorem 2.4.1, we can write :

$$\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = \frac{1}{b-a}\int_a^b \left(\hat{\boldsymbol{\beta}}_1(z) - \boldsymbol{\beta}_1(z)\right) dz = \frac{1}{b-a}\boldsymbol{A}^{-1}\int_a^b \hat{\boldsymbol{b}}(z)dz,$$

where $\boldsymbol{A}$ and $\hat{\boldsymbol{b}}(z)$ are given in Theorem 2.4.1 and (2.7.1), respectively. Then, as a direct consequence of Corollary 2.4.2, we have that $|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1| = O_p(n^{-1/2})$. Following the same arguments as in the proof of Theorem 2.4.1 one can also show the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}_1$, by showing that $n^{1/2}\left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right)$ converges in distribution to a normal distribution.

However, a question that still needs investigation is whether $\boldsymbol{\beta}_1(z)$ is time-dependent or not. For this we will not only need the point-wise convergence of $\hat{\boldsymbol{\beta}}(z)$, the estimator of $\boldsymbol{\beta}(z)$, for each fixed time z, as shown in Chapter 2, but also the convergence of the process $\hat{\boldsymbol{\beta}}(z)$, $z \in [a,b]$. To establish the convergence of that process, one can adapt, for example, the techniques in Tian et al. (2005). Once the convergence of the process is established, the first idea would be to individually check the graphical representation of $\hat{\boldsymbol{\beta}}_1(z)$ and of its confidence bands to see whether a horizontal line is enclosed by the confidence bands in each plot. The second idea is to use a more objective and efficient way by considering the following process :

$$\Gamma(t) = n^{1/2}\int_a^t \left(\hat{\boldsymbol{\beta}}_1(z) - \hat{\boldsymbol{\beta}}_1\right) dz, \tag{5.0.3}$$

where $t \in [a,b]$, $\hat{\boldsymbol{\beta}}_1(z)$ and $\hat{\boldsymbol{\beta}}_1$ is defined in (2.2.1) and (5.0.2), respectively. Then the limiting distribution of this process should be studied and used in the same way as in Tian et al. (2005) to check the hypothesis of interest.

Another point that may be considered is to consider model (5.0.1) with an unknown $\phi(u)$, also known as GLM with missing link. To the best of our knowledge, little references exist in the literature for such a setup. When data are censored and the coefficients are constant, Wang et al. (2007) and Muggeo and Ferrara (2008) proposed some methods do deal with this case. It should be interesting to try to investigate the ideas in these papers to the case where the coefficients are time-dependent or when we have a mixture of coefficients (some constants, some time dependent) and also to the case of truncated data.

Another interesting problem related to the estimation of model (5.0.1) would be to consider a different estimation procedure of $\boldsymbol{\beta}(z)$. For uncensored data, equation (1.3.2) is equivalent to solving

$$\int \left\{ 1_{\{y \leq z\}} - \pi(\boldsymbol{\beta}^{'}(z)\boldsymbol{x}) \right\} \frac{\pi_\phi(\boldsymbol{\beta}^{'}(z)x)x}{\pi(\boldsymbol{\beta}^{'}(z)x)(1-\pi(\boldsymbol{\beta}^{'}(z)x))} d\tilde{F}(\boldsymbol{x}, y) = 0 \tag{5.0.4}$$

where $\pi = \phi^{-1}$, $\pi_\phi(\boldsymbol{\beta}^{'}(z)\boldsymbol{x}) = \dfrac{\partial \pi(\boldsymbol{\beta}^{'}(z)\boldsymbol{x})}{\partial(\boldsymbol{\beta}^{'}(z)\boldsymbol{x})}$ and $\tilde{F}(\boldsymbol{x}, y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_{i1} \leq x_1, \ldots, X_{ip} \leq x_p, Y_i \leq y\}}$. When Y is subject to censoring, a natural extension of equation (5.0.4) is to replace $\tilde{F}(\boldsymbol{x}, y)$ by an estimator of the bivariate distribution of $(\boldsymbol{X}, Y)$. Different estimators of such a distribution exist in the literature. For example Stute (1993, 1996) proposes an estimator which is valid for any number p of discrete and continuous covariates (his estimator does not require smoothing), but assumes that $P(Y \leq C|\boldsymbol{X}, Y) = P(Y \leq C|Y)$ which is a strong assumption. Another estimator is the one proposed by Akritas (1994) which is valid for a one-dimensional continuous covariate and requires smoothing. When data are subject to left or right truncation an estimator of the bivariate distribution is given by Gürler (1996, 1997), while Gürler and Gijbels (1997) proposed an estimator in the case where both left-truncation and right-censoring are present. It would be then interesting to compare from both theoretical and practical point of view the above method with the one proposed in Chapter 2.

Chapter 6

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.

Akritas, M. G. and LaValley, M. (1997). Statistical analysis with incomplete data :a selective review. In : *Handbook of statistics* (eds. G. S. Maddala and C.R. Rao). Elsevier Science, 15, 551-632.

Amir, C. and McKeague, I.W. (2000). Identifying a time-dependent covariate effect in the additive risk model. *unpublished manuscript availabale 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.

Andersen, P.K., Borgan, O., Gill, R., and Keiding, N. (1992). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

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.

Bennett, S. (1983). Analysis of survival data by the proportional odds models. *Statistics in Medicine*, 2, 273-277.

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

Breslow, N.E. and Day, N.E. (1980). *Statistical Models in Cancer Research, 1, The Design and Analysis of Case-Control Studies*. Lyon : IARC.

Buckley, J.D. (1984). Additive and multiplicative models for relative survival rates. *Biometrics*, 40, 51-62.

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.

Casariego-Vales, E., Pita-Fernández, S., Rigueiro-Veloso, M.T., Pértega-Díaz, S., Rabuñal-Rey, R., García-Rodeja, M.E. and Álvárez-cervela, L. (2001). Supervivencia en 2334 pacientes con cáncer gástrico y factores que modifican el prognóstico. *Med. Clin. (Barc)*, 117, 361-365.

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

Chen, Y.Q. (2001). Accelerated Hazards Regression Model and Its Adequacy for Censored Survival Data. *Biometrics*, 57, 853-860.

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

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

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

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, New York.

Cox, D.R. (1979). A note on the graphical analysis of survival data. *Biometrika*, 66, 188-190.

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

Dabrowska, D. M. and Doksum, K. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 15, 1-24.

de Jong, P. (1987). A Central Limit Theorem for Generalized Quadratic Forms. *Prob. Th. Rel. Fields*, 75, 261-277.

Dette, H., Neumeyer, N. and Pilz, K.F. (2006). A simple nonparametric estimator of a monotone regression function. *Bernoulli*, 12, 469-490.

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

Fan, J., Gijbels, I. and King, M. (1997). Local likelihood and local partial likelihood in hazard regression. *Annals of Statistics*, 25, 1661-1690.

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

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

Gamerman, D. (1991). Dynamic Bayesian methods for survival data. *Applied Statistics*, 40, 63-79.

Gürler, U. (1996). Bivariate Estimation with Right-Truncated Data. *Journal of the American Statistical Association*, 91, 1152-1165.

Gürler, U. (1997). Bivariate distribution and hazard functions when a component is randomly truncated. *Journal of multivariate analysis*, 60, 20-47.

Gürler, U. and Gijbels, I. (1997). A Bivariate Distribution Function Estimator and Its Variance under Left Truncation and Right Censoring. (Disscusion Paper DR9702 available at http ://www.stat.ucl.ac.be/publications).

Gill, R.D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika*, 74, 289-300.

González-Manteiga, W. and Cadarso-Suárez, Z. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics*, 4, 65-78.

Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81, 515-526.

Grigoletto, M. and Akritas, M.G. (1999). Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics*, 55, 1177-1187.

Haberman, S.J. (1977) - Maximum likelihood estimates in exponential response model. *Annals of statistics*, 5, 815-841.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21, 1926-1947.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficients models (with discussion). *Journal of The Royal Statistical Society-Series B*, 55, 757-796.

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.

Jin, Z., Lin, D.Y. and Ying, Z. (2006). On least-squares regression with censored data. *Biometrika*, 93, 147-161.

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

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd Edition.

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

Kardaun, O. (1983). Statistical Analysis of Male Larynx-Cancer Patients-A Case Study. *Statistical Nederlandica*, 37, 103-126.

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

Kay, R. (1977). Proportional hazards regression models and the analysis of censored survival data. *Applied Statistics*, 26, 227-237.

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.

La Valley, M.P. and Akritas, M.G. (2005). Extensions of the Lynden-Bell-Woodroofe method for truncated data. *Journal of Nonparametric Statistics*, 17, 643-663.

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).

Lin, D.Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81, 61-71.

Lynden-Bell, D. (1971) - A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Nat. Roy. Astr. Soc.*, 155, 95-118.

Martinussen, T., Scheike, T. H. and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. *Scandinavian Journal of Statistics*, 29, 59-77.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, New York.

McKeague, I. W. and Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *Annals of Statistics*, 18, 1172-1187.

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.

Moreau, T., O'Quigley, J., and Mesbah, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Annals of Statistics*, 3, 599-616.

Moreau, T., O'Quigley, J., and Lellouch, J. (1986). On D. Schoenfeld's approach for testing the proportional hazards assumption. *Biometrika*, 73, 513-515.

Muggeo, V.M.R. and Ferrara, G. (2008). Fitting generalized linear models with unspecified link function : A P-spline approach. *Computational Statistics and data analysis*, 52, 2529-2537.

Murphy, S.A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scandinavian Journal of Statistics*, 20, 35-50.

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.

Pettitt, A.N. (1982). Parametric tests for agreement amongst groups of judges. *Biometrika*, 69, 365-375.

Pierce, D.A. and Preston, D.L. (1984). Hazard function modelling for dose-response analysis of cancer incidence in A-bomb survivor data. In *Atomic Bomb Survivor Data : Utilization and Analysis*, Ed R.L. Prentice and D.J. Thompson, 51-66. Philadelphia : SIAM.

Pocock, S.J., Gore, S.M. and Kerr, G.R. (1982). Long term survival analysis : the curability of breast cancer. *Statistics in Medicine*, 1, 93-104.

Rabuñal, R.R., Salvador, F.P., Rigueiro, M.T.V., Casariego, E.V.L, Pértega, S.D., García-Rodeja, E. and Abraira, V. (2004). Presentation and Prognosis of Gastric Cancer in Patients Aged 80 Years and Older. *World. J. Surg.*, 28, 155-159.

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.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shen, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*, 85, 165-177.

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.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45, 89-103.

Stute, W. (1996). Distributional Convergence under Random Censorship when Covariables are Present. *Scandinavian Journal of Statistics*, 23, 461-471.

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. To appear in *The Annals of the Institute of Statistical Mathematics*, available on-line at http ://dx.doi.org/10.1007/s10463-008-0187-z.

Teodorescu, B. and Van Keilegom, I. (2008). Goodness-of-fit Test for Generalized Conditional Linear Models under Left Truncation and Right Censoring. *submitted.*

Tian, L., Zucker, D. and Wei, L.J. (2005). On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American Statistical Association*, 100, 172-183.

Thomas, D.C. (1986). Use of auxiliary information in fitting nonproportional hazards models. In *Modern Statistical Methods in Chronic Disease Epidemiology*, Ed. S.H. Moolgavkar and R.L. Prentice, 197-210. New York : Wiley.

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.

Wedderburn, R.W.M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, 63, 27-32.

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.

Wang, Y., He, S., Zhu, L. and Yuen, K.C. (2007). Asymptotics for a censored generalized linear model with unknown link function. *Probability theory and related Fields*, 138, 235-267.

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.