UNIVERSITE CATHOLIQUE DE LOUVAIN

INSTITUT DE STATISTIQUE

Mean preservation in censored regression using preliminary nonparametric smoothing

Membres du jury:

Prof. I. Gijbels
(Katholieke Universiteit Leuven)
Prof. W. González Manteiga
(Universidad de Santiago de Compostela)
Prof. Ph. Lambert
Prof. J.-M. Rolin
Prof. I. Van Keilegom (promotrice)
Prof. N. Veraverbeke
(Limburgs Universitair Centrum)

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

Louvain-la-Neuve

Août 2005

Avant toute entrée en matière, il est certaines personnes que je tiens à remercier pour leur contribution, de près ou de loin, à la réalisation de cette thèse.

Ten eerst wil ik Ingrid Van Keilegom, mijn promotor, bedanken voor haar beschikbaarheid. Met een ondersteunde aandacht heeft ze altijd naar mij geluisterd en mij in mijn onderzoek bijgestaan met haar relevante opmerkingen en deskundige raad.

While attending conferences, I have met researchers working on closely related topics. I am grateful we could raise interesting discussions leading to valuable input for my work. I namely think to Prof. M. Akritas (Penn State University, U.S.A) who kindly accepted my invitation to come and discuss Survival Analysis at the Young Researcher Day. This provided me food for though for further research.

Graag wil ik ook Prof. I. Gijbels bedanken voor haar bijdrage tot mijn wetenschappelijke opleiding. In this respect, I also want to thank Prof N. Veraverbeke, Prof. W.G. Manteiga, Prof. Ph. Lambert, Jury members, and Prof. J.-M. Rollin, Jury President, for the attention worthiness they grant to this research work.

Mes collègues assistants et doctorants ont aussi contribué au bon déroulement de cette thèse en me permettant de partager avec eux non seulement mon envie d'explorer toujours plus les domaines scientifiques mais aussi mon engouement à, modestement, les transmettre à nos étudiants.

Je souhaite également remercier mes parents pour leur soutien durant mon parcours académique. Bien que parfois difficile à comprendre, ils ont toujours respecté mon profond investissement dans ce travail de recherche et m'ont permis de le mener jusqu'à terme.

Enfin, j'adresse un merci tout particulier à ma compagne, Françoise, qui, malgré de nombreuses concessions, a toujours respecté et profondément soutenu mon travail.

Cédric Heuchenne
Louvain-la-Neuve, 18 août 2005.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Some examples

While examining a data set, it may happen that some of them are in some way incorrect or incomplete. Among the root causes for incompleteness, one of the most studied is "censoring". Censoring often occurs in medical surveys when we analyse for instance the survival time of a patient who has received a given treatment. During the observation period of the patient, some unexpected events can arise : the patient might decide to leave the study, die due to another cause than the disease from which he suffers or the study itself could be stopped.

In all those cases, we only observe a lower bound on the survival time of the patient. This is what is referred to as a right censored data point. Censoring not only happens when we study the lifetime of a person but also when we try to determine the failure time of an engine or the number of times a material can be submitted to the same constraint before breaking and so on. This problem finds indeed a large interest in many scientific contexts and is therefore covered by lots of research.

When conducting such experiments, some additional measures to the survival time are usually available. Coming back to our medical example, we may for instance know the blood pressure associated to each patient. Our interest in this thesis will be to study the

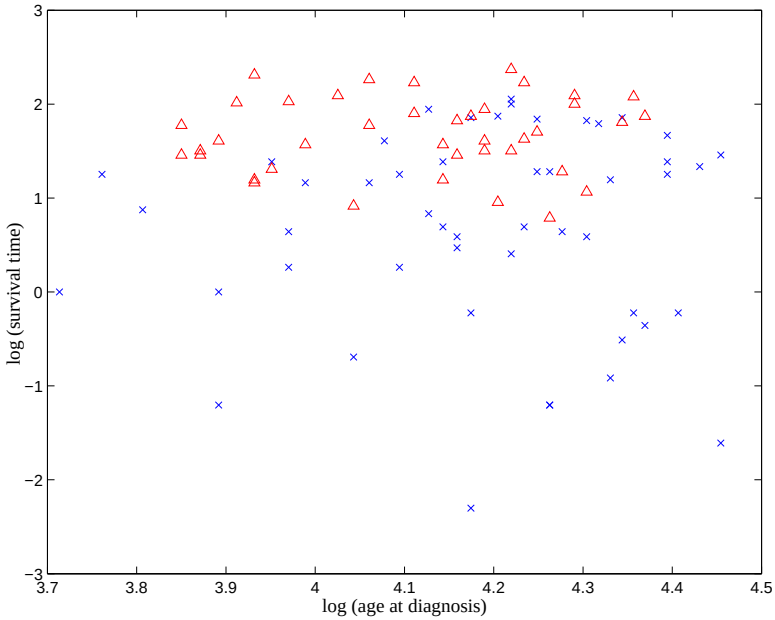


Figure 1.1.1: *Larynx cancer data. Scatter plot of the logarithm of the survival time versus the logarithm of the age at diagnosis. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*.*

relation between the response (lifetime) and the predictor (blood pressure in this case). This typically fits to a regression context taking into account the fact that the response values are possibly censored. The methods that will be developed will enable to deal with this kind of problems and will be applied to three concrete situations that are described below.

The first one is regarding 90 male larynx cancer patients, diagnosed and treated during the period 1970-1978 in a peripheral hospital in the Netherlands (see Kardaun (1983) for more details). The variable of interest is the time interval (in years) between first treatment and death of the patient. At the end of the study (1 March 1981), 40 patients were alive, and their survival time was therefore censored. We are interested in studying the relationship between the survival time and the age of the patient at diagnosis (in years). The available data points are illustrated in Figure 1.1.1.

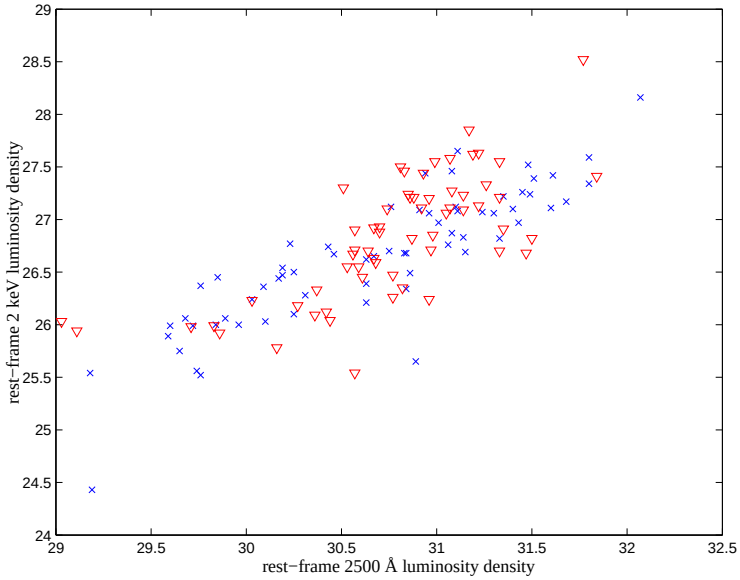


Figure 1.1.2: *Quasars data. Scatter plot of the rest-frame 2 keV luminosity density versus rest-frame 2500 Å luminosity density. Uncensored data points are given by* $\times$*, and censored observations by* $\bigtriangledown$*.*

The second data set comes from a study of quasars in astronomy. To date, many studies have focused on the dependence on luminosity and redshift of quasar ultraviolet-to-X-ray spectral energy distributions (characterized by means of the spectral index $\alpha_{ox} = 0.384 \log(L_{2\ keV}/L_{2500\ \mathring{A}})$, where $l_{uv} = \log L_{2500\ \mathring{A}}$ and $l_x = \log L_{2\ keV}$ denote the rest-frame 2500 $\mathring{A}$ and 2 keV luminosity densities) (see Vignali, Brandt and Schneider (2003)). This enables to obtain information and to validate the proposed mechanism driving quasar broad-band emission (accretion disk onto a super-massive black hole). Due to technical constraints of the used instruments, only upper bounds on 69 of the 137 values of l_x are observed. Note that observations are censored to the left here, instead of to the right. However, by inverting the direction of the time axis, results concerning right censored data can be translated into their left censoring-counterparts. Our results will therefore be applicable to left censored data as well (which are often encountered in astronomy). This data set is represented in Figure 1.1.2.

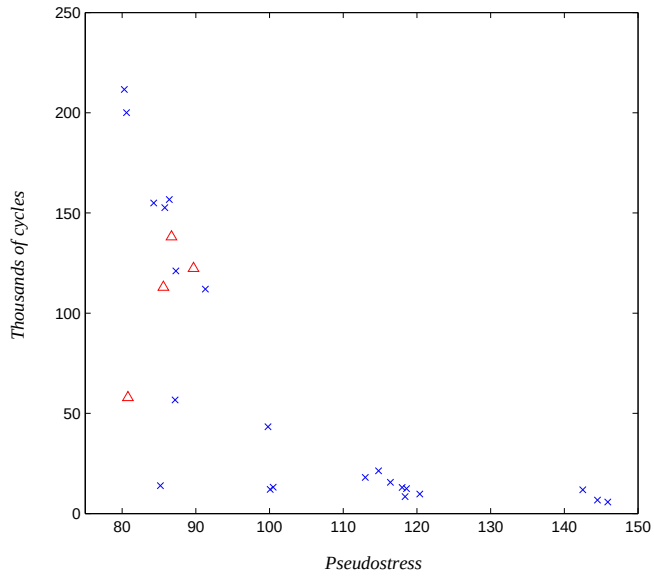


Figure 1.1.3: *Fatigue life data. Scatter plot of low-cycle superalloy fatigue life versus pseudostress for specimens of a nickel-base superalloy. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*.*

The relationship between fatigue life of metal, ceramic or composite materials and applied stress is an important input to design-for-reliability processes. This is motivated by the need to develop and present quantitative fatigue-life information used in the design of jet engines. For example, according to the air speed that enters an aircraft engine, the fan, the compressor and the turbine rotate at different speeds and therefore are submitted to different stresses. We present here a third set with low-cycle fatigue life data for a strain-controlled test on 26 cylindrical specimens of a nickel-base superalloy. The data were originally described and analysed in Nelson (1984) and can also be found in Meeker and Escobar (1998). Figure 1.1.3 shows the log of the number of cycles before failure against the pseudostress (Young's modulus times strain). Four censored data are observed; a data point is censored if failure occurs in the radius, weld or threads (censoring coming from impurities or vacuums) or if no failure at all occurs.

The objective of this thesis is to provide new estimation procedures which in some way outperform the existing ones. For each method, we want to study the asymptotic performance of our estimators and examine their finite sample behaviour through simulations.

There are many approaches in the statistical literature to describe the effect of a covariate on the lifetime when censored data are present.
First,when no assumption on the relation between the response and the covariate is made, a completely nonparametric approach can be followed. An estimator for the conditional distribution of the response given the covariate has been studied by Beran (1981). Based on this estimator, Doksum and Yandell (1982) studied the conditional mean and median. Akritas (1994) proposed a bivariate estimator of the response and the covariate and Fan and Gijbels (1994) used local polynomial techniques to obtain an estimator of the conditional mean (see also Section 1.7 for a larger overview).

Second, some assumptions on the way on which the response depends on the covariate can be made. Lots of work has been carried out in this context and we refer to e.g. Andersen et al. (1993) or Akritas and La Valley (1997) for a detailed survey. Among the most popular techniques lies the accelerated failure time model. It takes the form

$$S(w|\mathbf{X}) = S_0(\varphi(\mathbf{X})w), \tag{1.1.1}$$

where $S(w|\mathbf{X})$ is the survival function of W (survival time) given the covariate vector $\mathbf{X}$. S_0 refers to this survival function under standard conditions (for instance $\mathbf{X} = 0$) and $\varphi(\mathbf{X})$, the acceleration factor, is some unknown positive function of the covariate vector such that, under the standard conditions $\mathbf{X} = 0$, $\varphi(0) = 1$. In many cases, the assumption of parametric forms $\varphi(\mathbf{X};\beta)$ and $S_{0\theta}(\varphi(\mathbf{X})w)$ for $\varphi(\mathbf{X})$ and $S_0(\cdot)$ is taken. $S_{0\theta}(\cdot)$ is allowed to belong to a parametric class of distributions of parameter θ and a common choice for $\varphi(\mathbf{X};\beta)$ is $\exp(\beta'\mathbf{X})$.

An equivalent representation for (1.1.1) is

$$Y = \log W = -\log\varphi(\mathbf{X}) + \varepsilon, \tag{1.1.2}$$

where ε is a random variable independent of $\mathbf{X}$. Of course, if $\varphi(\mathbf{X};\beta) = \exp(-\beta'\mathbf{X})$, the accelerated failure time model is equivalent to a polynomial regression model where ε is independent of $\mathbf{X}$.

Another well-known model is the proportional hazards model of Cox (1972). It is written as

$$\lambda(w|\mathbf{X}) = \lambda_0(w)\psi(\mathbf{X}),$$

where $\lambda(w|\mathbf{X})$ denotes the hazard function of the response given the covariate vector, $\lambda_0(w)$ is the unknown baseline hazard function and the unknown function $\psi(\mathbf{X})$ is such that $\psi(0) = 1$ under the standard conditions $\mathbf{X} = 0$. As for the accelerated time failure model, $\lambda_0(\cdot)$ can be replaced by $\lambda_{0\theta}(\cdot)$ which belongs to a parametric class of hazards functions and the function $\psi(\mathbf{X})$ can be parametrized as $\psi(\mathbf{X};\beta)$. Here again, one of the most used form for $\psi(\mathbf{X};\beta)$ is $e^{\beta'\mathbf{X}}$.

A generalization of the proportional hazards model is called the frailty model. It is motivated by the need to include the possibility of correlated observations in a study. It is also considered as a way to incorporate unexplained variation among the subjects under study, caused by e.g. unmeasured variables. For those reasons, its form is given by

$$\lambda_i(w|\mathbf{X}) = V_i\lambda_0(w)\psi(\mathbf{X}),$$

where V_i is a random effect generated from a parametric distribution on the real line and corresponding to a group i of individuals.

In a common proportional hazards model, the hazard functions at different levels of $\mathbf{X}$ are proportional. It's possible, however, that the hazard functions are parallel or more generally, that they can be made parallel by a known nonlinear transformation h of λ. Thus, we may consider additive hazards regression models of the form

$$h(\lambda(w|\mathbf{X})) = \phi(\mathbf{X}) + h(\lambda_0(w)),$$

where $\phi(0) = 0$. Parametric versions are obtained by, for instance, writing $\phi(\mathbf{X};\beta') = \beta'\mathbf{X}$ and by taking one of the parametric families for $\lambda_0(\cdot)$.

Heteroscedastic regression models are defined as

$$h(W) = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon, \tag{1.1.3}$$

where h is a known monotone transformation of the survival time W, m is an unknown location function, σ is an unknown scale function representing possible heteroscedasticity

and the error variable ε is independent of $\mathbf{X}$. Extensive research using this model has been investigated for particular situations. In the literature, m and σ can be estimated in a parametric or nonparametric way. In the parametric case, m can be e.g. a linear, polynomial or nonlinear function of the covariate. The models considered can be heteroscedastic or simply restricted to the homoscedastic case ($\sigma(\mathbf{X}) = \sigma$ independent of $\mathbf{X}$). Since it is largely used and reduces to the accelerated failure time model (1.1.2) when the error terms are homoscedastic and h is the log-transformation, this model (with one-dimensional covariate X) will also be considered all along this thesis.

In the first part, the conditional mean of the response Y given the covariate X is assumed to have a polynomial form. We therefore propose a new procedure to estimate the regression parameters. This method is based on a reconstruction of the censored data points obeying a criterion of preservation of mean. Using a preliminary smoothing based on (1.1.3) allows to introduce information about the whole model when reconstructing the data points. This is particularly interesting if there are regions in the covariate space where censoring is heavy.

Then, this procedure is extended to the case where the conditional mean is allowed to have a nonlinear form. This method is interesting from a practical point of view since, in the censored nonlinear context, few estimators are provided and the applications are various and numerous.

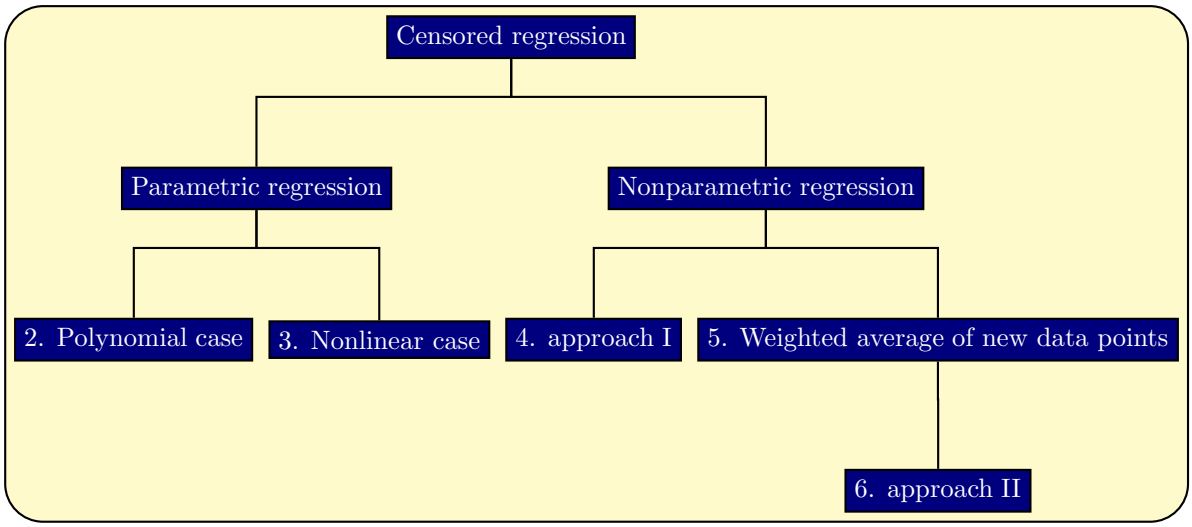
In the second part, we extend the study to any kind of conditional location function (e.g. conditional mean, truncated mean, quartiles ...). Moreover, no parametric form on the conditional location function is assumed anymore. In this context, we propose two new nonparametric estimation procedures. The first one consists in replacing the usual nonparametric estimator of the distribution function (Stone (1977)) for complete data by a suitable estimator for censoring in the expression of the location function. The estimation is improved by the fact that the method allows to tranfer tail information from regions of light censoring to regions of heavy censoring.

The second new nonparametric procedure is based on a different idea, closer to the technique proposed for the polynomial case. It constructs new 'synthetic' data points using model (1.1.3) and estimates a conditional location by using a weighted average of those new data points. Here again, a preservation of mean criterion is used but it varies according to the function we want to estimate. As a consequence, that method requires important

theoretical results about the weighted average of general synthetic data points.

For each estimator proposed in this thesis, the objective is twofold. First, we want to study its asymptotic performance by establishing properties like e.g. consistency, asymptotic normality... Second, we want to examine its finite sample performance through simulations and comparisons with some other existing methods. Finally, applications to real data sets will allow to illustrate the studied techniques and to highlight some domains where such estimators are useful.

Diagram of the thesis



1.2 Nonparametric estimation in survival analysis

Let $Y_1, \ldots, Y_n$ be n i.i.d. random variables such as survival times or a monotone transformation of them with unknown distribution $F(t) = P(Y_1 \leq t)$. Suppose those random variables are possibly right censored by $C_1, \ldots, C_n$ n i.i.d. random variables. For each i, C_i is independent of Y_i and their distribution is denoted $G(t) = P(C_1 \leq t)$. Then, in the right random censorship model, we observe for each i, the pair (Z_i, Δ_i), where $Z_i = Y_i \wedge C_i$ and $\Delta_i = I(Y_i \leq C_i)$. The most widely used estimator of the distribution function F is the Kaplan-Meier estimator (1958):

$$F_n(t) = 1 - \left\{ \prod_{Z_{(i)} \leq t} (1 - \frac{1}{n-i+1})^{\Delta_{(i)}} \right\} I(t < Z_{(n)}), \tag{1.2.1}$$

where $Z_{(1)}, \ldots, Z_{(n)}$ are the ordered Z_i and $\Delta_{(1)}, \ldots, \Delta_{(n)}$ are the corresponding indicators. This estimator is a step function that jumps at the uncensored data points and for which the jumps are all positive. In case of no censoring (i.e. $Z_i = Y_i$ for all i), the estimator reduces to the empirical distribution function. If $t \in I$, $t_* = \inf\{t : t \in I\}$ and $t^* = \sup\{t : t \in I\}$, $F_n(t_*) = 0$ and $F_n(t^*) = 1$. However, the last indicator on the right hand side of (1.2.1) can be deleted such that $F_n(t)$ can be defined (or not defined) in a different way for $t \geq Z_{(n)}$ ($F_n(t^*) \leq 1$). In fact, under independence between Y and C, consistency of $F_n(t)$ can only be proved for $t \leq \tau_H = \tau_G \wedge \tau_F$, where $H(t) = P(Z_1 \leq t)$ and $\tau_H = \inf\{t : H(t) = 1\}$. Foldes and Rejtö (1981) showed the strong uniform consistency of the Kaplan-Meier estimator while Lo and Singh (1986) obtained an asymptotic representation which decomposes $F_n(t) - F(t)$ in an average of i.i.d. terms and a remainder term of lower order. From this representation, the weak convergence (and in particular the asymptotic normality) of $F_n(t)$ can be easily derived . If the sample contains tied observations, consider the uncensored observations to occur just before the censored observations. The Kaplan-Meier estimator then generalizes to :

$$F_n(t) = 1 - \left\{ \prod_{Z'_{(i)} \leq t} (1 - \frac{d_i}{n_i})^{\Delta'_{(i)}} \right\} I(t < Z'_{(r)}), \tag{1.2.2}$$

where $Z'_{(1)}, \ldots, Z'_{(r)}$ are the distinct ordered Z_i, $\Delta'_{(1)}, \ldots, \Delta'_{(r)}$ are the corresponding indicators, n_i the number of subjects at risk at time $Z'_{(i)}-$ and d_i the number of deaths or failures at time $Z'_{(i)}$. Note that the generalization of (1.2.1) to (1.2.2) is very simple. For simplicity and as this extension can be applied similarly for all the estimators treated in this thesis, we won't consider tied observations in the sequel.

1.3 Nonparametric estimation in regression

Suppose $Y_1, \ldots, Y_n$ are n i.i.d. random variables corresponding to $X_1, \ldots, X_n$, n i.i.d covariates with distribution $F_X(x) = P(X_1 \leq x)$. Let $F(t|x) = P(Y_1 \leq t|X_1 = x)$, a continuous function in x, be the conditional distribution of the response given the covariate. The usual nonparametric kernel estimation of a conditional distribution function $F(t|x)$ uses the following idea. A window (possibly infinite) around x in the covariate space is constructed such that only responses corresponding to covariates lying in this window are used.It is clear that observations Y_i with X_i close to x are more informative than observations with X_i distant from x. Therefore, the estimation of the conditional distribution function considers responses for which covariates lie in the window mentioned above, giving more weight to observations close to x. This leads to the Stone estimator (1977)

$$\hat{F}(t|x) = \sum_{i=1}^{n} W_i(x, a_n) I(Y_i \leq t), \tag{1.3.1}$$

where

$$W_i(x, a_n) = \frac{K\left(\frac{x - X_i}{a_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{a_n}\right)}, \tag{1.3.2}$$

are the Nadaraya-Watson weights, K is a known probability density function, called kernel, and $\{a_n\}$ is a sequence of positive constants tending to 0 as $n \to \infty$, called the bandwidth sequence. Note that this estimator extends the empirical distribution function since it replaces n^{-1} by $K(\frac{x - X_i}{a_n})$ depending on the position of X_i with respect to x. Strong uniform consistency of the Stone estimator can be seen as a particular case of Theorem 3.2 of Härdle,

Janssen and Serfling (1988) or Theorem 5.2.3 of this thesis. Asymptotic normality can be found in Aerts, Janssen and Veraverbeke (1994).

1.4 Nonparametric estimation in censored regression

Now, suppose $Y_1, \ldots, Y_n$ are possibly right censored by $C_1, \ldots, C_n$ n i.i.d. random variables with distribution function $G(t|x) = P(C_1 \leq t | X = x)$. The observed random variable for the covariate X_i is therefore the pair (Z_i, Δ_i), $i = 1, \ldots, n$, with $Z_i = Y_i \wedge C_i$ and $\Delta_i = I(Y_i \leq C_i)$. We will now assume independence of Y_i and C_i conditionally on X_i. Beran (1981) was the first to study the estimation of the conditional distribution function in the presence of censored data. His estimator can be considered as an extension to censored data of the Stone estimator and a generalization of the Kaplan-Meier estimator to the local case (i.e. to estimate a conditional distribution $F(t|x)$ using a window around $X = x$). It is written as (in the case of no ties)

$$\tilde{F}(t|x) = 1 - \prod_{Z_i \leq t, \Delta_i = 1} \Big\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)} \Big\} I(t < Z_{(n)}). \tag{1.4.1}$$

If $W_i(x, a_n)$ equals to n^{-1} for all $i = 1, \ldots, n$, the Beran (1981) estimator reduces to the Kaplan-Meier estimator. Clearly, the estimator (1.4.1) only jumps at the uncensored data points with jump at Z_j that equals to

$$W'_j(x, a_n) = \prod_{Z_i < Z_j, \Delta_i = 1} \Big\{ 1 - \frac{W_i(x, a_n)}{\sum_{k=1}^n I(Z_k \geq Z_i) W_k(x, a_n)} \Big\} \Big\{ \frac{W_j(x, a_n)}{\sum_{k=1}^n I(Z_k \geq Z_j) W_k(x, a_n)} \Big\}^{\Delta_j},$$

for $Z_j < Z_{(n)}$ and

$$W'_n(x, a_n) = \prod_{i=1}^n \Big\{ 1 - \frac{W_i(x, a_n)}{\sum_{k=1}^n I(Z_k \geq Z_i) W_k(x, a_n)} \Big\}^{\Delta_i},$$

for $Z_j = Z_{(n)}$, so that $\tilde{F}(t|x)$ can also be expressed as $\tilde{F}(t|x) = \sum_{i=1}^n W'_i(x, a_n) I(Z_i \leq t)$. In case of no censoring, $W'_i = W_i$ and $\tilde{F}(t|x)$ becomes the Stone estimator. The indicator $I(t < Z_{(n)})$ in (1.4.1) can be deleted as for the Kaplan-Meier estimator. Indeed, here again, if $t \in I$, $t_* = \inf\{t : t \in I\}$ and $t^* = \sup\{t : t \in I\}$, $\tilde{F}(t_*|x) = 0$ but $\tilde{F}(t^*|x)$ can

be smaller than 1 : under the conditional independence between Y and C given $X = x$, consistency proofs are only obtained for $t \leq \tau_{H(\cdot|x)} = \tau_{F(\cdot|x)} \wedge \tau_{G(\cdot|x)}$, where $H(t|x) = P(Z_1 \leq t|X = x)$. First, this estimator has been proposed by Beran (1981) but was further studied by Dabrowska (1987,1989,1992a,b), McKeague and Utikal (1990), Akritas (1994) and González Manteiga and Cadarso Suárez (1994). The Beran estimator can be seen as a local constant estimator and can therefore be transformed in a local linear estimator (Li and Doss (1995)). This is motivated by the fact that, in the uncensored case, local constant estimation of the regression function has larger bias near the endpoints of the covariate region than local linear estimation. In the fixed design case, strong uniform consistency of the Beran (1981) estimator was proved by Van Keilegom and Veraverbeke (1996) while an asymptotic representation and the weak convergence were obtained by Van Keilegom and Veraverbeke (1997a,1997b).

1.5 Estimation in polynomial censored regression models

Suppose the random vector (X, Y) satisfies the polynomial regression model

$$Y = \beta_0 + \beta_1 X + \ldots + \beta_p X^p + \sigma(X)\varepsilon, \tag{1.5.1}$$

where $\sigma^2(X) = \mathrm{Var}(Y|X)$, and the error term ε (with unknown distribution $F_\varepsilon(\cdot)$) is independent of X and has zero mean. We suppose that Y is subject to random right censoring, i.e. instead of observing Y we only observe (Z, Δ), where $Z = \min(Y, C)$, $\Delta = I(Y \leq C)$ and the random variable C represents the censoring time, which is independent of Y, conditionally on X. Let $(Y_i, C_i, X_i, Z_i, \Delta_i)$ $(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z, Δ) and let $V = (X, Z, \Delta)$ denote the vector of observed random variables. The objective is to estimate the regression parameters $\beta_0, \ldots, \beta_p$, providing an extension to censored data of the classical least squares method.

Two main ideas have been proposed in the literature for extending the least squares estimators to censored data. The first one was to substitute censored data versions in place of empirical distributions for complete data yielding the uncensored least squares estimators. The second idea was to construct new ('synthetic') data points using appropriate estimators

for censored data and subsequently introduce them into an ordinary sum of squares. To the first class of extensions of the ordinary least squares estimators belong e.g. the estimators of Miller (1976), Stute (1993, 1996), Akritas (1994) and Van Keilegom and Akritas (2000). The second group includes the estimators of Buckley and James (1979) (which was further studied by Ritov (1990), Lai and Ying (1991a) and Lin and Wei (1992a)), Koul Susarla and Van Ryzin (1981) (Fygenson and Zhou (1994) considered a slight modification of this estimator), Leurgans (1987) (also studied by Zhou (1992a)) and Akritas (1996).

In Chapter 2, we propose a new procedure which is inspired by the method of Buckley and James (1979). The idea of this method is as follows. Consider for simplicity the case where $p = 1$, and suppose that $\sigma(X) \equiv 1$. Then,

$$E(Y_i^*|X_i) = \beta_0 + \beta_1 X_i,$$

where $Y_i^* = E(Y_i|Z_i, \Delta_i, X_i) = Y_i\Delta_i + E(Y_i|Y_i > C_i, X_i)(1 - \Delta_i)$. The idea of Buckley and James (1979) is to write

$$E(Y_i|Y_i > C_i, X_i) = \beta_1 X_i + \frac{1}{1 - F_{\beta_1}(Z_i - \beta_1 X_i)} \int_{Z_i - \beta_1 X_i}^{\infty} y \, dF_{\beta_1}(y)$$

and next to estimate Y_i^* by the 'synthetic' data points

$$\hat{Y}_i^*(\beta_1) = Y_i\Delta_i + \Big\{\beta_1 X_i + \frac{1}{1 - \hat{F}_{\beta_1}(Z_i - \beta_1 X_i)} \int_{Z_i - \beta_1 X_i}^{\infty} y \, d\hat{F}_{\beta_1}(y)\Big\}(1 - \Delta_i),$$

where $F_{\beta_1}(y)$ is the distribution of $Y - \beta_1 X$ and $\hat{F}_{\beta_1}(y)$ is the Kaplan-Meier (1958) estimator of $F_{\beta_1}(y)$ based on $(Z_i - \beta_1 X_i, \Delta_i)$ $(i = 1, \ldots, n)$. Next, Buckley and James (1979) estimate the parameters (β_0, β_1) from the normal equations :

$$\begin{cases} \sum_{i=1}^{n} (\hat{Y}_i^*(\beta_1) - \beta_0 - \beta_1 X_i) = 0, \\ \sum_{i=1}^{n} (\hat{Y}_i^*(\beta_1) - \beta_0 - \beta_1 X_i) X_i = 0. \end{cases} \tag{1.5.2}$$

A solution to these equations can be found in an iterative way. Ritov (1990) and Lai and Ying (1991a) obtained the asymptotic properties of a (slightly modified) version of this estimator.

Although this estimator behaves usually well in practice, there are a number of disadvantages. First, the iterative procedure suffers in certain cases from convergence problems which lead to unstable solutions or no solution at all. Second, the estimation method restricts to homoscedastic models, while in practice the data often follow a heteroscedastic model.

In light of these drawbacks, we propose in this thesis a variant of the Buckley-James procedure, which does not suffer from the above disadvantages. The idea is to estimate $E(Y_i|Y_i > C_i, X_i)$ (and hence Y_i^*) in a nonparametric way using the heteroscedastic regression model

$$Y = \bar{m}(X) + \bar{\sigma}(X)\bar{\varepsilon}, \tag{1.5.3}$$

where $\bar{m}(\cdot)$ is a general unknown location function, $\bar{\sigma}(\cdot)$ is a general unknown scale function, and $\bar{\varepsilon}$ is independent of X. This model generalizes (1.5.1) and $E(Y_i|Y_i > C_i, X_i)$ can therefore be rewritten

$$E(Y_i|Y_i > C_i, X_i) = \bar{m}(X_i) + \frac{\bar{\sigma}(X_i)}{1 - F_e(\bar{E}_i)} \int_{\bar{E}_i}^{\infty} y \, dF_e(y), \tag{1.5.4}$$

where $F_e(\cdot)$ is the distribution of $\frac{Y-\bar{m}(X)}{\bar{\sigma}(X)}$ and $\bar{E} = \frac{Z-\bar{m}(X)}{\bar{\sigma}(X)}$. Particular choices $m^0(\cdot)$ and $\sigma^0(\cdot)$ for $\bar{m}(\cdot)$ and $\bar{\sigma}(\cdot)$ will be introduced in Section 2.1. The preliminary estimation of specified $m^0(\cdot)$ and $\sigma^0(\cdot)$ is done using kernel smoothing and $F_e(\cdot)$ is estimated with all the residuals $E_i, i = 1, \ldots n$. The whole method is achieved with an adaptively chosen bandwidth parameter. The advantage of this is that, contrary to the Buckley-James procedure, the so-obtained 'synthetic' data points do not depend on the unknown β-vector and hence the normal equations have an explicit (non-iterative) solution. As will be seen in the simulations (see Chapter 2), this leads to more stable solutions and hence to a smaller variance.

The method of Miller (1976) also suffers from convergence problems due to an iterative procedure and restricts to homoscedastic models. Moreover, a sufficient but restrictive condition to ensure the consistency of the slope estimator is that the censoring distribution shifts with the regression line just as the distribution of the dependent variable does. Stute (1993, 1996) and Akritas (1994) have the disadvantage that their asymptotic bias increases as the censoring in the upper tails increases. Moreover, the assumptions used in the asymptotic theory of Stute's (1993,1996) estimator are too restrictive and often unrealistic (see

also Section 1.6). For homoscedastic models, Koul, Susarla and Van Ryzin (1981) and Leurgans (1987) construct 'synthetic' data points without using any specific relation between the response and the covariate. Moreover, the structure of the new data points of Koul, Susarla and Van Ryzin involves a large variance of the slope estimator. Akritas (1996) also considers homoscedastic models and replaces each (censored or uncensored) data point by a local completely nonparametric estimation of its location function.

Apart from least squares methods, some other techniques have been proposed in the statistical literature for fitting linear regression models with censored data. M-estimators or rank procedures have been used by e.g. Tsiatis (1990), Wei, Ying and Lin (1990), Lai and Ying (1991b, 1994), Zhou (1992b), Ying (1993) or Yang (1997) to obtain more robust estimators. A number of other approaches have also been proposed by LeBlanc and Crowley (1995), Lin and Wei (1992b), Fygenson and Ritov (1994), Ying, Jung and Wei (1995)...

In the next chapter, we will describe the estimation procedure in detail. An asymptotic representation for the estimators of the regression parameters is established. As a consequence, the asymptotic normality of those estimators follows. Next, we carry out a simulation study, in which the new procedure is compared with the Buckley-James method. Simulations show that the proposed estimators have usually smaller variance and smaller mean squared error than the Buckley-James estimators. Finally, the two data sets on larynx cancer and on spectral energy distributions of quasars are analysed by means of the two methods. The results of Chapter 2 can be found in Heuchenne and Van Keilegom (2004).

1.6 Estimation in nonlinear censored regression models

Consider the heteroscedastic regression model

$$Y = m_{\theta_0}(X) + \sigma(X)\varepsilon, \tag{1.6.1}$$

where $\sigma^2(\cdot) = \mathrm{Var}(Y|\cdot)$, $m_{\theta_0}(\cdot) = E(Y|\cdot)$ is the regression curve, known up to a parameter vector $\theta \in \Theta$ with true unknown value θ_0, Θ is a compact subset of $\mathbb{R}^d$, and the error term ε (with unknown distribution $F_\varepsilon(\cdot)$) is independent of the (one-dimensional) covariate X. As in Section 1.5, Y is subject to random right censoring and C is independent of

Y, conditionally on X. Let $(Y_i, C_i, X_i, Z_i, \Delta_i)$ $(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z, Δ).

We are interested in the estimation of the parameter vector θ_0. Here also, we consider extensions to censored data of the classical least squares procedure but, here, for nonlinear regression. When the regression function is polynomial, this estimation problem has been studied extensively in the literature (see Section 1.5). However, when the regression curve belongs to some parametric, but nonlinear family of regression functions, much less research has been developed. Stute (1999) proposed the following estimation procedure, which is an extension of his earlier paper (Stute (1993)) on linear regression : minimize

$$\sum_{i=1}^{n} W_{in}\{Z_{(i)} - m_\theta(X_{(i)})\}^2 \tag{1.6.2}$$

with respect to θ, where

$$W_{in} = \frac{\Delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\Delta_{(j)}}$$

is the jump size of the bivariate empirical distribution function $\hat{F}(x, y)$ proposed by Stute (1993) :

$$\hat{F}(x, y) = \sum_{i=1}^{n} W_{in} I(X_{(i)} \leq x, Z_{(i)} \leq y), \tag{1.6.3}$$

$Z_{(1)}, \ldots, Z_{(n)}$ are the order statistics of $Z_1, \ldots, Z_n$, and $X_{(i)}$ and $\Delta_{(i)}$ are the corresponding covariate and censoring indicator. The method is very easy to implement in practice and, unlike many of the estimation procedures for polynomial regression, it is not based on a, sometimes delicate, tuning or bandwidth parameter. A major drawback of this estimation procedure however, is that it assumes that

1. Y and C are independent (unconditionally on X) and that
2. $P(Y \leq C|X, Y) = P(Y \leq C|Y)$, which is satisfied when e.g. C is independent of X.

Both assumptions are often violated in practice. Another drawback lies in the little amount of information used in the estimation procedure. Indeed, since the relation (1.6.1) between

Y and X is known, this should be used, at least partially, when constructing an estimator of the bivariate distribution like (1.6.3).

To overcome those drawbacks, we propose, in Chapter 3, to extend the estimation procedure of Chapter 2 to the nonlinear case. New data points are therefore constructed preserving the conditional mean $m_{\theta_0}(X)$ ($E[Y^*|X] = m_{\theta_0}(X)$) and θ_0 is then estimated by minimizing the least squares criterion (for completely observed data) applied to those new data points.

After having explained some details about the estimation procedure, consistency and asymptotic normality (following from an asymptotic representation) of the proposed estimator are established. A simulation study is then carried out, in which the new procedure is compared with the method of Stute (1999). In most cases, the new estimation procedure enables to outperform Stute's (1999) method. Finally, the method is applied to the analysis of the relationship between fatigue life of metal and applied stress. The results of Chapter 3 can be found in Heuchenne and Van Keilegom (2005a).

1.7 Estimation in nonparametric censored regression

Consider a random vector (X, Y), where X is a one-dimensional covariate and Y is possibly right censored by a random variable C (independent of Y conditionally on X). Let $(Y_i, C_i, X_i, Z_i, \Delta_i)$ $(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z, Δ). Suppose Y follows the heteroscedastic model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.7.1}$$

where $m(X)$ and $\sigma(X)$ are some unknown but smooth location and scale functions and the error term ε is independent of X. So, we assume that the conditional distribution of Y given X depends on X only via its first and second conditional moment. Model (1.7.1) has been studied extensively in the literature on censored data; see e.g. Fan and Gijbels (1994), Van Keilegom and Akritas (1999), Einmahl and Van Keilegom (2004), Neumeyer et al. (2004), Chen, Dahl and Kahn (2005).

It is well known that any location function $m(x)$ that involves the right tail of the conditional distribution $F(\cdot|x) = P(Y \leq \cdot|X = x)$ of Y given $X = x$ (like the conditional mean $E(Y|X = x) = \int y\, dF(y|x)$) cannot be estimated in a consistent way in a completely nonparametric model, due to the presence of right censoring. Indeed, as seen in Section 1.4, the completely nonparametric (kernel) estimator $\tilde{F}(\cdot|x)$ of Beran (1981) is inconsistent in the right tail.

In Chapter 4, we propose an estimator which enables to reduce considerably the inconsistency problem in the right tails of the estimated conditional distribution functions. This is achieved by using the weak model assumption (1.7.1). The method we propose applies to any L-functional of the type (see e.g. Serfling (1980), page 265) :

$$m(x) = a_0 \int_0^1 F^{-1}(s|x)L(s)\, ds + \sum_{j=1}^{k} a_j F^{-1}(s_j|x), \tag{1.7.2}$$

where $F^{-1}(s|x) = \inf\{y; F(y|x) \geq s\}$ is the quantile function of Y given x, $L(s)$ is a given weight function satisfying $\int_0^1 L(s)ds = 1$, $k \geq 0$, $a_0, \ldots, a_k$ are real numbers such that $\sum_{j=0}^{k} a_j = 1$, and $0 \leq s_1, \ldots, s_k \leq 1$. This definition of $m(x)$ includes a very broad class of common location functions. For example, when $L \equiv 1$, $a_0 = 1$ and $k = 0$, $m(x)$ equals the conditional mean and when $a_0 = 0, k = 1, a_1 = 1$ and $s_1 = 1/2$, we obtain the conditional median.

The proposed method is as follows. First, we estimate the conditional distribution $F(y|x)$ under model (1.7.1). This estimator must be consistent for a large range of values of y (larger than the range obtained with a completely nonparametric estimator of Beran, 1981). Second, we plug-in the obtained estimator in (1.7.2). To estimate $F(y|x)$, note that under model (1.7.1), $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$ is independent of X for any location function $m^0(X)$ and scale function $\sigma^0(X)$ (for a formal definition of location and scale functions, see Section 2.1). Hence,

$$F(y|x) = P\Big(\varepsilon^0 \leq \frac{y - m^0(x)}{\sigma^0(x)}\Big|X = x\Big) = F_\varepsilon^0\Big(\frac{y - m^0(x)}{\sigma^0(x)}\Big), \tag{1.7.3}$$

where F_ε^0 is the distribution of ε^0. The idea is now to choose m^0 and σ^0 in such a way that they can be estimated consistently, i.e. choose location and scale functions that do not make use of the right tail of the distribution of Y given X. We then estimate $F(y|x)$ by

replacing $m^0(\cdot)$, $\sigma^0(\cdot)$ and $F_\varepsilon^0(\cdot)$ by appropriate estimators. It is easy to see that, provided there is a region of the covariate space where censoring is light, the so-obtained estimator of $F(\cdot|x)$ behaves well in the right tail (see Van Keilegom and Akritas, 1999). Hence, the estimator of $m(x)$ based on the latter estimator of $F(\cdot|x)$ will outperform the completely nonparametric estimator.

The estimation of the conditional quantile or mean function with censored data has been studied extensively in the literature. Dabrowska (1987, 1992b), Van Keilegom and Veraverbeke (1997b, 1998), Chen, Dahl and Kahn (2005), among others, studied the nonparametric estimation of the conditional quantile function, whereas Powell (1986), Buchinski and Hahn (1998) and Portnoy (2003) estimated this function under the assumption of a parametric model. For the estimation of the conditional mean function, Doksum and Yandell (1982), Dabrowska (1987), Fan and Gijbels (1994), Kim and Truong (1998) and Cai and Hong (2003) used a nonparametric approach, whereas a large number of other papers, including e.g. Buckley and James (1979), Akritas (1994), Heuchenne and Van Keilegom (2004) assumed a polynomial model for the regression function.

Chapter 4 starts with some details about the estimation procedure. Next, the uniform consistency of the proposed estimator is established and an asymptotic representation is provided. This involves the asymptotic normality of the estimator for a fixed value of the covariate. A simulation study is then carried out in order to compare the finite sample behaviour of the new estimator with the corresponding completely nonparametric estimator of Beran (1981). Finally, the two estimators are used to analyse the data set on spectral energy distributions of quasars (see Section 1.1). The results of Chapter 4 can be found in Heuchenne and Van Keilegom (2005b).

1.8 Mean preservation in nonparametric censored regression

In Chapter 6, we will propose a second new estimator of any L-functional of the type (1.7.2) when censored data are present. This one is also based on model assumption (1.7.1) but at a lower level than the one obtained in Chapter 4. Moreover, it can be seen as a weighted average of 'synthetic' data points where the weights are of the Nadaraya-Watson

type (1.3.2). For this reason, asymptotic properties of a weighted average of general functions of the data points are established in Chapter 5.

In some situations, the estimator proposed in Chapter 4 suffers from a problem of stability due to the estimation procedure. This problem will be largely described by simulations in Chapter 6. In fact, it will be shown that a too extensive use of the heteroscedastic model (global information) reduces the amount of local information. That can sometimes deteriorate the estimation of $m(\cdot)$ along the whole support of the covariate. A possible alternative is therefore to reduce the use of model (1.7.1) in order to preserve at least uncensored local information. However, we also need to preserve the estimation of $F_{\varepsilon}^{0}(\cdot)$ since it behaves well in the right tails.

The idea of the proposed method is as follows. First, we estimate a given function of the true unknown survival time by making use of model (1.7.1), and then we approximate $m(x)$ in (1.7.2) by using a kernel estimator based on this estimated function of the survival time. More precisely, note that $m(x)$ can be written as

$$m(x) = a_0 E[YL(F(Y|x))|x] + \sum_{j=1}^{k} a_j F^{-1}(s_j|x),$$

and that $F(y|x) = E[I(Y \leq y)|x]$. Let $\phi_1(y|x) = yL(F(y|x))$ and $\phi_{2t}(y|x) = \phi_2(y|x) = I(y \leq t)$ for fixed t. The idea is now to replace $E[\phi_j(Y|x)|x]$ $(j = 1, 2)$ by a kernel estimator of the type $\sum_{i=1}^{n} W_i(x, a_n)\phi_j^*(Z_i, \Delta_i|x)$, where $W_i(x, a_n)$ are the Nadaraya-Watson weights (1.3.2) and $\phi_j^*(Z_i, \Delta_i|x)$, $i = 1, \ldots, n$, are chosen in such a way that $E[\phi_j^*(Z_i, \Delta_i|x)|x] = E[\phi_j(Y_i|x)|x]$. It is easy to check that this preservation of means is obtained for

$$\begin{aligned}\phi_j^*(z, \delta|x) &= \phi_j(z|x)\delta + E[\phi_j(Y|x)|Y > z, x](1 - \delta) \\ &= \phi_j(z|x)\delta + \frac{1}{1 - F(z|x)} \int_z^{+\infty} \phi_j(y|x) dF(y|x)(1 - \delta), \end{aligned} \tag{1.8.1}$$

where $F(y|x)$ will be estimated using model (1.7.1). Therefore, the method uses model (1.7.1) only in the second term on the right hand side of the expression above and does not use it in the construction of the estimator itself. So, in a sense, it is little sensitive to the validity of model (1.7.1), and it can be expected that it works well, even in situations where model (1.7.1) does not hold.

In Chapter 5, two asymptotic results for the weighted sum of general functions of the data points are stated. First, we introduce some popular examples (for complete and censored data) for which those results are applicable. Next, we show the strong uniform consistency and establish a strong uniform modulus of continuity result for this kernel estimator.

Chapter 6 starts by describing the estimation procedure in detail. Next, we state the asymptotic properties of the estimators of $m(\cdot)$ and $F(\cdot|\cdot)$. The uniform consistency of the estimators of $m(\cdot)$ and $F(\cdot|\cdot)$ and a uniform modulus of continuity result for the estimator of $F(\cdot|\cdot)$ are mainly obtained by using the asymptotic theory developed in Chapter 5. Asymptotic representations (involving asymptotic normalities) for the proposed estimators are also established. We then achieve a simulation study, in which the two new estimators of $m(\cdot)$ and the corresponding completely nonparametric estimator are compared. Finally, we add the new estimation procedure to the analysis of the data set on spectral energy distributions of quasars. The results of Chapter 5 and 6 can be found in Heuchenne (2005) and Heuchenne and Van Keilegom (2005c).

Chapter 2

Polynomial regression with censored data based on preliminary nonparametric estimation

We will focus in this chapter on the estimation of the regression parameters when the response is possibly right censored and model (1.5.1) is assumed. As already mentioned in Section 1.5, we propose a new method which is applicable to the heteroscedastic case and enables to avoid iterative procedures. It consists of a usual least squares minimization problem using new data points obtained from a nonparametric preliminary estimation based on model (1.5.3). The proposed method will be described in Section 2.1 as well as the assumptions under which the main results are valid. An asymptotic representation of the proposed estimator from which the asymptotic normality follows will be derived in Section 2.2. Section 2.3 shows via a simulation study how the preliminary kernel smoothing enables (through the choice of a bandwidth parameter) to outperform the Buckley-James procedure. Larynx cancer and quasars data will then be analysed in Section 2.4 and the estimation of the distribution of the proposed estimator will be obtained by means of a bootstrap procedure. Finally, a result on the mean of the absolute residuals (subject to censoring), which is needed in the proof of Theorem 2.2.1, will be established in the appendix of this chapter. The content of this chapter can also be found in Heuchenne and Van Keilegom (2004).

2.1 Notations and description of the method

As already outlined in Section 1.5, the proposed method is divided in two steps : first to estimate the unknown survival times of the censored observations, and second to apply the standard least squares procedure on the so-obtained artificial data points. The first step is based on the definition

$$Y_i^* = Y_i\Delta_i + E[Y_i|Y_i > C_i, X_i](1-\Delta_i),$$

for which

$$E(Y_i|X_i) = E(Y_i^*|X_i) = \beta_0 + \beta_1 X + \ldots + \beta_p X^p.$$

Hence, we can work in the sequel with the variable Y_i^* instead of with Y_i. In order to estimate Y_i^* for a censored observation, we first need to introduce a number of notations.

Let $m^0(\cdot)$ be any location function and $\sigma^0(\cdot)$ be any scale function, meaning that $m^0(x) = T(F(\cdot|x))$ and $\sigma^0(x) = S(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x))$, for all $a \geq 0$ and $b \in I\!R$ ($F_{aY+b}(\cdot|x)$ denotes here the conditional distribution of $aY+b$, given $X = x$). Let $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$. Then, it can be easily seen that if model (1.5.1) holds (i.e. ε is independent of X), then ε^0 is also independent of X.

Define $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H(y) = P(Z \leq y)$, $H_\delta(y|x) = P(Z \leq y, \Delta = \delta|x)$, $F_X(x) = P(X \leq x)$, $F_\varepsilon^0(y) = P(\varepsilon^0 \leq y)$, $S_\varepsilon^0(y) = 1 - F_\varepsilon^0(y)$, for $E^0 = (Z - m^0(X))/\sigma^0(X)$, we denote $H_\varepsilon^0(y) = P(E^0 \leq y)$, $H_{\varepsilon\delta}^0(y) = P(E^0 \leq y, \Delta = \delta)$, $H_\varepsilon^0(y|x) = P(E^0 \leq y|x)$, $H_{\varepsilon\delta}^0(y|x) = P(E^0 \leq y, \Delta = \delta|x)$ ($\delta = 0, 1$) and for $C^0 = (C - m^0(X))/\sigma^0(X)$, we denote $G_\varepsilon^0(y) = P(C^0 \leq y)$. The probability density functions of the distributions defined above will be denoted with lower case letters, and R_X will denote the support of the variable X.

It is easily seen that

$$Y_i^* = Y_i\Delta_i + [m^0(X_i) + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^0)} \int_{E_i^0}^{\infty} y\, dF_\varepsilon^0(y)](1-\Delta_i)$$

for any location function $m^0(\cdot)$ and scale function $\sigma^0(\cdot)$. The idea is now to choose m^0 and σ^0 in such a way that they can be estimated consistently. As it is well known, the right tail of the distribution $F(y|\cdot)$ cannot be estimated in a consistent way due to the presence of right censoring. Therefore, we work with the following choices of m^0 and σ^0 :

$$m^0(x) = \int_0^1 F^{-1}(s|x)J(s)\,ds, \quad \sigma^{02}(x) = \int_0^1 F^{-1}(s|x)^2 J(s)\,ds - m^{02}(x), \quad (2.1.1)$$

where $J(s)$ is a given score function satisfying $\int_0^1 J(s)\,ds = 1$. When $J(s)$ is chosen appropriately (which means put to zero in the right tail, where the quantile function cannot be estimated in a consistent way due to the right censoring), $m^0(x)$ and $\sigma^0(x)$ can be estimated consistently. Now, replace the distribution $F(y|x)$ in (2.1.1) by the Beran (1981) estimator, defined by

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \le y, \Delta_i = 1} \Big\{1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \ge Z_i) W_j(x, a_n)}\Big\}, \quad (2.1.2)$$

where $W_i(x, a_n)$ are defined as in (1.3.2). Also define

$$\hat{m}^0(x) = \int_0^1 \tilde{F}^{-1}(s|x)J(s)\,ds, \quad \hat{\sigma}^{02}(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 J(s)\,ds - \hat{m}^{02}(x) \quad (2.1.3)$$

as estimators for $m^0(x)$ and $\sigma^{02}(x)$. Next, let

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}_{(i)}^0 \le y, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1}\right), \quad (2.1.4)$$

denote the Kaplan-Meier type estimator of F_ε^0, where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_{(i)}^0$ is the i-th order statistic of $\hat{E}_1^0, \ldots, \hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999). This leads to the following estimator of Y_i^* :

$$\hat{Y}_{Ti}^* = Y_i\Delta_i + \Big\{\hat{m}^0(X_i) + \frac{\hat{\sigma}^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} \int_{\hat{E}_i^{0T}}^{\hat{S}_{X_i}} y\, d\hat{F}_\varepsilon^0(y)\Big\}(1 - \Delta_i), \quad (2.1.5)$$

where $\hat{S}_{X_i} = (T_{X_i} - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_i^{0T} = \hat{E}_i^0 \wedge \hat{S}_{X_i}$, and for any x, $T_x \le T\sigma^0(x) + m^0(x)$, where $T < \tau_{H_\varepsilon^0}$ and $\tau_F = \inf\{y : F(y) = 1\}$ for any distribution F. Note that due to the

right censoring, we have to truncate the integral in the definition of $\hat{Y}^*_{Ti}$. However, when $\tau_{F^0_\varepsilon} \leq \tau_{G^0_\varepsilon}$, the bound $\hat{S}_{X_i}$ can be chosen arbitrarily close to $\tau_{F^0_\varepsilon}$ for n sufficiently large.

Finally, define the estimator of $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ by the usual least squares estimator based on the pairs $(X_i, \hat{Y}^*_{Ti})$ $(i = 1, \ldots, n)$ and denote this estimator by $\hat{\boldsymbol{\beta}}_T = (\hat{\beta}_{T0}, \ldots, \hat{\beta}_{Tp})'$. As it is clear from the definition of $\hat{Y}^*_{Ti}$, $\hat{\beta}_{T0}, \ldots, \hat{\beta}_{Tp}$ are actually estimating

$$\boldsymbol{\beta}_T = (\beta_{T0}, \ldots, \beta_{Tp})' = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}' E(Y^*_{Ti}|X_i)_{i=1}^n$$

(conditionally on $X_1, \ldots, X_n$), where the element (i, j) of the matrix $\mathcal{X}$ equals X_i^{j-1} $(i = 1, \ldots, n; j = 1, \ldots, p+1)$,

$$Y^*_{Ti} = Y_i\Delta_i + \Big\{m^0(X_i) + \frac{\sigma^0(X_i)}{1 - F^0_\varepsilon(E_i^{0T})} \int_{E_i^{0T}}^{S_{X_i}} y\, dF^0_\varepsilon(y)\Big\}(1 - \Delta_i),$$

$S_{X_i} = (T_{X_i} - m^0(X_i))/\sigma^0(X_i)$ and $E_i^{0T} = (Z_i \wedge T_{X_i} - m^0(X_i))/\sigma^0(X_i) = E_i^0 \wedge S_{X_i}$. As above, these coefficients $\beta_{T0}, \ldots, \beta_{Tp}$ can be made arbitrarily close to $\beta_0, \ldots, \beta_p$, provided $\tau_{F^0_\varepsilon} \leq \tau_{G^0_\varepsilon}$.

The following functions will be needed in the statement of the asymptotic results given in the next section :

$$\xi_\varepsilon(z, \delta, y) = (1 - F^0_\varepsilon(y)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH^0_{\varepsilon 1}(s)}{(1 - H^0_\varepsilon(s))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H^0_\varepsilon(z)} \right\},$$

$$\xi(z, \delta, y|x) = (1 - F(y|x)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H(z|x)} \right\},$$

$$\eta(z, \delta|x) = \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x))\, dv\, \sigma^0(x)^{-1},$$

$$\zeta(z, \delta|x) = \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) \frac{v - m^0(x)}{\sigma^0(x)}\, dv\, \sigma^0(x)^{-1},$$

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h^0_\varepsilon(s|x)}{(1 - H^0_\varepsilon(s))^2}\, dH^0_{\varepsilon 1}(s) + \int_{-\infty}^{y} \frac{d\, h^0_{\varepsilon 1}(s|x)}{1 - H^0_\varepsilon(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{sh^0_\varepsilon(s|x)}{(1 - H^0_\varepsilon(s))^2}\, dH^0_{\varepsilon 1}(s) + \int_{-\infty}^{y} \frac{d\,(sh^0_{\varepsilon 1}(s|x))}{1 - H^0_\varepsilon(s)},$$

$$\varphi(x,z,\delta,y)=\xi_\varepsilon\left(\frac{z-m^0(x)}{\sigma^0(x)},\delta,y\right)-S^0_\varepsilon(y)\eta(z,\delta|x)\gamma_1(y|x)-S^0_\varepsilon(y)\zeta(z,\delta|x)\gamma_2(y|x),$$

$$\alpha_i(v)=\frac{\int_v^{S_i}u\,dF^0_\varepsilon(u)}{1-F^0_\varepsilon(v)},$$

$$B_k(z,Z_j,\Delta_j|X_i)$$

$$=X_i^k f_X^{-1}(X_i)\sigma^0(X_i)\left\{\left[\alpha_i'(e_i^{0T}(z))-1+\frac{S_i f^0_\varepsilon(S_i)}{1-F^0_\varepsilon(e_i^{0T}(z))}\right]\eta(Z_j,\Delta_j|X_i)\right.$$

$$\left.+\left[e_i^{0T}(z)\alpha_i'(e_i^{0T}(z))-\alpha_i(e_i^{0T}(z))+\frac{S_i^2 f^0_\varepsilon(S_i)}{1-F^0_\varepsilon(e_i^{0T}(z))}\right]\zeta(Z_j,\Delta_j|X_i)\right\},$$

$k=0,\dots,p$, $i,j=1,\dots,n$, where $S_i=S_{X_i}$, $e_i^{0T}(z)=e_{X_i}^{0T}(z)$ and for any $x\in R_X$, $S_x=(T_x-m^0(x))/\sigma^0(x)$, $z_x=(z-m^0(x))/\sigma^0(x)$ and $e_x^{0T}(z)=z_x\wedge S_x$.

The assumptions required to prove the main results of Section 2.2 are enumerated hereafter. In order to formulate them, we will call for a few more notations. Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x\in R_X}(1-H(\tilde{T}_x|x))>0$. For a (sub)distribution function $L(y|x)$, we will use the notations $l(y|x)=L'(y|x)=(\partial/\partial y)L(y|x)$, $\dot{L}(y|x)=(\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives.

$(A1)(i)$ $na_n^4\to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1}\to\infty$ for some $\delta<1/2$.
(ii) R_X is a compact interval.
(iii) K is a symmetric density with compact support and K is twice continuously differentiable.
(iv) $\det(M)\neq 0$, where $M=(M_{jk})$ $(j,k=1,\dots,p+1)$, $M_{jk}=E(X^{j+k-2})$.

$(A2)(i)$ There exist $0\le s_0\le s_1\le 1$ such that $s_1\le\inf_x F(\tilde{T}_x|x)$, $s_0\le\inf\{s\in[0,1];J(s)\neq 0\}$, $s_1\ge\sup\{s\in[0,1];J(s)\neq 0\}$ and $\inf_{x\in R_X}\inf_{s_0\le s\le s_1}f(F^{-1}(s|x)|x)>0$.
(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds=1$ and $J(s)\ge 0$ for all $0\le s\le 1$.
(iii) The function $x\to T_x$ $(x\in R_X)$ is twice continuously differentiable.

$(A3)(i)$ F_X is three times continuously differentiable and $\inf_{x\in R_X}f_X(x)>0$.
(ii) m^0 and σ^0 are twice continuously differentiable and $\inf_{x\in R_X}\sigma^0(x)>0$.
(iii) $E[\varepsilon^{02}]<\infty$ and $E[E^{04}]<\infty$.

$(A4)(i)$ $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X$, $z < \tilde{T}_x$ and δ.
(ii) The first derivatives of $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ with respect to z are of bounded variation and the variation norms are uniformly bounded over all x.

$(A5)$ The function $y \to P(m^0(X) + e\sigma^0(X) \leq y)$ $(y \in I\!R)$ is differentiable for all $e \in I\!R$ and the derivative is uniformly bounded over all $e \in I\!R$.

$(A6)$ For $L(y|x) = H(y|x), H_1(y|x), H^0_\varepsilon(y|x)$ or $H^0_{\varepsilon 1}(y|x)$: $L'(y|x)$ is continuous in (x,y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$. The same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three, and $\sup_{x,y} |y^3 L'''(y|x)| < \infty$.

$(A7)(i)$ $\sup_{x,z} \int |B'_k(t,z,\delta|x)| h(t)\, dt < \infty$ $(k = 1,\ldots,d;\ \delta = 0,1)$.
(ii) $\sup_z \int \sup_x |B''_k(t,z,\delta|x)| h(t)\, dt < \infty$ $(k = 1,\ldots,d;\ \delta = 0,1)$, where $B_k^{'(')}(t,z,\delta|x)$ equals the first (second) derivative of $B_k(t,z,\delta|x)$ with respect to x when $t \neq T_x$ and equals 0 otherwise.

$(A8)$ For the density $f_{X|Z,\Delta}(x|z,\delta)$ of X given (Z,Δ), $\sup_{x,z} |f_{X|Z,\Delta}(x|z,\delta)| < \infty$, $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$ and $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$ $(\delta = 0,1)$, where $\dot{f}_{X|Z,\Delta}(x|z,\delta)$ $(\ddot{f}_{X|Z,\Delta}(x|z,\delta))$ denotes the first (second) derivative of $f_{X|Z,\Delta}(x|z,\delta)$ with respect to x

2.2 Asymptotic results

We start by developing an asymptotic representation for $\hat{\beta}_{Tj} - \beta_{Tj}$ $(j = 0,\ldots,p)$. This representation will be useful to obtain afterwards the asymptotic normality of the estimators.

Theorem 2.2.1 *Assume (A1)–(A8), Then,*

$$\begin{pmatrix} \hat{\beta}_{T0} - \beta_{T0} \\ \vdots \\ \hat{\beta}_{Tp} - \beta_{Tp} \end{pmatrix} = M^{-1} n^{-1} \sum_{i=1}^{n} \rho(X_i, Z_i, \Delta_i) + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

where $\rho = (\rho_0, \ldots, \rho_p)'$,

$$\begin{aligned} \rho_j(X_i, Z_i, \Delta_i) &= \int_{R_X} x^j \sigma^0(x) \int_{-\infty}^{+\infty} \left\{ \frac{\varphi(X_i, Z_i, \Delta_i, e_x^{0T}(z))}{(1 - F_\varepsilon^0(e_x^{0T}(z)))^2} \int_{e_x^{0T}(z)}^{S_x} u \, dF_\varepsilon^0(u) \right. \\ &\quad \left. + \frac{1}{1 - F_\varepsilon^0(e_x^{0T}(z))} \int_{e_x^{0T}(z)}^{S_x} u \, d\varphi(X_i, Z_i, \Delta_i, u) \right\} dH_0(z|x) dF_X(x) \\ &\quad + f_X(X_i) \int B_j(z, Z_i, \Delta_i | X_i) \, dH_0(z|X_i) + X_i^j (Y_{Ti}^* - E[Y_{Ti}^* | X_i]) \end{aligned}$$

$(j = 0, \ldots, p; i = 1, \ldots, n)$.

Proof. Let $\beta_T^* = (\beta_{T0}^*, \ldots, \beta_{Tp}^*)'$ be the least squares estimator obtained from the pairs (X_i, Y_{Ti}^*) $(i = 1, \ldots, n)$. We will first consider

$$\hat{\beta}_T - \beta_T^* = (n^{-1} \mathcal{X}' \mathcal{X})^{-1} n^{-1} \mathcal{X}' (\hat{\mathcal{Y}}^* - \mathcal{Y}^*),$$

where $\mathcal{Y}^* = (Y_{T1}^*, \ldots, Y_{Tn}^*)'$, and $\hat{\mathcal{Y}}^* = (\hat{Y}_{T1}^*, \ldots, \hat{Y}_{Tn}^*)'$. The $(k+1)^{th}$ element $(k = 0, \ldots, p)$ of the vector $n^{-1} \mathcal{X}' (\hat{\mathcal{Y}}^* - \mathcal{Y}^*)$ equals

$$\begin{aligned} & n^{-1} \sum_{\Delta_i = 0} X_i^k \left\{ [\hat{m}^0(X_i) - m^0(X_i)] \right. \\ & \left. + \frac{\hat{\sigma}^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} \int_{\hat{E}_i^{0T}}^{\hat{S}_i} u \, d\hat{F}_\varepsilon^0(u) - \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{E_i^{0T}}^{S_i} u \, dF_\varepsilon^0(u) \right\} \\ & = n^{-1} \sum_{\Delta_i = 0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\}, \end{aligned}$$

where $\hat{S}_i = \hat{S}_{X_i}$. The asymptotic representation given in Proposition 4.8 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA) yields

$$A_{1i} = -(n a_n)^{-1} f_X^{-1}(X_i) \sigma^0(X_i) \sum_{j=1}^{n} K(\frac{X_i - X_j}{a_n}) \eta(Z_j, \Delta_j | X_i) + o_P(n^{-1/2}),$$

uniformly in $i = 1, \ldots, n$. Next, write

$$n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} \tag{2.2.1}$$
$$= n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} I(E_i^0 \leq U_n) + n^{-1} \sum_{\Delta_i=0} X_i^k \{A_{1i} + A_{2i} + A_{3i}\} I(E_i^0 > U_n),$$

where $U_n < 0$ is defined by $U_n = -n^{1/2} a_n^{1+\gamma}$ for some $\gamma > 0$ to be determined later. First, let us show that the first sum of this expression is asymptotically negligible. Let V_n be the number of residuals E_i^0 that are less than or equal to U_n. Then, by the law of the iterated logarithm (see e.g. Serfling (1980), page 35),

$$V_n - n H_\varepsilon^0(U_n) \leq 2[H_\varepsilon^0(U_n)(1 - H_\varepsilon^0(U_n)) n \log\log n]^{1/2} \quad a.s.$$

Since $|U_n|^4 H_\varepsilon^0(U_n) \leq \int_{-\infty}^{U_n} |y|^4 \, dH_\varepsilon^0(y) \to 0$, it follows that $H_\varepsilon^0(U_n) \leq C_n |U_n|^{-4}$ for some sequence $C_n \to 0$. From this, we get $V_n = o(n|U_n|^{-4} + |U_n|^{-2} n^{1/2} (\log\log n)^{1/2})$ a.s. Next, $A_{1i} + A_{2i} + A_{3i}$ is bounded in probability, which follows from Lemma 2.5.1, the fact that $E|\varepsilon^0| < \infty$, the uniform consistency of $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ given by Proposition 4.5 in VKA (1999) and the consistency of $\sup_{x,z} |\hat{F}_\varepsilon^0(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x - m^0(x)}{\sigma^0(x)})|$ which is obtained as follows.

$$\begin{aligned}
&\hat{F}_\varepsilon^0(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x - m^0(x)}{\sigma^0(x)}) \\
&= \hat{F}_\varepsilon^0(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x - \hat{m}(x)}{\hat{\sigma}^0(x)}) \\
&\quad + F_\varepsilon^0(\frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x - m^0(x)}{\hat{\sigma}^0(x)}) \\
&\quad + F_\varepsilon^0(\frac{z \wedge T_x - m^0(x)}{\hat{\sigma}^0(x)}) - F_\varepsilon^0(\frac{z \wedge T_x - m^0(x)}{\sigma^0(x)}) \\
&= \alpha_n^1(z,x) + \alpha_n^2(z,x) + \alpha_n^3(z,x).
\end{aligned} \tag{2.2.2}$$

Using Corollary 3.2 in VKA (1999), $\sup_{x,z} |\alpha_n^1(z,x)|$ is $O_p(n^{-1/2})$. For the two other terms, we use two first order Taylor developments

$$\alpha_n^2(z,x) + \alpha_n^3(z,x) \quad = \quad -\frac{\hat{m}^0(x) - m^0(x)}{\hat{\sigma}^0(x)} f_\varepsilon^0(A_x) - \frac{\hat{\sigma}^0(x) - \sigma^0(x)}{\hat{\sigma}^0(x)} \frac{z \wedge T_x - m^0(x)}{\sigma^0(x)} f_\varepsilon^0(B_x),$$

for some A_x (B_x) between $\frac{z\wedge T_x - m^0(x)}{\hat\sigma^0(x)}$ and $\frac{z\wedge T_x - \hat m^0(x)}{\hat\sigma^0(x)}$ ($\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)}$ and $\frac{z\wedge T_x - m^0(x)}{\hat\sigma^0(x)}$). Using Proposition 4.5 of VKA (1999) and the fact that $\sup_e |e f_\varepsilon^0(e)| < +\infty$, $\alpha_n^2(z,x) + \alpha_n^3(z,x) = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ *a.s.* Therefore, the first term on the right hand side of (2.2.1) is $o_P(|U_n|^{-4}) = o_P(n^{-1/2})$ for γ small enough. We next consider the second term on the right hand side of (2.2.1). Write

$$\begin{aligned}
A_{2i} + A_{3i} = {} & \frac{\hat\sigma^0(X_i) - \sigma^0(X_i)}{1 - \hat F_\varepsilon^0(\hat E_i^{0T})} \int_{\hat E_i^{0T}}^{\hat S_i} u\, d\hat F_\varepsilon^0(u) \\
& + \sigma^0(X_i) \frac{\hat F_\varepsilon^0(\hat E_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - \hat F_\varepsilon^0(\hat E_i^{0T}))(1 - F_\varepsilon^0(E_i^{0T}))} \int_{\hat E_i^{0T}}^{\hat S_i} u\, d\hat F_\varepsilon^0(u) \\
& + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{\hat E_i^{0T}}^{E_i^{0T}} u\, d\hat F_\varepsilon^0(u) + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{E_i^{0T}}^{S_i} u\, d(\hat F_\varepsilon^0(u) - F_\varepsilon^0(u)) \\
& + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{S_i}^{\hat S_i} u\, d\hat F_\varepsilon^0(u) \\
= {} & \sum_{j=1}^{5} B_{ji}.
\end{aligned}$$

First consider

$$\begin{aligned}
\int_{\hat E_i^{0T}}^{\hat S_i} u\, d\hat F_\varepsilon^0(u) = {} & \int_{E_i^{0T}}^{S_i} u\, dF_\varepsilon^0(u) + \int_{E_i^{0T}}^{S_i} u\, d(\hat F_\varepsilon^0(u) - F_\varepsilon^0(u)) \\
& + \int_{\hat E_i^{0T}}^{E_i^{0T}} u\, d\hat F_\varepsilon^0(u) + \int_{S_i}^{\hat S_i} u\, d\hat F_\varepsilon^0(u),
\end{aligned} \tag{2.2.3}$$

which appears in B_{1i} and B_{2i}. By using integration by parts, the third term of (2.2.3) can be rewritten as

$$\begin{aligned}
& [E_i^{0T}(\hat F_\varepsilon^0(E_i^{0T}) - F_\varepsilon^0(E_i^{0T}))] + [E_i^{0T} F_\varepsilon^0(E_i^{0T}) - (\hat E_i^{0T}) F_\varepsilon^0(E_i^{0T})] \\
& + [(\hat E_i^{0T})(F_\varepsilon^0(E_i^{0T}) - \hat F_\varepsilon^0(\hat E_i^{0T}))] - \int_{\hat E_i^{0T}}^{E_i^{0T}} \hat F_\varepsilon^0(u) du.
\end{aligned} \tag{2.2.4}$$

By Corollary 3.2 in VKA (1999) and the order of U_n, the first term of (2.2.4) is $O_P(a_n^{1+\gamma})$, while from Proposition 4.5 in VKA (1999) it follows that the second and fourth terms are $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$. Using the fact that $\sup_{x,z} |\hat F_\varepsilon^0(\frac{z\wedge T_x - \hat m^0(x)}{\hat\sigma^0(x)}) - F_\varepsilon^0(\frac{z\wedge T_x - m^0(x)}{\sigma^0(x)})| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ yields that the order of the third term is $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$.

Hence, the third term of (2.2.3) is $O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$, uniformly in $i = 1, \ldots, n$. In a similar way, it can be shown that the second and fourth terms of (2.2.3) are of this order, which implies that

$$B_{1i} + B_{2i} = \frac{\hat{\sigma}^0(X_i) - \sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{E_i^{0T}}^{S_i} u \, dF_\varepsilon^0(u) + \sigma^0(X_i) \frac{\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - F_\varepsilon^0(E_i^{0T}))^2} \int_{E_i^{0T}}^{S_i} u \, dF_\varepsilon^0(u) + o_P(n^{-1/2}).$$

By applying the representation for $\hat{\sigma}^0(X_i) - \sigma^0(X_i)$ given by Proposition 4.9 in VKA (1999) and using the fact that

$$\begin{aligned} &\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T}) \\ &= (na_n)^{-1} \sum_{j=1}^n K(\frac{X_i - X_j}{a_n}) f_X^{-1}(X_i)[\eta(Z_j, \Delta_j | X_i) + \zeta(Z_j, \Delta_j | X_i) E_i^{0T}] f_\varepsilon^0(E_i^{0T}) \\ &\quad + n^{-1} \sum_{j=1}^n \varphi(X_j, Z_j, \Delta_j, E_i^{0T}) + o_P(n^{-1/2}), \end{aligned} \tag{2.2.5}$$

where this development is obtained after two Taylor expansions and by applying Theorem 3.1, Lemma B.1 and Propositions 4.8 and 4.9 in VKA (1999), $B_{1i} + B_{2i}$ can now be written as a sum of i.i.d. terms (up to the $o_P(n^{-1/2})$ remainder term). For B_{3i} write

$$\int_{\hat{E}_i^{0T}}^{E_i^{0T}} u \, d\hat{F}_\varepsilon^0(u) = \int_{\hat{E}_i^{0T}}^{E_i^{0T}} u \, dF_\varepsilon^0(u) + \int_{\hat{E}_i^{0T}}^{E_i^{0T}} u \, d(\hat{F}_\varepsilon^0(u) - F_\varepsilon^0(u)).$$

Through integration by parts of the second term of the above expression and combined use of Corollary 3.2 in VKA (1999) and the fact that $|\hat{E}_i^{0T} - E_i^{0T}| = O_P(a_n^{1/2+\gamma}(\log a_n^{-1})^{1/2})$, we obtain

$$E_i^{0T}[\hat{F}_\varepsilon^0(E_i^{0T}) - F_\varepsilon^0(E_i^{0T}) - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}) + F_\varepsilon^0(\hat{E}_i^{0T})] - \int_{\hat{E}_i^{0T}}^{E_i^{0T}} (\hat{F}_\varepsilon^0(u) - F_\varepsilon^0(u)) \, du + o_P(n^{-1/2}).$$

It is easy to see that the integral in this expression is also $o_P(n^{-1/2})$. As a consequence of Theorem 3.1 and Lemma B.1 in VKA (1999), the first term is $o_P(|E_i^{0T}| n^{-1/2})$. Hence,

$$\begin{aligned} B_{3i} &= -\frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \Big[\int_0^{\hat{E}_i^{0T}} u \, dF_\varepsilon^0(u) - \int_0^{E_i^{0T}} u \, dF_\varepsilon^0(u) \Big] + o_P(|E_i^{0T}| n^{-1/2}) \\ &= [\hat{m}^0(X_i) - m^0(X_i) + E_i^{0T}(\hat{\sigma}^0(X_i) - \sigma^0(X_i))] \frac{E_i^{0T} f_\varepsilon^0(E_i^{0T})}{1 - F_\varepsilon^0(E_i^{0T})} \\ &\quad + o_P(|E_i^{0T}| n^{-1/2}) + o_P(n^{-1/2}), \end{aligned}$$

using a Taylor expansion. Note that the term $o_P(n^{-1/2})$ in the above-mentioned expression is obtained from the fact that $\sup_z |z f_\varepsilon^0(z)| < \infty$ and $\sup_z |z^2 f_\varepsilon^{0\prime}(z)| < \infty$. Next, the term B_{4i} is given by

$$\frac{\sigma^0(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\Big\{S_i[\hat{F}_\varepsilon^0(S_i)-F_\varepsilon^0(S_i)] \\ -E_i^{0T}[\hat{F}_\varepsilon^0(E_i^{0T})-F_\varepsilon^0(E_i^{0T})]-\int_{E_i^{0T}}^{S_i}(\hat{F}_\varepsilon^0(u)-F_\varepsilon^0(u))\,du\Big\}.$$

Finally, the term B_{5i} is treated in the same way as the term B_{3i}, leading to

$$B_{5i} = -[\hat{m}^0(X_i)-m^0(X_i)+S_i(\hat{\sigma}^0(X_i)-\sigma^0(X_i))]\frac{S_i f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(E_i^{0T})} \\ +o_P(|S_i|n^{-1/2})+o_P(n^{-1/2}).$$

It now follows that the complete asymptotic representation for the $(k+1)^{th}$ component of $n^{-1}\mathcal{X}'(\hat{\mathcal{Y}}^*-\mathcal{Y}^*)$ can be written as

$$\begin{aligned}
&n^{-1}\sum_{\Delta_i=0} X_i^k I(E_i^0 > U_n) \\
&\times\Bigg\{\left[-\frac{E_i^{0T}f_\varepsilon^0(E_i^{0T})}{1-F_\varepsilon^0(E_i^{0T})}+\frac{f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^{S_i}u\,dF_\varepsilon^0(u)}{(1-F_\varepsilon^0(E_i^{0T}))^2}-1+\frac{S_i f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(E_i^{0T})}\right] \\
&\quad\times(na_n)^{-1}f_X^{-1}(X_i)\sigma^0(X_i)\sum_{j=1}^n K\Big(\frac{X_i-X_j}{a_n}\Big)\eta(Z_j,\Delta_j|X_i) \\
&\quad+\left[\frac{E_i^{0T}f_\varepsilon^0(E_i^{0T})\int_{E_i^{0T}}^{S_i}u\,dF_\varepsilon^0(u)}{(1-F_\varepsilon^0(E_i^{0T}))^2}-\frac{\int_{E_i^{0T}}^{S_i}u\,dF_\varepsilon^0(u)}{1-F_\varepsilon^0(E_i^{0T})}-\frac{(E_i^{0T})^2 f_\varepsilon^0(E_i^{0T})}{1-F_\varepsilon^0(E_i^{0T})}+\frac{S_i^2 f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(E_i^{0T})}\right] \\
&\quad\times(na_n)^{-1}f_X^{-1}(X_i)\sigma^0(X_i)\sum_{j=1}^n K\Big(\frac{X_i-X_j}{a_n}\Big)\zeta(Z_j,\Delta_j|X_i) \\
&+\sigma^0(X_i)\left[\frac{\int_{E_i^{0T}}^{S_i}u\,dF_\varepsilon^0(u)}{(1-F_\varepsilon^0(E_i^{0T}))^2}-\frac{E_i^{0T}}{1-F_\varepsilon^0(E_i^{0T})}\right]n^{-1}\sum_{j=1}^n\varphi(X_j,Z_j,\Delta_j,E_i^{0T}) \\
&+\frac{\sigma^0(X_i)S_i}{1-F_\varepsilon^0(E_i^{0T})}n^{-1}\sum_{j=1}^n\varphi(X_j,Z_j,\Delta_j,S_i) \\
&-\frac{\sigma^0(X_i)}{1-F_\varepsilon^0(E_i^{0T})}n^{-1}\sum_{j=1}^n\int_{E_i^{0T}}^{S_i}\varphi(X_j,Z_j,\Delta_j,u)du\Bigg\}+o_P(n^{-1/2}),
\end{aligned} \tag{2.2.6}$$

where use is made of the representation (2.2.5) and the representations for $\hat{F}^0_\varepsilon(\cdot), \hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ respectively given by Theorem 3.1, Propositions 4.8 and 4.9 in VKA (1999).
We can rewrite the sum of the first two terms of equation (2.2.6) as

$$(n^2 a_n)^{-1} \sum_{j \neq i} (1 - \Delta_i) I(E_i^0 > U_n) B_k(Z_i, Z_j, \Delta_j | X_i) K(\frac{X_i - X_j}{a_n}) + o_P(n^{-1/2}). \quad (2.2.7)$$

Using a similar development as for the first term of (2.2.1) can easily show that (2.2.7) can be written as

$$\begin{aligned}
&(n^2 a_n)^{-1} \sum_{j \neq i} (1 - \Delta_i) B_k(Z_i, Z_j, \Delta_j | X_i) K(\frac{X_i - X_j}{a_n}) + o_P(n^{-1/2}) \\
&= (n^2 a_n)^{-1} \sum_{j \neq i} \{A_k^*(V_i, V_j) + E[A_k(V_i, V_j)|V_i] + E[A_k(V_i, V_j)|V_j] - E[A_k(V_i, V_j)]\} \\
&\quad + o_P(n^{-1/2}) \\
&= T_1 + T_2 + T_3 + T_4 + o_P(n^{-1/2}),
\end{aligned}$$

where

$$A_k(V_i, V_j) = (1 - \Delta_i) B_k(Z_i, Z_j, \Delta_j | X_i) K(\frac{X_i - X_j}{a_n}),$$

$A_k^*(V_i, V_j) = A_k(V_i, V_j) - E[A_k(V_i, V_j)|V_i] - E[A_k(V_i, V_j)|V_j] + E[A_k(V_i, V_j)]$ and $V_i = (X_i, Z_i, \Delta_i)$. Consider

$$\begin{aligned}
&E[A_k(V_i, V_j)|V_i] \\
&= (1 - \Delta_i) \sum_{\delta=0,1} \int \int B_k(Z_i, z, \delta | X_i) K(\frac{X_i - x}{a_n}) h_\delta(z|x) f_X(x)\, dz\, dx \\
&= a_n (1 - \Delta_i) \sum_{\delta=0,1} \int \int B_k(Z_i, z, \delta | X_i) K(u) (h_\delta(z|X_i) - a_n u \dot{h}_\delta(z|X_i) + O(a_n^2)) \\
&\qquad \times (f_X(X_i) - a_n u f_X'(X_i) + O(a_n^2))\, dz\, du \\
&= a_n (1 - \Delta_i) f_X(X_i) \sum_{\delta=0,1} \int B_k(Z_i, z, \delta | X_i) h_\delta(z|X_i)\, dz + O(a_n^3) = O(a_n^3),
\end{aligned}$$

since $E[\eta(Z, \Delta|X)|X] = E[\zeta(Z, \Delta|X)|X] = 0$. Hence, we also have $E[A_k(V_i, V_j)] = O(a_n^3)$. In a similar way, using three Taylor expansions of order 2, we get

$$E[A_k(V_i, V_j)|V_j] = a_n f_X(X_j) \sum_{\delta=0,1} (1 - \delta) \int B_k(z, Z_j, \Delta_j | X_j)\, dH_\delta(z|X_j) + O(a_n^3).$$

It follows that

$$T_2 + T_3 + T_4 = n^{-1} \sum_{i=1}^{n} f_X(X_i) \int B_k(z, Z_i, \Delta_i | X_i)\, dH_0(z|X_i) + O(a_n^2).$$

Note that for T_1, $E[T_1] = 0$, resulting, by Chebyshev's inequality, in

$$\begin{aligned} & P(|T_1| > K(na_n)^{-1} E[A_k^*(V_1, V_2)^2]^{1/2}) \\ & \leq K^{-2}(na_n)^2 E[A_k^*(V_1, V_2)^2]^{-1} E[T_1^2] \\ & = K^{-2} n^{-2} E[A_k^*(V_1, V_2)^2]^{-1} \sum_{j \neq i} \sum_{m \neq l} E[A_k^*(V_i, V_j) A_k^*(V_l, V_m)]. \end{aligned} \tag{2.2.8}$$

Since $E[A_k^*(V_i, V_j)] = 0$, the terms for which $i, j \neq l, m$ are zero. The terms for which either i or j equals l or m and the other differs from l and m, are also zero, because, for example when $i = l$ and $j \neq m$,

$$E[A_k^*(V_i, V_j) E[A_k^*(V_i, V_m) | V_i, V_j]] = 0.$$

Thus, only the $2n(n-1)$ terms for which (i, j) equals (l, m) or (m, l) remain such that, (2.2.8) is bounded by $2K^{-2}$, which can be made arbitrarily small for K large enough. Since $A_k^*(V_1, V_2)$ is bounded by $K(\frac{X_1 - X_2}{a_n})C + O(a_n)$ for some constant $C > 0$, independent of X_1 and X_2, we have that

$$E[A_k^*(V_1, V_2)^2] \leq C^2 a_n \int f_X^2(x)\, dx \int K^2(u)\, du + O(a_n^2) = O(a_n).$$

It now follows that $T_1 = O_P(n^{-1} a_n^{-1/2}) = o_P(n^{-1/2})$. Let's consider then the third, fourth and fifth terms of (2.2.6). Their sum equals

$$\begin{aligned} & n^{-1} \sum_{\Delta_i = 0} I(E_i^0 > U_n) X_i^k \left\{ \frac{\int_{E_i^{0T}}^{S_i} u\, dF_\varepsilon^0(u)}{(1 - F_\varepsilon^0(E_i^{0T}))^2} n^{-1} \sigma^0(X_i) \sum_{j=1}^{n} \varphi(X_j, Z_j, \Delta_j, E_i^{0T}) \right. \\ & \left. + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})} \int_{E_i^{0T}}^{S_i} u\, n^{-1} \sum_{j=1}^{n} d\varphi(X_j, Z_j, \Delta_j, u) \right\} \\ & = n^{-2} \sum_{j \neq i} h_k(V_i, V_j) + o_P(n^{-1/2}), \end{aligned} \tag{2.2.9}$$

where

$$h_k(V_i,V_j) = (1-\Delta_i)X_i^k \left\{ \frac{\int_{E_i^{0T}}^{S_i} u\, dF_\varepsilon^0(u)}{(1-F_\varepsilon^0(E_i^{0T}))^2}\sigma^0(X_i)\varphi(X_j,Z_j,\Delta_j,E_i^{0T}) \right.$$
$$\left. + \frac{\sigma^0(X_i)}{1-F_\varepsilon^0(E_i^{0T})}\int_{E_i^{0T}}^{S_i} u\, d\varphi(X_j,Z_j,\Delta_j,u) \right\},$$

using arguments similar as before. Defining $h_k^*(V_i,V_j) = h_k(V_i,V_j) + h_k(V_j,V_i)$, (2.2.9) can be written as

$$\frac{n-1}{2n}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{j>i} h_k^*(V_i,V_j) + o_P(n^{-1/2}).$$

Using the Hájek-projection of a U-statistic on its conditional expectations (see e.g. Serfling (1980), page 189), this expression equals

$$n^{-1}\sum_{i=1}^n E[h_k^*(V_i,V_j)|V_i] + o_P(n^{-1/2})$$
$$= n^{-1}\sum_{i=1}^n \int x^k\sigma^0(x)\int \left\{ \frac{\varphi(X_i,Z_i,\Delta_i,e_x^{0T}(z))}{(1-F_\varepsilon^0(e_x^{0T}(z)))^2}\int_{e_x^{0T}(z)}^{S_x} u\, dF_\varepsilon^0(u) \right.$$
$$\left. + \frac{1}{1-F_\varepsilon^0(e_x^{0T}(z))}\int_{e_x^{0T}(z)}^{S_x} u\, d\varphi(X_i,Z_i,\Delta_i,u) \right\} dH_0(z|x)dF_X(x) + o_P(n^{-1/2}).$$

It remains to consider $\beta_T^* - \beta_T$, which equals

$$M^{-1}\begin{pmatrix} n^{-1}\sum_{i=1}^n (Y_{Ti}^* - E[Y_{Ti}^*|X_i]) \\ \vdots \\ n^{-1}\sum_{i=1}^n X_i^p(Y_{Ti}^* - E[Y_{Ti}^*|X_i]) \end{pmatrix} + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

using standard arguments. This finishes the proof.

Theorem 2.2.2 *Under the assumptions of Theorem 2.2.1,* $n^{1/2}(\hat\beta_{T0}-\beta_{T0},\ldots,\hat\beta_{Tp}-\beta_{Tp})' \xrightarrow{d} N(0,\Sigma)$*, where*

$$\Sigma = M^{-1}E[\rho(X,Z,\Delta)\rho'(X,Z,\Delta)]M^{-1}.$$

The proof of this result follows readily from Theorem 2.2.1.

2.3 Practical implementation and simulations

2.3.1 Practical implementation

The estimator $\hat{\boldsymbol{\beta}}_T$ depends on a number of parameters, namely on the bandwidth a_n, the score function J and the cut-off point T. All of them can be chosen in a data driven way. First, for the score function J, we recommend the choice

$$J(s) = b^{-1} I(0 \leq s \leq b)$$

$(0 \leq s \leq 1)$, where $b = \min_{1 \leq i \leq n} \tilde{F}(+\infty | X_i)$. Doing so, the region where the Beran estimators $\tilde{F}(\cdot|X_1), \ldots, \tilde{F}(\cdot|X_n)$ are inconsistent is not used and we exploit to a maximum the 'consistent' region. Note that this b is different from the theoretical $s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\}$ since it is chosen as random. In the same way, this slight discrepancy between random in practice and fixed theoretically also appears in the choice of the bandwidth parameter. However, this way of doing is very usual and improves the results in practice (see Akritas, Van Keilegom and Veraverbeke -2001-).

Depending on the desired criterion function, a number of adaptive procedures can be used to select the bandwidth a_n. Would the goal be to optimise the estimation of m^0 and σ^0, we then could use e.g. a bootstrap approach selecting the bandwidth that minimizes the estimated MSE of their estimators. However, it seems more appropriate here to choose the bandwidth according to the end goal, namely in order to optimise the estimators $\hat{\beta}_{T0}, \ldots, \hat{\beta}_{Tp}$. We would therefore recommend to choose the bandwidth by minimizing the least squares criterion function

$$\min_{a_n} \sum_{i=1}^{n} (\hat{Y}^*_{Ti}(a_n) - \hat{\beta}_{T0}(a_n) - \ldots - \hat{\beta}_{Tp}(a_n) X_i^p)^2, \tag{2.3.1}$$

over a grid of a_n-values. This idea has been used in other contexts as well, see e.g. Härdle, Hall and Ichimura (1993) where a similar principle is used in the context of single index models. Note that the argument a_n is added to $\hat{Y}^*_{Ti}$ and $\hat{\beta}_{Tj}$ $(i = 1, \ldots, n; j = 0, \ldots, p)$ in order to emphasize the dependence on a_n of these quantities. We will illustrate this procedure for selecting the bandwidth in the next subsection.

Finally, $\hat{S}_{X_i}$ $(i = 1, \ldots, n)$ can be chosen larger than (or equal to) the last order statistic $\hat{E}^0_{(n)}$ of the estimated residuals. In this way, all the Kaplan-Meier jumps of the integral (2.1.5) will be considered.

As an alternative to the normal approximation obtained in the previous section, a bootstrap procedure can be used to estimate the distribution of $\hat{\boldsymbol{\beta}}_T$. The procedure for censored data proposed by Li and Datta (2001) and which extends Efron's one (1981) to the nonparametric regression context, would fit this case. First, generate $X_1^*, \ldots, X_n^*$ i.i.d. from the empirical distribution of $X_1, \ldots, X_n$. Next, for each $i = 1, \ldots, n$, select at random a Y_i^* from the distribution $\tilde{F}(\cdot|X_i^*)$, and a C_i^* from $\tilde{G}(\cdot|X_i^*)$ (which is the Beran estimator of $G(\cdot|X_i^*)$ obtained by replacing Δ_i by $1 - \Delta_i$ in the expression for $\tilde{F}(\cdot|X_i^*)$). For the generation of these bootstrap data, we use a pilot bandwidth g_n asymptotically larger than the original a_n. Finally, let $Z_i^* = \min(Y_i^*, C_i^*)$ and $\Delta_i^* = I(Y_i^* \leq C_i^*)$. For each so-obtained resample $\{(X_i^*, Z_i^*, \Delta_i^*) : i = 1, \ldots, n\}$, calculate a bootstrap estimator of the regression parameters $\hat{\boldsymbol{\beta}}_T^{*b}$, $b = 1, \ldots, B$, where B is the number of resamples. These B resamples are thus used to approximate the full distribution of $\hat{\boldsymbol{\beta}}_T$. For instance, the variance of $\hat{\beta}_{Tj}$ is approximated by the estimated variance computed with the B bootstrap estimates $\hat{\beta}_{Tj}^{*b}$, $b = 1, \ldots, B$ (see the next section). Note that a more elaborated procedure could be used if one wants to take advantage of the validity of model (1.5.1). In this case, the survival times would be drawn under model (1.5.1) (instead of from a nonparametric estimator of their conditional distribution).

2.3.2 Simulations

In this section, we compare, by means of Monte Carlo simulations, the finite sample behaviour of the Buckley-James estimator with the estimator proposed in this paper. We are primarily interested in the behaviour of the bias and variance of the two estimators. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 500 simulations.

In the first setting, we generate i.i.d. data from the normal homoscedastic regression model

$$Y = \beta_0 + \beta_1 X + \sigma\varepsilon, \tag{2.3.2}$$

for various choices of β_0, β_1 and σ, where X has a uniform distribution on the unit interval and the error term ε is a standard normal random variable. The censoring variable C satisfies $C = \alpha_0 + \alpha_1 X + \sigma\varepsilon^*$, for certain choices of α_0 and α_1 and where ε^* has a standard normal distribution. We further assume that ε, ε^* and σ are independent of X, and that ε is independent of ε^*. It is easy to see that, under this model,

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{\alpha_0 - \beta_0 + (\alpha_1 - \beta_1)x}{\sqrt{2}\sigma}\Big).$$

For the weights appearing in the Beran estimator $\tilde{F}(y|x)$, we choose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. In order to improve the behaviour near the boundaries of the covariate space, we work with the boundary corrected kernels proposed by Müller and Wang (1994). As a consequence of the fact that these boundary corrected kernels can become negative, the Beran estimator decreases at certain time points. In these cases, the estimator is redefined as being constant until it starts increasing again. It is clear that (A1) (iii) is not satisfied when kernels are possibly negative. However, simulations clearly showed that the decrease of bias induced by using boundary kernels leads to a decrease of the resulting mean squared error.

For the bandwidth sequence a_n, we select the minimizer of (2.3.1) among a grid of 20 possible values between 0 and 1. For small values of a_n, the window $[x - a_n, x + a_n]$ may, for some points x, not contain any X_i $(i = 1, \ldots, n)$ for which the corresponding Y_i is uncensored, resulting in impossible estimation of $F(\cdot|x)$. In such cases, we need to enlarge the window so that it contains at least one uncensored data point in its interior. It may also happen that the bandwidth a_n, at a point x, is larger than the distance from x to both the left and right endpoint of the interval. In such cases, the bandwidth is redefined as the maximum of these two distances.

In a number of situations, the iterative Buckley-James method does not converge, but oscillates around two or more values. The estimator is then defined as the average of these values.

Table 2.1, hereafter displayed, summarizes the simulation results for different values of $\alpha_0, \alpha_1, \beta_0, \beta_1$ and σ. For fixed values of β_0, β_1 and σ, the values of α_0 and α_1 are chosen in such a way that some variation in the censoring probability curves is obtained (different proportions of censoring, censoring probability curves that do or do not wiggle a lot, ...). In particular, the proportion of censoring is computed as the average of $P(\Delta = 0|X = x)$ for an equispaced grid of values of x. For β_0, β_1 and σ fixed, α_0 and α_1 are chosen such that this proportion is small (about 30%) for the first simulation and large (about 60%) for the second simulation (see Table 2.1). Table 2.1 shows that, in general, the Buckley-James estimator has a larger variance but a smaller bias than the newly proposed estimator. In most cases the effect of the bias on the mean squared error is however small (compared to the variance). As a consequence, the new estimator has in most cases a smaller mean squared error than the Buckley-James estimator. This can be explained as follows. In the new method, all the new censored data points depend on the bandwidth parameter. Therefore, this smoothing parameter gives the possibility to fine-tune the new estimation procedure. While looking at the structure of those new data points, we observe that a change of the bandwidth parameter does not involve too large variations in the resulting new data set. This is not the case for the Buckley-James method since extreme values of the parameters can actually deteriorate artificial data points. From this, we can interpret the obtained results. First, the new estimator has a larger bias than the Buckley-James estimator because of the smoothing method used to construct the new data set. This inherent bias is larger than the bias obtained using a parametric model, but the contribution of this bias to the mean squared error is in most cases small. Second, the smaller variance of the new method can also be explained by the preliminary smoothing since this one allows to obtain more stable data sets. Moreover, the Buckley-James estimator suffers in certain cases from instability problems that are inherent to this method, as explained in Section 1.5.

β_0	β_1		$\hat{\beta}_0$			$\hat{\beta}_1$		
α_0	α_1	σ^2	Bias	Var	MSE	Bias	Var	MSE
0	1		-.004	.022	.022	-.009	.068	.069
0.6	0.85	0.5	.005	.021	.021	-.019	.065	.066
0	1		-.013	.026	.026	-.011	.084	.084
0.27	0.45	0.5	-.009	.024	.024	-.043	.075	.077
0	1		-.006	.041	.041	-.015	.141	.141
1.5	-0.5	1	.002	.040	.040	-.052	.135	.137
0	1		-.018	.050	.050	-.013	.169	.169
0.6	-0.2	1	-.008	.047	.047	-.074	.153	.158
0	5		-.004	.021	.021	-.011	.069	.069
1	4.1	0.5	.008	.021	.021	-.050	.067	.069
0	5		-.013	.025	.025	-.006	.088	.088
0.5	4	0.5	.011	.025	.025	-.079	.086	.092
0	5		-.006	.042	.042	-.014	.138	.138
1.3	3.9	1	.009	.041	.041	-.067	.130	.135
0	5		-.015	.047	.047	-.004	.186	.186
1	3	1	.033	.047	.048	-.170	.171	.200

Table 2.1: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (2.3.2). For β_0, β_1 and σ fixed, the proportion of censoring is about 30% (60%) in the first (second) simulation.*

The characteristics of the third and fourth simulations of Table 2.1 are also represented in Figure 2.3.1, displayed on next page. In addition, we observe that the discrepancies between biases and variabilities seem to be more pronounced when the proportion of censoring is larger. For each method, the mean line (obtained with the averages of each parameter) lies between two lines computed as follows. The upper (lower) line is constructed by using the 95% (5%) quantiles of the empirical distributions of the slope and intercept parameters. Those distributions are computed with the 2×500 estimated parameters obtained from the simulations. In this way, at least 80 % of the estimated lines lie between the upper and lower lines.

Variability of pointwise estimations is then represented in Figure 2.3.2. For each method, the mean line lies between two curves which contain at each value of the covariate (a grid of 100 points) 90% of the estimated conditional means (by taking for each value of the grid the 5% and 95% quantiles of the empirical distribution constructed with the estimated conditional means). In the same way, the median curve is also added to show symmetric behaviour of the distributions of the conditional means. Similar differences between pointwise biases and variabilities are observed especially when the proportion of censoring is larger.

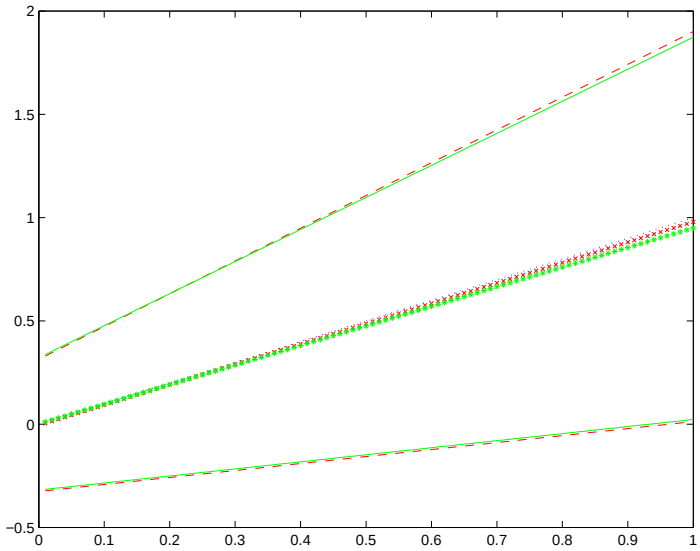
Next, suppose that Y and C are distributed according to

$$Y|X = x \sim Weibull(\exp[-d(\gamma_0 + \gamma_1 x + \gamma_2 x^2)], d),$$
$$C|X = x \sim Weibull(\exp[-d(\alpha_0 + \alpha_1 x + \alpha_2 x^2)], d),$$

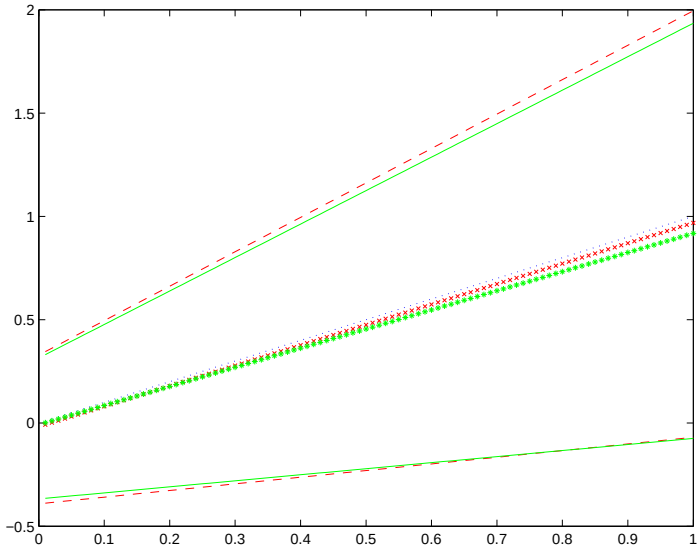
and are independent conditionally on X. The covariate X is uniformly distributed on $[0, 1]$. It is easy to check that

$$\log Y|X = x \sim F(y|x) = 1 - \exp(-\exp[d(y - \gamma_0 - \gamma_1 x - \gamma_2 x^2)]), \qquad (2.3.3)$$
$$\log C|X = x \sim G(y|x) = 1 - \exp(-\exp[d(y - \alpha_0 - \alpha_1 x - \alpha_2 x^2)]).$$

It follows that $\log Y$ has, conditionally on $X = x$, an extreme value distribution and hence $E(\log Y|X = x) = -D/d + \gamma_0 + \gamma_1 x + \gamma_2 x^2 = \beta_0 + \beta_1 + \beta_2 x^2$ and $Var(\log Y|X = x) = \pi^2/(6d^2)$, where $\beta_0 = -D/d + \gamma_0$, $\beta_1 = \gamma_1$, $\beta_2 = \gamma_2$ and $D = 0.5772$ is the Euler constant. It easily follows that if $m(x) = E(\log Y|x)$ and $\sigma^2(x) = \mathrm{Var}(\log Y|x)$, then $P(\varepsilon \leq y|x) = P((\log Y - m(x))/\sigma(x) \leq y|x) = 1 - \exp(-\exp(y\pi/\sqrt{6} - D))$. Since this expression is

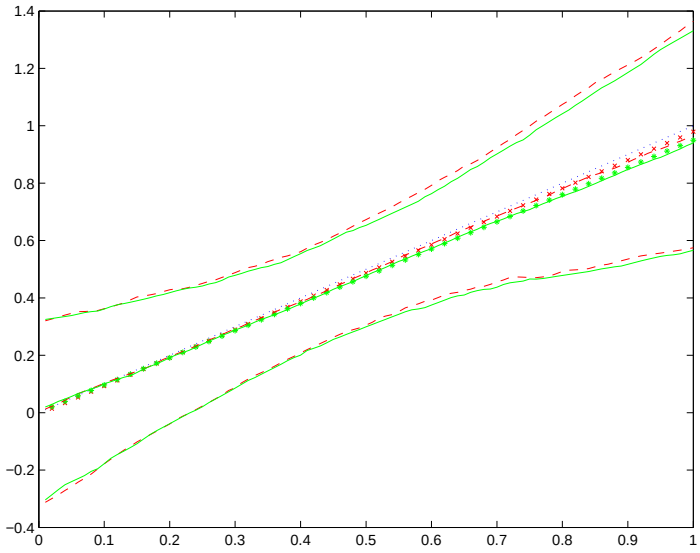


(a)

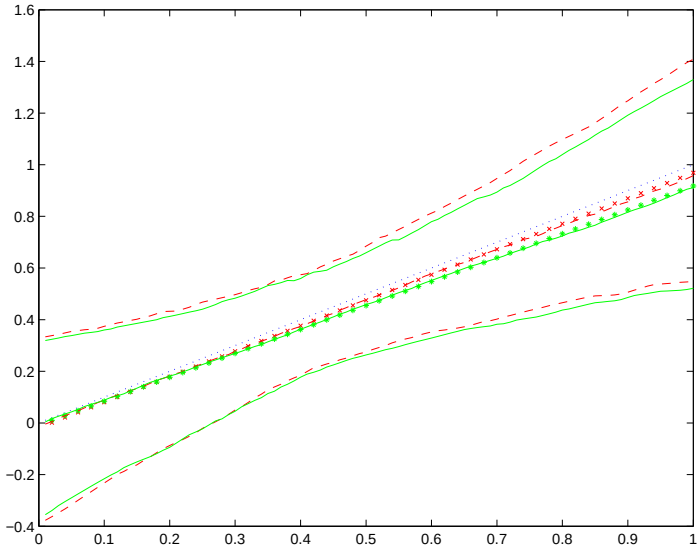


(b)

Figure 2.3.1: Comparison of biases and variabilities between the Buckley-James and the new line. The solid, respectively, dashed lines are obtained from the new, respectively, Buckley-James method. Dots indicate the true line and $*$, respectively, $\times$ represent the mean line for the new, respectively, Buckley-James method. (a) Third simulation of 2.1; (b) Fourth simulation of Table 2.1.



(a)



(b)

Figure 2.3.2: Comparison of pointwise biases and variabilities of the Buckley-James method with the new one. The solid, respectively, dashed curves are obtained from the new, respectively, Buckley-James method. Dots indicate the true line and $*$, respectively, $\times$ represent the mean line for the new, respectively, Buckley-James method. (a) Third simulation of 2.1; (b) Fourth simulation of Table 2.1.

independent of x, model (1.5.1) holds. Further, with $a_x = \exp(-d(\gamma_0 + \gamma_1 x + \gamma_2 x^2))$ and $b_x = \exp(-d(\alpha_0 + \alpha_1 x + \alpha_2 x^2))$, the conditional censoring probability curve is given by $P(\Delta = 0|X = x) = b_x/(a_x + b_x)$.

The bias, variance and mean squared error of the new and the Buckley-James estimator for 16 sets of parameters are given in Tables 2.2 and 2.3. γ_0, γ_1, γ_2, α_0, α_1, α_2 and d are chosen as in the first setting.

γ_0	γ_1	γ_2		$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$		
α_0	α_1	α_2	d	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
7.6	1	1		.013	.069	.069	-.053	1.66	1.66	.043	1.62	1.62
8.7	-0.2	1	5/3	-.014	.064	.064	.172	1.45	1.48	-.248	1.35	1.41
7.6	1	1		.013	.085	.085	-.064	2.32	2.32	.067	2.41	2.41
8.2	-0.2	1	5/3	-.032	.073	.074	.323	1.72	1.82	-.452	1.60	1.81
7.6	1	1		.022	.200	.201	-.088	4.70	4.71	.061	4.50	4.50
9	-0.2	1	1	-.014	.182	.183	.174	4.07	4.10	-.257	3.78	3.85
7.6	1	1		.027	.255	.255	-.117	6.43	6.45	.100	6.36	6.37
8.2	-0.2	1	1	-.046	.205	.207	.381	4.60	4.75	-.507	4.27	4.53
7.6	5	1		.013	.070	.070	-.054	1.66	1.66	.040	1.60	1.60
8.6	4	1	5/3	-.004	.069	.069	.121	1.56	1.57	-.185	1.44	1.48
7.6	5	1		.014	.087	.087	-.072	2.31	2.31	.069	2.35	2.35
8.1	4	1	5/3	-.006	.090	.090	.212	2.17	2.22	-.316	2.05	2.15
7.6	5	1		.020	.202	.203	-.083	4.73	4.73	.058	4.50	4.50
8.9	4	1	1	-.007	.190	.190	.162	4.27	4.29	-.252	3.94	4.01
7.6	5	1		.027	.264	.265	-.115	6.49	6.51	.099	6.32	6.33
8.1	4	1	1	-.025	.236	.237	.357	5.28	5.41	-.513	4.79	5.06

Table 2.2: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (2.3.3). For γ_0, γ_1, γ_2 and d fixed, the proportion of censoring is about 30% (60%) in the first (second) simulation.*

γ_0	γ_1	γ_2		$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\beta}_2$		
α_0	α_1	α_2	d	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
6.7	5	5		.017	.127	.127	-.064	2.97	2.97	.044	2.84	2.84
7.9	4	5	5/4	-.001	.126	.126	.131	2.84	2.86	-.188	2.63	2.66
6.7	5	5		.021	.161	.161	-.093	4.09	4.10	.082	4.06	4.06
7.2	4	5	5/4	-.007	.166	.167	.288	3.94	4.02	-.370	3.70	3.84
6.7	5	5		.044	.840	.842	-.170	19.0	19.1	.120	17.8	17.8
8.9	4	5	0.5	-.022	.789	.789	.333	17.4	17.5	-.463	15.9	16.1
6.7	5	5		.076	1.14	1.15	-.303	25.9	25.9	.246	24.2	24.2
7.2	4	5	0.5	-.069	.971	.976	.782	21.2	21.8	-1.00	19.2	20.2
6.7	1	5		.016	.085	.086	-.064	1.83	1.84	.047	1.68	1.68
7	2	4	5/3	-.002	.081	.081	.103	1.69	1.70	-.142	1.51	1.53
6.7	1	5		.031	.127	.128	-.128	2.62	2.63	.104	2.37	2.38
6.5	2	4	5/3	-.006	.110	.110	.236	2.16	2.21	-.307	1.90	1.99
6.7	1	5		.027	.332	.332	-.103	7.48	7.49	.073	7.04	7.04
7.9	2	3	0.8	-.033	.305	.306	.336	6.63	6.75	-.441	6.04	6.24
6.7	1	5		.054	.463	.466	-.241	10.8	10.8	.212	10.5	10.5
6.8	2	3	0.8	-.077	.377	.383	.727	8.05	8.58	-.928	7.28	8.14

Table 2.3: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (2.3.3). For γ_0, γ_1, γ_2 and d fixed, the proportion of censoring is about 30% (60%) in the first (second) simulation.*

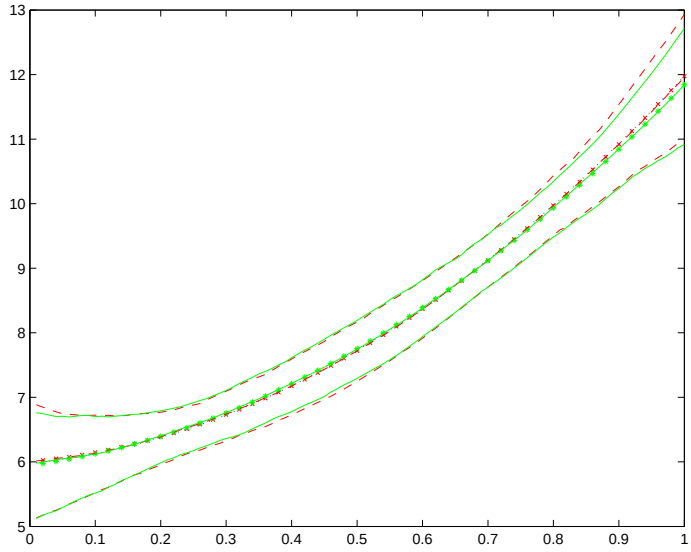
The results are similar (but even more pronounced) than in Table 2.1 : in most cases, the new estimator has a slightly larger bias, but a much smaller variance, which leads to a substantial smaller mean squared error compared to the Buckley-James estimator. As could have been expected from the new data points structure, increasing the number of parameters is an advantage for the new method since it continues to construct sufficiently stable curves while the Buckley-James procedure still enlarges its set of possible solutions. Note also that, when censoring becomes heavier, the increase in the mean squared error is in most cases more limited for the new method than for the Buckley-James' one (see Tables 2.1, 2.2 and 2.3). This is due to a larger increase of flexibility within the new data sets for the Buckley-James method.

Those facts are also reflected in Figure 2.3.3 (see next page) when analysing the behaviour of pointwise estimations. Figure 2.3.3 is constructed as Figure 2.3.2 for the two last simulations of Table 2.3.

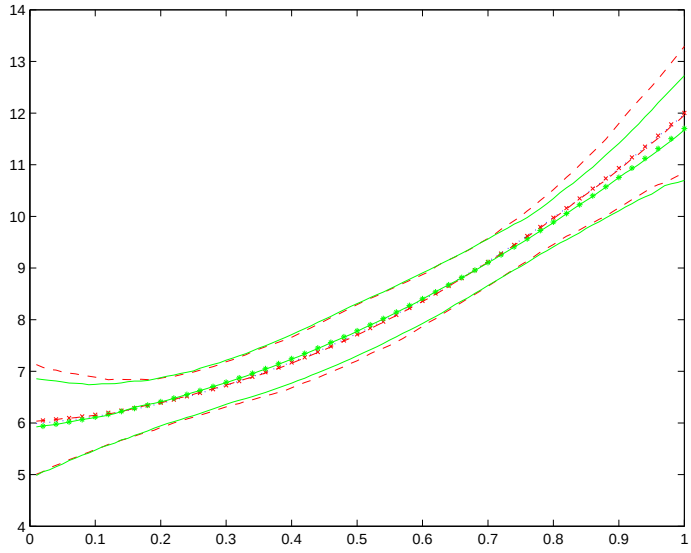
Other choices of parameters lead to similar results.

				$\hat{\beta}_0$			$\hat{\beta}_1$		
α_0	α_1	ϱ	γ	Bias	Var	MSE	Bias	Var	MSE
0.7	9.85	1	1	.085	.007	.014	.181	.049	.082
				.063	.006	.010	.135	.048	.066
1.5	9.5	2	2	.166	.027	.054	.366	.197	.331
				.100	.023	.033	-.266	.190	.260
2.4	10	4	3	.230	.061	.114	.475	.466	.692
				.119	.052	.066	-.326	.444	.550
2.6	10	4	5	.465	.172	.388	.967	1.20	2.13
				.201	.136	.177	-.569	1.16	1.48

Table 2.4: *Results for the Buckley-James estimator (first line) and the new estimator (second line) for model (2.3.4).*



(a)



(b)

Figure 2.3.3: Comparison of pointwise biases and variabilities of the Buckley-James method with the new one. The solid, respectively, dashed curves are obtained from the new, respectively, Buckley-James method. Dots indicate the true curve and $*$, respectively, $\times$ represent the mean curve for the new, respectively, Buckley-James method. (a) Seventh simulation of Table 2.3; (b) Last simulation of Table 2.3.

The final setting we consider is a normal heteroscedastic regression model

$$Y = \beta_0 + \beta_1 X + \gamma X \varepsilon, \tag{2.3.4}$$

with $\beta_0 = 0$, $\beta_1 = 10$. X has a uniform distribution on $[0, 1]$, ε has a standard normal distribution, and γ equals $1, 2, 3$ or 5. The censoring variable is given by $C = \alpha_0 + \alpha_1 X + \varrho\varepsilon^*$, where ε^* has a standard normal distribution. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. As the Buckley-James estimator is limited to homoscedastic models, we keep on using the same estimator as the one used before, while the new estimator is now taking the heteroscedasticity into account. Therefore, we expect the Buckley-James estimator to behave poorly when there is a lot of heteroscedasticity in the model. This is indeed confirmed by the results in Table 2.4 hereunder, which shows deteriorating results for the Buckley-James estimator for increasing values of γ.

2.4 Data analysis

We will illustrate the proposed method on the medical and the astronomical data set (see Section 1.1, Figures 1.1.1 and 1.1.2). We are interested in studying the relationship between $Y = \log(\text{survival time})$ and $X = \log(\text{age of the patient at diagnosis (in years)})$. The data shown in Figure 1.1.1 suggest that a linear model might be appropriate :

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{2.4.1}$$

where ε and X are independent and $E(\varepsilon) = 0$. The Buckley-James algorithm and the new method yield respectively the values -1.03 and -0.97 for the slope parameter and 5.64 and 5.39 for the intercept parameter. It was observed that the Buckley-James method does not converge to a single value of the slope parameter, but oscillates between three values. The estimator was then defined as the average of these values. For the new method, boundary corrected kernels are used. The bandwidth is selected from a grid of 16 bandwidths, according to the method described in Section 2.3.1. From Figure 2.4.4, it is clear that the regression lines (and also the new data points) obtained from the Buckley-James method and the new method are very close to each other.

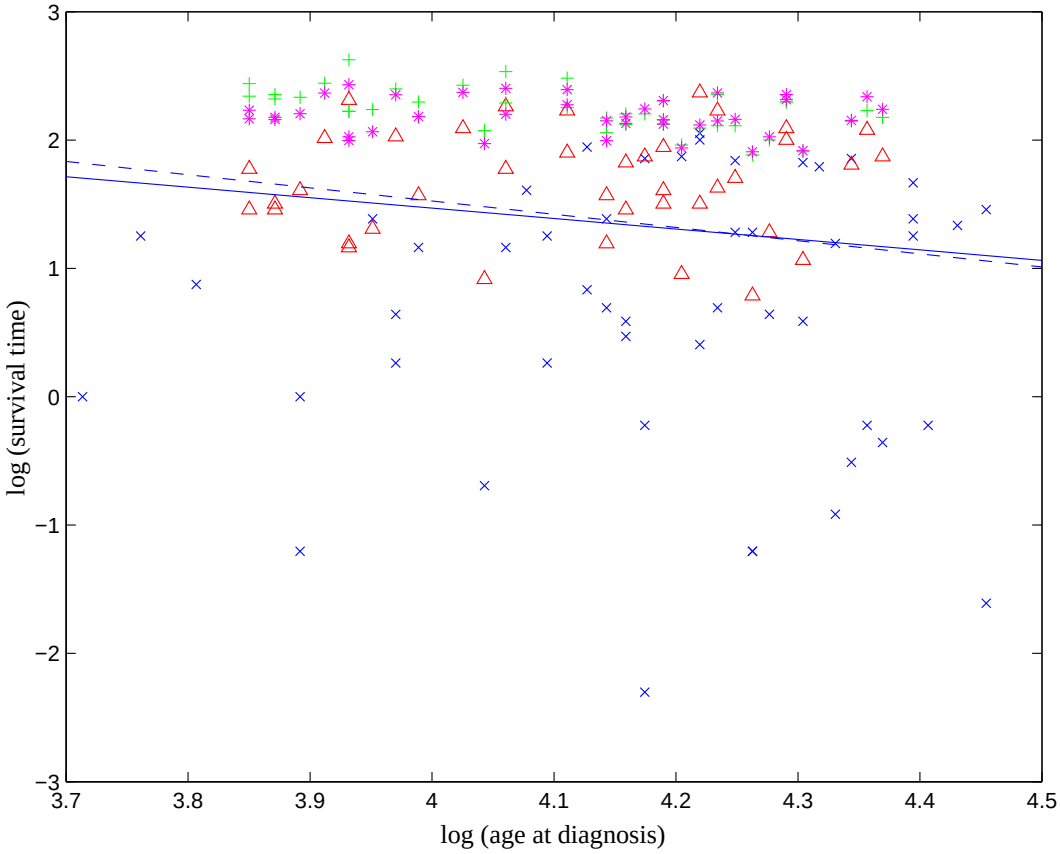


Figure 2.4.4: *Linear regression for the larynx cancer data. The solid, respectively, dashed line represents the estimated regression line for the new, respectively, Buckley-James method. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*. The new data points obtained from the new method are represented by* $*$*, while those of the Buckley-James method are symbolised by* $+$*.*

By using the bootstrap method explained in Section 2.3.1, the variance of the slope of the new method is given by 1.05 while this amounts to 18.32 for the intercept. Confidence intervals obtained from the percentile bootstrap method are $[-3.10, 0.92]$ for the slope and $[-2.65, 14.00]$ for the intercept. The intervals obtained from the normal approximation are very similar, which suggests that the asymptotic normality result is accurate here.

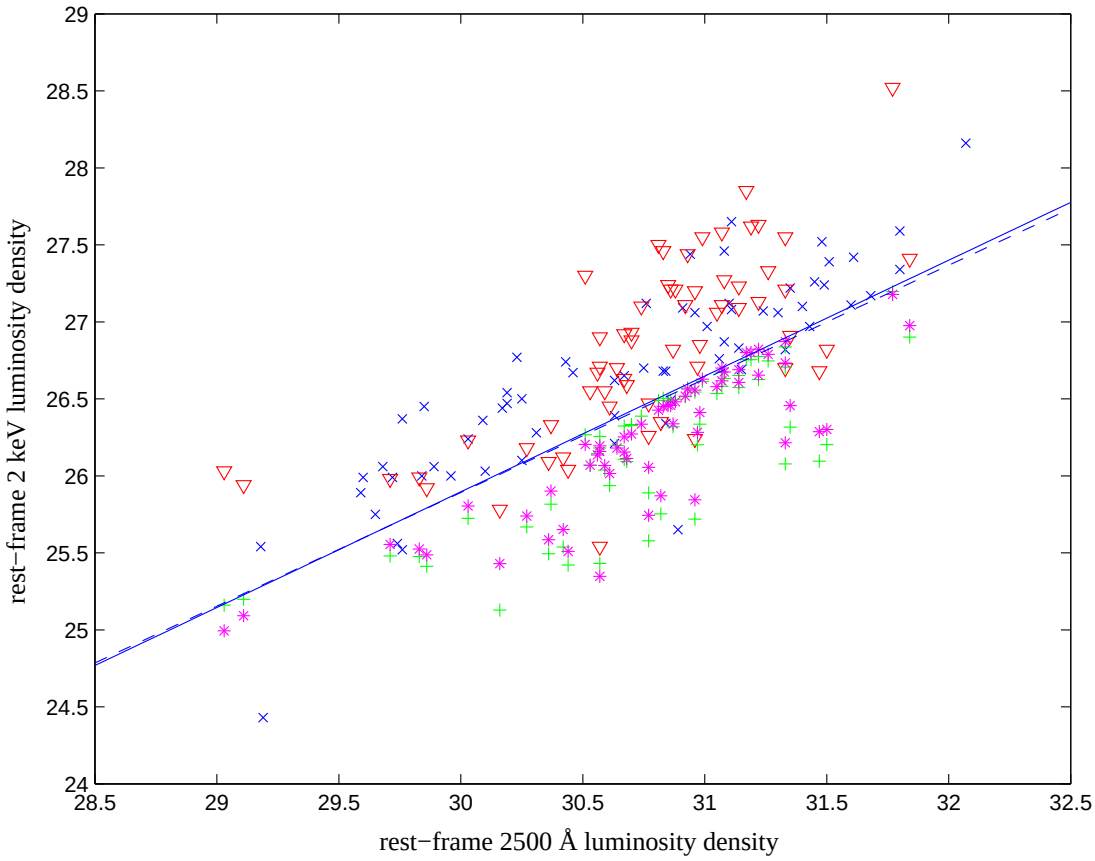


Figure 2.4.5: *Linear regression for the quasars data. The solid, respectively, dashed line represents the estimated regression line for the new, respectively, Buckley-James method. Uncensored data points are given by* $\times$*, and censored observations by* $\bigtriangledown$*. The new data points obtained from the new method are represented by* $*$*, while those of the Buckley-James method are symbolised by* $+$*.*

Next, we analyse the study of quasars in astronomy. Right-censored data points are obtained by replacing the left censored $l_{x,i}$ by $Z_i = (\max_{j:j=1,\ldots,137}(l_{x,j}) - l_{x,i})$, $i = 1, \ldots, 137$. We show in Figure 2.4.5 hereabove the results of the regression of l_x on l_{uv} for both the new and the Buckley-James algorithm, assuming that model (2.4.1) is valid (where the latter is again obtained by taking the average of the values around which it oscillates).

We observe a big similarity between the two regression lines. For both methods there is a strong correlation between the two variables. The slope and intercept are respectively 0.75 and 3.48 for the new method and 0.74 and 3.76 for the Buckley-James method. The variance of the slope and intercept for the new method equal 0.006 and 5.68 respectively, while the percentile bootstrap confidence intervals are given by $[0.52, 0.83]$ and $[0.98, 10.57]$ respectively.

2.5 Appendix

Lemma 2.5.1 *Assume (A1) (i)–(iii), (A2) (i), (ii), (A3) (ii), F_X is twice continuously differentiable , $\inf_{x\in R_X} f_X(x) > 0$, for $L(y|x) = H(y|x)$ or $H_1(y|x)$, $L(y|x)$ is continuous, $\dot{L}(y|x)$ and $\ddot{L}(y|x)$ exist and are continuous in (x, y), $\sup_{x,y} |y\dot{L}(y|x)| < \infty$ and $\sup_{x,y} |y^2\ddot{L}(y|x)| < \infty$, $h^0_\varepsilon(y|x)$ exists and is continuous in (x, y), $\sup_{x,y} |yh^0_\varepsilon(y|x)| < \infty$ and $E[|\varepsilon^0|] < \infty$. Then, for any $T < \tau_{H^0_\varepsilon}$,*

$$\int_{-\infty}^{T} |u|\, d\hat{F}^0_\varepsilon(u)$$

is bounded in probability.

Proof. We have for $T < \tau_{H^0_\varepsilon}$,

$$\int_{-\infty}^{T} |u| d\hat{F}^0_\varepsilon(u) = \sum_{\Delta_i=1} |\frac{Y_i - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}|\, W_i\, I(\frac{Y_i - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)} \le T),$$

where W_i are the Kaplan-Meier jumps of $\hat{F}^0_\varepsilon$. First, let us show that the jumps of $\hat{F}^0_\varepsilon$ are uniformly $O_P(n^{-1})$. It is easy to see that the jump of $\hat{F}^0_\varepsilon$ at the j-th order statistic of $\hat{E}^0_i = (Y_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$ $(i = 1, \ldots, n)$ is bounded by $(n - j + 1)^{-1} \le (n - a + 1)^{-1}$, where a is the number of $\hat{E}^0_i$'s smaller than or equal to T. From Proposition A.3 in VKA (1999) we know that

$$\hat{H}^0_\varepsilon(T) = H^0_\varepsilon(T) + o_P(1),$$

where $\hat{H}^0_\varepsilon$ is the empirical distribution function of the $\hat{E}^0_i$'s. Thus, $a = nH^0_\varepsilon(T) + o_P(n)$ and $(n-a+1)^{-1} = O_P(n^{-1})$. It follows that, for n sufficiently large,

$$\begin{aligned}\sum_{\Delta_i=1} |\frac{Y_i - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}| W_i I(\frac{Y_i - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)} \leq T) &\leq O_P(n^{-1}) \sum_{\Delta_i=1} |\frac{Y_i - m^0(X_i)}{\sigma^0(X_i)}| + o_P(1) \\ &= O_P(1) \int |y| d\tilde{H}^0_{\varepsilon 1}(y) + o_P(1) \\ &= O_P(1)[\int |y| dH^0_{\varepsilon 1}(y) + o_P(1)] + o_P(1) \\ &= O_P(1),\end{aligned}$$

where $\tilde{H}^0_{\varepsilon 1}(y) = n^{-1} \sum_{i=1}^n I(\frac{Y_i - m^0(X_i)}{\sigma^0(X_i)} \leq y, \Delta_i = 1)$, and provided that $\int |y| dH^0_{\varepsilon 1} = E[|\varepsilon^0| I(\Delta = 1)] \leq E[|\varepsilon^0|] < \infty$.

Chapter 3

Nonlinear regression with censored data

In the previous chapter, we considered models of the type (1.5.1) and we introduced specifications of the heteroscedastic regression model (1.5.3) to use an optimal amount of information when constructing new data points. Using more information (like a parametric regression model) would have lead to an iterative procedure. In this chapter, we will consider models of the type (1.6.1). In this context, introducing a parametric model in the new data points (as in Buckley and James (1979)) or in the weights of the estimated distribution (as in Miller (1976)) has never been achieved in a general way. So, Stute (1999) doesn't use any model in the bivariate Kaplan-Meier (1958) weights W_{in} in (1.6.2). In the same way, other methods introduced in the polynomial case could be extended to the nonlinear case but would suffer from the same disadvantage. Therefore, the use of the heteroscedastic regression model (1.5.3) when constructing artificial data points is even more highlighted.

In Section 3.1, we will describe an extension to the nonlinear case of the method proposed in Chapter 2 and add some assumptions under which the main results are valid. The consistency and an asymptotic representation of the proposed estimator (from which follows its asymptotic normality) will be derived is Section 3.2. A simulation study will then be carried out in Section 3.3. As in Section 2.3.2, the dependency of the new data points on

the model (1.5.3) enables (in particular through the choice of a bandwidth parameter) to outperform Stute's (1999) procedure. Finally, fatigue life data are analysed in Section 3.4 and the distribution of the proposed estimator is obtained by a bootstrap procedure. The content of this chapter can also be found in Heuchenne and Van Keilegom (2005a).

3.1 Notations and description of the method

In this chapter, we will use notations of Chapter 2. Since the model (1.6.1) is a particular case of (1.5.3), the new data points can be computed as in (2.1.5). Next, they are introduced into the least squares problem

$$\min_{\theta\in\Theta}\sum_{i=1}^{n}[\hat{Y}_{Ti}^{*}-m_{\theta}(X_i)]^2. \tag{3.1.1}$$

In order to focus on the primary issues, we assume the existence of a well-defined minimizer of (3.1.1). The solution of this problem can be obtained using an (iterative) procedure for nonlinear minimization problems, like e.g. a Newton-Raphson procedure. Denote a minimizer of (3.1.1) by $\hat{\theta}_n^T = (\hat{\theta}_{n1}^T,\ldots,\hat{\theta}_{nd}^T)'$. As it is clear from the definition of $\hat{Y}_{Ti}^{*}$, $\hat{\theta}_{n1}^T,\ldots,\hat{\theta}_{nd}^T$ are actually estimating the unique $\theta_0^T = (\theta_{01}^T,\ldots,\theta_{0d}^T)'$ which minimizes $E[\{E(Y_T^{*}|X)-m_{\theta}(X)\}^2]$ (see hypothesis (A10) below). These coefficients $\theta_{01}^T,\ldots,\theta_{0d}^T$ can be made arbitrarily close to $\theta_{01},\ldots,\theta_{0d}$, provided $\tau_{F_\varepsilon^0}\leq\tau_{G_\varepsilon^0}$.

The functions $\xi_\varepsilon(z,\delta,y)$, $\xi(z,\delta,y)$, $\eta(z,\delta|x)$, $\zeta(z,\delta|x)$, $\gamma_1(y|x)$, $\gamma_2(y|x)$, $\varphi(x,z,\delta,y)$, $\alpha_i(v)$ of Section 2.1 will also be needed in the statement of the asymptotic results given in the next section. Moreover, the function $B_k(z,Z_j,\Delta_j|X_i)$ is generalized to the nonlinear case such that

$$\begin{aligned}
&B_k(z,Z_j,\Delta_j|X_i)\\
&\quad=\frac{\partial m_{\theta_0^T}(X_i)}{\partial\theta_k}f_X^{-1}(X_i)\sigma^0(X_i)\left\{\left[\alpha_i'(e_i^{0T}(z))-1+\frac{S_i f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(e_i^{0T}(z))}\right]\eta(Z_j,\Delta_j|X_i)\right.\\
&\qquad\left.+\left[e_i^{0T}(z)\alpha_i'(e_i^{0T}(z))-\alpha_i(e_i^{0T}(z))+\frac{S_i^2 f_\varepsilon^0(S_i)}{1-F_\varepsilon^0(e_i^{0T}(z))}\right]\zeta(Z_j,\Delta_j|X_i)\right\},
\end{aligned}$$

$k=1,\ldots,d,\ i,j=1,\ldots,n.$

The assumptions we need for the proofs of the main results of Section 3.2 are listed below.

$(A2) - A(8)$ mentioned in Section 2.1.

$(A1)(i) - (iii)$ mentioned in Section 2.1.
$(iv)'$ Ω (defined in Theorem 3.2.2) is non-singular.

$(A9)$ Θ is compact and θ_0^T is an interior point of Θ. All partial derivatives of $m_\theta(x)$ with respect to the components of θ up to order three exist and are continuous in (x, θ) for all x and θ.

$(A10)$ The function $E[\{E(Y_T^*|X) - m_\theta(X)\}^2]$ has a unique minimum in $\theta = \theta_0^T$.

3.2 Asymptotic results

We start by showing the convergence in probability of $\hat{\theta}_n^T$ and of the least squares criterion function. This will allow us to develop an asymptotic representation for $\hat{\theta}_{nj}^T - \theta_{0j}^T$ $(j = 1, \ldots, d)$, which in turn will give rise to the asymptotic normality of these estimators.

Theorem 3.2.1 *Assume (A1) (i)–(iii), (A2) (i), (ii), (A3) (i), (ii), (A4) (i), (A6), (A10) and $m_\theta(x)$ is continuous in (x, θ). Let $S_n(\theta) = \frac{1}{n}\sum_{i=1}^n [\hat{Y}_{Ti}^* - m_\theta(X_i)]^2$. Then,*

$$\hat{\theta}_n^T - \theta_0^T = o_P(1),$$

and

$$S_n(\hat{\theta}_n^T) = E[\sigma^{02}(X) Var(\varepsilon_T^{0*}|X)] + E[\{E(Y_T^*|X) - m_{\theta_0^T}(X)\}^2] + o_P(1),$$

where

$$\varepsilon_T^{0*} = \varepsilon^0 \Delta + \frac{1}{1 - F_\varepsilon^0(E^{0T})} \int_{E^{0T}}^{S_X} u dF_\varepsilon^0(u)(1 - \Delta).$$

Proof. We prove the consistency of $\hat{\theta}_n^T$ by verifying the conditions of Theorem 5.7 in van der Vaart (1998), page 45). From the definition of $\hat{\theta}_n^T$ and condition $(A10)$, it follows that it suffices to show that

$$\sup_{\theta} |S_n(\theta) - S_0(\theta)| \to_P 0, \tag{3.2.1}$$

where $S_0(\theta) = E[\sigma^{02}(X)Var(\varepsilon_T^{0*}|X)] + E[\{E(Y_T^*|X) - m_\theta(X)\}^2]$. The second statement of Theorem 3.2.1 then follows immediately from (3.2.1) together with the consistency of $\hat{\theta}_n^T$. To prove (3.2.1) we write

$$\begin{aligned} S_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_{Ti}^* - Y_{Ti}^*)^2 + \frac{1}{n}\sum_{i=1}^{n}(Y_{Ti}^* - m_\theta(X_i))^2 + \frac{2}{n}\sum_{i=1}^{n}(\hat{Y}_{Ti}^* - Y_{Ti}^*)(Y_{Ti}^* - m_\theta(X_i)) \\ &= S_{n1} + S_{n2}(\theta) + S_{n3}(\theta). \end{aligned}$$

In order to treat S_{n1} and $S_{n3}(\theta)$, we first consider the difference

$$\begin{aligned} \hat{Y}_{Ti}^* - Y_{Ti}^* = \Bigg\{ & [\hat{m}^0(X_i) - m^0(X_i)] + \frac{\hat{\sigma}^0(X_i) - \sigma^0(X_i)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} \int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u) \\ & + \sigma^0(X_i)\frac{\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}))(1 - F_\varepsilon^0(E_i^{0T}))} \int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u) \\ & + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}\Big[\int_{\hat{E}_i^{0T}}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u) - \int_{E_i^{0T}}^{S_i} u\, dF_\varepsilon^0(u)\Big] \Bigg\} (1 - \Delta_i) \\ = & \left\{ V_{n1}^i + (V_{n2}^i + V_{n3}^i)V_{n4}^i + \frac{\sigma^0(X_i)}{1 - F_\varepsilon^0(E_i^{0T})}(V_{n4}^i - V_4^i) \right\} (1 - \Delta_i), \end{aligned}$$

where $\hat{S}_i = \hat{S}_{X_i}$. First, we treat the difference $V_{n4}^i - V_4^i$.

$$\begin{aligned} V_{n4}^i - V_4^i &= \int_{\hat{E}_i^{0T}}^{E_i^{0T}} u\, d\hat{F}_\varepsilon^0(u) + \int_{E_i^{0T}}^{S_i} u\, d(\hat{F}_\varepsilon^0(u) - F_\varepsilon^0(u)) + \int_{S_i}^{\hat{S}_i} u\, d\hat{F}_\varepsilon^0(u) \\ &= W_{n1}^i + W_{n2}^i + W_{n3}^i. \end{aligned}$$

By using integration by parts, we can write

$$\begin{aligned} W_{n1}^i = & E_i^{0T}[\hat{F}_\varepsilon^0(E_i^{0T}) - F_\varepsilon^0(E_i^{0T})] + [E_i^{0T}F_\varepsilon^0(E_i^{0T}) - \hat{E}_i^{0T}F_\varepsilon^0(E_i^{0T})] \\ & + \hat{E}_i^{0T}[F_\varepsilon^0(E_i^{0T}) - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})] - \int_{\hat{E}_i^{0T}}^{E_i^{0T}} \hat{F}_\varepsilon^0(u)du. \end{aligned} \tag{3.2.2}$$

By Corollary 3.2 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA (1999), as in Section 2.2), the first term of (3.2.2) is $|E_i^{0T}|O_P(n^{-1/2})$, while from Proposition 4.5 in the same paper, it follows that the second and fourth terms are $|E_i^{0T}|O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. The third term is $|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ using Proposition 4.5 of VKA (1999) and the fact that

$$\sup_{x,z}\Big|\hat{F}_\varepsilon^0\Big\{\frac{z\wedge T_x-\hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big\}-F_\varepsilon^0\Big\{\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)}\Big\}\Big|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

This is shown in Theorem 2.2.1 (see (2.2.2)). Hence, $W_{n1}^i=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. In a similar way it can be shown that $W_{n3}^i=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. Next,

$$\begin{aligned}W_{n2}^i&=S_i[\hat{F}_\varepsilon^0(S_i)-F_\varepsilon^0(S_i)]-E_i^{0T}[\hat{F}_\varepsilon^0(E_i^{0T})-F_\varepsilon^0(E_i^{0T})]-\int_{E_i^{0T}}^{S_i}(\hat{F}_\varepsilon^0(u)-F_\varepsilon^0(u))\,du\\&=|E_i^{0T}|O_P(n^{-1/2}).\end{aligned}$$

Therefore, $V_{n4}^i-V_4^i=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. Finally, $V_{n1}^i+(V_{n2}^i+V_{n3}^i)V_{n4}^i=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$, using Proposition 4.5 in VKA (1999) and the uniform consistency of $\hat{F}_\varepsilon^0(\hat{E}_i^{0T})$, such that

$$\hat{Y}_{Ti}^*-Y_{Ti}^*=|E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}). \tag{3.2.3}$$

Therefore, $|S_{n1}|=O_P((na_n)^{-1}\log a_n^{-1})$. In the same way, using (3.2.3) and the continuity of $m_\theta(x)$ on $R_X\times\Theta$, $\sup_\theta|S_{n3}(\theta)|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. For $S_{n2}(\theta)$, write

$$\begin{aligned}S_{n2}(\theta)&=\frac{1}{n}\sum_{i=1}^n\{Y_{Ti}^*-E(Y_{Ti}^*|X_i)\}^2+\frac{1}{n}\sum_{i=1}^n\{E(Y_{Ti}^*|X_i)-m_\theta(X_i)\}^2\\&\quad+\frac{2}{n}\sum_{i=1}^n\{Y_{Ti}^*-E(Y_{Ti}^*|X_i)\}\{E(Y_{Ti}^*|X_i)-m_\theta(X_i)\}=S_{n21}+S_{n22}(\theta)+S_{n23}(\theta).\end{aligned}$$

Since $E[\varepsilon^{02}]<\infty$, it is easily seen that

$$S_{n21}=E[\sigma^{02}(X)\{\varepsilon_{0T}^*-E(\varepsilon_{0T}^*|X)\}^2]+o(1)\quad a.s.$$

For $S_{n22}(\theta)$ and $S_{n23}(\theta)$ we use Theorem 2 of Jennrich (1969). The function g in this theorem is given by $g_\theta(x)=\{E(Y_T^*|x)-m_\theta(x)\}^2$ for $S_{n22}(\theta)$ and

$$\begin{aligned}g_\theta(z,\delta,x)=&\left[z\delta+(1-\delta)\left\{m^0(x)+\frac{\sigma^0(x)}{1-F_\varepsilon^0(\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)})}\int_{\frac{z\wedge T_x-m^0(x)}{\sigma^0(x)}}^{\frac{T_x-m^0(x)}{\sigma^0(x)}}u dF_\varepsilon^0(u)\right\}-E(Y_T^*|x)\right]\\&\times\{E(Y_T^*|x)-m_\theta(x)\}\end{aligned}$$

for $S_{n23}(\theta)$. Since $E|\varepsilon^0| < \infty$, $|g_\theta(x)| \le C_1$ and $|g_\theta(z,\delta,x)| \le h(z,\delta) = C_2 z\delta + C_3$ for some $C_1, C_2, C_3 > 0$, and for all (z,δ,x) and θ, where h is integrable with respect to the joint distribution of (z,δ,x). From this,

$$\sup_{\theta\in\Theta} |S_n(\theta) - E[\sigma^{02}(X)Var(\varepsilon^*_{0T}|X)] - E[\{E(Y^*_T|X) - m_\theta(X)\}^2]| = o_P(1). \qquad (3.2.4)$$

This finishes the proof.

Theorem 3.2.2 *Assume (A1) (i)–(iii), (A1) (iv)', (A2)–(A10). Then,*

$$\hat\theta^T_n - \theta^T_0 \ = \Omega^{-1} n^{-1} \sum_{i=1}^n \pi(X_i, Z_i, \Delta_i) + \begin{pmatrix} o_P(n^{-1/2}) \\ \vdots \\ o_P(n^{-1/2}) \end{pmatrix},$$

where $\Omega = (\Omega_{jk})$ $(j,k = 1,\dots,d)$,

$$\Omega_{jk} = E\left[\frac{\partial m_{\theta^T_0}(X)}{\partial\theta_j}\frac{\partial m_{\theta^T_0}(X)}{\partial\theta_k} - \{Y^*_T - m_{\theta^T_0}(X)\}\frac{\partial^2 m_{\theta^T_0}(X)}{\partial\theta_j\partial\theta_k}\right],$$

$\pi = (\pi_1,\dots,\pi_d)'$,

$$\begin{aligned}\pi_j(X_i,Z_i,\Delta_i) \ = \ & \int_{R_X} \frac{\partial m_{\theta^T_0}(x)}{\partial\theta_j}\sigma^0(x)\int\left\{\frac{\varphi\{X_i,Z_i,\Delta_i,e^{0T}_x(z)\}}{[1-F^0_\varepsilon\{e^{0T}_x(z)\}]^2}\int_{e^{0T}_x(z)}^{S_x} u\,dF^0_\varepsilon(u)\right. \\ & \left. + \frac{1}{1-F^0_\varepsilon\{e^{0T}_x(z)\}}\int_{e^{0T}_x(z)}^{S_x} u\,d\varphi(X_i,Z_i,\Delta_i,u)\right\} dH_0(z|x)\,dF_X(x) \\ & + f_X(X_i)\int B_j(z,Z_i,\Delta_i|X_i)\,dH_0(z|X_i) + \frac{\partial m_{\theta^T_0}(X_i)}{\partial\theta_j}(Y^*_{Ti} - m_{\theta^T_0}(X_i))\end{aligned}$$

$(j = 1,\dots,d; i = 1,\dots,n)$.

Proof. For some θ_{1n} between $\hat\theta^T_n$ and θ^T_0

$$\hat\theta^T_n - \theta^T_0 = -\left\{\frac{\partial^2 S_n(\theta_{1n})}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial S_n(\theta^T_0)}{\partial\theta} = -R_1^{-1}R_2.$$

We have

$$R_2 = -\frac{2}{n}\sum_{i=1}^n(\hat Y^*_{T_i} - Y^*_{T_i})\frac{\partial m_{\theta^T_0}(X_i)}{\partial\theta} - \frac{2}{n}\sum_{i=1}^n\{Y^*_{T_i} - m_{\theta^T_0}(X_i)\}\frac{\partial m_{\theta^T_0}(X_i)}{\partial\theta} = R_{21} + R_{22},$$

such that R_{22} is a sum of i.i.d. random variables with zero mean (by definition of θ_0^T). For each component j of R_{21}, we use a technique similar to Theorem 2.2.1 to obtain an asymptotic representation. So we obtain

$$\begin{aligned} R_{21j} = -\frac{2}{n}\sum_{i=1}^{n}\int_{R_X}\frac{\partial m_{\theta_0^T}(x)}{\partial\theta_j}\sigma(x)\int_{-\infty}^{+\infty}\Bigg\{\frac{\varphi(X_i,Z_i,\Delta_i,e_x^{0T}(z))}{\{1-F_\varepsilon^0(e_x^{0T}(z))\}^2}\int_{e_x^{0T}(z)}^{S_x} u\,dF_\varepsilon^0(u) \\ +\frac{1}{1-F_\varepsilon^0(e_x^{0T}(z))}\int_{e_x^{0T}(z)}^{S_x} u\,d\varphi(X_i,Z_i,\Delta_i,u)\Bigg\}dH_0(z|x)dF_X(x) \\ +f_X(X_i)\int_{-\infty}^{\infty}B_j(z,Z_i,\Delta_i|X_i)\,dH_0(z|X_i)+o_P(n^{-1/2}), \end{aligned} \tag{3.2.5}$$

$(j=1,\ldots,d; i=1,\ldots,n)$. For R_1, we write

$$\begin{aligned} R_1 = -\frac{2}{n}\Bigg\{\sum_{i=1}^{n}(\hat{Y}_{T_i}^*-Y_{T_i}^*)\frac{\partial^2 m_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'}+\sum_{i=1}^{n}(Y_{T_i}^*-m_{\theta_{1n}}(X_i))\frac{\partial^2 m_{\theta_{1n}}(X_i)}{\partial\theta\partial\theta'} \\ -\sum_{i=1}^{n}\Big(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial\theta}\Big)\Big(\frac{\partial m_{\theta_{1n}}(X_i)}{\partial\theta'}\Big)\Bigg\} = R_{11}+R_{12}+R_{13}. \end{aligned}$$

Using assumption (A9), the fact that $|\hat{Y}_{T_i}^*-Y_{T_i}^*| = |E_i^{0T}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ (see the proof of Theorem 3.2.1), we have that $R_{11}=o_P(1)$. Again using condition (A9),

$$\begin{aligned} R_1 &= \frac{2}{n}\sum_{i=1}^{n}\frac{\partial m_{\theta_0^T}(X_i)}{\partial\theta}\Big(\frac{\partial m_{\theta_0^T}(X_i)}{\partial\theta}\Big)'-\frac{2}{n}\sum_{i=1}^{n}\{Y_{T_i}^*-m_{\theta_0^T}(X_i)\}\frac{\partial^2 m_{\theta_0^T}(X_i)}{\partial\theta\partial\theta'}+o_P(1) \\ &= 2E\Big[\frac{\partial m_{\theta_0^T}(X)}{\partial\theta}\Big(\frac{\partial m_{\theta_0^T}(X)}{\partial\theta}\Big)'-\{Y_T^*-m_{\theta_0^T}(X)\}\frac{\partial^2 m_{\theta_0^T}(X)}{\partial\theta\partial\theta'}\Big]+o_P(1) \\ &= 2\Omega+o_P(1). \end{aligned}$$

The result now follows.

Theorem 3.2.3 *Under the assumptions of Theorem 3.2.2,* $n^{1/2}(\hat{\theta}_n^T-\theta_0^T)\xrightarrow{d}N(0,\Sigma)$, *where*

$$\Sigma=\Omega^{-1}E[\pi(X,Z,\Delta)\pi'(X,Z,\Delta)]\Omega^{-1}.$$

The proof of this result follows readily from Theorem 3.2.2.

Remark 3.2.4 (Homoscedastic model) Note that when model (1.5.1) or (1.6.1) is homoscedastic, the representation in Theorem 2.2.1 or in Theorem 3.2.2 simplifies respectively.

In fact, it is easily seen that the function ζ equals zero in that case.

Remark 3.2.5 (Polynomial model) Note that when model (1.6.1) is polynomial, the representation in Theorem 3.2.2 reduces to the one obtained in Theorem 2.2.1 since

$$n^{-1}\sum_{i=1}^{n}\frac{\partial m_{\theta_0^T}(X_i)}{\partial \theta_j}[E(Y_{Ti}^*|X_i) - m_{\theta_0^T}(X_i)] = 0$$

in that case.

Remark 3.2.6 (Extensions) The estimation procedure and the methodology used to obtain the results of this section could be used as a basis for a number of more general models. For instance, the extension to situations where the covariate is subject to censoring could be considered (in that case the Beran estimator will need to be replaced by e.g. the estimator proposed in Van Keilegom (2003)).

Remark 3.2.7 (Multiple regression) The method cannot be extended to multiple regression problems due to the bandwidth assumption (A1) (i) required in the proofs of VKA (1999). Indeed, for a two dimensional covariate, the assumption $na_n^{5+2\delta}(\log a_n^{-1})^{-1} \to \infty$ is needed in Proposition 4.7 of VKA (1999) such that (A1) (i) is not satisfied. This flaw will also be observed in the two nonparametric procedures of Chapters 4 and 6. However, it would be interesting to extend the obtained results to semiparametric regression models, like partial linear or single index models.

3.3 Practical implementation and simulations

The proposed estimator can be easily implemented in practice. The parameters on which the estimator depends, can all be chosen in an adaptive way. The finite sample performance of $\hat{\theta}_n^T$ for these adaptively chosen parameters is illustrated in this section. For the choices of J, $\hat{S}_{X_i}$ and the resampling scheme, we refer to Section 2.3.1. The choice of the bandwidth

parameter can be carried out through the minimization of the function

$$\sum_{i=1}^{n}\{\hat{Y}^*_{Ti}(a_n) - m_{\hat{\theta}^T_n(a_n)}(X_i)\}^2, \tag{3.3.1}$$

over a specific grid of values of the smoothing parameter a_n. Here also the argument a_n is added to $\hat{Y}^*_{Ti}$ and $\hat{\theta}^T_n$, $i = 1, \ldots, n$, in order to highlight the dependence on a_n of these quantities.

We compare now the finite sample behaviour of Stute's (1999) estimator with the newly proposed estimator. We are primarily interested in the behaviour of the bias and variance of the two estimators. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 250 simulations.

In the first setting, we generate i.i.d. data from the normal homoscedastic regression model

$$Y = \frac{\theta_0}{11}\sin(\theta_1 X^2) + \sigma\varepsilon, \tag{3.3.2}$$

where $\theta_0 = \theta_1 = 11$, $\sigma^2 = 0.5$ or 1, X has a uniform distribution on the unit interval and the error term ε is a standard normal random variable. The censoring variable C satisfies $C = (\alpha_0/11)\sin(\alpha_1 X^2) + \sigma\varepsilon^*$, for certain choices of α_0 and α_1 and where ε^* has a standard normal distribution. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. Under this model,

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{(\alpha_0/11)\sin(\alpha_1 x^2) - (\theta_0/11)\sin(\theta_1 x^2)}{\sqrt{2}\sigma}\Big).$$

We compare here the new method for homoscedastic errors, with Stute's method. Note that the conditions of the latter method, which are outlined in Section 1.6 (i.e. Y and C are independent and $P(Y \leq C|X, Y) = P(Y \leq C|Y)$) are not satisfied for this model.

We work with a biquadratic kernel function $K(x) = (15/16)(1-x^2)^2 I(|x| \leq 1)$ and with the boundary corrected kernels of Müller and Wang (1994). When the Beran estimator decreases, it is redefined as being constant until it starts increasing again.

For the two methods, the Levenberg-Marquardt algorithm (Levenberg (1944) and Marquardt (1963)) is used to solve equations (1.6.2) and (3.1.1) (for a fixed value of the bandwidth parameter).

The bandwidth a_n is selected by minimizing expression (3.3.1) over a grid of 20 possible bandwidths between 0 and 1. If the window $[x - a_n, x + a_n]$ at a point x does not contain any X_i $(i = 1, \ldots, n)$ for which the corresponding Y_i is uncensored, we enlarge this window such that it contains at least one uncensored data point in its interior. If the bandwidth a_n at a point x is larger than the distance from x to both the left and right endpoint of the interval, the bandwidth is redefined as the maximum of these two distances.

Table 3.1 summarizes the simulation results for different values of α_0, α_1 and σ (values of α_0 and α_1, for fixed θ_0, θ_1 and σ, chosen such that some variation in the censoring probability curves is obtained). The proportion of censoring (in % and denoted CP in the tables) is computed as the average of $P(\Delta = 0|x)$ for an equispaced grid of values of x.

α_0	α_1	$\hat{\theta}_0$			$\hat{\theta}_1$		
σ^2	CP	Bias	Var	MSE	Bias	Var	MSE
24	1.6	-0.56	4.340	4.652	-0.06	0.209	0.212
1	34.37	-0.36	2.808	2.940	0.034	0.116	0.117
11	11	1.271	3.669	5.285	0.160	0.240	0.266
1	50	-0.58	3.046	3.377	0.043	0.169	0.171
11	12	1.258	3.909	5.493	0.350	0.284	0.406
1	51.12	-0.61	3.175	3.544	0.023	0.198	0.199
24	1.6	-0.49	2.043	2.285	-0.04	0.081	0.083
0.5	32.73	-0.16	1.660	1.686	0.021	0.052	0.052
11	11	0.701	1.799	2.290	0.136	0.120	0.138
0.5	50	-0.44	1.641	1.836	0.027	0.088	0.089
11	12	0.637	2.078	2.484	0.300	0.152	0.242
0.5	51.52	-0.46	1.581	1.794	-0.01	0.113	0.113

Table 3.1: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.2).*

In the second setting, we generate i.i.d. data from the normal homoscedastic regression model

$$Y = \frac{5}{4}\exp(\theta_0 X + \theta_1 X^2) + \sigma\varepsilon, \tag{3.3.3}$$

where $\theta_0 = 0.8$ and $\theta_1 = 1$. The other quantities in model (3.3.3) are chosen as in (3.3.2). The censoring variable C satisfies $C = \alpha_0 \exp(\alpha_1 X + \alpha_2 X^2) + \sigma\varepsilon^*$, with the same characteristics as in the first setting. Table 3.2 summarizes the simulation results for different values $\alpha_0, \alpha_1, \alpha_2$ and σ chosen as for Table 3.1. Again we compare the new method for homoscedastic errors with Stute's estimator, whose assumptions on the survival and censoring variables are not satisfied here.

α_0	α_1	α_2	$\hat{\theta}_0$			$\hat{\theta}_1$		
σ^2	CP		Bias	Var	MSE	Bias	Var	MSE
1.25	1.1	1	-0.52	0.081	0.349	0.549	0.104	0.406
1	33.98		0.002	0.077	0.077	-0.01	0.098	0.098
1.25	0.8	1	-0.71	0.142	0.646	0.711	0.173	0.678
1	50		0.026	0.088	0.088	-0.05	0.111	0.113
1.25	0.2	1.65	-0.91	0.173	0.995	0.912	0.208	1.040
1	55.22		0.033	0.108	0.109	-0.05	0.132	0.135
1.25	1.05	1	-0.36	0.038	0.164	0.376	0.049	0.190
0.5	32.27		0.011	0.039	0.039	-0.02	0.049	0.050
1.25	0.8	1	-0.53	0.064	0.345	0.524	0.081	0.355
0.5	50		0.030	0.045	0.046	-0.04	0.057	0.059
1.25	0.25	1.65	-0.66	0.077	0.516	0.672	0.094	0.546
0.5	53.62		0.046	0.056	0.058	-0.06	0.069	0.072

Table 3.2: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.3).*

In the third setting we consider a normal heteroscedastic regression model

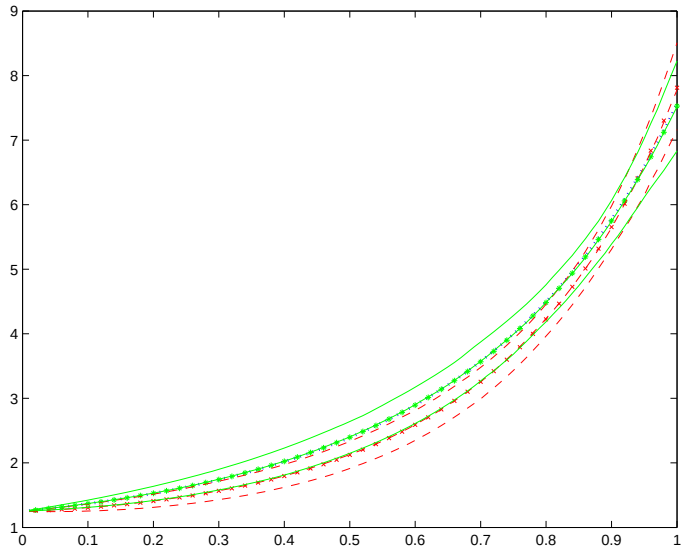
$$Y = \theta_0 X + \sin(\theta_1 X) + \gamma X \varepsilon, \tag{3.3.4}$$

with $\theta_0 = 1$, $\theta_1 = 10$, X has a uniform distribution on $[0, 1]$, ε has a standard normal distribution, and γ equals $1, 2, 3$ or 4. The censoring variable is given by $C = \alpha_0 X + \sin(\alpha_1 X) + \gamma\varepsilon^*$, where ε^* has a standard normal distribution. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. The assumptions of Stute's method outlined in Section 1.6 are again not satisfied here. Since the model is heteroscedastic, we estimate here the scale function $\sigma^0(\cdot)$. Table 3.3 summarizes the simulation results for increasing values of γ, approximately constant proportion of censoring and approximately the same shape of censoring probability curve.

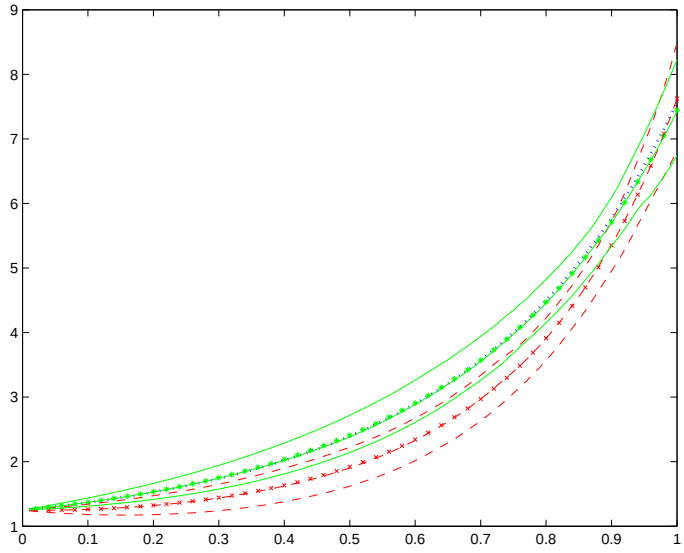
α_0	α_1	$\hat{\theta}_0$			$\hat{\theta}_1$		
γ	CP	Bias	Var	MSE	Bias	Var	MSE
1.2	11	-0.12	0.044	0.059	0.279	0.090	0.168
1	49.54	-0.05	0.037	0.039	0.038	0.066	0.068
1.2	12	-0.17	0.171	0.200	0.278	1.091	1.168
2	50.84	-0.17	0.124	0.154	0.022	0.694	0.694
1.2	13.7	-0.15	0.416	0.438	-0.72	4.166	4.678
3	50.33	-0.27	0.286	0.358	-0.16	1.560	1.589
1.2	14.3	-0.16	0.669	0.695	-1.14	5.227	6.518
4	49.99	-0.31	0.478	0.572	-0.42	3.025	3.205

Table 3.3: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.4).*

Tables 3.1 till 3.3 show that the new method outperforms Stute's estimator when the restrictive conditions of Stute's procedure are not satisfied. This is especially reflected in the bias, which is in most cases quite large in comparison with the new method. Figure 3.3.1 illustrates similar trends for pointwise estimations : when the proportion of censoring increases, the new method keeps stable curves whereas curves of Stute deteriorate more



(a)



(b)

Figure 3.3.1: Comparison of pointwise biases and variabilities of Stute's method with the new one. The solid, respectively, dashed curves are obtained from the new, respectively, Stute's method. Dots indicate the true curve and $*$, respectively, $\times$ represent the mean curve for the new, respectively, Stute's method. (a) First simulation of Table 3.2; (b) Third simulation of Table 3.2.

and more, especially for the distance with the true curve. Note that this graph is constructed as Figure 2.3.2 except that the mean curve is the average at each value of the covariate of the 250 estimated conditional means.

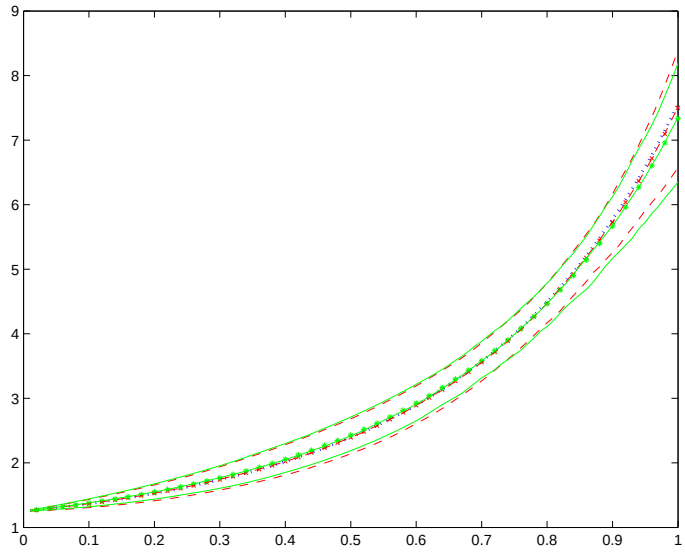
Let us now consider a homoscedastic model, in which C and X are independent :

$$Y = \theta_0 \exp(\theta_1 X + \theta_2 X^2) + \sigma\varepsilon, \tag{3.3.5}$$

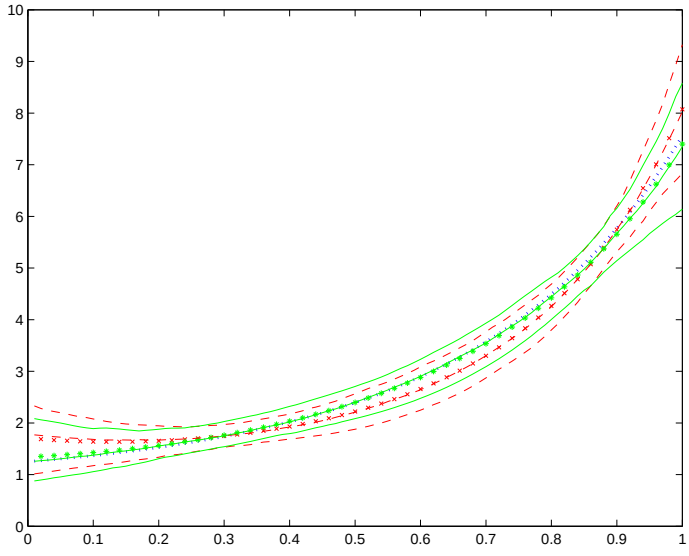
where $\theta_0 = 1.25$, $\theta_1 = 0.8$, $\theta_2 = 1$, and X, ε and σ are as in model (3.3.2). The censoring variable C satisfies $C = \alpha_0 + \rho\varepsilon^*$ for some α_0 and ρ, where ε^* is the same as before. This model equals model (3.3.3) except that the censoring variable is independent of X here. It is easily seen that Stute's assumptions are satisfied in that case. The results in Tables 3.4 and 3.5 show more moderated results. For two parameters to estimate, Stute's method behaves slightly better than the new one while when the new method can fit three parameters, its tuning ability (thanks to the bandwidth parameter in the preliminary smoothing) allows to outperform Stute's method. This is also reflected in Figure 3.3.2.

α_0	ρ	$\hat{\theta}_1$			$\hat{\theta}_2$		
σ^2	CP	Bias	Var	MSE	Bias	Var	MSE
7.5	10	0.003	0.098	0.098	-0.01	0.131	0.131
1	33.04	0.086	0.095	0.103	-0.12	0.130	0.144
2.7	15	-0.01	0.139	0.139	-0.00	0.186	0.186
1	50.87	0.162	0.137	0.164	-0.22	0.192	0.240
7	17	0.009	0.055	0.055	-0.02	0.073	0.073
0.5	40.82	0.087	0.056	0.063	-0.12	0.077	0.091
0	13	0.012	0.084	0.085	-0.02	0.118	0.119
0.5	59.12	0.173	0.081	0.111	-0.24	0.119	0.175

Table 3.4: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.5).*



(a)



(b)

Figure 3.3.2: Comparison of pointwise biases and variabilities of Buckley-James method with the new one. The solid, respectively, dashed curves are obtained from the new, respectively, Stute's method. Dots indicate the true curve and $*$, respectively, $\times$ represent the mean curve for the new, respectively, Stute's method. (a) First simulation of Table 3.4; (b) First simulation of Table 3.5.

α_0	ρ	$\hat{\theta}_0$			$\hat{\theta}_1$			$\hat{\theta}_2$		
σ^2	CP	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
7.5	10	0.457	0.201	0.410	-1.20	1.328	2.768	0.987	0.904	1.879
1	33.04	0.090	0.140	0.148	-0.08	1.082	1.089	0.022	0.742	0.743
2.7	15	0.452	0.189	0.393	-1.23	1.213	2.716	1.022	0.862	1.906
1	50.87	0.045	0.157	0.159	0.160	1.291	1.317	-0.22	0.927	0.975
7	17	0.493	0.160	0.403	-1.34	1.117	2.921	1.114	0.789	2.031
0.5	40.82	0.071	0.100	0.105	-0.03	0.794	0.795	-0.03	0.561	0.562
0	13	0.359	0.136	0.265	-0.28	1.218	1.297	-0.14	0.934	0.952
0.5	59.12	-0.01	0.084	0.084	0.315	0.823	0.922	-0.37	0.613	0.748

Table 3.5: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.5).*

Table 3.6 displayed hereafter shows the relative performance of the new estimator and the one of Stute for small samples. Let $n = 30$ and consider the model

$$Y = \theta_0 \exp(\theta_1 X + \theta_2 X^2) + (\gamma X + 0.1)\varepsilon, \tag{3.3.6}$$

where $\theta_0 = 1.25$, $\theta_1 = 0.8$ $\theta_2 = 1$, and $\gamma = 1, 2, 3$ or 4. The other quantities of (3.3.6) are chosen as in (3.3.2) and the censoring variable C satisfies $C = \alpha_0 \exp(\alpha_1 X + \alpha_2 X^2) + \gamma\varepsilon^*$ for some $\alpha_0, \alpha_1, \alpha_2$ and ε^* chosen as in the other models. Table 3.6 shows that also for small samples the new method performs well. This can also be explained by an increasing of tuning ability for the new method when reducing the sample size.

Finally, other simulations (not reported here) show that models different from the ones considered here, lead to similar simulation results : whenever the restrictive conditions of Stute's method are satisfied, his method can in some cases outperform the new one, whereas the new method behaves considerably better than Stute's estimator in situations where these conditions are not satisfied. As in Chapter 2, this is highly due to the fine-tuning of the new estimation procedure. The dependence of the new data points on a bandwidth a_n can thus be considered as a considerable advantage for the new method.

α_0	α_1	α_2	$\hat{\theta}_0$			$\hat{\theta}_1$			$\hat{\theta}_2$		
σ^2	γ	CP	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
1.25	1.1	1	0.544	0.231	0.526	-1.51	1.233	3.513	1.244	0.923	2.471
1	1	32.77	0.232	0.148	0.202	-0.49	1.307	1.544	0.359	1.043	1.172
1.25	1.4	0.9	0.516	0.351	0.617	-1.62	2.560	5.175	1.400	2.137	4.096
1	2	33	0.242	0.270	0.328	-0.47	2.665	2.882	0.340	2.227	2.343
1.25	1.6	0.9	0.457	0.690	0.899	-1.49	6.397	8.628	1.384	5.343	7.259
1	3	33.06	0.266	0.450	0.521	-0.39	5.736	5.887	0.268	4.860	4.932
1.25	1.8	0.9	0.427	1.013	1.196	-1.47	8.124	10.28	1.489	7.377	9.593
1	4	32.65	0.276	0.749	0.825	-0.18	11.20	11.23	0.118	9.176	9.190

Table 3.6: *Results for the Stute estimator (first line) and the new estimator (second line) for model (3.3.6) and* $n = 30$.

3.4 Data analysis

In this section, we illustrate the method proposed in Section 3.1 on the set of low-cycle fatigue life data (see Section 1.1, Figure 1.1.3) and we compare it with Stute's estimator. This data set is interesting for two main reasons. First, from the censoring mechanism, it may be reasonable to think that censoring depends on pseudostress. So, the assumptions of Stute's procedure (i.e. Y independent of C and $P(Y \leq C|X,Y) = P(Y \leq C|Y)$) are possibly not satisfied. Second, the data seem to follow a heteroscedastic model such that the estimation of the scale function $\sigma^0(x)$ will be needed.

A model often used in the literature (see Pascual and Meeker (1997)) is given by

$$\log Y = \beta_0 + \beta_1 \log(X - \gamma) + \sigma(X)\varepsilon \qquad (X > \gamma), \tag{3.4.1}$$

where ε is independent of X and $E[\varepsilon] = 0$. Unlike Pascual and Meeker (1997), we do not impose any parametric form for $\sigma(\cdot)$ and $F_\varepsilon(\cdot)$. This has obviously the advantage of being more robust and flexible, although Pascual and Meeker's parametric model for $\sigma(\cdot)$ and $F_\varepsilon(\cdot)$ will be more efficient in some particular situations.

Note that our method does not require the estimation of the variance function $\sigma^2(\cdot)$ (since we work with $\sigma^{02}(\cdot)$ in the estimation procedure). It would however be possible to estimate $\sigma^2(\cdot)$ by using the following idea, in analogy with the ideas that will be developed in Chapters 5 and 6 for the nonparametric estimation of the mean function : create artificial data points obeying a specific preservation of mean criterion adapted to the estimation of $\sigma^2(\cdot)$ and estimate the variance based on these new synthetic data. This estimator of $\sigma^2(\cdot)$ can then also be used to validate the parametric form of $\sigma^2(\cdot)$ used in Pascual and Meeker (1997).

The parameter γ in model (3.4.1) can be interpreted as a fatigue limit parameter i.e. specimens tested below this fatigue limit level of stress will never fail. Therefore, this parameter is constrained to be positive. The data set is then analysed by means of Stute's procedure and the new method. For the computation of the variance of the least squares estimators, an asymptotic formula is used for Stute's method while for the new method we make use of a bootstrap approximation (remind Section 2.3.1). Finally, confidence intervals are constructed for each parameter. A graph of the estimated curves is given in Figure 3.4.3 (see next page).

The results obtained by Stute's method are $\hat{\beta}_{0s} = 11.0474$, $\hat{\beta}_{1s} = -2.1315$, $\hat{\gamma}_s = 65.2730$ with estimated variances 13.6611; 0.6587; 120.5972 respectively, while the new method provides $\hat{\beta}_{0n} = 9.2432$, $\hat{\beta}_{1n} = -1.7221$, $\hat{\gamma}_n = 71.1797$ with estimated variances 9.3389; 0.4100; 116.4890 respectively. Confidence intervals for Stute's method are $\beta_0 \in [3.8031,\ 18.2917]$, $\beta_1 \in [-3.7222,\ -0.5408]$, and $\gamma \in [43.7489,\ 86.7971]$. For the new method, confidence intervals based on the normal approximation are $\beta_0 \in [3.6009;\ 15.5802]$, $\beta_1 \in [-3.0564;\ -0.5463]$ and $\gamma \in [48.9448;\ 91.2534]$, and intervals obtained with the bootstrap percentile method are $\beta_0 \in [6.7985;\ 19.0904]$, $\beta_1 \in [-3.7492;\ -1.1587]$ and $\gamma \in [33.4215;\ 77.9497]$.

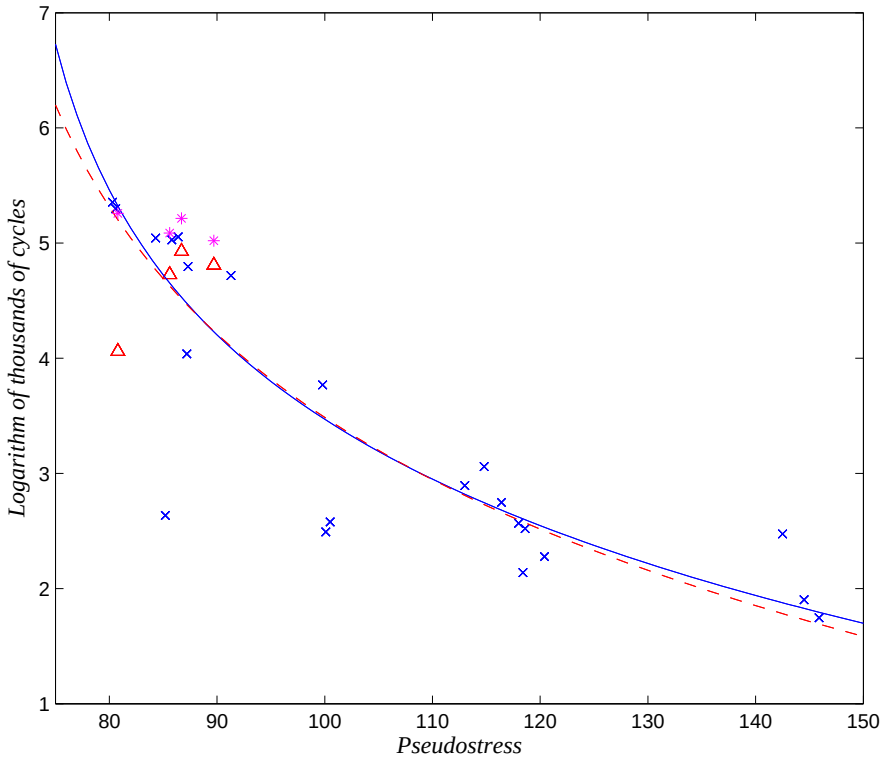


Figure 3.4.3: *Nonlinear regression for the fatigue data. The solid, respectively, dashed line represents the estimated regression curve for the new method, respectively, Stute's method. Uncensored data points are given by* $\times$*, and censored observations by* $\triangle$*. The new data points obtained from the new method are represented by* $*$*.*

Chapter 4

Estimation in nonparametric location-scale regression models with censored data

In Chapters 2 and 3, we considered the estimation of the conditional mean of the response given the covariate in the parametric models (1.5.1) and (1.6.1). Since those models are particular cases of model (1.5.3), we could introduce the nonparametric preliminary smoothing into the artificial data points without requiring any assumption on the model. In the completely nonparametric context, estimation of the conditional mean (or more generally the conditional location) is achieved by using local information coming from a window around the covariate. This procedure involves substantial consistency problems when censored data are present, especially if there are regions of the covariate space that contain a large amount of censored data. This will be the main motivation of this chapter. Under the weak model assumption (1.7.1), the idea will be to provide global information to the estimation procedure in order to reduce its consistency problems.

In Section 4.1, the estimation procedure outlined in Section 1.7 will be described in detail and the assumptions needed for the asymptotic results will be mentioned. Uniform consistency and asymptotic normality (following from an asymptotic representation) of the

proposed estimator will then be stated in Section 4.2. The contribution of provided global information will be illustrated and discussed in Section 4.3 through a simulation study. We will compare the proposed estimator with the estimation of the corresponding conditional location based on the completely nonparametric estimator of Beran (1981). Next, several estimated conditional locations will be used to analyse the quasars data (remind Section 1.1). Finally, the appendix of this chapter will contain some results on the location of the residuals estimated under (1.5.3). These will be required in the proofs of Section 4.2. All the results of this chapter can be found in Heuchenne and Van Keilegom (2005b).

4.1 Notations and description of the method

Before explaining the method in detail, let us introduce some more notations. Let $F_\varepsilon(y) = P(\varepsilon \leq y)$ and $S_\varepsilon(y) = 1 - F_\varepsilon(y)$ denote the distribution and survival function of $\varepsilon = (Y - m(X))/\sigma(X)$, m and σ are the location and scale functions of interest. Next, for $E = (Z - m(X))/\sigma(X)$ define $H_\varepsilon(y) = P(E \leq y)$, $H_{\varepsilon\delta}(y) = P(E \leq y, \Delta = \delta)$, $H_\varepsilon(y|x) = P(E \leq y|x)$ and $H_{\varepsilon\delta}(y|x) = P(E \leq y, \Delta = \delta|x)$ $(\delta = 0, 1)$.

As explained in Section 1.7, the idea of our approach is to first estimate $F(\cdot|x)$ under model (1.7.1) and then to plug-in this estimator in the formula of $m(x)$ given in (1.7.2). In order to estimate $F(\cdot|x)$, use is made of equation (1.7.3). $m^0(\cdot)$ and $\sigma^0(\cdot)$ in (1.7.3) are defined and chosen as in Section 2.1. Since they depend themselves also on $F(\cdot|x)$, we estimate $F(\cdot|x)$ by means of the completely nonparametric kernel estimator of Beran (1981). This yields

$$\hat{m}^0(x) = \int_0^1 \tilde{F}^{-1}(s|x) J(s)\, ds$$

and

$$\hat{\sigma}^{02}(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 J(s)\, ds - \hat{m}^{02}(x)$$

as estimators for $m^0(x)$ and $\sigma^{02}(x)$, where the score function J will be chosen in such a way that $\tilde{F}(\cdot|x)$ is consistent on the support of J. Next, estimate the residual distribution F_ε^0

by

$$\hat{F}_{\varepsilon}^{0}(y) = 1 - \prod_{\hat{E}_{(i)}^{0} \leq y, \Delta_{(i)}=1} \left(1 - \frac{1}{n-i+1}\right) , \tag{4.1.1}$$

where $\hat{E}_{i}^{0}$, $\hat{E}_{(i)}^{0}$ and $\Delta_{(i)}$ are defined as in (2.1.4). $F(y|x)$ is therefore estimated by

$$\hat{F}_{1}(y|x) = \hat{F}_{\varepsilon}^{0}\Big(\frac{y - \hat{m}^{0}(x)}{\hat{\sigma}^{0}(x)}\Big). \tag{4.1.2}$$

Finally, define

$$\hat{m}^{T}(x) = a_0 \int_{-\infty}^{\hat{T}_x} yL(\hat{F}_1(y|x))\, d\hat{F}_1(y|x) + \sum_{j=1}^{k} a_j [\hat{F}_1^{-1}(s_j|x) \wedge \hat{T}_x], \tag{4.1.3}$$

where $\hat{T}_x = T\hat{\sigma}^{0}(x) + \hat{m}^{0}(x)$ and $T < \tau_{H_{\varepsilon}^{0}}$. As it is clear from (4.1.3), $\hat{m}^{T}(x)$ is actually estimating

$$m^{T}(x) = a_0 \int_{-\infty}^{T_x} yL(F(y|x))\, dF(y|x) + \sum_{j=1}^{k} a_j [F^{-1}(s_j|x) \wedge T_x],$$

where $T_x = T\sigma^{0}(x) + m^{0}(x)$, which can be made arbitrarily close to $m(x)$, provided $\tau_{F_{\varepsilon}^{0}} \leq \tau_{G_{\varepsilon}^{0}}$.

For sake of comparison, the completely nonparametric estimator of $m(x)$ is given by

$$\tilde{m}^{T}(x) = a_0 \int_{-\infty}^{\tilde{T}_x} yL(\tilde{F}(y|x))\, d\tilde{F}(y|x) + \sum_{j=1}^{k} a_j [\tilde{F}^{-1}(s_j|x) \wedge \tilde{T}_x], \tag{4.1.4}$$

where $\tilde{T}_x < \tau_{H(\cdot|x)}$. Note that we truncate at $\tilde{T}_x$, because of the inconsistency of $\tilde{F}(y|x)$ for $y > \tilde{T}_x$ (see e.g. Van Keilegom and Veraverbeke, 1997b).

Note that in the definition of $\hat{m}^{T}(x)$ we have to truncate at the point $\hat{T}_x$ due to the presence of right censoring. However, T_x is always greater than or equal to the truncation point $\tilde{T}_x$ used in the definition of $\tilde{m}^{T}(x)$, and the difference between the two truncation points can be substantial, especially when the censoring proportion is not uniform over x. Indeed, when there exists a region in the interval R_X of 'light' censoring, then the estimator $\hat{F}_{\varepsilon}^{0}$ of the error distribution remains consistent up to far in the right tail (and hence T_x will be large), whereas $\tilde{T}_x$ completely depends on the censoring proportion at the point x.

In heavy censored regions, $\tilde{T}_x$ can therefore be quite small. This is the main motivation for using $\hat{m}^T(x)$ instead of the completely nonparametric estimator $\tilde{m}^T(x)$.

The functions $\xi(z,\delta,y|x)$, $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$ of Section 2.1 are again needed in the statement of the asymptotic results given in Section 4.2. Moreover, the function B which appears in the asymptotic representation of Theorem 4.2.2 is defined by

$$B(z,\delta|x) = -f_X^{-1}(x)\sigma^0(x)\left\{\Big[a_0\int_0^{F_\varepsilon^0(T)} L(s)ds + \sum_{j=1}^k a_j\Big]\eta(z,\delta|x) \right.$$
$$\left. + \Big[a_0\int_0^{F_\varepsilon^0(T)} (F_\varepsilon^0)^{-1}(s)L(s)ds + \sum_{j=1}^k a_j((F_\varepsilon^0)^{-1}(s_j)\wedge T)\Big]\zeta(z,\delta|x)\right\}.$$

The assumptions we need to prove the main results of Section 4.2 are listed below.

$(A1)(i)-(iii)$, $(A2)(i)-(ii)$, $(A4)(i)$ and $(A8)$ mentioned in Section 2.1.

$(A3)(i)$ mentioned in Section 2.1.
$(ii)'$ m^0 and σ^0 are three times continuously differentiable and $\inf_{x\in R_X}\sigma^0(x) > 0$.
$(iii)'$ $E[\varepsilon^{02}] < \infty$ and $E|E^0| < \infty$.

$(A6)'$ For $L(y|x) = H(y|x), H_1(y|x), H_\varepsilon^0(y|x)$ or $H_{\varepsilon 1}^0(y|x)$: $L'(y|x)$ is continuous in (x,y) and $\sup_{x,y}|y^2L'(y|x)| < \infty$, the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three.

$(A11)(i)$ Let $s_\alpha < F_\varepsilon^0(T)$ and s_β be such that $0 < s_\alpha < s_j < s_\beta < 1$ for all $j = 1,\ldots,k$, and let $Q = [s_\alpha, s_\beta \wedge F_\varepsilon^0(T)]$. Then, $\inf_{s\in Q} f_\varepsilon^0((F_\varepsilon^0)^{-1}(s)) > 0$.
(ii) L is three times continuously differentiable, $\int_0^1 L(s)ds = 1$, $L(s) \geq 0$ for all $0 \leq s \leq 1$.

4.2 Asymptotic results

In this section, we show the consistency of $\hat{m}^T(x)$ uniformly over x. We also develop an asymptotic representation for $\hat{m}^T(x) - m^T(x)$, which is useful for obtaining afterwards the asymptotic normality.

Theorem 4.2.1 *Assume (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i), m^0 and σ^0 are twice continuously differentiable and* $\inf_{x\in R_X}\sigma^0(x)>0$, *(A3) (iii)', (A4) (i), (A6)', (A11) (i) and L continuously differentiable with* $\int_0^1 L(s)ds=1$, $L(s)\geq 0$ *for all* $0\leq s\leq 1$. *Then,*

$$\sup_{x\in R_X}|\hat{m}^T(x)-m^T(x)|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Proof. Write for any $x\in R_X$,

$$\begin{aligned}
&\hat{m}^T(x)-m^T(x)\\
&=a_0\hat{m}^0(x)\left\{\int_{-\infty}^T L(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)-\int_{-\infty}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e)\right\}\\
&\quad+\{\hat{m}^0(x)-m^0(x)\}\Big\{a_0\int_{-\infty}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e)+\sum_{j=1}^k a_j\Big\}\\
&\quad+a_0\hat{\sigma}^0(x)\left\{\int_{-\infty}^T eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)-\int_{-\infty}^T eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)\right\}\\
&\quad+\{\hat{\sigma}^0(x)-\sigma^0(x)\}\left\{a_0\int_{-\infty}^T eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)+\sum_{j=1}^k a_j((F_\varepsilon^0)^{-1}(s_j)\wedge T)\right\}\\
&\quad+\hat{\sigma}^0(x)\left[\sum_{j=1}^k a_j\Big\{(\hat{F}_\varepsilon^0)^{-1}(s_j)\wedge T-(F_\varepsilon^0)^{-1}(s_j)\Big\}I(s_j\leq\hat{F}_\varepsilon^0(T),s_j\leq F_\varepsilon^0(T))\right.\\
&\quad+\sum_{j=1}^k a_j\Big\{T-(F_\varepsilon^0)^{-1}(s_j)\Big\}I(\hat{F}_\varepsilon^0(T)<s_j\leq F_\varepsilon^0(T))\\
&\quad\left.+\sum_{j=1}^k a_j\Big\{(\hat{F}_\varepsilon^0)^{-1}(s_j)\wedge T-T\Big\}I(F_\varepsilon^0(T)<s_j\leq\hat{F}_\varepsilon^0(T))\right]\\
&=\sum_{\ell=1}^7 A_\ell(x).
\end{aligned}$$

Since $E|\varepsilon^0|<\infty$, $\sup_x|A_2(x)|$ and $\sup_x|A_4(x)|$ are $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ using Proposition 4.5 in Van Keilegom and Akritas (1999) (hereafter again abbreviated by VKA). From Corollary 3.2 in VKA (1999) and Theorem 1 in Doss and Gill (1992), we obtain that $\sup_{s\in Q}|(\hat{F}_\varepsilon^0)^{-1}(s)-(F_\varepsilon^0)^{-1}(s)|=O_P(n^{-1/2})$ and hence $\sup_x|A_5(x)|=O_P(n^{-1/2})$. For $\sup_x|A_3(x)|$, we use Lemma 4.5.3. In a similar way, it can be shown that $\sup_x|A_1(x)|$ is of negligible order. Finally, $A_6(x)$ and $A_7(x)$ are uniformly negligible using Corollary 3.2 in

VKA (1999).

Theorem 4.2.2 *Assume (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i), (ii)', (iii)', (A4) (i), (A6)', (A8), (A11) and* $\sup_e |e^3(f_\varepsilon^0)''(e)| < \infty$. *Then, for any* $x \in R_X$,

$$\hat{m}^T(x) - m^T(x) = (na_n)^{-1} \sum_{i=1}^{n} K(\frac{x - X_i}{a_n}) B(Z_i, \Delta_i|x) + R_n(x),$$

where $\sup\{|R_n(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$ *and* $B(z, \delta|x)$ *is defined in Section 4.1.*

Proof. We use the same decomposition of $\hat{m}^T(x) - m^T(x)$ as in the proof of Theorem 4.2.1. Using Propositions 4.8 and 4.9 in VKA (1999) and the fact that $E|\varepsilon^0| < \infty$, we obtain that

$$A_2(x) = -\Big[a_0 \int_0^{F_\varepsilon^0(T)} L(s)ds + \sum_{j=1}^{k} a_j\Big](na_n)^{-1} f_X^{-1}(x)\sigma^0(x) \sum_{i=1}^{n} K\Big(\frac{x - X_i}{a_n}\Big)\eta(Z_i, \Delta_i|x) + R_n(x),$$

and

$$\begin{aligned} A_4(x) &= -(na_n)^{-1} f_X^{-1}(x)\sigma^0(x)\Big\{a_0 \int_0^{F_\varepsilon^0(T)} (F_\varepsilon^0)^{-1}(s)L(s)ds \\ &\quad + \sum_{j=1}^{k} a_j((F_\varepsilon^0)^{-1}(s_j) \wedge T)\Big\} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{a_n}\Big)\zeta(Z_i, \Delta_i|x) + R_n(x), \end{aligned}$$

where $R_n(x) = O_P((na_n)^{-3/4}(\log n)^{3/4})$. For $A_3(x)$ (and similarly for $A_1(x)$) we use Lemma 4.5.4. The remaining terms $A_5(x)$, $A_6(x)$ and $A_7(x)$ are $o_P((na_n)^{-1/2})$, as shown in the proof of Theorem 4.2.1. Therefore,

$$\hat{m}^T(x) - m^T(x) = (na_n)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{a_n}\Big) B(Z_i, \Delta_i|x) + o_P((na_n)^{-1/2}).$$

Theorem 4.2.3 *Under the assumptions of Theorem 4.2.2,*

$$(na_n)^{1/2}(\hat{m}^T(x) - m^T(x)) \xrightarrow{d} N(0, s^2(x)),$$

where

$$s^2(x) = \int K^2(u)du \sum_{\delta=0,1} \int B^2(z, \delta|x) f_X(x)\, dH_\delta(z|x).$$

Proof. The result is obtained by using Lyapounov's Theorem. It's easy to check that the Lyapounov ratio is $O((na_n)^{-1/2})$.

Remark 4.2.4 (Choice of the bandwidth) A first way to compute the bandwidth parameter a_n would be to calculate the asymptotic bias and variance obtained from the moments of $R_n(x)$ in the representation of Theorem 4.2.2. However, this task seems fastidious since $R_n(x)$ highly depends on representations in theorems of VKA (1999) for which the moments of the lower order terms are not calculated. Therefore, the bootstrap procedure of Section 2.3.1 is preferred. For each resample $\{(X_i^{j*}, Z_i^{j*}, \Delta_i^{j*}) : i = 1, \ldots, n\}$, $j = 1, \ldots, B$ for some large B, let $\hat{m}_{a_n}^{*jT}(x)$ be the estimator of $m^T(x)$ obtained by using bandwidth a_n. From this, the integrated mean squared error $\int E[\hat{m}^T(x) - m^T(x)]^2\, dx$ can be approximated by

$$IMSE^*(a_n) = B^{-1} \sum_{j=1}^{B} \int [\hat{m}_{a_n}^{*jT}(x) - \hat{m}_{g_n}^{T}(x)]^2\, dx,$$

where $\hat{m}_{g_n}^T(x)$ is the estimator of $m^T(x)$ obtained with the initial sample and the pilot bandwidth g_n. We now select the value of a_n that minimizes $IMSE^*(a_n)$. The same bootstrap procedure can also be used to approximate the distribution of $\hat{m}^T(x)$, instead of using the above asymptotic distribution which might be hard to estimate in practice.

Remark 4.2.5 (Scale estimation) A similar idea as the one developed above to estimate $m(x)$ can be used to estimate any scale function $\sigma(x)$. Indeed, the principle to use equation (1.7.3) in order to better estimate the right tail of the distribution $F(y|x)$ can also be applied in the construction of an estimator of $\sigma(x)$. Define

$$\begin{aligned}\hat{\sigma}^{T2}(x) = a_0^2 &\left\{ \int_{-\infty}^{\hat{T}_x} y^2 J(\hat{F}(y|x))\, d\hat{F}(y|x) - \hat{m}^{T2}(x) \right\} \\ &+ \sum_{j=1}^{k} a_j^2 \left\{ \int_{-\infty}^{\hat{T}_x} \rho_j(y - \hat{F}^{-1}(s_j|x) \wedge \hat{T}_x)\, d\hat{F}(y|x) \right\}^2,\end{aligned}$$

where $\rho_j(u) = s_j u I(u \geq 0) + (s_j - 1) u I(u < 0)$. The asymptotic results for $\hat{\sigma}^{T2}(x)$ can be obtained along the same lines as for the estimator $\hat{m}^T(x)$.

Remark 4.2.6 (Homoscedastic model) Note that when model (1.7.1) is homoscedastic and we estimate σ^0 by a global estimator $\hat{\sigma}^0$, the representation in Theorem 4.2.2 simplifies. In fact, it is easily seen that the function $\zeta(z, \delta|x)$ in the definition of $B(z, \delta|x)$ equals zero in that case.

Remark 4.2.7 (Implementation) The estimator $\hat{m}^T(x)$ is easy to implement in practice, and the parameters on which it depends (namely the truncation point T, the bandwidth a_n and the score function J) can be chosen in a data driven way. In Remark 4.2.4, we already explained how to choose the bandwidth a_n by means of a bootstrap procedure. The truncation point T can be taken equal to the largest residual $\hat{E}^0_{(n)}$ and the function J can be chosen as in Section 2.3.1.

4.3 Simulations

In this section, we compare the finite sample behaviour of the completely nonparametric location estimator $\tilde{m}^T(x)$ with the location estimator $\hat{m}^T(x)$ proposed in this paper by means of Monte Carlo simulations. We are interested in the behaviour of the integrated mean squared error of the estimators, defined by $IMSE = \int E[\hat{m}^T(x) - m(x)]^2 \, dx$ for $\hat{m}^T(x)$ and similarly for $\tilde{m}^T(x)$. The simulations are carried out for samples of size $n = 100$ and the results are obtained by using 250 simulations.

In the first setting, we generate i.i.d. data from the normal homoscedastic regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \sigma\varepsilon, \tag{4.3.1}$$

for various choices of $\beta_0, \beta_1, \beta_2, \beta_3$ and σ, where X has a uniform distribution on the interval $[0, 3]$, and the error term ε is a normal random variable with zero mean and variance 1. The censoring variable C satisfies $C = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \sigma\varepsilon^*$, for certain choices of α_0, α_1, α_2, and α_3, where ε^* has a normal distribution with zero mean and variance 1. We further assume that ε, ε^* and σ are independent of X and that ε is independent of ε^*.

Under this model, we have

$$P(\Delta = 0|X = x) = 1 - \Phi\Big(\frac{\alpha_0 - \beta_0 + (\alpha_1 - \beta_1)x + (\alpha_2 - \beta_2)x^2 + (\alpha_3 - \beta_3)x^3}{\sqrt{2}\sigma}\Big).$$

For the weights that appear in the Beran estimator $\tilde{F}(y|x)$, we choose a biquadratic kernel function $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. For the bandwidth sequence a_n, we select for each estimator the minimizer of an approximated $IMSE$ among a grid of 20 possible values of a_n between 0 and 3. This $IMSE$ is computed as follows. For each a_n and each simulation, we compute an integrated squared error (ISE) using the true parameters of the model (4.3.1) and we obtain the approximated $IMSE$ for each a_n by averaging those ISE over the 250 simulations. A bootstrap technique for computing the smoothing parameter is proposed in Section 4.2, but for simulations it is too computationally intensive. When needed, we enlarge the window $[x - a_n, x + a_n]$ such that it contains at least one X_i (in its interior) for which the corresponding Y_i is uncensored. When the bandwidth a_n at a point x is larger than the distance from x to both the left and right endpoint of the interval, it is redefined as the maximum of these two distances. Finally, we work with $J(s) = I(s \leq b)/b$, where $b = \min_{1 \leq i \leq n} \tilde{F}(+\infty|X_i)$. For the estimators $\tilde{F}(y|x)$ and $\hat{F}^0_\varepsilon(y)$, the last data point or the last residual is often censored. In this case, this point is redefined as uncensored.

We compare the two methods for four different locations : the conditional mean, the conditional truncated mean ($L(s) = (1/0.9)I(0.05 < s \leq 0.95)$), the conditional median and conditional third quartile.

Tables 4.1, 4.2 and 4.3 summarize the simulation results for different values of the parameters $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, $\beta_0, \beta_1, \beta_2, \beta_3$ and σ. For fixed values of $\beta_0, \beta_1, \beta_2, \beta_3$ and σ, the values of $\alpha_0, \alpha_1, \alpha_2$ and α_3 are chosen in such a way that some variation in the censoring probability curves is obtained. The proportion of censoring (in % and denoted by CP in the tables) is computed like in the previous chapters.

The tables show that, in general, $\hat{m}^T(x)$ has a smaller $IMSE$ than $\tilde{m}^T(x)$ for each of the four considered location functions. The higher the quantile, or the smaller the support of L, the worse the estimation. The new method resists however better. The simulations can be explained as follows. The most important problem of the Beran estimator is its consistency in the right tail : this is mainly due to the fact that it is a local estimator.

In regions with a large proportion of censored data, the Beran estimator therefore behaves badly. The other estimator also has this problem but at a lower degree : it uses a global estimator of the distribution of the residuals. The inconsistency problems arise thus in the right tail of a global distribution. On the other hand, the new approach is based on the estimation of $m^0(\cdot)$ and $\sigma^0(\cdot)$. The score function J in these functions is determined by $\min_{1\leq i\leq n} \tilde{F}(+\infty|X_i)$. When censoring is heavy, this value can be small. In that case, the estimators $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ will be quite variable and unstable.

The results of Tables 4.1, 4.2 and 4.3 show that the relative performance of the two methods depends on the shape of the regression function and the amount of censoring. In fact, when the regression function is relatively flat, the optimal bandwidth will be quite large. Hence, there will be little difference between the local and global estimators. Table 4.1 summarizes the results for this kind of regression functions. When the regression function becomes more and more wiggly, the merits of the proposed method become clearer (see Tables 4.2 and 4.3). In Tables 4.2 and 4.3, the models are more wiggly, leading to smaller bandwidth parameters and hence the advantages of the new estimator in comparison with the completely nonparametric estimator become more and more transparent. The nice

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
0	0.4	0	0	37.1	0.326	0.331	0.349	0.404
-0.4	1	-0.05	0	0.5	0.320	0.322	0.336	0.365
0	0.4	0	0	38.2	0.357	0.361	0.381	0.429
0.3	0.4	0	0	0.5	0.355	0.356	0.369	0.395
0	0.4	0	0	58.8	0.390	0.396	0.454	0.569
0.24	0	0	0.02	0.5	0.390	0.388	0.408	0.507
0	0.4	0	0	71.1	0.394	0.414	0.507	0.718
-0.3	0	0	0.05	0.5	0.384	0.390	0.445	0.586

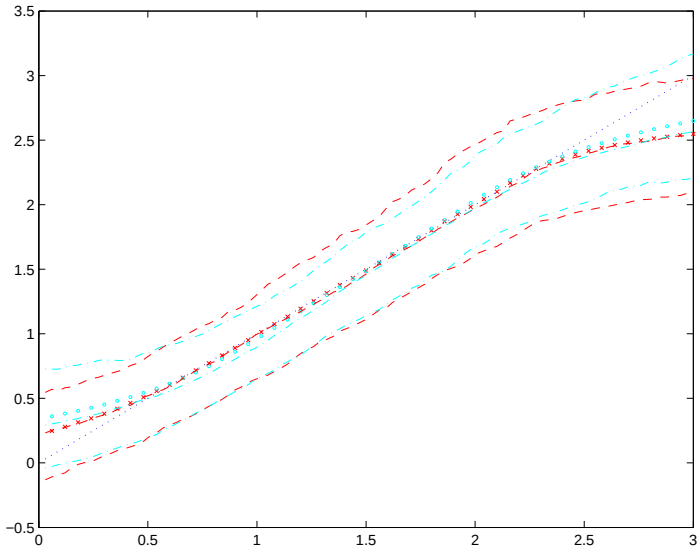
Table 4.1: *Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (4.3.1) with large optimal bandwidth a_n.*

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
0	1	0	0	35.5	1.759	1.765	1.802	2.148
2	0	-0.2	0.09	0.5	1.749	1.747	1.762	1.772
0	1	0	0	38.2	1.333	1.347	1.392	1.604
0.3	1	0	0	0.5	1.299	1.303	1.319	1.354
0	1	0	0	58.0	1.631	1.681	1.862	1.926
0.5	0.13	0.2	0	0.5	1.517	1.525	1.547	1.676
0	1	0	0	72.0	1.760	1.832	2.091	2.015
0	0.4	0.1	0	0.5	1.618	1.626	1.698	1.824

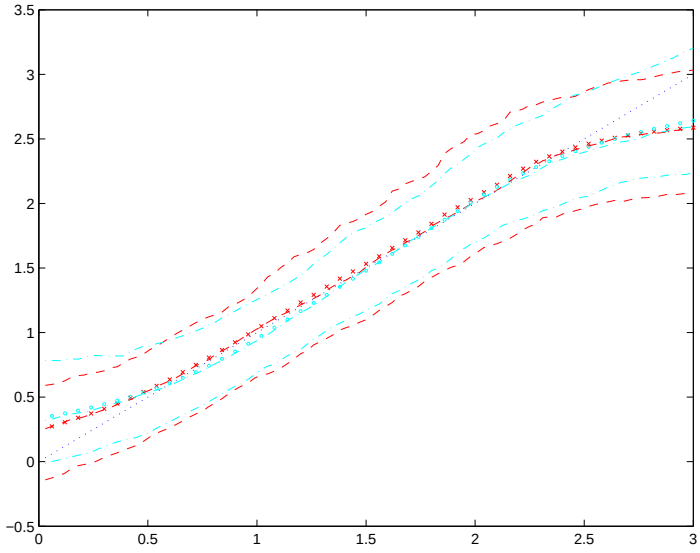
Table 4.2: *Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (4.3.1) with moderately large optimal bandwidth a_n.*

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
4	-7.5	6	-1.3	31.7	1.139	1.159	1.260	1.570
3.5	-7.45	7	-1.6	0.5	1.081	1.085	1.100	1.165
4	-7.5	6	-1.3	38.2	1.047	1.066	1.161	1.513
4.3	-7.5	6	-1.3	0.5	1.030	1.034	1.043	1.111
4	-7.5	6	-1.3	51.3	1.251	1.314	1.508	1.559
3.2	-7.6	7	-1.6	0.5	1.142	1.158	1.188	1.315
4	-7.5	6	-1.3	56.4	1.336	1.392	1.553	2.043
3	-7.6	7	-1.6	1	1.296	1.321	1.391	1.620

Table 4.3: *Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (4.3.1) with small optimal bandwidth a_n.*

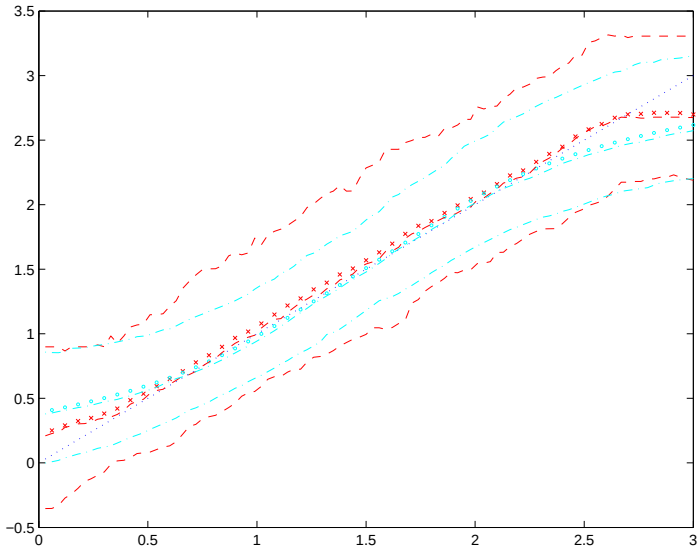


(a)

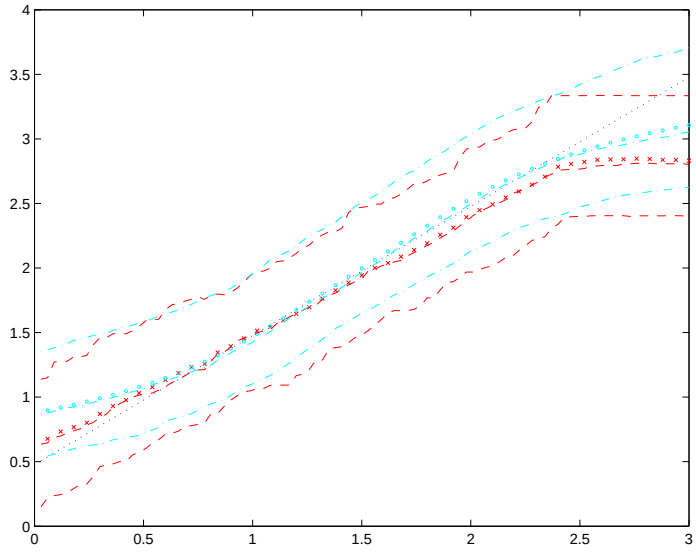


(b)

Figure 4.3.1: Comparison of pointwise biases and variabilities between the new nonparametric and the Beran method. The dashed, respectively, dashed dotted curves are obtained from the Beran, respectively, new nonparametric method. Dots indicate the true curve and $\times$, respectively, $\circ$ represent the mean curve for the Beran, respectively, new method. (a) Conditional mean for the last simulation of Table 4.2; (b) Conditional truncated mean for the last simulation of Table 4.2.



(a)



(b)

Figure 4.3.2: Comparison of pointwise biases and variabilities between the new nonparametric and the Beran method. The dashed, respectively, dashed dotted curves are obtained from the Beran, respectively, new nonparametric method. Dots indicate the true curve and $\times$, respectively, $\circ$ represent the mean curve for the Beran, respectively, new method. (a) Conditional first quartile for the last simulation of Table 4.2; (b) Conditional third quartile for the last simulation of Table 4.2.

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	γ^2	mean	trunc. mean	median	3^{rd} quartile
0	0.4	0	0	58.2	0.365	0.377	0.425	0.957
-0.1	0	0	0.1	0.1	0.338	0.347	0.335	0.943
0	1	6	-4	48.9	0.621	0.631	0.638	0.950
0.5	1	-5	9	1	0.570	0.566	0.557	0.866
0	1	6	-4	56.8	1.040	1.066	1.152	2.546
0.5	0.8	-6	8.5	5	1.032	1.032	1.069	2.161

Table 4.4: *Results for $\tilde{m}^T(x)$ (first line) and $\hat{m}^T(x)$ (second line) for model (4.3.2). R_X is $[0,3]$ for the first model and $[0,1]$ for the two other ones.*

behaviour of the new estimator is also reflected in Figures 4.3.1 and 4.3.2 where the graphs are constructed in the same way as in Figures 3.3.1 and 3.3.2.

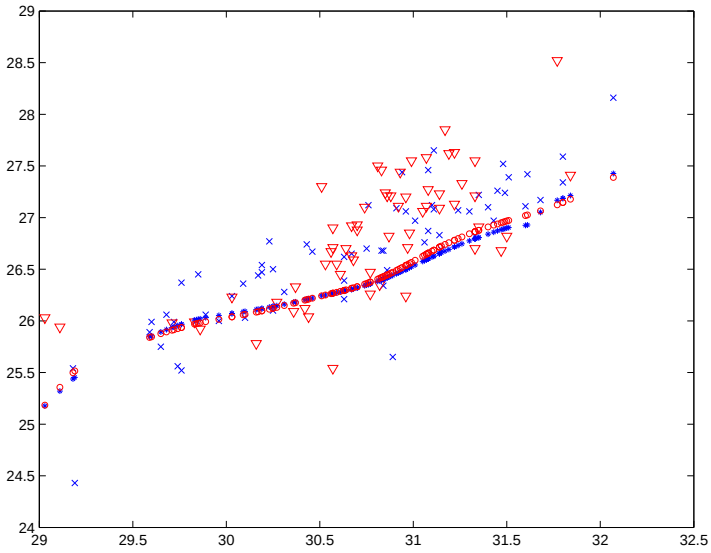
The final setting we consider is a normal heteroscedastic regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + (\gamma X + 0.1)\varepsilon, \tag{4.3.2}$$

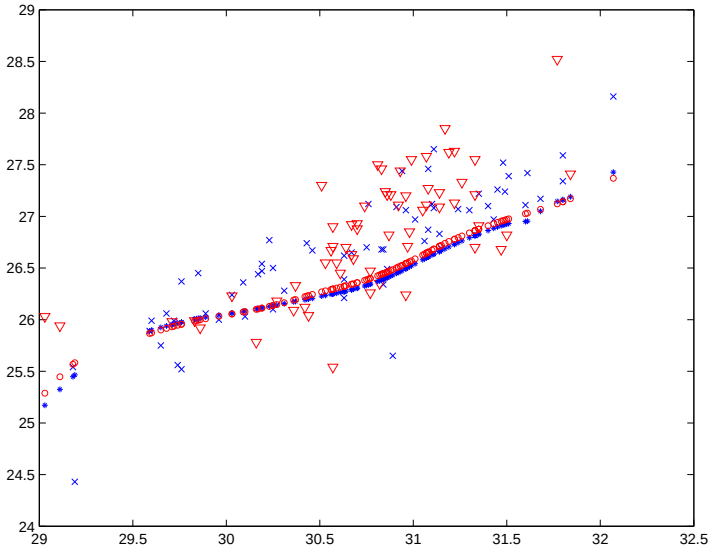
where X has a uniform distribution on $[0,1]$ or on $[0,3]$, and ε has a normal distribution with zero mean and variance equal to one. The censoring variable is given by $C = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \gamma\varepsilon^*$, where ε^* has a normal distribution with zero mean and variance equal to one. We further assume that ε and ε^* are independent of X, and that ε is independent of ε^*. The variance of Y given X is now supposed to be unknown. The results are in Table 4.4. Not surprisingly, one can show that when the degree of heteroscedasticity is small, the gain in precision of $\hat{m}^T(x)$ with respect to $\tilde{m}^T(x)$ is relatively small, since $\hat{m}^T(x)$ looses some precision due to the estimation of the scale function $\sigma^0(x)$. However, the estimator $\hat{m}^T(x)$ still outperforms $\tilde{m}^T(x)$ for all models and all location functions considered.

4.4 Data analysis

We apply the proposed method on the study of quasars in astronomy (see Section 1.1). We show in Figures 4.4.3 and 4.4.4 the results of regression of l_x on l_{uv} for the new estimator $\hat{m}^T(x)$ and the completely nonparametric estimator $\tilde{m}^T(x)$. The bandwidth is selected from a grid of 18 bandwidths, according to the method described in Remark 4.2.4. The selected bandwidth parameter is approximately the same for each method (around 0.75). For the conditional mean, truncated mean and median, we observe a strong linear relation between the two variables for both methods. This suggests to fit a linear model to these data (as done in Section 2.4). For the first quartile, this relation is not so obvious for $\tilde{m}^T(x)$, while the new estimator again suggests to choose a linear model. Note that, contrary to the simulation section, we focus here on the first and not the third quartile. This is because for left censored data, the first quartile is harder to estimate, and hence it interests us more.

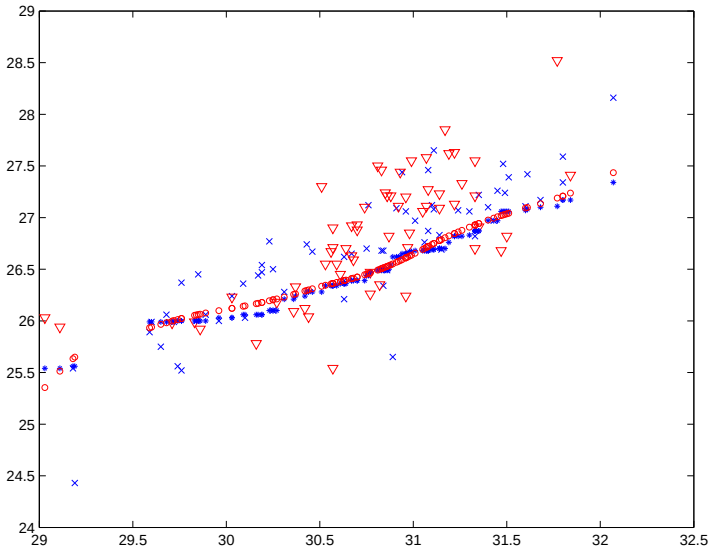


(a)

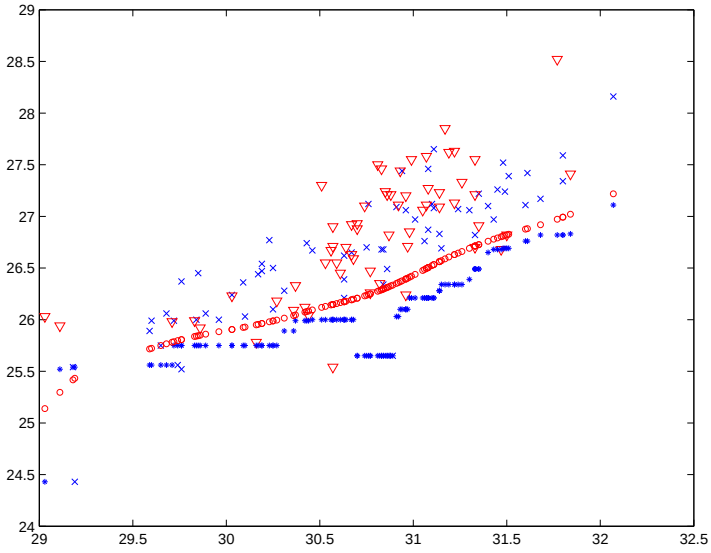


(b)

Figure 4.4.3: Regression curve estimation for the quasars data. The estimators $\tilde{m}^T(x)$ and $\hat{m}^T(x)$ are indicated by $*$ and $\circ$ respectively. Uncensored data points are represented by $\times$, and (left) censored observations by $\bigtriangledown$. (a) Conditional mean; (b) Conditional truncated mean (5 percent of truncation at both sides).



(a)



(b)

Figure 4.4.4: Regression curve estimation for the quasars data. The estimators $\tilde{m}^T(x)$ and $\hat{m}^T(x)$ are indicated by $*$ and $\circ$ respectively. Uncensored data points are represented by $\times$, and (left) censored observations by $\bigtriangledown$. (a) Conditional median; (b) Conditional first quartile.

4.5 Appendix : Some more results on the location of the residuals in a heteroscedastic regression model.

In this section, we denote $T^i = \frac{T_{X_i} - m^0(X_i)}{\sigma^0(X_i)} = T$ for all i, $i = 1, \ldots, n$, $\hat{T}^i = \frac{T_{X_i} - \hat{m}^0(X_i)}{\hat{\sigma}^0(X_i)}$, $E_i^{0T} = E_i^0 \wedge T$ and $\hat{E}_i^{0T} = \hat{E}_i^0 \wedge \hat{T}^i$.

Lemma 4.5.1 *Assume (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i), (ii)', (iii)', (A4) (i), (A6)', (A8), (A11) (ii), and* $\sup_e |e^3 (f_\varepsilon^0)''(e)| < \infty$. *Then,*

$$n^{-1} \sum_{i=1}^{n} \left\{ \hat{E}_i^0 L(\hat{F}_\varepsilon^0(\hat{E}_i^0)) I(\hat{E}_i^0 \le \hat{T}^i) I(\Delta_i = 1) + \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(\hat{F}_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(\Delta_i = 0) \right\}$$
$$- \int_{-\infty}^{T} eL(F_\varepsilon^0(e))\, dF_\varepsilon^0(e) = o_P((na_n)^{-1/2}).$$

Proof. We first consider

$$n^{-1} \sum_{i=1}^{n} \{ \hat{E}_i^0 L(\hat{F}_\varepsilon^0(\hat{E}_i^0)) I(\hat{E}_i^0 \le \hat{T}^i) - E_i^0 L(F_\varepsilon^0(E_i^0)) I(E_i^0 \le T) \} I(\Delta_i = 1) \tag{4.5.1}$$
$$+ n^{-1} \sum_{i=1}^{n} \Big\{ \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(\hat{F}_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} - \frac{\int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e))\, dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_i^{0T})} \Big\} I(\Delta_i = 0) = A_1 + A_2.$$

Using Corollary 3.2 and Proposition 4.5 in VKA (1999), the differentiability of L and the fact that $E|\varepsilon^0| < \infty$, we have

$$A_1 = n^{-1} \sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1) \{ (\hat{E}_i^0 - E_i^0) L(F_\varepsilon^0(\hat{E}_i^0))$$
$$+ E_i^0 [L(F_\varepsilon^0(\hat{E}_i^0)) - L(F_\varepsilon^0(E_i^0))] \} + O_P(n^{-1/2}).$$

Next, using Proposition 4.5 in VKA (1999), and the fact that $\sup_y |y^2 f_\varepsilon^{0\prime}(y)| < \infty$ and $\sup_y |y f_\varepsilon^0(y)| < \infty$,

$$F_\varepsilon^0(\hat{E}_i^0) - F_\varepsilon^0(E_i^0)$$
$$= (\hat{E}_i^0 - E_i^0) f_\varepsilon^0(E_i^0) + o_P((na_n)^{-1/2})$$
$$= -\frac{\hat{m}^0(X_i) - m^0(X_i)}{\sigma^0(X_i)} f_\varepsilon^0(E_i^0) - \frac{\hat{\sigma}^0(X_i) - \sigma^0(X_i)}{\sigma^0(X_i)} E_i^0 f_\varepsilon^0(E_i^0) + o_P((na_n)^{-1/2}). \tag{4.5.2}$$

From this, the fact that L is twice continuously differentiable and that $E|\varepsilon^0| < \infty$, A_1 can be rewritten as

$$\begin{aligned} A_1 &= n^{-1}\sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1)(\hat{E}_i^0 - E_i^0)[L(F_\varepsilon^0(E_i^0)) + L'(F_\varepsilon^0(E_i^0))E_i^0 f_\varepsilon^0(E_i^0)] \\ &\quad + o_P((na_n)^{-1/2}) \\ &= (n^2 a_n)^{-1}\sum_{i=1}^{n} I(Z_i \le T_{X_i}) I(\Delta_i = 1) f_X^{-1}(X_i)[L(F_\varepsilon^0(E_i^0)) + L'(F_\varepsilon^0(E_i^0))E_i^0 f_\varepsilon^0(E_i^0)] \\ &\quad \times \Big\{ \sum_{j=1}^{n} K\Big(\frac{X_i - X_j}{a_n}\Big)[\eta(Z_j, \Delta_j | X_i) + \zeta(Z_j, \Delta_j | X_i) E_i^0] \Big\}, \end{aligned} \tag{4.5.3}$$

where the last equality follows from Propositions 4.8 and 4.9 in VKA (1999). Next, we treat the term A_2. Using Corollary 3.2 in VKA (1999), Lemma 2.5.1 and the uniform consistency of $\hat{m}^0$ and $\hat{\sigma}^0$ in (4.5.2) (see Proposition 4.5 in VKA (1999)), we have

$$\begin{aligned} A_2 &= n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\Big\{ \frac{\hat{F}_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T}))(1 - F_\varepsilon^0(E_i^{0T}))} \int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(F_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(e) \\ &\quad + \frac{1}{1 - F_\varepsilon^0(E_i^{0T})}\Big[\int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(F_\varepsilon^0(e))\, d\hat{F}_\varepsilon^0(e) - \int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e))\, dF_\varepsilon^0(e) \Big] \Big\} + o_P((na_n)^{-1/2}) \\ &= n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\{A_{21i} + A_{22i} + A_{23i}\} + o_P((na_n)^{-1/2}). \end{aligned} \tag{4.5.4}$$

For A_{21i}, we write

$$\begin{aligned} \int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e) &= \int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e)) dF_\varepsilon^0(e) + \int_{T}^{\hat{T}^i} eL(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e) \\ &\quad + \int_{\hat{E}_i^{0T}}^{E_i^{0T}} eL(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e) + \int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e)) d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e)) \\ &= B_{1i} + B_{2i} + B_{3i} + B_{4i}. \end{aligned}$$

Easy calculations show that the three last terms of this expression are $|E_i^{0T}|$ $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ uniformly in i, such that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0) A_{21i} &= n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0) \frac{F_\varepsilon^0(\hat{E}_i^{0T}) - F_\varepsilon^0(E_i^{0T})}{(1 - F_\varepsilon^0(E_i^{0T}))^2} \int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e))\, dF_\varepsilon^0(e) \\ &\quad + o_P((na_n)^{-1/2}), \end{aligned} \tag{4.5.5}$$

using the fact that $E|E^{0T}| < \infty$. Next,

$$n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\{A_{22i} + A_{23i}\} = n^{-1}\sum_{i=1}^{n} I(\Delta_i = 0)\frac{(B_{2i} + B_{3i} + B_{4i})}{1 - F_\varepsilon^0(E_i^{0T})}.$$

B_{4i} is $|E_i^{0T}|o_P((na_n)^{-1/2})$ uniformly in i. For B_{3i}, we write

$$\begin{aligned}
&\int_{\hat{E}_i^{0T}}^{E_i^{0T}} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e) + \int_{\hat{E}_i^{0T}}^{E_i^{0T}} eL(F_\varepsilon^0(e))d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e)) \\
&= -\Big\{\int_0^{\hat{E}_i^{0T}} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e) - \int_0^{E_i^{0T}} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)\Big\} + |E_i^{0T}|o_P((na_n)^{-1/2}) \\
&= -E_i^{0T}L(F_\varepsilon^0(E_i^{0T}))f_\varepsilon^0(E_i^{0T})[\hat{E}_i^{0T} - E_i^{0T}] + |E_i^{0T}|o_P((na_n)^{-1/2}).
\end{aligned}$$

The last equality is obtained using Proposition 4.5 in VKA (1999), the fact that L is continuously differentiable, that $\sup_e |ef_\varepsilon^0(e)| < \infty$ and that $\sup_e |e^2 f_\varepsilon^{0\prime}(e)| < \infty$. A similar expression is found for B_{2i}. This together with (4.5.5), (4.5.4), (4.5.3) and (4.5.1) leads to

$$A_1 + A_2 = (n^2 a_n)^{-1}\sum_{i\neq j} B_0(X_i, Z_i, \Delta_i, Z_j, \Delta_j)K\Big(\frac{X_i - X_j}{a_n}\Big) + o_P((na_n)^{-1/2}), \quad (4.5.6)$$

where

$$\begin{aligned}
B_0(X_i, Z_i, \Delta_i, Z_j, \Delta_j) = f_X^{-1}(X_i)\Big[&I(\Delta_i = 1, Z_i \le T_{X_i})M'(E_i^0)\gamma_{ij}(E_i^0) \\
&+ I(\Delta_i = 0)\left\{\frac{\int_{E_i^{0T}}^{T} M(e)dF_\varepsilon^0(e)}{(1 - F_\varepsilon^0(E_i^{0T}))^2} - \frac{M(E_i^{0T})}{1 - F_\varepsilon^0(E_i^{0T})}\right\} f_\varepsilon^0(E_i^{0T})\gamma_{ij}(E_i^{0T}) \\
&+ I(\Delta_i = 0)\frac{M(T)}{1 - F_\varepsilon^0(E_i^{0T})} f_\varepsilon^0(T)\gamma_{ij}(T)\Big],
\end{aligned}$$

$M(e) = eL(F_\varepsilon^0(e))$ and $\gamma_{ij}(e) = \eta(Z_j, \Delta_j|X_i) + e\zeta(Z_j, \Delta_j|X_i)$.
The expression (4.5.6) is then treated as (2.2.7) in Theorem 2.2.1 such that

$$A_1 + A_2 = n^{-1}\sum_{j=1}^{n} f_X(X_j)\int \sum_{\delta=0,1} B_0(X_j, z, \delta, Z_j, \Delta_j)\, dH_\delta(z|X_j) + O(a_n^2) = O_P(n^{-1/2}),$$

since it is a sum of i.i.d. random variables with zero mean. The result now follows since it is easily seen that (using $E[\varepsilon^{02}] < \infty$)

$$n^{-1}\sum_{i=1}^{n}\Big\{E_i^0 L(F_\varepsilon^0(E_i^0))I(E_i^0 \le T)I(\Delta_i = 1) + \frac{\int_{E_i^{0T}}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_i^{0T})} I(\Delta_i = 0)\Big\}$$

$$-\int_{-\infty}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)| = O_P(n^{-1/2}).$$

Remark 4.5.2 *A weaker version of Lemma 4.5.1 can be obtained under less restrictive conditions. In fact, it can be easily seen that if (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i),* m^0 *and* σ^0 *are twice continuously differentiable and* $\inf_{x\in R_X}\sigma^0(x) > 0$*, (A3) (iii)', (A4) (i), (A6)' hold and if* L *is continuously differentiable,* $\int_0^1 L(s)ds = 1$ *and* $L(s) \geq 0$ *for all* $0 \leq s \leq 1$*, then the expression on the left hand side in Lemma 4.5.1 is* $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$.

Lemma 4.5.3 *Assume (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i),* m^0 *and* σ^0 *are twice continuously differentiable and* $\inf_{x\in R_X}\sigma^0(x) > 0$*, (A3) (iii)', (A4) (i), (A6)',* L *is continuously differentiable,* $\int_0^1 L(s)ds = 1$ *and* $L(s) \geq 0$ *for all* $0 \leq s \leq 1$*. Then,*

$$\int_{-\infty}^{T} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) - \int_{-\infty}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e) = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Proof. By Lemma 4.5.1, it suffices to prove that

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\Big\{\hat{E}_i^0 L(\hat{F}_\varepsilon^0(\hat{E}_i^0))I(\hat{E}_i^0 \leq T) - \hat{E}_i^0 L(\hat{F}_\varepsilon^0(\hat{E}_i^0))I(\hat{E}_i^0 \leq \hat{T}^i)\Big\}I(\Delta_i = 1)\\
&+n^{-1}\sum_{i=1}^{n}\left\{\frac{\int_{\hat{E}_i^0\wedge T}^{T} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^0\wedge T)} - \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})}\right\}I(\Delta_i = 0)\\
&= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}). \qquad (4.5.7)
\end{aligned}$$

The left hand side of (4.5.7) can be written as

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\Big\{\hat{E}_i^0 L(\hat{F}_\varepsilon^0(\hat{E}_i^0))[I(\hat{E}_i^0 \leq T) - I(\hat{E}_i^0 \leq \hat{T}^i)]\Big\}I(\Delta_i = 1)\\
&+n^{-1}\sum_{i=1}^{n}\left\{\frac{\int_{\hat{E}_i^0\wedge T}^{T} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^0\wedge T)}I(\hat{E}_i^0 \leq T, \hat{E}_i^0 > \hat{T}^i)\right.\\
&\quad - \frac{\int_{\hat{E}_i^{0T}}^{\hat{T}^i} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})}I(\hat{E}_i^0 > T, \hat{E}_i^0 \leq \hat{T}^i)\\
&\quad \left.+\frac{\int_{\hat{T}^i}^{T} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_i^{0T})}I(\hat{E}_i^0 \leq T, \hat{E}_i^0 \leq \hat{T}^i)\right\}I(\Delta_i = 0). \qquad (4.5.8)
\end{aligned}$$

Using classical arguments, the three last terms in the above expression are $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ and the first one can be rewritten as

$$n^{-1}\sum_{i=1}^{n} E_i^0 L(F_\varepsilon^0(E_i^0))[I(\hat{E}_i^0 \le T) - I(\hat{E}_i^0 \le \hat{T}^i)]I(\Delta_i = 1) + O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),$$

since $E|E^0| < \infty$. Using arguments similar to those used in Lemma A.1 in VKA (1999), we find that

$$n^{-1}\sum_{i=1}^{n}\Big\{E_i^0 L(F_\varepsilon^0(E_i^0))\Delta_i[I(\hat{E}_i^0 \le T) - I(E_i^0 \le T)] - E[E^0 L(F_\varepsilon^0(E^0))\Delta I(\hat{E}^0 \le T)|\mathcal{X}_n]$$
$$+E[E^0 L(F_\varepsilon^0(E^0))\Delta I(E^0 \le T)]\Big\} = o_P(n^{-1/2}), \tag{4.5.9}$$

where $E[\cdot|\mathcal{X}_n]$ is the mean conditional on the data (X_j, Z_j, Δ_j), $j = 1, \ldots, n$. Finally, since

$$\begin{aligned} &E[E^0 L(F_\varepsilon^0(E^0))\Delta I(\hat{E}^0 \le T)|\mathcal{X}_n] - E[E^0 L(F_\varepsilon^0(E^0))\Delta I(E^0 \le T)] \\ &= \int_{R_X}\int_{T}^{\frac{T\hat{\sigma}^0(x)+\hat{m}^0(x)-m^0(x)}{\sigma^0(x)}} eL(F_\varepsilon^0(e))h_{e1}(e|x)f_X(x)\,de\,dx \\ &= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}), \end{aligned} \tag{4.5.10}$$

it follows that the first term of (4.5.8) is also $O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$.

The next lemma is a refinement of Lemma 4.5.3, obtained under somewhat stronger conditions.

Lemma 4.5.4 *Assume (A1) (i)–(iii), (A2) (i)–(ii), (A3) (i), (ii)', (iii)', (A4) (i), (A6)', (A8), (A11) (ii), and* $\sup_e |e^3(f_\varepsilon^0)''(e)| < \infty$. *Then,*

$$\int_{-\infty}^{T} eL(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) - \int_{-\infty}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e) = o_P((na_n)^{-1/2}).$$

Proof. Similarly as in the proof of Lemma 4.5.3, we will prove the lemma by showing that the four terms of (4.5.8) are of the stated order. First, we treat the first term of (4.5.8). It can be written as

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n} I(\Delta_i = 1)\Big\{E_i^0 L(F_\varepsilon^0(E_i^0))[I(\hat{E}_i^0 \le T) - I(\hat{E}_i^0 \le \hat{T}^i)] \\ &+E_i^0[L(\hat{F}_\varepsilon^0(\hat{E}_i^0)) - L(F_\varepsilon^0(E_i^0))][I(\hat{T}^i < \hat{E}_i^0 \le T) - I(T < \hat{E}_i^0 \le \hat{T}^i)] \\ &+(\hat{E}_i^0 - E_i^0)L(\hat{F}_\varepsilon^0(\hat{E}_i^0))[I(\hat{T}^i < \hat{E}_i^0 \le T) - I(T < \hat{E}_i^0 \le \hat{T}^i)]\Big\}. \end{aligned} \tag{4.5.11}$$

Note that $|\hat{E}_i^0 - E_i^0| I(\hat{T}^i < \hat{E}_i^0 \leq T) = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ uniformly in i by Proposition 4.5 in VKA (1999). When $\hat{E}_i^0 \leq T$ it holds that $E_i^0 \leq T\hat{\sigma}^0(X_i)/\sigma^0(X_i) + [\hat{m}^0(X_i) - m^0(X_i)]/\sigma^0(X_i) \leq T + V$, where $V = [\inf_x \sigma^0(x)]^{-1}[\sup_x |\hat{m}^0(x) - m^0(x)| + \sup_x |\hat{\sigma}^0(x) - \sigma^0(x)|] = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ and hence the third term of (4.5.11) is bounded by

$$O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, n^{-1} \sum_{i=1}^{n} \{I(T < E_i^0 \leq T + V) + I(T - V < E_i^0 \leq T)\}$$
$$= O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})\, \{[\tilde{H}_\varepsilon^0(T + V) - \tilde{H}_\varepsilon^0(T)] + [\tilde{H}_\varepsilon^0(T) - \tilde{H}_\varepsilon^0(T - V)]\},$$

where $\tilde{H}_\varepsilon^0(\cdot)$ is the empirical distribution of E_i^0, $i = 1, \ldots, n$. Using the fact that $\tilde{H}_\varepsilon^0(y) - H_\varepsilon^0(y) = O_P(n^{-1/2})$ uniformly in y, the above term is $o_P(n^{-1/2})$. The second term of (4.5.11) and the second and third terms of (4.5.8) are treated similarly. In the same way, the last term of (4.5.8) becomes

$$n^{-1} \sum_{i=1}^{n} I(\Delta_i = 0) \frac{\int_{\hat{T}^i}^{T} eL(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_i^{0T})} I(E_i^0 \leq T) + o_P((na_n)^{-1/2}). \tag{4.5.12}$$

Next, using classical arguments, (4.5.12) is written

$$n^{-1} \sum_{i=1}^{n} I(\Delta_i = 0) \frac{\int_{\hat{T}^i}^{T} eL(F_\varepsilon^0(e)) dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_i^{0T})} I(E_i^0 \leq T) + O_P(n^{-1/2})$$
$$= -(n^2 a_n)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} I(E_i^0 \leq T) B_{01}(X_i, Z_i, \Delta_i, Z_j, \Delta_j) K\Big(\frac{X_i - X_j}{a_n}\Big) + O_P(n^{-1/2}),$$

where

$$B_{01}(X_i, Z_i, \Delta_i, Z_j, \Delta_j) = I(\Delta_i = 0) f_X^{-1}(X_i) \frac{TL(F_\varepsilon^0(T)) f_\varepsilon^0(T)}{1 - F_\varepsilon^0(E_i^{0T})} [\eta(Z_j, \Delta_j | X_i) + T\zeta(Z_j, \Delta_j | X_i)].$$

Treating the function B_{01} in a similar way as the function B_0 in Lemma 4.5.1 or the function B_k in Theorem 2.2.1, we find that the above expression equals

$$-n^{-1} \sum_{i=1}^{n} f_X(X_i) \int_{-\infty}^{T_{X_i}} B_{01}(X_i, z, 0, Z_i, \Delta_i)\, dH_0(z|X_i) + O(a_n^2) = O_P(n^{-1/2}).$$

Finally, together with (4.5.9) and (4.5.10), the first term of (4.5.11) becomes using a Taylor development and Propositions 4.8 and 4.9 in VKA (1999),

$$\int_{R_X} TL(F_\varepsilon^0(T)) h_{e1}(T|x) \Big\{ T \frac{\hat{\sigma}^0(x) - \sigma^0(x)}{\sigma^0(x)} + \frac{\hat{m}^0(x) - m^0(x)}{\sigma^0(x)} \Big\} f_X(x) dx + o_P(n^{-1/2})$$
$$= (na_n)^{-1} \sum_{j=1}^{n} \int_{R_X} W(Z_j, \Delta_j | x) K\Big(\frac{x - X_j}{a_n}\Big) + o_P(n^{-1/2}), \tag{4.5.13}$$

where $W(Z_j, \Delta_j|x) = -TL(F_\varepsilon^0(T))h_{e1}(T|x)\{T\zeta(Z_j, \Delta_j|x) + \eta(Z_j, \Delta_j|x)\}$. Using three Taylor developments of order two for $\zeta(Z_j, \Delta_j|x), \eta(Z_j, \Delta_j|x)$ and $h_{e1}(T|x)$ around X_j, we obtain using condition $(A4)(i)$, that (4.5.13) equals

$$n^{-1}\sum_{j=1}^{n} W(Z_j, \Delta_j|X_j) + o_P(n^{-1/2}), \tag{4.5.14}$$

which is a sum of i.i.d. random variables with zero mean and hence it is $O_P(n^{-1/2})$. This finishes the proof.

Chapter 5

Strong uniform consistency of the estimated conditional mean of a conditional synthetic random variable.

In Chapters 2 and 3, we studied least squares estimators for parametric regression problems. Those estimators are based on the construction of synthetic data points which obey the preservation of mean criterion $E[Y_i^*|X_i] = E[Y_i|X_i]$, $i = 1, \ldots, n$. In Chapter 4, the objective was to estimate different conditional location functions. This was achieved by directly improving the estimation of the conditional distribution function in the formula of the conditional location function. In this respect, this method does not entail any construction of new data points that preserve a certain conditional mean (in fact, it can be easily seen that, in this case, a preservation of mean criterion exists but applies to the residuals). However, for some reasons that will be explained in Chapter 6, a classical method for complete data which uses new versions of the data points can be very useful to estimate a conditional location function nonparametrically. In Chapter 6, we will propose different ways to transform data according to the function to estimate. This way of doing enlarges

beyond survival analysis and applies to complete data with specific new versions of the data points (see example 5.1.1). For those reasons, we will establish in this chapter a general strong uniform consistency result of the Nadaraya-Watson (1964) estimator for general new data points. This result generalizes the one of Härdle, Janssen and Serfling (1988). Moreover, the strong uniform consistency of a modulus of continuity will be obtained for this estimator. Those two results will be largely used in Chapter 6.

In Section 5.1, we will present some examples for which the results established in Sections 5.2 and 5.3 apply. Those results can also be found in Heuchenne (2005).

5.1 Basic formulation and examples

Let $\{\Gamma_t, t \in I\}$ be a family of real valued measurable functions on R and suppose we want to estimate

$$E[\Gamma_t(Z, \Delta|x)|x] = \sum_{\delta=0,1} \int \Gamma_t(z, \delta|x) dH_\delta(z|x), \tag{5.1.1}$$

where I is a possibly infinite or degenerate interval in R. A natural estimator for this conditional mean is given by

$$\frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n})\Gamma_t(Z_i, \Delta_i|x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})}. \tag{5.1.2}$$

In the case $\Gamma_t(Z, \Delta|x) = Z$, this estimator reduces to the usual Nadaraya-Watson (1964) estimator and in the case $\Gamma_t(Z, \Delta|x) = I(Z \leq t)$, we obtain the Stone (1977) estimator defined by (1.3.1). The objective of Section 5.2 is to provide the almost sure convergence of (5.1.2) uniformly in x, t with the rate $(na_n)^{-1/2}(\log n)^{1/2}$. Now, suppose $s, t \in I$ with $|t - s| \leq d_n$. In Section 5.3, we aim to obtain the almost sure convergence of the modulus of continuity based on (5.1.2) uniformly in x, s, t, $|t - s| \leq d_n$, with the rate $(na_n)^{-1/2}(\log n)^{1/2} d_n^{1/2}$. The useful purpose of these results is illustrated below for some typical examples.

Example 5.1.1 (Nonparametric estimation of conditional location and scale functions for complete data)

Suppose we have location and scale estimators given by

$$\hat{m}_{ST}(x) = \int_0^1 \hat{F}^{-1}(s|x)L(s)\,ds, \quad \hat{\sigma}^2_{ST}(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 L(s)\,ds - \hat{m}^2_{ST}(x), \tag{5.1.3}$$

where $\hat{F}(\cdot|\cdot)$ is the Stone estimator given by (1.3.1), and $L(s)$ is a given score function satisfying $\int_0^1 L(s)ds = 1$. If the objective is to estimate

$$\int_0^1 F^{-1}(s|x)L(s)\ ds \tag{5.1.4}$$

and

$$\int_0^1 F^{-1}(s|x)^2 L(s)\ ds, \tag{5.1.5}$$

it is clear that $\Gamma_{t1}(Z,\Delta|x) = YL(F(Y|x))$ for (5.1.4) and $\Gamma_{t2}(Z,\Delta|x) = Y^2L(F(Y|x))$ for (5.1.5) as $E[\Gamma_{ti}(Z,\Delta|x)|x]$ equals the function to estimate (5.1.4) for $i = 1$ and (5.1.5) for $i = 2$. Since the data points $\Gamma_{ti}(Z,\Delta|x)$ depend themselves on $F(Y|x)$, they are estimated by $YL(\hat{F}(Y|x))$ and $Y^2L(\hat{F}(Y|x))$ so that the classical Nadaraya-Watson estimator based on those data points corresponds to (5.1.3).

Note that when $L(s) = I(0 \le s \le 1)$, $\hat{m}_{ST}(x)$ and $\hat{\sigma}^2_{ST}(x)$ reduce to estimators of the conditional mean and conditional variance. Theorem 5.2.3 of the next section thus enables to prove at the same time the strong uniform consistency of estimators of any location and scale functions defined by the score function L. This is achieved in two steps : first, an application of Theorem 5.2.3 for data points $I(Y_i \le t)$ $(i = 1,\ldots,n)$ in order to delete the Stone estimators in the expressions $YL(\hat{F}(Y|x))$ and $Y^2L(\hat{F}(Y|x))$, and second, an application of the same theorem on the functions $\Gamma_{t1}(Z,\Delta|x)$ and $\Gamma_{t2}(Z,\Delta|x)$.

Example 5.1.2 (Nonparametric estimation of conditional location and scale functions for censored data)

Suppose we have location and scale estimators given by

$$\hat{m}_B(x) = \int_{-\infty}^{\tilde{T}} yL(\tilde{F}(y|x))\, d\tilde{F}(y|x), \tag{5.1.6}$$

and

$$\hat{\sigma}_B^2(x) = \int_{-\infty}^{\tilde{T}} y^2 L(\tilde{F}(y|x))\, d\tilde{F}(y|x) - \hat{m}_B^2(x), \tag{5.1.7}$$

where $\tilde{F}(\cdot|\cdot)$ is the Beran estimator given by (2.1.2), and $L(s)$ is a given score function satisfying $\int_0^1 L(s)ds = 1$. In order to avoid consistency problems in the right tails of the Beran estimators, $\tilde{T}$ is chosen smaller than $\inf_x \tau_{H(\cdot|x)}$. Seeing that the objective is to estimate $E[YI(Y \leq \tilde{T})L(Y|x)|x]$ and $E[Y^2 I(Y \leq \tilde{T})L(Y|x)|x]$ with an estimator of the Nadaraya-Watson type, we rewrite (5.1.6) and (5.1.7) as

$$\hat{m}_B(x) = \frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n})\hat{\Gamma}_{t3}(Z_i, \Delta_i|x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})}, \tag{5.1.8}$$

and

$$\hat{\sigma}_B^2(x) = \frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n})\hat{\Gamma}_{t4}(Z_i, \Delta_i|x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})} - \hat{m}_B^2(x), \tag{5.1.9}$$

where

$$\hat{\Gamma}_{t3}(Z_i, \Delta_i|x) = Z_i I(Z_i \leq \tilde{T})L(\tilde{F}(Z_i|x))\Delta_i + \frac{\int_{Z_i\wedge\tilde{T}}^{\tilde{T}} yL(\tilde{F}(y|x))d\tilde{F}(y|x)}{1-\tilde{F}(Z_i\wedge\tilde{T}|x)}(1-\Delta_i),$$

and

$$\hat{\Gamma}_{t4}(Z_i, \Delta_i|x) = Z_i^2 I(Z_i \leq \tilde{T})L(\tilde{F}(Z_i|x))\Delta_i + \frac{\int_{Z_i\wedge\tilde{T}}^{\tilde{T}} y^2 L(\tilde{F}(y|x))d\tilde{F}(y|x)}{1-\tilde{F}(Z_i\wedge\tilde{T}|x)}(1-\Delta_i).$$

$\hat{\Gamma}_{t3}(Z, \Delta|x)$ and $\hat{\Gamma}_{t4}(Z, \Delta|x)$ actually estimate

$$\Gamma_{t3}(Z, \Delta|x) = ZI(Z \leq \tilde{T})L(F(Z|x))\Delta + \frac{\int_{Z\wedge\tilde{T}}^{\tilde{T}} yL(F(y|x))dF(y|x)}{1-F(Z\wedge\tilde{T}|x)}(1-\Delta),$$

and

$$\Gamma_{t4}(Z,\Delta|x) = Z^2 I(Z \leq \tilde{T}) L(F(Z|x))\Delta + \frac{\int_{Z\wedge\tilde{T}}^{\tilde{T}} y^2 L(F(y|x))dF(y|x)}{1-F(Z\wedge\tilde{T}|x)}(1-\Delta)$$

respectively. It is easy to check that

$$E[\Gamma_{t3}(Z,\Delta|x)|x] = E[YI(Y \leq \tilde{T})L(Y|x)|x],$$

and

$$E[\Gamma_{t4}(Z,\Delta|x)|x] = E[Y^2 I(Y \leq \tilde{T})L(Y|x)|x].$$

As for the complete data case, Theorem 5.2.3 enables to prove the strong uniform consistency of estimators of any location and scale functions (truncated by $\tilde{T}$) defined by the score function L. In this respect, Theorem 5.2.3 generalizes Proposition 4.5 of Van Keilegom and Akritas (1999) since the strong uniform consistency of location and scale estimators of this paper is established under assumption (A2) (i) (mentioned in Section 2.1). $\hat{m}_B(x)$ and $\hat{\sigma}_B^2(x)$ can therefore be considered as a substitute to the preliminary $\hat{m}^0(x)$ and $\hat{\sigma}^0(x)$ used all along this thesis.

Note that in order to use Theorem 5.2.3 with the functions $\Gamma_{t3}(Z,\Delta|x)$ and $\Gamma_{t4}(Z,\Delta|x)$, we first need to delete the Beran estimators that appear in $\hat{\Gamma}_{t3}(Z,\Delta|x)$ and $\hat{\Gamma}_{t4}(Z,\Delta|x)$. This can be done by using Proposition 4.3 of Van Keilegom and Akritas (1999).

Example 5.1.3 (Estimation of a conditional distribution function under the heteroscedastic model)

As it will be explained into more details in the next chapter, it is sometimes useful to construct estimators which only use the heteroscedastic model (1.7.1) to complete censored information. This is not the case for the estimators of Chapter 4 based on (4.1.2) since the conditional distribution of the response variable is directly replaced by the distribution of the residuals based on model (1.7.1). In Chapter 6, we will therefore study an estimator of the distribution function which preserves local uncensored information. This estimator is a weighted sum of data points $\hat{\Gamma}_{t5}(Z_i,\Delta_i|x)$, $i = 1,\ldots,n$, that approximate

$$\Gamma_{t5}(Z_i, \Delta_i|x) = I(Z_i \leq t)\Delta_i + \frac{F_\varepsilon^0(T_t^x) - F_\varepsilon^0(E_{ix}^{0T_t})}{1 - F_\varepsilon^0(E_{ix}^{0T})}(1 - \Delta_i), \tag{5.1.10}$$

where $E_{ix}^{0T_t} = \frac{Z_i \wedge T_x \wedge t - m^0(x)}{\sigma^0(x)}$, $E_{ix}^{0T} = \frac{Z_i - m^0(x)}{\sigma^0(x)} \wedge T$, $T_t^x = \frac{T_x \wedge t - m^0(x)}{\sigma^0(x)}$. T_x is defined as in Section 4.1 and $T < \tau_{H_\varepsilon^0}$. We refer the reader to the next chapter for complete description and explanation of this estimator.

Strong uniform consistency and modulus of continuity proofs are achieved in three steps. First, we consider new data points $\gamma_t(Z_i, \Delta_i|x)$, $i = 1, \ldots n$, $t \in I$, $x \in R_X$, and kernels that are defined by indicators. Second, we combine those data points to obtain the $\Gamma_t(Z_i, \Delta_i|x)$ used in this section and we sum indicators to construct kernels of step-function form. Third, by using a number of indicators that tends to infinity in the step-function kernel, we show the announced results for usual smooth kernels. The assumptions we need for the proofs of Sections 5.2 and 5.3 are listed below.

$(A1)(ii)$, $(A8)$ mentioned in Section 2.1.

$(B1)$ $\gamma_t(\cdot, \cdot|\cdot)$ is Lipshitz on R_X uniformly in $t \in I$:

$$\sup_{|x - x_j| \leq d,\ x, x_j \in R_X} \sup_{t \in I} |\gamma_t(z, \delta|x) - \gamma_t(z, \delta|x_j)| \leq L_0(z, \delta|x_j)d$$

where $L_0(\cdot, \cdot|\cdot)$ is a (positive) function independent of t such that $E[L_0(Z, \Delta|x)^6] \leq L_6 < \infty$ for all $x \in R_X$.

$(B2)$ $0 \leq \gamma_t(z, \delta|x) \leq \gamma_{t'}(z, \delta|x)$, $t < t' \in I$, for all x, z and $\delta = 0, 1$.

$(B3)$ $g(t|x) = E[\gamma_t(Z, \Delta|x)]$ is a continuous function of t in I for all x.

$(B4)$ The limit functions $\gamma_{t^*} = \lim_{t \to t^*} \gamma_t$ and $\gamma_{t_*} = \lim_{t \to t_*} \gamma_t$ exist and are finite a.s. (w.r.t. $H(z)$) for all x, where $t_* = \inf\{t : t \in I\}$, $t_* \geq -\infty$ and $t^* = \sup\{t : t \in I\}$, $t^* \leq \infty$.

$(B5)$ For all x, $E[\gamma_{t^*}(Z, \Delta|x)^{6\lambda}] \leq M_{6\lambda} < \infty$ for some λ, $2 < \lambda < \infty$; in the case $\lambda = \infty$, $\sup_{x,z,\delta} |\gamma_{t^*}(z, \delta|x)| < \infty$.

(*B*6) Let $\{c_n\}$ a nonnegative sequence satisfy (i) $0 \leq c_n \to 0$, (ii) $\Delta_n = nc_n/\log n \to \infty$, (*iii*) $c_n^{-1} \leq (n/\log n)^{1-2/\lambda}$, for λ as in (B5).

(*B*7)(*i*) $F_X(x)$ is differentiable with respect to x.
(*ii*) $H_\delta(x,z)$, the joint distribution of $x \in R_X,\ z \in R,\ \delta = 0,1$, is differentiable with respect to (x,z) (with derivative with respect x and z denoted $h_\delta(x,z)$).
(*iii*) $H_\delta(z)$, the joint distribution of z and δ, $z \in R$, $\delta = 0,1$, is differentiable with respect to z (with derivative denoted $h_\delta(z)$).

(*B*8) Define new data points as $\Gamma_t(z,\delta|x) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(z,\delta|x), \quad z \in R,\ t \in I,\ x \in R_X$, $\delta = 0,1$, with fixed and finite $i_0, q_1, \ldots, q_{i_0}$ and with families $\{\gamma_{ti},\ t \in I\}$, $1 \leq i \leq i_0$, satisfying assumptions (B1)-(B5), with common λ in (B5).

(*B*9)(*i*) Consider kernel sequences of step-function form, $K_n(u) = \sum_{j=1}^{j_n} m_{nj} I(-b_{nj} \leq u \leq b_{nj}),\ u \in R$, with $\{j_n\}$, $\{m_{nj}\}$, $\{b_{nj}\}$ nonnegative constants such that $|2\sum_{j=1}^{j_n} m_{nj} b_{nj} - 1| = O(\max(\Delta_n^{-1/2}, a_n^2))$, with $j_n = O(n^s),\ s > 0$ and $\Delta_n = na_n/\log n$.
(*ii*) $\sup_n \sum_{j=1}^{j_n} m_{nj} b_{nj}^{1/2} < \infty$.
(*iii*) $\sup_n \sum_{j=1}^{j_n} m_{nj} b_{nj}^3 < \infty$.

(*B*10) Let $\{c_n\}$ and $\{d_n\}$ two nonnegative sequences satisfy (i) $0 \leq c_n, d_n \to 0$, (*ii*) $\Delta_n = nc_n/\log n \to \infty$, (*iii*) $c_n^{-1} \leq d_n(n/\log n)^{1-2/\lambda}$ for λ as in (B5).

(*B*11) The data points $\gamma_t(Z_i, \Delta_i|x)$, $t \in I,\ x \in R_X,\ i = 1, \ldots, n$, have the following mean-Lipshitz properties : (*i*) $\sup_{\{x \in R_X, |t-s| \leq d_n, s,t \in I, d_n \to 0\}} |E[\gamma_t(Z,\Delta|x) - \gamma_s(Z,\Delta|x)]| \leq C_L d_n$, (*ii*) $\sup_{\{x \in R_X, |t-s| \leq d_n, s,t \in I, d_n \to 0\}} E[(\gamma_t(Z,\Delta|x) - \gamma_s(Z,\Delta|x))^2] \leq C_{L_2} d_n$, for n sufficiently large.

(*B*12)(*i*) $-$ (*iii*) Consider kernel sequences of the same form and with the same assumptions as in (B9) except that $|2\sum_{j=1}^{j_n} m_{nj} b_{nj} - 1| = O(\max(\Delta_n^{-1/2} d_n^{-1/2}, a_n^2))$ in (B9) (i).

5.2 Strong uniform consistency of the weighted average of artificial data points

We start by showing two preliminary results which will lead to the strong uniform consistency of the estimator (5.1.2).

Proposition 5.2.1 *Assume (A1) (ii), (A8), (B6) and (B7). Then,*

$$P(M_{0n}(c_n) > C_0\Delta_n^{-1/2} + C_1c_n^2) = O(n^{-2}),$$

where

$$\begin{aligned} M_{0n}(c_n) \quad &= \sup_{x\in R_X}\sup_{t\in I}|\frac{1}{2nc_n}\sum_{i=1}^{n}\gamma_t(Z_i,\Delta_i|x)I(x-c_n<X_i\leq x+c_n) \\ &\quad -\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x)h_\delta(x,z)dz|, \end{aligned}$$

$\gamma_t(z,\delta|x),\ t\in I,\ x\in R_X,\ z\in R,\ \delta=0,1,$ *satisfy assumptions (B1)-(B5).*

Proof. Let $f_n=\Delta_n^{-1/2}c_n$. We have

$$\begin{aligned} M_{0n}(c_n) \ &= \sup_{x\in R_X}\sup_{t\in I}|\frac{1}{2nc_n}\sum_{i=1}^{n}\gamma_t(Z_i,\Delta_i|x)I(x-c_n<X_i\leq x+c_n) \\ &\quad -\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x)h_\delta(s,z)dzds| \\ &\quad +\sup_{x\in R_X}\sup_{t\in I}|\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x)h_\delta(s,z)dzds \\ &\quad -\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x)h_\delta(x,z)dz| \\ &= \sup_{x\in R_X}\sup_{t\in I}|M_{1tn}(x)|+\sup_{x\in R_X}\sup_{t\in I}|M_{2tn}(x)|. \end{aligned}$$

First, we treat the term $M_{2tn}(x)$. It is given by

$$\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x)\left\{\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}h_\delta(s,z)ds-h_\delta(x,z)\right\}dz.$$

Using two Taylor developments of order three around x, we get

$$\frac{1}{2c_n}\int_{x-c_n}^{x+c_n} h_\delta(s,z)ds - h_\delta(x,z) = (c_n^2/12)[\ddot{f}_{X|Z,\Delta}(\theta_1|z,\delta) + \ddot{f}_{X|Z,\Delta}(\theta_2|z,\delta)]h_\delta(z),$$

where θ_1 (θ_2) is between $x+c$ and x (x and $x-c$). Since $\sup_{x,z}|\ddot{f}_{X|Z,\Delta}(x|z,\delta)| < \infty$ ($\delta = 0,1$) and $\sup_{\{x\in R_X,\ t\in I\}} E[\gamma_t(Z,\Delta|x)] < \infty$ with $\gamma_t(Z,\Delta|x) \geq 0$,

$$\sup_{\{x\in R_X,\ t\in I\}} |M_{2tn}(x)| \leq C_1 c_n^2. \tag{5.2.1}$$

Let L_X the length of R_X and divide R_X into $[\frac{2L_X}{f_n}]$ intervals of length smaller or equal to f_n ([x] denotes the integer part of x). Denote $x_0 = \inf\{x : x \in R_X\}$, I_X the set of points $\{x_k = x_0 + k[\frac{2L_X}{f_n}]^{-1}L_X,\ \ 1 \leq k \leq [\frac{2L_X}{f_n}] - 1 = L_X^n\}$ and $x_{L_X^n+1} = \sup\{x : x \in R_X\}$ which limit the intervals. Using the Lipshitz condition (B1), we can rewrite for $1 \leq j \leq L_X^n$,

$$\begin{aligned}
&\sup_{x\in R_X}\sup_{t\in I}|M_{1tn}(x)| \qquad (5.2.2)\\
&\leq \max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\sup_{t\in I}|\frac{1}{2nc_n}\sum_{i=1}^n \gamma_t(Z_i,\Delta_i|x_j)I(x-c_n < X_i \leq x+c_n)\\
&\quad -\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x_j)h_\delta(s,z)dzds| + \max_{x_j\in I_X}(E[L_0(Z,\Delta|x_j)] + |V_n(x_j)|)\Delta_n^{-1/2}\\
&\quad = (2c_n)^{-1}\max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\sup_{t\in I}|M_{3tn}(x_j,x)| + \max_{x_j\in I_X}(E[L_0(Z,\Delta|x_j)] + |V_n(x_j)|)\Delta_n^{-1/2},
\end{aligned}$$

where $V_n(x_j) = (1/2n)\sum_{i=1}^n L_0(Z_i,\Delta_i|x_j) - (E[L_0(Z,\Delta|x_j)]/2)$. We have

$$\begin{aligned}
&P(\max_{x_j\in I_X}(E[L_0(Z,\Delta|x_j)] + |V_n(x_j)|) > 2C_2)\\
&\leq \sum_j\{P(|2V_n(x_j)| > 2C_2) + P(E[L_0(Z,\Delta|x_j)] > C_2)\},
\end{aligned}$$

where the second term on the right hand side of the above expression is zero when $C_2 > L_6^{1/6}$. For the first term, we use an extension of Chebyshev's inequality :

$$\begin{aligned}
&P(|n^{-1}\sum_{i=1}^n L_0(Z_i,\Delta_i|x_j) - E[L_0(Z,\Delta|x_j)]| > 2C_2)\\
&\leq \frac{1}{(2nC_2)^6}E[\{\sum_{i=1}^n(L_0(Z_i,\Delta_i|x_j) - E[L_0(Z,\Delta|x_j)])\}^6] = O(n^{-3}),
\end{aligned}$$

since $E[L_0(Z,\Delta|x_j)^6] \leq L_6 < \infty$. Then, with $L_X^n O(n^{-3}) = o(n^{-2})$,

$$P(\max_{x_j \in I_X}(E[L_0(Z,\Delta|x_j)] + |V_n(x_j)|) > 2C_2) = o(n^{-2}),$$

for which, using Borel-Cantelli Lemma, we obtain

$$V_n = \max_{x_j \in I_X}(E[L_0(Z,\Delta|x_j)] + |V_n(x_j)|) = O(1)\ a.s.$$

To treat the first term on the right hand side of (5.2.2), we introduce some notations. Let

$$G_{tn}(x_j,x) = n^{-1}\sum_{i=1}^{n}\gamma_t(Z_i,\Delta_i|x_j)I(X_i \leq x),$$

and

$$G_t(x_j,x) = E[G_{tn}(x_j,x)] = \int_{x_0}^{x}\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_t(z,\delta|x_j)h_\delta(s,z)dzds.$$

Therefore,

$$\begin{aligned}|M_{3tn}(x_j,x)| &= |G_{tn}(x_j,x+c_n) - G_{tn}(x_j,x-c_n) - [G_t(x_j,x+c_n) - G_t(x_j,x-c_n)]| \\ &\leq 2\sup_{|z|\leq c_n}|G_{tn}(x_j,x+z) - G_{tn}(x_j,x) - [G_t(x_j,x+z) - G_t(x_j,x)]| \\ &= 2M_{4tn}(x_j,x,c_n). \end{aligned} \tag{5.2.3}$$

By conditions $(B2)-(B5)$, the functions $g(t|x_j)$, $j = 1,\ldots,L_X^n$, are nondecreasing, continuous in t with finite limits $g(t^*|x_j)$ and $g(t_*|x_j)$ as $t \to t^*$ and t_*. Divide I in $O(f_n^{-2})$ intervals such that $|g(t_1|x_j) - g(t_*|x_j)| \leq f_n$, $|g(t_{k+1}|x_j) - g(t_k|x_j)| \leq f_n$, for $k = 1,\ldots N_{nj}-1$, $|g(t_*|x_j) - g(t_{N_{nj}}|x_j)| \leq f_n$, $j = 1,\ldots,L_X^n$. Let I_{nj} denote the set $\{t_*, t_1, \ldots, t_{N_{nj}}, t^*\}$ and I_{nj}^* the set $\{(t_*,t_1),(t_1,t_2),\ldots,(t_{N_{nj}},t^*)\}$. Clearly, the cardinality N_{nj} of I_{nj}^* is bounded by

$$N_{nj} \leq \frac{2(g(t^*|x_j) - g(t_*|x_j))}{f_n}.$$

Also, for fixed j,x,z,n, the functions $G_{tn}(x_j,x+z) - G_{tn}(x_j,x)$ and $G_t(x_j,x+z) - G_t(x_j,x)$ are monotone in t and have finite limits as $t \to t_*, t^*$. We therefore have

$$\begin{aligned}&|G_{tn}(x_j,x+z) - G_{tn}(x_j,x) - [G_t(x_j,x+z) - G_t(x_j,x)]| \\ &\leq \max_{t\in I_{nj}}|G_{tn}(x_j,x+z) - G_{tn}(x_j,x) - [G_t(x_j,x+z) - G_t(x_j,x)]| \\ &\quad + \max_{(s,t)\in I_{nj}^*}|G_t(x_j,x+z) - G_s(x_j,x+z) - [G_t(x_j,x) - G_s(x_j,x)]|. \end{aligned} \tag{5.2.4}$$

It is easily shown that the right hand side of the above expression is bounded by

$$\begin{aligned}
&2\int_{R_X}\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x_j))-\gamma_s(z,\delta|x_j))h_\delta(s,z)dzds\\
&=2\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x_j))-\gamma_s(z,\delta|x_j))h_\delta(z)dz\\
&\le 2(g(t|x_j)-g(s|x_j))\le 2f_n,
\end{aligned}$$

using monotonicity of γ_t with respect to t. Therefore,

$$\max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\sup_{t\in I}M_{4tn}(x_j,x,c_n)\le\max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\max_{t\in I_{nj}}M_{4tn}(x_j,x,c_n)+2f_n. \tag{5.2.5}$$

Let

$$M_{5tn}(x_j,x,z)=|G_{tn}(x_j,x+z)-G_{tn}(x_j,x)-[G_t(x_j,x+z)-G_t(x_j,x)]|.$$

We have

$$\begin{aligned}
&\max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\max_{t\in I_{nj}}\sup_{|z|\le c_n}M_{5tn}(x_j,x,z) \qquad (5.2.6)\\
&\le\max\Big(\max_{j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\max_{t\in I_{nj}}\sup_{x+z\in[x_{j-1},x_{j+1}]}M_{5tn}(x_j,x,z),\\
&\qquad\max_{x_j\in I_X\setminus\{x_{L_X^n}\}}\sup_{x\in[x_j,x_{j+1}]}\max_{t\in I_{nj}}\sup_{x+z\in[x_{j+1},x+c_n]}M_{5tn}(x_j,x,z),\\
&\qquad\max_{x_j\in I_X\setminus\{x_1\}}\sup_{x\in[x_j,x_{j+1}]}\max_{t\in I_{nj}}\sup_{x+z\in[x-c_n,x_{j-1}]}M_{5tn}(x_j,x,z),\\
&\qquad\max_{x_j\in I_X\setminus\{x_{L_X^n}\}}\sup_{x\in[x_{j-1},x_j]}\max_{t\in I_{nj}}\sup_{x+z\in[x_{j+1},x+c_n]}M_{5tn}(x_j,x,z),\\
&\qquad\max_{x_j\in I_X\setminus\{x_1\}}\sup_{x\in[x_{j-1},x_j]}\max_{t\in I_{nj}}\sup_{x+z\in[x-c_n,x_{j-1}]}M_{5tn}(x_j,x,z)\Big).
\end{aligned}$$

Introducing $G_{tn}(x_j,x_{j+k})$ and $G_t(x_j,x_{j+k})$ for $k=-1,0$ or 1, it is easily shown that

$$\begin{aligned}
&\max_{x_j\in I_X}\sup_{x\in[x_{j-1},x_{j+1}]}\max_{t\in I_{nj}}M_{4tn}(x_j,x,c_n)\\
&\le 2\max_{x_j\in I_X}\max_{t\in I_{nj}}\max_{k\in\{-1,0,1\}}\sup_{|z|\le c_n}M_{5tn}(x_j,x_{j+k},z).
\end{aligned}$$

Now, put $Q_n=M_\lambda f_n^{-1/(\lambda-1)}$, where, for all x, $(E[(\gamma_{t^*}(Z,\Delta|x))^\lambda])^{1/\lambda}\le M_\lambda<\infty$ for some λ, $2<\lambda<\infty$. In the case $\lambda=\infty$, M_∞ denotes then $\sup_{x,z,\delta}|\gamma_{t^*}(z,\delta|x)|$. Also, put

$$H_{tn}(x_j,x)=n^{-1}\sum_{i=1}^{n}\gamma_t(Z_i,\Delta_i|x_j)I(\gamma_t(Z_i,\Delta_i|x_j)\le Q_n)I(X_i\le x),$$

and define $M_{6tn}(x_j, x, z)$ by substitution of H_{tn} for G_{tn} and $E[H_{tn}]$ for G_t in $M_{5tn}(x_j, x, z)$. That yields

$$\sup_{x\in R_X}\sup_{t\in I}|M_{1tn}(x)| \le 2c_n^{-1}\max_{x_j\in I_X}\max_{t\in I_{nj}}\max_{k\in\{-1,0,1\}}\sup_{|z|\le c_n} M_{6tn}(x_j, x_{j+k}, z) + 2f_n c_n^{-1}(1 + V_n/2 + W_n + \theta_n),$$

where

$$W_n = f_n^{-1}\max_{x_j\in I_X}\max_{t\in I_{nj}}\max_{k\in\{-1,0,1\}}\sup_{|z|\le c_n} M_{7tn}(x_j, x_{j+k}, z),$$

$$M_{7tn}(x_j, x, z) = |G_{tn}(x_j, x+z) - G_{tn}(x_j, x) - [H_{tn}(x_j, x+z) - H_{tn}(x_j, x)]|,$$

$$\theta_n = f_n^{-1}\max_{x_j\in I_X}\max_{t\in I_{nj}}\max_{k\in\{-1,0,1\}}\sup_{|z|\le c_n} M_{8t}(x_j, x_{j+k}, z),$$

and

$$M_{8t}(x_j, x, z) = |G_t(x_j, x+z) - G_t(x_j, x) - [E[H_{tn}(x_j, x+z)] - E[H_{tn}(x_j, x)]]|.$$

Using (B2), (B4) and the fact that $f_n^{-1} = (Q_n/M_\lambda)^{\lambda-1}$, we have

$$\begin{aligned} M_\lambda^{\lambda-1}W_n &\le Q_n^{\lambda-1}\max_{x_j\in I_X} n^{-1}\sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i|x_j)I(\gamma_{t^*}(Z_i, \Delta_i|x_j) > Q_n) \\ &\le \max_{x_j\in I_X} n^{-1}\sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i|x_j)^\lambda, \end{aligned}$$

if $\lambda < \infty$ and $W_n = 0$ if $\lambda = \infty$. Next, for $\lambda < \infty$,

$$\begin{aligned} &P(\max_{x_j\in I_X} n^{-1}\sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i|x_j)^\lambda > C_3) \\ &\le \sum_j\{P(n^{-1}\sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i|x_j)^\lambda - E[\gamma_{t^*}(Z, \Delta|x_j)^\lambda] > C_3/2) \\ &\qquad + P(E[\gamma_{t^*}(Z, \Delta|x_j)^\lambda] > C_3/2)\}, \end{aligned}$$

where the second term on the right hand side of the above expression is zero when $C_3/2 > M_\lambda^\lambda$. For the first term, we also use the extension of Chebyshev's inequality :

$$P(n^{-1}\sum_{i=1}^{n}\gamma_{t^*}(Z_i,\Delta_i|x_j)^\lambda - E[\gamma_{t^*}(Z,\Delta|x_j)^\lambda] > C_3/2)$$
$$\leq \frac{64}{(nC_3)^6}E[\{\sum_{i=1}^{n}(\gamma_{t^*}(Z_i,\Delta_i|x_j)^\lambda - E[\gamma_{t^*}(Z,\Delta|x_j)^\lambda])\}^6] = O(n^{-3}),$$

since $E[\gamma_{t^*}(Z,\Delta|x_j)^{6\lambda}] \leq M_{6\lambda} < \infty$. Then, with $L_X^n O(n^{-3}) = o(n^{-2})$,

$$P(\max_{x_j\in I_X} n^{-1}\sum_{i=1}^{n}\gamma_{t^*}(Z_i,\Delta_i|x_j)^\lambda > C_3) = o(n^{-2}),$$

for which, using Borel-Cantelli Lemma, we obtain $W_n = O(1)$ *a.s.* We also see that

$$\theta_n \leq \frac{\max_{x_j\in R_X} E[\gamma_{t^*}(Z,\Delta|x_j)^\lambda]}{M_\lambda^{\lambda-1}} \leq M_\lambda,$$

if $\lambda < \infty$ and $\theta_n = 0$ if $\lambda = \infty$.

Now, define

$$w_n = [\frac{2Q_n c_n}{f_n} + 1]$$

and

$$\eta_{njkr} = x_{j+k} + \frac{rc_n}{w_n}, \quad for\ r = -w_n, -w_n+1, \ldots, w_n.$$

Defining

$$\xi_{ntjkr} = |H_{tn}(x_j,\eta_{njkr}) - H_{tn}(x_j,x_{j+k}) - [E[H_{tn}(x_j,\eta_{njkr})] - E[H_{tn}(x_j,x_{j+k})]]|,$$

we have

$$\sup_{|z|\leq c_n} M_{6tn}(x_j,x_{j+k},z) \leq \max_{-w_n\leq r\leq w_n}\xi_{ntjkr}$$
$$+ \max_{-w_n\leq r\leq w_n-1}|E[H_{tn}(x_j,\eta_{njk(r+1)})] - E[H_{tn}(x_j,\eta_{njkr})]|.$$

The second term of the right hand side of the above expression is bounded by

$$Q_n \max_{-w_n\leq r\leq w_n-1}\int_{\eta_{njkr}}^{\eta_{njk(r+1)}}\sum_{\delta=0,1}\int_{-\infty}^{\infty}h_\delta(s,z)dzds$$
$$\leq Q_n \max_{-w_n\leq r\leq w_n-1}\int_{\eta_{njkr}}^{\eta_{njk(r+1)}}f_X(s)ds \leq Q_n C_4(\eta_{njk(r+1)} - \eta_{njkr}) \leq C_4 f_n/2,$$

where C_4 is the Lipshitz constant of $F_X(\cdot)$. The goal is therefore to calculate

$$\begin{aligned}
&P(\sup_{x\in R_X}\sup_{t\in I}|M_{1tn}(x)| > C_0\Delta_n^{-1/2})\\
&\leq P(2\max_{j,t,k,r}\xi_{ntjkr} + f_n(2W_n + V_n) > (C_0 - 2 - C_4 - 2M_\lambda)f_n)\\
&\leq \sum_{j,t,k,r} P(\xi_{ntjkr} > (1/6)(C_0 - 2 - C_4 - 2M_\lambda)f_n)\\
&\quad + P(W_n > (1/6)(C_0 - 2 - C_4 - 2M_\lambda))\\
&\quad + P(V_n > (1/3)(C_0 - 2 - C_4 - 2M_\lambda)),
\end{aligned}$$

where C_0, C_2 and C_3 have to satisfy $2M_\lambda < (C_3/M_\lambda^{\lambda-1}) = (1/6)(C_0 - C_4 - 2M_\lambda - 2)$ and $L_6^{1/6} < C_2 = (1/6)(C_0 - C_4 - 2M_\lambda - 2)$. Therefore, we only have to treat the first term on the right hand side of the above expression. Defining $C_0' = (1/6)(C_0 - 2 - C_4 - 2M_\lambda)$, we have by Bernstein's inequality,

$$P(\xi_{ntjkr} > C_0' f_n) \leq 2\exp(-\nu_{tjknr}),$$

where

$$\nu_{tjknr} = \frac{C_0'^2 n^2 f_n^2}{2n\sigma_{tjknr}^2 + \frac{2}{3}nC_0' f_n Q_n},$$

and $\sigma_{tjknr} = Var[D_{tjknr}]$, where

$$D_{tjknr} = \gamma_t(Z,\Delta|x_j)I(\gamma_t(Z,\Delta|x_j) \leq Q_n)(I(X \leq \eta_{njkr}) - I(X \leq x_{j+k})).$$

We have

$$\begin{aligned}
\sigma_{tjknr}^2 \leq E[D_{tjknr}^2] \quad &\leq \int_{x_{j+k}\wedge\eta_{njkr}}^{\eta_{njkr}\vee x_{j+k}} \sum_{\delta=0,1}\int_{-\infty}^{\infty} \gamma_t^2(z,\delta|x_j)I(\gamma_t(z,\delta|x_j) \leq Q_n)h_\delta(s,z)dzds\\
&\leq C_5 M_\lambda^2 c_n, \qquad (5.2.7)
\end{aligned}$$

using (B5), condition (A8) and where $C_5 = \max_\delta \sup_{x,z}|f_{X|Z,\Delta}(x|z,\delta)|$. Using (B6) (iii),

$$Q_n f_n = M_\lambda f_n^{(\lambda-2)/(\lambda-1)} = M_\lambda(\frac{c_n \log n}{n})^{(\lambda-2)/2(\lambda-1)} \leq M_\lambda c_n. \qquad (5.2.8)$$

We thus have by (5.2.7) and (5.2.8),

$$\nu_{tjknr} \geq C_0'' \log n,$$

with

$$C_0'' = \frac{C_0'^2}{2M_\lambda(C_5 M_\lambda + \frac{1}{3}C_0')} > \max\left(\frac{6}{3C_5+2}, \frac{3L_6^{1/3}}{M_\lambda(6M_\lambda C_5 + 2L_6^{1/6})}\right).$$

Therefore,

$$\sum_{j,t,k,r} P(\xi_{ntjkr} > C_0' f_n) \leq 6 \sum_{j=1}^{L_X^n} (N_{nj}+2)(2w_n+1) n^{-C_0''}, \tag{5.2.9}$$

where C_0'' has to be chosen large enough so that the right hand side of (5.2.9) tends to zero sufficiently fast. Thus, the highest order term on the right hand side of (5.2.9) is

$$96\left(\frac{L_X M_\lambda Q_n c_n}{f_n^3}\right) n^{-C_0''}, \tag{5.2.10}$$

where

$$N_{nj} \leq \frac{2g(t^*|x_j)}{f_n} \leq 2\frac{M_\lambda}{f_n}$$

and

$$\frac{Q_n c_n}{f_n} = M_\lambda c_n \left(\frac{n}{c_n \log n}\right)^{\frac{\lambda}{2(\lambda-1)}}.$$

Using (B6) (iii), (5.2.10) is bounded by

$$L_\lambda \left(\frac{n}{c_n \log n}\right)^{\frac{3\lambda-2}{2(\lambda-1)}} c_n n^{-C_0''} \leq L_\lambda \left(\frac{n}{\log n}\right)^2 n^{-C_0''},$$

where $L_\lambda = 96 M_\lambda^2 L_X$. Therefore, choosing $C_0'' \geq 4$ allows to write

$$P\Big(\sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| > C_0 \Delta_n^{-1/2}\Big) = O(n^{-2}). \tag{5.2.11}$$

By (5.2.1) and (5.2.11), we finally obtain

$$P\Big(\sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x) + M_{2tn}(x)| > C_0 \Delta_n^{-1/2} + C_1 c_n^2\Big) = O(n^{-2}). \tag{5.2.12}$$

Proposition 5.2.2 *Assume (A1) (ii), (A8), (B7)-(B9).* a_n *satisfies* (i) $a_n B_n \to 0$, (ii) $na_n b_n/\log n \to \infty$ *and* (iii) $a_n^{-1} \leq b_n(n/\log n)^{1-2/\lambda}$, *where* $b_n = \min_{j\leq j_n} b_{nj}$, $B_n = \max_{j\leq j_n} b_{nj}$ *and* λ *is given as in (B5). Then,*

$$\sup_{x\in R_X}\sup_{t\in I}|d_{tn}(x) - d_t(x)| = O(\max(\Delta_n^{-1/2}, a_n^2)) \ a.s.,$$

where

$$d_{tn}(x) = \frac{1}{na_n}\sum_{i=1}^{n}\Gamma_t(Z_i,\Delta_i|x)K_n(\frac{x-X_i}{a_n}),$$

and

$$d_t(x) = \sum_{\delta=0,1}\int_{-\infty}^{\infty}\Gamma_t(z,\delta|x)h_\delta(x,z)dz.$$

Proof. Write

$$\sup_{x\in R_X}\sup_{t\in I}|d_{tn}(x) - d_t(x)| \leq \sum_{i=1}^{i_0}|q_i|(S_{ni}^{(1)} + S_{ni}^{(2)}),$$

where

$$S_{ni}^{(1)} = 2\sum_{j=1}^{j_n} m_{nj}b_{nj}\sup_{x\in R_X}\sup_{t\in I}|M_{1tn}^{(i,j)}(x) + M_{2tn}^{(i,j)}(x)|, \tag{5.2.13}$$

$$\begin{aligned} M_{1tn}^{(i,j)}(x) + M_{2tn}^{(i,j)}(x) = &\sup_{x\in R_X}\sup_{t\in I}|\frac{1}{2nc_{nj}}\sum_{k=1}^{n}\gamma_{ti}(Z_k,\Delta_k|x)I(x-c_{nj} < X_k \leq x+c_{nj}) \\ &- \sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_{ti}(z,\delta|x)h_\delta(x,z)dz|, \end{aligned}$$

$c_{nj} = a_n b_{nj}$ and

$$\begin{aligned} S_{ni}^{(2)} &= \sup_{x\in R_X}\sup_{t\in I}|(2\sum_{j=1}^{j_n} m_{nj}b_{nj} - 1)\sum_{\delta=0,1}\int_{-\infty}^{\infty}\gamma_{ti}(z,\delta|x)h_\delta(x,z)dz| \\ &\leq C_5 M_\lambda|(2\sum_{j=1}^{j_n} m_{nj}b_{nj} - 1)|. \end{aligned}$$

Next, define

$$\varepsilon_n = 2C_0\Delta_n^{-1/2}\sum_{j=1}^{j_n} m_{nj}b_{nj}^{1/2} + 2C_1 a_n^2 \sum_{j=1}^{j_n} m_{nj}b_{nj}^3.$$

Then,

$$P(S_{ni}^{(1)} > \varepsilon_n) \leq \sum_{j=1}^{j_n} P(\sup_{x\in R_X}\sup_{t\in I}|M_{1tn}^{(i,j)}(x) + M_{2tn}^{(i,j)}(x)| > C_0\Delta_{nj}^{-1/2} + C_1 c_{nj}^2),$$

with $\Delta_{nj} = \Delta_n b_{nj}$ and $c_{nj} = a_n b_{nj}$. By using (5.2.12), we thus obtain

$$P(S_{ni}^{(1)} > \varepsilon_n) \leq O(j_n n^{-2}).$$

For $s < 1$ and using the Borel-Cantelli Lemma, we obtain

$$S_{ni}^{(1)} = O(\varepsilon_n)\ a.s.,$$

for which $\varepsilon_n = O(\max(\Delta_n^{-1/2}, a_n^2))$.

Theorem 5.2.3 *Assume (A1) (ii), (A8), (B7), (B8). For the sequence a_n, we suppose (i) $a_n \to 0$, (ii) $na_n^{5/2}/\log n \to \infty$, (iii) $a_n^{-5/2} \leq (n/\log n)^{1-2/\lambda}$, where λ is given as in (B5), and (iv) $na_n^4 \to 0$. K is a symmetric kernel with bounded support, bounded first derivative and $\int K(u)du = 1$. Then,*

$$\sup_{x\in R_X}\sup_{t\in I}|d_{tn}(x) - d_t(x)| = O(\Delta_n^{-1/2})\ a.s.,$$

where $d_{tn}(x)$ and $d_t(x)$ are defined with kernel K and $\Delta_n = na_n/\log n$. Moreover, if $\inf_{x\in R_X}|f_X(x)| > 0$,

$$\sup_{x\in R_X}\sup_{t\in I}|\frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n})\Gamma_t(Z_i,\Delta_i|x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})} - \frac{d_t(x)}{f_X(x)}| = O(\Delta_n^{-1/2})\ a.s.$$

Proof. Let $b_{nj} = ja_n^{3/2}$ and $m_{nj} = K(ja_n^{3/2}) - K((j+1)a_n^{3/2})$ in (B9). (B9) (i) becomes

$$|(2\sum_{j=1}^{j_n} m_{nj}b_{nj} - 1)| \leq \int |K_n(u) - K(u)|du \leq Ca_n^{3/2},$$

for some $C > 0$. (B9) (ii) and (B9) (iii) become

$$\sup_n a_n^{9/4} \sum_{j=1}^{j_n} j^{1/2} |K'(\theta_{nj})| < \infty$$

and

$$\sup_n a_n^{6} \sum_{j=1}^{j_n} j^{3} |K'(\theta_{nj})| < \infty,$$

where θ_{nj} is between $j a_n^{3/2}$ and $(j+1)a_n^{3/2}$. Therefore, we can choose $0 < s < 1$ such that $j_n = O(a_n^{-3/2})$. Next, if we denote $d_{tn}(x, K)$, $d_{tn}(x)$ using kernel K, it is clear that

$$\begin{aligned}
&\sup_{x \in R_X} \sup_{t \in I} |d_{tn}(x, K) - d_{tn}(x, K_n)| \\
&\leq \sup_{x \in R_X} \sup_{t \in I} \left(\frac{D a_n^{3/2}}{n a_n} \sum_{i=1}^{n} |\Gamma_t(Z_i, \Delta_i | x)| I(x - a_n < X_i \leq x + a_n) \right) = O(a_n^{3/2}) \ a.s.,
\end{aligned}$$

for some constant $D > 0$, a kernel support equal to $[-1, 1]$ and where Proposition 5.2.1 is used with $c_n = a_n$ (with $c_n = (L/2)a_n$ if L is the length of the support).
Finally, write

$$\begin{aligned}
\sup_{x \in R_X} \sup_{t \in I} \left| \frac{\sum_{i=1}^n K(\frac{x - X_i}{a_n}) \Gamma_t(Z_i, \Delta_i | x)}{\sum_{i=1}^n K(\frac{x - X_i}{a_n})} - \frac{d_t(x)}{f_X(x)} \right| &\leq \sup_{x \in R_X} \sup_{t \in I} \left| \frac{d_{tn}(x) - d_t(x)}{f_{nX}(x)} \right| \\
&+ \sup_{x \in R_X} \sup_{t \in I} \left| \frac{d_t(x)(f_X(x) - f_{nX}(x))}{f_X(x) f_{nX}(x)} \right|.
\end{aligned}$$

If we use the fact that $\inf_{x \in R_X} |f_X(x)| > 0$ in addition to the obtained result for $d_{tn}(x, K)$, the two terms on the right hand side of the above expression are $O(\Delta_n^{-1/2})$ $a.s.$ since $\sup_{x \in R_X} \sup_{t \in I} |d_t(x)|$ is bounded (using the definition (B8) of the points $\Gamma(\cdot, \cdot|\cdot)$).

Remark 5.2.4 (density estimator) If we denote $f_{nX}(x) = (1/n a_n) \sum_{i=1}^n K(\frac{x - X_i}{a_n})$, the classical density estimator, we have using Proposition 5.2.3 with $\Gamma(Z_i, \Delta_i | x) = 1$ that $\sup_{x \in R_X} |f_{nX}(x) - f_X(x)| = O(\Delta_n^{-1/2})$ $a.s.$, since $\sup_{x \in R_X} |f_X(x)| < \infty$.

Remark 5.2.5 (moment conditions) For a number of artificial data points, the moment conditions in (B1) and (B5) are not used. Indeed, those data points can often be of the form

$\gamma_{t^*}(Z_i, \Delta_i|x) \leq \gamma^*_{t^*}(Z_i, \Delta_i)$ and such that $L_0(Z_i, \Delta_i|x) \leq L^*_0(Z_i, \Delta_i)$. In this case, strong law of large numbers can be immediately used with $(1/n)\sum_{i=1}^n \gamma^{*\lambda}_{t^*}(Z_i, \Delta_i) - E[\gamma^{*\lambda}_{t^*}(Z, \Delta)]$ and $(1/n)\sum_{i=1}^n L^*_0(Z_i, \Delta_i) - E[L^*_0(Z, \Delta)]$. The terms $V_n\Delta_{nj}^{-1/2}$ and $2W_n\Delta_{nj}^{-1/2}$ can then be treated outside Proposition 5.2.1 and be directly introduced in (5.2.13) such that the final result of Theorem 5.2.3 is preserved.

Remark 5.2.6 (boundary effects) The degree of smoothing of $f_{X|Z,\Delta}(x|z,\delta)$ allows via (A8) to obtain the artificial order $O(\Delta_n^{-1/2})$ near the boundaries of R_X. If we suppose for instance the weaker condition

$$\sup_{|x-x'|\leq d,\ x,x'\in R_X} \sup_{t\in I} \Big| \sum_{\delta=0,1} \int \gamma_t(z,\delta|x)(h_\delta(x',z) - h_\delta(x,z))dz \Big| \leq Cd,$$

instead of A(8), then the more realistic rate $O(a_n)$ can be obtained near the boundaries.

Remark 5.2.7 (bandwidth assumptions) The bandwidth parameter a_n could tend to zero more slowly. Indeed, the condition $na_n^4 \to 0$ of Theorem 5.2.3 can be written with another power on a_n. By example, if $na_n^5(\log n)^{-1} = O(1)$, Theorem 5.2.3 also holds with $b_{nj} = ja_n^2$ if $na_n^3/\log n \to \infty$ and $a_n^{-3} \leq (n/\log n)^{1-2/\lambda}$.

Remark 5.2.8 (artificial data representation) The representation

$$\Gamma_t(z,\delta|x) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(z,\delta|x),$$

needed in the above proofs, requires nonnegative $\gamma_{ti}(z,\delta|x)$, $i = 1,\ldots,i_0$. This assumption is not restrictive since any random variable X with real values can be represented by $X = \max(X,0) - (-\min(X,0))$, where the two terms of this difference are nonnegative.

5.3 Modulus of continuity for the weighted average of conditional synthetic data points

The development of this section is similar to Section 5.2. The strong uniform consistency of the modulus of continuity is established via two preliminary results.

Proposition 5.3.1 *Assume (A1) (ii), (A8), (B7) and (B10). Then,*

$$P(M_{9n}(c_n) > C_{m0}\Delta_n^{-1/2}d_n^{1/2} + C_{m1}c_n^2 d_n) = O(n^{-2}),$$

where

$$\begin{aligned} M_{9n}(c_n) \;=\; & \sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}\Big|\frac{1}{2nc_n}\sum_{i=1}^{n}(\gamma_t(Z_i,\Delta_i|x)-\gamma_s(Z_i,\Delta_i|x))I(x-c_n<X_i\le x+c_n) \\ & -\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x)-\gamma_s(z,\delta|x))h_\delta(x,z)dz\Big|, \end{aligned}$$

$\gamma_t(z,\delta|x),\ t\in I,\ x\in R_X,\ z\in R,\ \delta=0,1,$ *satisfy assumptions (B1)-(B5) and (B11).*

Proof. Let

$$\begin{aligned} M_{9n}(c_n) \;=\; & \sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}\Big|\frac{1}{2nc_n}\sum_{i=1}^{n}(\gamma_t(Z_i,\Delta_i|x) \\ & \qquad -\gamma_s(Z_i,\Delta_i|x))I(x-c_n<X_i\le x+c_n) \\ & -\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x)-\gamma_s(z,\delta|x))h_\delta(s,z)dzds\Big| \\ & +\sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}\Big|\frac{1}{2c_n}\int_{x-c_n}^{x+c_n}\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x)-\gamma_s(z,\delta|x))h_\delta(s,z)dzds \\ & -\sum_{\delta=0,1}\int_{-\infty}^{\infty}(\gamma_t(z,\delta|x)-\gamma_s(z,\delta|x))h_\delta(x,z)dz\Big| \\ =\; & \sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}|M_{10stn}(x)|+\sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}|M_{11stn}(x)|. \end{aligned}$$

First, $M_{11stn}(x)$ is treated as $M_{2tn}(x)$ in Proposition 5.2.1 such that

$$\sup_{\{x\in R_X,|t-s|\le d_n,\ s,t\in I\}}|M_{11stn}(x)|\le C_{m1}d_nc_n^2, \tag{5.3.1}$$

using (B11).

Divide R_X into $[\frac{2L_X}{f_n d_n^{1/2}}]$ intervals of length smaller or equal to $f_n d_n^{1/2}$. Denote J_X the set of points $\{x_k = x_0 + k[\frac{2L_X}{f_n d_n^{1/2}}]^{-1} L_X, \quad 1 \le k \le [\frac{2L_X}{f_n d_n^{1/2}}] - 1 = L_X^{d^n}\}$ and $x_{L_X^{d^n}+1} = \sup\{x : x \in R_X\}$ which limit the intervals. $M_{10stn}(x)$ is treated like (5.2.2) in Proposition 5.2.1, where $\gamma_t(\cdot,\cdot|\cdot)$ is replaced by $\gamma_t(\cdot,\cdot|\cdot) - \gamma_s(\cdot,\cdot|\cdot)$, $V_n(x_j)$ by $V_n^d(x_j) = (1/n)\sum_{i=1}^n L_0(Z_i, \Delta_i|x_j) - E[L_0(Z, \Delta|x_j)]$ and V_n by $V_n^d = \max_{x_j \in J_X}(2E[L_0(Z,\Delta|x_j)] + |V_n^d(x_j)|)$. Using Chebyshev's inequality, $P(V_n^d > C_{m2}) = o(n^{-2})$ with C_{m2} chosen larger than $4L_6^{1/6}$. A development similar to (5.2.3) and (5.2.6) is used to obtain

$$\begin{aligned} &\sup_{x \in R_X} \sup_{|t-s| \le d_n,\ s,t \in I} |M_{10stn}(x)| \qquad (5.3.2)\\ &\le 2c_n^{-1} \max_{x_j \in J_X} \max_{|t-s| \le d_n,\ s,t \in I} \max_{k \in \{-1,0,1\}} \sup_{|z| \le c_n} M_{12stn}(x_j, x_{j+k}, z) + V_n^d \Delta_n^{-1/2} d_n^{1/2}, \end{aligned}$$

where

$$\begin{aligned} M_{12stn}(x_j, x, z) &= |G_{stn}(x_j, x+z) - G_{stn}(x_j, x) - [G_{st}(x_j, x+z) - G_{st}(x_j, x)]|,\\ G_{stn}(x_j, x) &= n^{-1}\sum_{i=1}^n (\gamma_t(Z_i, \Delta_i|x_j) - \gamma_s(Z_i, \Delta_i|x_j)) I(X_i \le x), \end{aligned}$$

and

$$G_{st}(x_j, x) = E[G_{stn}(x_j, x)] = \int_{x_0}^x \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta|x_j) - \gamma_s(z, \delta|x_j)) h_\delta(s, z) dz ds.$$

Partition I into $O(f_n^{-1} d_n^{-3/2})$ intervals such that for each x_j, $j = 0, \ldots, L_X^{d^n} + 2$, $g(t^*|x_j) - g(t_*|x_j)$ is divided into $m_j = [\frac{g(t^*|x_j) - g(t_*|x_j)}{C_L d_n}]$ intervals of length $C_{m3}(x_j) C_L d_n$, $1 \le C_{m3}(x_j) \le 2$ ($|g(t_*|x_j) - g(t_1|x_j)| = C_{m3}(x_j) C_L d_n$, $|g(t_{\alpha+1}|x_j) - g(t_\alpha|x_j)| = C_{m3}(x_j) C_L d_n$, $\alpha = 1, \ldots, m_j - 2$, $|g(t^*|x_j) - g(t_{m_j-1}|x_j)| = C_{m3}(x_j) C_L d_n$, $t_0 = t_*$, $t_{m_j} = t^*$ for all j). Let $I_{j\alpha} = [g(t_{\alpha-1}|x_j), g(t_{\alpha+1}|x_j)]$, $\alpha = 1, \ldots, m_j - 1$. For each $s, t \in I$ with $|t - s| \le d_n$, there exists an interval $I_{j\alpha}$ such that $g(s|x_j), g(t|x_j) \in I_{j\alpha}$. Partition each $I_{j\alpha}$ by a grid $g(t_{\alpha\beta}|x_j) = g(t_\alpha|x_j) + \beta \frac{C_{m3}(x_j) C_L d_n}{p_n}$, $\beta = -p_n, \ldots, p_n$, where $p_n = [\Delta_n^{1/2} d_n^{1/2} + 1]$. Using (A8), (B4), (B5) and the monotonicity of $\gamma_t(Z, \Delta|x)$, (5.3.2) is majorized by

$$\begin{aligned} &2c_n^{-1} \max_{x_j \in J_X} \max_{k \in \{-1,0,1\}} \max_{1 \le \alpha \le m_j - 1} \max_{-p_n \le \beta, \zeta \le p_n} \sup_{|z| \le c_n} M_{12 t_{\alpha\zeta} t_{\alpha\beta} n}(x_j, x_{j+k}, z) \qquad (5.3.3)\\ &+4c_n^{-1} \max_{x_j \in J_X} \max_{1 \le \alpha \le m_j - 1} \max_{-p_n \le \beta \le p_n - 1} C_5 c_n |g(t_{\alpha(\beta+1)}|x_j) - g(t_{\alpha\beta}|x_j)| + V_n^d \Delta_n^{-1/2} d_n^{1/2}, \end{aligned}$$

where C_5 is defined as in (5.2.7) and the second term of the above expression equals $4C_5\frac{C_{m3}(x_j)C_L d_n}{p_n} \leq C_{m4}\Delta_n^{-1/2}d_n^{1/2}$, where $C_{m4} = 8C_5C_L$. Now, put $T_n = M_\lambda(f_n^{-1}d_n^{-1/2})^{\frac{1}{\lambda-1}}$. Define M_λ for $2 < \lambda \leq \infty$ as in the proof of Proposition 5.2.1, $H_{stn}(x_j, x)$ by

$$n^{-1}\sum_{i=1}^{n}(\gamma_t(Z_i,\Delta_i|x_j) - \gamma_s(Z_i,\Delta_i|x_j))I(|\gamma_t(Z_i,\Delta_i|x_j) - \gamma_s(Z_i,\Delta_i|x_j)| \leq T_n)I(X_i \leq x),$$

and $M_{13stn}(x_j, x, z)$ by substitution of H_{stn} for G_{stn} and $E[H_{stn}]$ for G_{st}. (5.3.3) is then majorized by

$$\begin{aligned}&2c_n^{-1}\max_{x_j\in J_X}\max_{k\in\{-1,0,1\}}\max_{1\leq\alpha\leq m_j-1}\max_{-p_n\leq\beta,\zeta\leq p_n}\sup_{|z|\leq c_n} M_{13t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, z)\\ &+2c_n^{-1}f_n d_n^{1/2}(W_n^d + \theta_n^d) + (V_n^d + C_{m4})\Delta_n^{-1/2}d_n^{1/2},\end{aligned} \tag{5.3.4}$$

where W_n^d and θ_n^d are defined similarly to W_n and θ_n in the proof of Proposition 5.2.1. It is easy to check that $P(2W_n^d > C_{m5}) = o(n^{-2})$ and $2\theta_n^d < C_{m6}$, where C_{m5} and C_{m6} are chosen such that $C_{m5} > 2^{\lambda+2}M_\lambda$ and $C_{m6} = 2^{\lambda+1}M_\lambda$.
Next, consider

$$\kappa_{njkr} = x_{j+k} + \frac{rc_n}{p_n}, \quad for\ r = -p_n, -p_n+1, \ldots, p_n.$$

Define $\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}$ by

$$|H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j,\kappa_{njkr}) - H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j,x_{j+k}) - [E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j,\kappa_{njkr})] - E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j,x_{j+k})]]|.$$

For fixed $j, k, \alpha, \beta, \zeta, n$, $H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}+z) - H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k})$ and $E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}+z)] - E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k})]$ are monotone with respect to z and have finite limits in $x_{j+k}+c_n$ and $x_{j+k} - c_n$. Therefore,

$$\begin{aligned}&\sup_{|z|\leq c_n} M_{13t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, z)\\ &\leq \max_{-p_n\leq r\leq p_n}\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}\\ &\quad+\max_{-p_n\leq r\leq p_n-1}|E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}+\frac{(r+1)c_n}{p_n})] - E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}+\frac{rc_n}{p_n})]|,\end{aligned}$$

where the second term on the right hand side of the above expression is bounded by

$$4C_LC_5c_n\Delta_n^{-1/2}d_n^{1/2}.$$

Therefore, (5.3.4) is majorized by

$$2c_n^{-1}\max_{x_j\in J_X}\max_{k\in\{-1,0,1\}}\max_{1\le\alpha\le m_j-1}\max_{-p_n\le\beta,\zeta,r\le p_n}\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}$$
$$+(2W_n^d+V_n^d+C_{m4}+C_{m6}+C_{m7})\Delta_n^{-1/2}d_n^{1/2}, \tag{5.3.5}$$

where $C_{m7}=8C_LC_5$.
Next,

$$P(\sup_{x\in R_X}\sup_{|t-s|\le d_n,\ s,t\in I}|M_{10stn}(x)|>C_{m0}\Delta_n^{-1/2}d_n^{1/2}) \tag{5.3.6}$$
$$\le\sum_{j,k,\alpha,\beta,\zeta,r}P(\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}>C'_{m0}f_nd_n^{1/2})+P(W_n^d>C'_{m0})+P(V_n^d>2C'_{m0}),$$

where $C'_{m0}=(1/6)(C_{m0}-16C_5C_L-2^{\lambda+1}M_\lambda)$ and C_{m0}, C_{m2} and C_{m5} have to satisfy

$$\max(2^{\lambda+1}M_\lambda,2L_6^{1/6})<C'_{m0}=C_{m2}/2=C_{m5}/2.$$

By Bernstein's inequality,

$$P(\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}>C'_{m0}f_nd_n^{1/2})\le 2\exp(-\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}),$$

where

$$\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}=\frac{C'^2_{m0}n^2f_n^2d_n}{2n\sigma^2_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}+\frac{2}{3}nC'_{m0}f_nd_n^{1/2}T_n},$$

$\sigma^2_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}=Var[\Omega_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}]$ and

$$\begin{aligned}\Omega_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}\quad&=(\gamma_{t_{\alpha\beta}}(Z,\Delta|x_j)-\gamma_{t_{\alpha\zeta}}(Z,\Delta|x_j))\\&\times I(|\gamma_{t_{\alpha\beta}}(Z,\Delta|x_j)-\gamma_{t_{\alpha\zeta}}(Z,\Delta|x_j)|\le T_n)(I(X\le\kappa_{njkr})-I(X\le x_{j+k})).\end{aligned}$$

Using (B11) (ii), $\sigma^2_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}\le C_{L_2}C_5c_nd_n$, and (B10) (iii), $T_n\Delta_n^{-1/2}\le d_n^{1/2}$. Therefore,

$$\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}\ge C''_{m0}\log n,$$

with

$$C''_{m0}=\frac{C'^2_{m0}}{2(C_{L_2}C_5+\frac{1}{3}C'_{m0})}.$$

Finally,

$$\sum_{j,k,\alpha,\beta,\zeta,r} P(\mathcal{M}_{nt_{\alpha\zeta}t_{\alpha\beta}jkr} > C'_{m0} f_n d_n^{1/2}) \leq 2 \sum_{j=1}^{L_X^{d^n}} \sum_{k=-1}^{1} \sum_{\alpha=1}^{m_j-1} \sum_{-p_n \leq \beta,\zeta,r \leq p_n} n^{-C''_{m0}},$$

for which the highest order term on the right hand side is

$$96 \frac{L_X M_\lambda}{C_L} \Delta_n^2 c_n^{-1} n^{-C''_{m0}} \leq 96 \frac{L_X M_\lambda}{C_L} \frac{n^2}{(\log n)^2} n^{-C''_{m0}}.$$

Choosing C''_{m0} sufficiently large finishes the proof.

Proposition 5.3.2 *Assume (A1) (ii), (A8), (B7), (B8), (B11), (B12). a_n and d_n satisfy (i) $a_n B_n \to 0$, $d_n \to 0$, (ii) $n a_n b_n / \log n \to \infty$ and (iii) $a_n^{-1} \leq b_n d_n (n/\log n)^{1-2/\lambda}$, where $b_n = \min_{j \leq j_n} b_{nj}$, $B_n = \max_{j \leq j_n} b_{nj}$, and λ is given as in (B5). Then,*

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n,\ s,t \in I} |d_{stn}(x) - d_{st}(x)| = O(\max(\Delta_n^{-1/2} d_n^{1/2}, a_n^2 d_n)) \ a.s.,$$

where

$$d_{stn}(x) = \frac{1}{n a_n} \sum_{i=1}^{n} (\Gamma_t(Z_i, \Delta_i | x) - \Gamma_s(Z_i, \Delta_i | x)) K_n(\frac{x - X_i}{a_n}),$$

$$d_{st}(x) = \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\Gamma_t(z, \delta | x) - \Gamma_s(z, \delta | x)) h_\delta(x, z) dz.$$

Proof. The proof is along the same lines as the proof of Proposition 5.2.2.

Theorem 5.3.3 *Assume (A1) (ii), (A8), (B7), (B8), (B11). a_n and d_n satisfy (i) $a_n \to 0$, $d_n \to 0$, (ii) $n a_n^{5/2} d_n^{-1/2} / \log n \to \infty$, (iii) $a_n^{-5/2} \leq d_n^{1/2} (n/\log n)^{1-2/\lambda}$, where λ is given as in (B5), (iv) $\log n / n a_n d_n = O(1)$ and (v) $n a_n^4 \to 0$. K is a symmetric kernel with bounded support, bounded first derivative and $\int K(u) du = 1$. Then,*

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n,\ s,t \in I} |d_{stn}(x) - d_{st}(x)| = O(\Delta_n^{-1/2} d_n^{1/2}) \ a.s.,$$

where $d_{stn}(x)$ and $d_{st}(x)$ are defined with kernel K and $\Delta_n = n a_n / \log n$. Moreover, if $n a_n^{5/2} / \log n \to \infty$ and $\inf_{x \in R_X} |f_X(x)| > 0$,

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n,\ s,t \in I} |\frac{\sum_{i=1}^n K(\frac{x - X_i}{a_n}) \Gamma_{ts}(Z_i, \Delta_i | x)}{\sum_{i=1}^n K(\frac{x - X_i}{a_n})} - \frac{d_{st}(x)}{f_X(x)}| = O(\Delta_n^{-1/2} d_n^{1/2}) \ a.s.,$$

where $\Gamma_{ts}(Z_i, \Delta_i|x) = \Gamma_t(Z_i, \Delta_i|x) - \Gamma_s(Z_i, \Delta_i|x)$.

Proof. Let $b_{nj} = ja_n^{3/2}d_n^{-1/2}$ and $m_{nj} = K(ja_n^{3/2}d_n^{-1/2}) - K((j+1)a_n^{3/2}d_n^{-1/2})$ in (B12). (B12) (i) becomes

$$|(2\sum_{j=1}^{j_n} m_{nj}b_{nj} - 1)| \leq \int |K_n(u) - K(u)|du \leq Ca_n^{3/2}d_n^{-1/2},$$

for some $C > 0$. (B12) (ii) and (B12) (iii) are easily satisfied using $j_n = O(a_n^{-3/2}d_n^{1/2})$ such that s can then be chosen between 0 and 1. Next, let denote $d_{stn}(x, K)$, $d_{stn}(x)$ using kernel K. It is clear that

$$\begin{aligned}
&\sup_{x\in R_X} \sup_{|t-s|\leq d_n,\ s,t\in I} |d_{stn}(x, K) - d_{stn}(x, K_n)| \\
&\leq \sup_{x\in R_X} \sup_{|t-s|\leq d_n,\ s,t\in I} (\frac{Da_n^{3/2}d_n^{-1/2}}{na_n} \sum_{i=1}^{n} |\Gamma_{ts}(Z_i, \Delta_i|x)|I(x - a_n < X_i \leq x + a_n)) \\
&= O(a_n^{3/2}d_n^{1/2}) \ a.s.,
\end{aligned}$$

for which we use Proposition 5.3.1 with $c_n = a_n$ (for a kernel support equal to $[-1, 1]$).

Remark 5.3.4 (bandwidth assumptions) If $na_n^5(\log n)^{-1} = O(1)$, Theorem 5.3.3 also holds with $b_{nj} = ja_n^2d_n^{-1/2}$ if $na_n^3/\log n \to \infty$ and $a_n^{-3} \leq d_n^{1/2}(n/\log n)^{1-2/\lambda}$.

Chapter 6

Mean preservation in nonparametric regression with censored data

In Chapter 4, we studied a first estimator of general L-functionals based on model (1.7.1). This estimator outperforms the completely nonparametric estimator in most cases since it allows to reduce considerably the inconsistency problem in the right tails of the estimated conditional distribution functions. However, it may happen that this estimator suffers from some stability flaws due to the preliminary nonparametric estimation of $m^0(\cdot)$ and $\sigma^0(\cdot)$. Indeed, the estimator (4.1.3) can be rewritten as

$$\begin{aligned}\hat{m}^T(x) \; &= a_0\{\hat{m}^0(x)\int_{-\infty}^{T} L(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) + \hat{\sigma}^0(x)\int_{-\infty}^{T} eL(\hat{F}_\varepsilon^0(e))\,d\hat{F}_\varepsilon^0(e)\} \\ &\quad + \sum_{j=1}^{k} a_j\{\hat{m}^0(x) + \hat{\sigma}^0(x)[(\hat{F}_\varepsilon^0)^{-1}(s_j) \wedge T]\},\end{aligned}$$

so that it highly depends on the preliminary location and scale estimations. In Section 4.2, these estimations were using the score function J chosen as

$$J(s) = b^{-1}I(0 \le s \le b) \;\; (0 \le s \le 1),$$

where $b = \min_{1\leq i\leq n} \tilde{F}(+\infty|X_i)$. Since $\tilde{F}(\cdot|X_1), \ldots, \tilde{F}(\cdot|X_n)$ are Beran estimators, their typical inconsistency problems may involve small values of b and therefore variability and instability of the estimations in the whole covariate space.

In light of this drawback, a possible alternative can be proposed as outlined in Section 1.8. Its objective is thus twofold : we want an estimator which uses a maximal amount of local information (i.e. which deletes the effect of b) but which preserves the nice properties of the estimated residuals distribution in the right tails as well.

In Section 6.1, the details of the estimation procedure and the assumptions needed for the main results of this chapter are given. In Section 6.2, uniform consistency and asymptotic normality of our second conditional location estimator are stated. The asymptotic normality follows from an asymptotic representation which is also established in that section. Moreover, the same asymptotic properties and a modulus of continuity result are obtained for the estimator of the corresponding conditional distribution function. Indeed, these results are required in the proofs of the asymptotic properties of the conditional location function. In Section 6.3, a simulation study is carried out in which the estimators of Chapters 4 and 6 as well as the corresponding completely nonparametric estimator are compared. Finally, we add the new estimator of Chapter 6 to the analysis of quasars data. The results of this chapter are also available in Heuchenne and Van Keilegom (2005c).

6.1 Notations and description of the method

As outlined in the introduction, the objective is first to estimate the new (functions of the) data points (1.8.1) and then, to approximate $m(\cdot)$ by a Nadaraya-Watson estimator based on these data. As in Chapter 4, we estimate $F(y|x)$ by

$$\hat{F}_1(y|x) = \hat{F}_\varepsilon^0\Big(\frac{y - \hat{m}^0(x)}{\hat{\sigma}^0(x)}\Big), \tag{6.1.1}$$

where $\hat{F}_\varepsilon^0(\cdot)$, $\hat{m}^0(x)$ and $\hat{\sigma}^0(x)$ follow the same estimation procedure as in Chapter 4. Now, let $\hat{\phi}_1(y|x) = yL(\hat{F}_1(y \wedge T_x|x))$ and $\hat{\phi}_{2t}(y|x) = \hat{\phi}_2(y|x) = I(y \leq t)$, and let

$$\hat{\phi}_j^*(z, \delta|x) = \hat{\phi}_j(z|x)\delta + \frac{1}{1 - \hat{F}_1(z \wedge T_x|x)} \int_{z\wedge T_x}^{T_x} \hat{\phi}_j(y|x) d\hat{F}_1(y|x)(1 - \delta), \tag{6.1.2}$$

where $T_x = T\sigma^0(x) + m^0(x)$ and $T < \tau_{H^0_\varepsilon}$. Note that we have to truncate the integral at T_x in the above definition. However, when $\tau_{F^0_\varepsilon} \leq \tau_{G^0_\varepsilon}$, T can be chosen arbitrarily close to $\tau_{F^0_\varepsilon}$. The estimator of $m(x)$ is now defined by

$$\begin{aligned}
&\hat{m}_1^T(x) \\
&= a_0 \sum_{i=1}^n W_i(x, a_n) \hat{\phi}_1^*(Z_i, \Delta_i | x) + \sum_{j=1}^k a_j [\hat{F}_{\phi_2}^{-1}(s_j|x) \wedge T_x] \\
&= a_0 \sum_{i=1}^n W_i(x, a_n) \Big[Y_i L(\hat{F}_1(Y_i \wedge T_x | x)) \Delta_i \qquad (6.1.3) \\
&\quad + \frac{1}{1 - \hat{F}_1(C_i \wedge T_x | x)} \int_{C_i \wedge T_x}^{T_x} y L(\hat{F}_1(y|x)) d\hat{F}_1(y|x)(1 - \Delta_i) \Big] + \sum_{j=1}^k a_j [\hat{F}_{\phi_2}^{-1}(s_j|x) \wedge T_x],
\end{aligned}$$

where

$$\begin{aligned}
&\hat{F}_{\phi_2}(t|x) \\
&= \sum_{i=1}^n W_i(x, a_n) \hat{\phi}_{2t}^*(Z_i, \Delta_i | x) \qquad (6.1.4) \\
&= \sum_{i=1}^n W_i(x, a_n) \left[I(Y_i \leq t) \Delta_i + \frac{1}{1 - \hat{F}_1(C_i \wedge T_x | x)} \int_{C_i \wedge T_x}^{T_x} I(y \leq t) d\hat{F}_1(y|x)(1 - \Delta_i) \right].
\end{aligned}$$

Note that $\hat{m}_1^T(x)$ is actually estimating

$$m_1^T(x) = a_0 E[\tilde{\phi}_1^*(Z, \Delta | x) | x] + \sum_{j=1}^k a_j [F_{\phi_2}^{-1}(s_j|x) \wedge T_x],$$

where

$$\tilde{\phi}_j^*(z, \delta | x) = \tilde{\phi}_j(z|x)\delta + \frac{1}{1 - F(z \wedge T_x | x)} \int_{z \wedge T_x}^{T_x} \tilde{\phi}_j(y|x) dF(y|x)(1 - \delta),$$

$\tilde{\phi}_1(y|x) = yL(F(y \wedge T_x|x))$, $\tilde{\phi}_{2t}(y|x) = \phi_{2t}(y|x)$ and $F_{\phi_2}(t|x) = E[\tilde{\phi}_{2t}^*(Z, \Delta|x)|x]$. As before, $m_1^T(x)$ and $F_{\phi_2}(t|x)$ can be made arbitrarily close to $m(x)$ and $F(t|x)$ respectively, provided $\tau_{F^0_\varepsilon} \leq \tau_{G^0_\varepsilon}$.

The functions we need in the asymptotic representations of Section 6.2 are $\xi(z, \delta, y|x)$, $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ of Section 2.1 and $h_{x,y}(z, \delta)$, $A(t, z, \delta|x)$ and $B_1(z, \delta|x)$ that are defined below.

$$h_{x,y}(z,\delta) = \Big[\eta(z,\delta|x) + \zeta(z,\delta|x)\frac{y-m^0(x)}{\sigma^0(x)}\Big] f_\varepsilon^0\big(\frac{y-m^0(x)}{\sigma^0(x)}\big) f_X^{-1}(x),$$

$$A(t,z,\delta|x)$$

$$= E\left[I(\Delta=0)\{h_{x,Z_x^T}(z,\delta)\frac{F(T_x\wedge t|x)-F(Z_x^T\wedge t|x)}{(1-F(Z_x^T|x))^2} + \frac{h_{x,T_x\wedge t}(z,\delta)+h_{x,Z_x^T\wedge t}(z,\delta)}{1-F(Z_x^T|x)}\}\right]$$

$$+f_X^{-1}(x)[\tilde\phi_{2t}^*(z,\delta|x) - E\{\tilde\phi_{2t}^*(Z,\Delta|x)|x\}],$$

$$B_1(z,\delta|x) = a_0 E\Big[I(\Delta=1)ZL'(F(Z_x^T|x))h_{x,Z_x^T}(z,\delta)$$

$$+I(\Delta=0)h_{x,Z_x^T}(z,\delta)\Big\{\frac{\int_{Z_x^T}^{T_x} M(y|x)\,dF(y|x)}{(1-F(Z_x^T|x))^2} - \frac{M(Z_x^T|x)}{1-F(Z_x^T|x)}\Big\}$$

$$+I(\Delta=0)h_{x,T_x}(z,\delta)\frac{M(T_x|x)}{1-F(Z_x^T|x)} - I(\Delta=0)\frac{\int_{Z_x^T}^{T_x} h_{x,y}(z,\delta)L(F(y|x))\,dy}{1-F(Z_x^T|x)}\Big]$$

$$+a_0 f_X^{-1}(x)[\tilde\phi_1^*(z,\delta|x) - E\{\tilde\phi_1^*(Z,\Delta|x)|x\}]$$

$$-\sum_{j=1}^{k}\frac{a_j A(F^{-1}(s_j|x),z,\delta|x)I(s_j\le F(T_x|x))}{f(F^{-1}(s_j|x)|x)},$$

where $Z_x^T = Z\wedge T_x$ and $M(y|x) = yL(F(y|x))$.

In the proofs of Section 6.2, we will also use the abbreviated notations $\hat T^x = (T_x - \hat m^0(x))/\hat\sigma^0(x)$, $T^x = (T_x - m^0(x))/\sigma^0(x) = T$, $\hat E_{ix}^0 = (Z_i - \hat m^0(x))/\hat\sigma^0(x)$, $E_{ix}^0 = (Z_i - m^0(x))/\sigma^0(x)$, $\hat E_{ix}^{0T} = \hat E_{ix}^0\wedge\hat T^x$, $E_{ix}^{0T} = E_{ix}^0\wedge T$, $\hat E_{ix}^{0T_t} = (Z_i\wedge T_x\wedge t - \hat m^0(x))/\hat\sigma^0(x)$, $E_{ix}^{0T_t} = (Z_i\wedge T_x\wedge t - m^0(x))/\sigma^0(x)$, $\hat T_t^x = (T_x\wedge t - \hat m^0(x))/\hat\sigma^0(x)$ and $T_t^x = (T_x\wedge t - m^0(x))/\sigma^0(x)$. The following hypothesis are also needed for the proofs of Section 6.2.

$(A1)(i)-(iii)$, $(A2)(i)-(ii)$, $(A4)(i)$ and $(A8)$ mentioned in Section 2.1.

$(A3)(i)-(ii)$ mentioned in Section 2.1.

$(iii)''$ $E|Z|^\lambda < \infty$, with $\lambda \ge (12+8\delta)/(1+4\delta)$ and δ chosen as in $(A1)(i)$.

$(A6)'$ mentioned in Section 4.1.

$(A11)(i)$ mentioned in Section 4.1.

$(A11)(ii)'$ L is twice continuously differentiable, $\int_0^1 L(s)ds = 1$ and $L(s) \geq 0$ for all $0 \leq s \leq 1$.

6.2 Asymptotic results

First, we give the three main results concerning the estimator $\hat{m}_1^T(x)$ proposed in Section 6.1, and then we state the four announced asymptotic properties of the estimator $\hat{F}_{\phi_2}(t|x)$ defined in (6.1.4). The last modulus of continuity result is secondary, as the obtained rate is not optimal. Indeed, its use in the other proofs doesn't require a better rate.

6.2.1 L-functionals results

Theorem 6.2.1 *Assume (A1) (i)-(iii), (A2) (i)-(ii), (A3) (i)-(ii), (A3) (iii)", (A4) (i), (A6)', (A8), (A11) (i), L is continuously differentiable,* $\int_0^1 L(s)ds = 1$ *and* $L(s) \geq 0$ *for all* $0 \leq s \leq 1$. *Then,*

$$\sup_{x \in R_X} |\hat{m}_1^T(x) - m_1^T(x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Proof. Consider the expression $\hat{m}_1^T(x) - m_1^T(x) = \Omega_1(x) + \Omega_2(x)$, where

$$\begin{aligned}\Omega_1(x) &= a_0 \sum_{i=1}^n W_i(x, a_n)[\hat{\phi}_1^*(Z_i, \Delta_i|x) - \tilde{\phi}_1^*(Z_i, \Delta_i|x)] \\ &\quad + a_0 \sum_{i=1}^n W_i(x, a_n)[\tilde{\phi}_1^*(Z_i, \Delta_i|x) - E\{\tilde{\phi}_1^*(Z, \Delta|x)|x\}] \\ &= \Omega_{11}(x) + \Omega_{12}(x),\end{aligned}$$

and

$$\Omega_2(x) = \sum_{j=1}^k a_j(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x) - \sum_{j=1}^k a_j(F_{\phi 2}^{-1}(s_j|x) \wedge T_x).$$

First, we treat $\Omega_{11}(x)$.

$$\begin{aligned}\Omega_{11}(x) = a_0 \sum_{i=1}^{n} W_i(x,a_n) \Big\{ & I(\Delta_i = 1) Y_i [L(\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T})) - L(F_\varepsilon^0(E_{ix}^{0T}))] \\ & I(\Delta_i = 0) \Big[\frac{\int_{\hat{E}_{ix}^{0T}}^{\hat{T}^x} (\hat{m}^0(x) + \hat{\sigma}^0(x)e) L(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{1 - \hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T})} \\ & - \frac{\int_{E_{ix}^{0T}}^{T} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} \Big] \Big\} \\ = a_0 \sum_{i=1}^{n} & W_i(x,a_n) \left\{ A_{1i}(x) + A_{2i}(x) \right\}.\end{aligned}$$

Using the fact that $\sup_{x,z} \left| \hat{F}_\varepsilon^0 \left\{ \frac{z \wedge T_x - \hat{m}^0(x)}{\hat{\sigma}^0(x)} \right\} - F_\varepsilon^0 \left\{ \frac{z \wedge T_x - m^0(x)}{\sigma^0(x)} \right\} \right| = O_P((na_n)^{-1/2} (\log a_n^{-1})^{1/2})$ (see (2.2.2) in Theorem 2.2.1) and $E[|Z|] < \infty$, we have

$$\sup_x |\sum_{i=1}^{n} W_i(x,a_n) A_{1i}(x)| = O_P((na_n)^{-1/2} (\log(a_n)^{-1})^{1/2}).$$

Next, write

$$\begin{aligned} A_{2i}(x) \; = I(\Delta_i = 0) \Bigg\{ & \frac{(\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})) \int_{\hat{E}_{ix}^{0T}}^{\hat{T}^x} (\hat{m}^0(x) + \hat{\sigma}^0(x)e) L(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{(1 - \hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}))(1 - F_\varepsilon^0(E_{ix}^{0T}))} \\ & + \frac{\int_{\hat{E}_{ix}^{0T}}^{E_{ix}^{0T}} (\hat{m}^0(x) + \hat{\sigma}^0(x)e) L(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} + \frac{\int_{T}^{\hat{T}^x} (\hat{m}^0(x) + \hat{\sigma}^0(x)e) L(\hat{F}_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} \\ & + \frac{1}{1 - F_\varepsilon^0(E_{ix}^{0T})} \int_{E_{ix}^{0T}}^{T} [(\hat{m}^0(x) - m^0(x)) + (\hat{\sigma}^0(x) - \sigma^0(x))e] \\ & \qquad \times [L(\hat{F}_\varepsilon^0(e)) - L(F_\varepsilon^0(e))] d\hat{F}_\varepsilon^0(e) \\ & + \frac{1}{1 - F_\varepsilon^0(E_{ix}^{0T})} \int_{E_{ix}^{0T}}^{T} [(\hat{m}^0(x) - m^0(x)) + (\hat{\sigma}^0(x) - \sigma^0(x))e] L(F_\varepsilon^0(e)) d\hat{F}_\varepsilon^0(e) \\ & + \frac{1}{1 - F_\varepsilon^0(E_{ix}^{0T})} \int_{E_{ix}^{0T}}^{T} (m^0(x) + \sigma^0(x)e) [L(\hat{F}_\varepsilon^0(e)) - L(F_\varepsilon^0(e))] d\hat{F}_\varepsilon^0(e) \\ & + \frac{1}{1 - F_\varepsilon^0(E_{ix}^{0T})} \int_{E_{ix}^{0T}}^{T} (m^0(x) + \sigma^0(x)e) L(F_\varepsilon^0(e)) d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e)) \Bigg\} \\ = I(\Delta_i = 0) & \sum_{j=1}^{7} B_{ji}. \end{aligned} \tag{6.2.1}$$

Using Corollary 3.2 and Proposition 4.5 of Van Keilegom and Akritas (1999) (hereafter denoted by VKA), the above-mentioned uniform consistency of $\hat{F}^0_\varepsilon(\cdot)$, the continuous differentiability of L and the fact that $\sup_e |ef^0_\varepsilon(e)| < \infty$, it is easy to check that $A_{2i}(x) = |E^{0T}_{ix}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ such that

$$\sup_x |\sum_{i=1}^n W_i(x, a_n)A_{2i}(x)| = O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

We also have by Theorem 5.2.3

$$\sup_{x\in R_X} |\Omega_{12}(x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \ a.s.,$$

since $E[|Z|^\lambda] < \infty$. Next, we treat $\Omega_2(x)$.
First, we show that $\sup_{x\in R_X} |\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x| = O_P(1)$. Define

$$\xi_\alpha = \inf_{x\in R_X}(m^0(x) + \sigma^0(x)(F^0_\varepsilon)^{-1}(s_\alpha)) = \inf_{x\in R_X} F^{-1}_{\phi_2}(s_\alpha|x).$$

We have

$$\begin{aligned}
&P(\inf_{x\in R_X}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x) < \xi_\alpha) \qquad (6.2.2)\\
&\le P(\sup_{x\in R_X} |F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)| \ge s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)\\
&\le P(\sup_{x\in R_X} |F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) - \hat{F}_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x)| \ge (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2)\\
&\quad + P(\sup_{x\in R_X} |\hat{F}_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)| \ge (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2).
\end{aligned}$$

Using Theorem 6.2.4, the first term on the right hand side of (6.2.2) tends to zero. For the second term on the right hand side of (6.2.2), write

$$\begin{aligned}
&P(\sup_{x\in R_X} |D_{1j}(x)| \ge \varepsilon_{j\alpha})\\
&\le P(D \sup_{x\in R_X} I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x)) \ge \varepsilon_{j\alpha}/4)\\
&\quad + P(D \sup_{x\in R_X} I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x)) \ge \varepsilon_{j\alpha}/4)\\
&\quad + P(\sup_{x\in R_X} (|\hat{F}_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x)|x) - s_j|I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x))) \ge \varepsilon_{j\alpha}/4)\\
&\quad + P(\sup_{x\in R_X} (|\hat{F}_{\phi 2}(T_x|x) - F_{\phi 2}(T_x|x)|I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x))) \ge \varepsilon_{j\alpha}/4)\\
&\quad = D_2 + D_3 + D_4 + D_5,
\end{aligned}$$

where

$$D_{1j}(x) = \hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x),$$

$\varepsilon_{j\alpha} = (s_j \wedge F_{\phi 2}(T_x|x) - s_\alpha)/2$ and $D = \max_j(\sup_{x\in R_X} |D_{1j}(x)|)$. D_2 and D_3 are bounded by

$$P(F_{\phi 2}(T_x|x) < s_j \leq \hat{F}_{\phi 2}(T_x|x) \ for\ at\ least\ one\ value\ of\ x) \to 0$$

and

$$P(\hat{F}_{\phi 2}(T_x|x) < s_j < F_{\phi 2}(T_x|x) \ for\ at\ least\ one\ value\ of\ x) \to 0,$$

$j = 1, \ldots, k$, respectively. Note that $s_j \leq F_{\phi 2}(T_x|x)$ has been replaced by $s_j < F_{\phi 2}(T_x|x)$ in the expression above since T_x can always be chosen such that $s_j \neq F_{\phi 2}(T_x|x)$ due to assumption (A11) (i). So, the two expressions above are obtained using Theorem 6.2.4. D_4 is bounded by

$$P(\sup_{x\in R_X} \sup_{-\infty<y<\infty} |\hat{F}_{\phi 2}(y|x) - \hat{F}_{\phi 2}(y - |x)| \geq \varepsilon_{j\alpha}/4),$$

for which Theorem 6.2.7 is used. For D_5, Theorem 6.2.4 is also used. Since

$$\inf_{x\in R_X}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x) \geq \inf_{x\in R_X}(F_{\phi 2}^{-1}(s_\alpha|x)) = \xi_\alpha,$$

we have

$$\sup_{x\in R_X} |\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x - F_{\phi 2}^{-1}(s_j|x) \wedge T_x| = \sup_{x\in R_X} |D_{6j}(x)| = O_P(1).$$

$\Omega_2(x)$ is therefore rewritten as

$$\begin{aligned}\Omega_2(x) \quad &= \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j \leq \hat{F}_{\phi 2}(T_x|x), s_j > F_{\phi 2}(T_x|x)) \\ &\quad + \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j \leq \hat{F}_{\phi 2}(T_x|x), s_j \leq F_{\phi 2}(T_x|x)) \\ &\quad + \sum_{j=1}^{k} a_j D_{6j}(x) I(s_j > \hat{F}_{\phi 2}(T_x|x), s_j \leq F_{\phi 2}(T_x|x)), \qquad (6.2.3)\end{aligned}$$

where the suprema of the first and third terms are negligible, using the same arguments as for D_2 and D_3. Note that when $s_j > F^0_\varepsilon(T)$ for all $j,\ j = 1\ldots,k$, only the first term of (6.2.3) is considered and treated with Theorem 6.2.4. Next, $\sup_{x\in R_X} |\Omega_2(x)|$ is now bounded by

$$\begin{aligned}
&\sum_{j=1}^{k} |a_j| \sup_{x\in R_X} (|D_{6j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) \in Q)) \\
&+\sum_{j=1}^{k} |a_j| \sup_{x\in R_X} (|D_{6j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) \notin Q)) \\
&+O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \\
&= D_7 + D_8 + O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).
\end{aligned}$$

D_8 is negligible. Indeed, for $s_j \le F_{\phi 2}(T_x|x)$, we only need to prove that

$$\begin{aligned}
&P(F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) < s_\alpha,\ \textit{for at least one value of } x) \\
&+P(F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) > s_\beta \wedge F_{\phi 2}(T_x|x),\ \textit{for at least one value of } x)
\end{aligned}$$

tends to zero. The first term of this expression is treated as in (6.2.2). The second term is bounded by

$$\begin{aligned}
&P(\sup_{x\in R_X} |F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) - \hat{F}_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x)| > \varepsilon_{j\beta}/2) \\
&P(\sup_{x\in R_X} |\hat{F}_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) - s_j \wedge F_{\phi 2}(T_x|x)| > \varepsilon_{j\beta}/2),
\end{aligned}$$

where $\varepsilon_{j\beta} = s_\beta \wedge F_{\phi 2}(T_x|x) - s_j \wedge F_{\phi 2}(T_x|x) > 0$ and both terms of this expression are treated with Theorem 6.2.4 and 6.2.7 respectively. Now, we define

$$\begin{aligned}
D_{9j}(x) \ &= F^{-1}(\max(F(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x), s_\alpha) \wedge (s_\beta \wedge F^0_\varepsilon(T))|x) \\
&\quad -F^{-1}(F(F^{-1}_{\phi 2}(s_j|x) \wedge T_x|x)|x),
\end{aligned}$$

such that

$$D_7 = \sum_{j=1}^{k} |a_j| \sup_{x\in R_X} (|D_{9j}(x)| I(s_j \le \hat{F}_{\phi 2}(T_x|x), s_j \le F_{\phi 2}(T_x|x), F_{\phi 2}(\hat{F}^{-1}_{\phi 2}(s_j|x) \wedge T_x|x) \in Q)).$$

Therefore, using a Taylor development

$$D_{9j}(x) \leq \frac{1}{f(F^{-1}(\theta_{jx}|x)|x)}|F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)|,$$

where θ_{jx} is between $\max(F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x), s_\alpha) \wedge (s_\beta \wedge F_\varepsilon^0(T))$ and $F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)$. Finally, the desired order is obtained with a successive application of Theorems 6.2.4 and 6.2.7.

Theorem 6.2.2 *Assume (A1) (i)-(iii), (A2) (i)-(ii), (A3) (i)-(ii), (A3) (iii)", (A4) (i), (A6)', (A8), (A11) (i) and (A11) (ii)'. Then,*

$$\hat{m}_1^T(x) - m_1^T(x) = \frac{1}{na_n}\sum_{i=1}^n K\Big(\frac{x - X_i}{a_n}\Big)B_1(Z_i, \Delta_i|x) + R_n(x),$$

where $\sup\{|R_n(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$.

Proof. First, consider $\Omega_{11}(x)$.

$$\sum_{i=1}^n W_i(x, a_n)A_{1i}(x) \tag{6.2.4}$$
$$= \sum_{i=1}^n W_i(x, a_n)I(\Delta_i = 1)Y_i L'(F_\varepsilon^0(E_{ix}^{0T}))(\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})) + o_P((na_n)^{-1/2}),$$

using the uniform consistency of $\hat{F}_\varepsilon^0(\cdot)$ as in (2.2.2) and a second order Taylor expansion. Using similar arguments as in (4.5.2) for $\hat{E}_i^0$ and E_i^0 replaced by $\hat{E}_{ix}^{0T}$ and E_{ix}^{0T} respectively, the asymptotic representation for $\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T})$ is given by

$$(na_n)^{-1}\sum_{j=1}^n K(\frac{x - X_j}{a_n})h_{x,Z_i \wedge T_x}(Z_j, \Delta_j) + o_P((na_n)^{-1/2}), \tag{6.2.5}$$

where use is made of Propositions 4.8 and 4.9 of VKA (1999). Next, consider the expression $\int_{\hat{E}_{ix}^{0T}}^{\hat{T}^x}(\hat{m}^0(x) + \hat{\sigma}^0(x)e)L(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e)$ which appears in the term B_{1i} of (6.2.1). We have

$$\int_{\hat{E}_{ix}^{0T}}^{\hat{T}^x} \hat{m}^0(x)L(\hat{F}_\varepsilon^0(e))d\hat{F}_\varepsilon^0(e) \tag{6.2.6}$$
$$= m^0(x)\left\{\int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e) + \int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e))d(\hat{F}_\varepsilon^0(e) - F_\varepsilon^0(e))\right\}$$
$$+O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),$$

using uniform consistency of $\hat{m}^0(\cdot)$ and $\hat{F}^0_\varepsilon(\cdot)$. By using integration by parts, the second term is $|E^{0T}_{ix}|O_P(n^{-1/2})$. In the same way,

$$\begin{aligned}
&\int_{\hat{E}^{0T}_{ix}}^{\hat{T}^x} \hat{\sigma}^0(x) eL(\hat{F}^0_\varepsilon(e))d\hat{F}^0_\varepsilon(e) \qquad (6.2.7)\\
&= \sigma^0(x)\left\{\int_{E^{0T}_{ix}}^{T} eL(F^0_\varepsilon(e))dF^0_\varepsilon(e) + \int_{E^{0T}_{ix}}^{T} eL(F^0_\varepsilon(e))d(\hat{F}^0_\varepsilon(e) - F^0_\varepsilon(e))\right.\\
&\quad \left. + \int_{\hat{E}^{0T}_{ix}}^{E^{0T}_{ix}} eL(\hat{F}^0_\varepsilon(e))d\hat{F}^0_\varepsilon(e) + \int_{T}^{\hat{T}^x} eL(\hat{F}^0_\varepsilon(e))d\hat{F}^0_\varepsilon(e)\right\} + |E^{0T}_{ix}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).
\end{aligned}$$

Using the fact that $\sup_e |ef^0_\varepsilon(e)| < \infty$, it is easily shown that the second, third and fourth terms are $|E^{0T}_{ix}|O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$. From this, we conclude

$$\begin{aligned}
&\sum_{i=1}^{n} W_i(x, a_n)I(\Delta_i = 0)B_{1i} \qquad (6.2.8)\\
&= \sum_{i=1}^{n} W_i(x, a_n)I(\Delta_i = 0) \times\\
&\quad \left\{\frac{(\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}) - F^0_\varepsilon(E^{0T}_{ix}))[m^0(x)\int_{E^{0T}_{ix}}^{T} L(F^0_\varepsilon(e))dF^0_\varepsilon(e) + \sigma^0(x)\int_{E^{0T}_{ix}}^{T} eL(F^0_\varepsilon(e))dF^0_\varepsilon(e)]}{(1 - F^0_\varepsilon(E^{0T}_{ix}))^2}\right\}\\
&\quad + o_P((na_n)^{-1/2}),
\end{aligned}$$

where the representation (6.2.5) will be used for $\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}) - F^0_\varepsilon(E^{0T}_{ix})$. Next, consider the expression B_{2i}. Easy calculations show that

$$B_{2i} = \frac{\int_{\hat{E}^{0T}_{ix}}^{E^{0T}_{ix}} (m^0(x) + \sigma^0(x)e)L(F^0_\varepsilon(e))d\hat{F}^0_\varepsilon(e)}{1 - F^0_\varepsilon(E^{0T}_{ix})} + |E^{0T}_{ix}|o_P((na_n)^{-1/2}).$$

We have

$$\begin{aligned}
&\int_{\hat{E}^{0T}_{ix}}^{E^{0T}_{ix}} (m^0(x) + \sigma^0(x)e)L(F^0_\varepsilon(e))d\hat{F}^0_\varepsilon(e)\\
&= \int_{\hat{E}^{0T}_{ix}}^{E^{0T}_{ix}} (m^0(x) + \sigma^0(x)e)L(F^0_\varepsilon(e))dF^0_\varepsilon(e)\\
&\quad + \int_{\hat{E}^{0T}_{ix}}^{E^{0T}_{ix}} (m^0(x) + \sigma^0(x)e)L(F^0_\varepsilon(e))d(\hat{F}^0_\varepsilon(e) - F^0_\varepsilon(e)).
\end{aligned}$$

The second term on the right hand side of the equation above is $|E_{ix}^{0T}|O_P(n^{-1/2})$, which follows, using integration by parts, from Corollary 3.2 and proposition 4.5 of VKA (1999) and the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$. Hence,

$$\begin{aligned} B_{2i} &= \frac{\int_0^{E_{ix}^{0T}} (m^0(x) + \sigma^0(x)e)L(F_\varepsilon^0(e))dF_\varepsilon^0(e) - \int_0^{\hat{E}_{ix}^{0T}} (m^0(x) + \sigma^0(x)e)L(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})} \\ &\quad + |E_{ix}^{0T}|o_P((na_n)^{-1/2}) \\ &= \frac{[\hat{m}^0(x) - m^0(x) + E_{ix}^{0T}(\hat{\sigma}^0(x) - \sigma^0(x))]}{\sigma^0(x)(1 - F_\varepsilon^0(E_{ix}^{0T}))}(m^0(x) + \sigma^0(x)E_{ix}^{0T})L(F_\varepsilon^0(E_{ix}^{0T}))f_\varepsilon^0(E_{ix}^{0T}) \\ &\quad + |E_{ix}^{0T}|o_P((na_n)^{-1/2}), \end{aligned}$$

using a Taylor expansion. Note that the term $|E_{ix}^{0T}|o_P((na_n)^{-1/2})$ in the expression above is obtained from the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$ and $\sup_e |e^2 f_\varepsilon^{0'}(e)| < \infty$. A similar expression for B_{3i} is obtained such that

$$\begin{aligned} &\sum_{i=1}^n W_i(x, a_n)I(\Delta_i = 0)(B_{2i} + B_{3i}) \qquad (6.2.9) \\ &= \sum_{i=1}^n W_i(x, a_n)I(\Delta_i = 0) \times \\ &\quad \left\{ \frac{[\hat{m}^0(x) - m^0(x) + E_{ix}^{0T}(\hat{\sigma}^0(x) - \sigma^0(x))]}{\sigma^0(x)(1 - F_\varepsilon^0(E_{ix}^{0T}))}(m^0(x) + \sigma^0(x)E_{ix}^{0T})L(F_\varepsilon^0(E_{ix}^{0T}))f_\varepsilon^0(E_{ix}^{0T}) \right. \\ &\quad \left. + \frac{[m^0(x) - \hat{m}^0(x) + T(\sigma^0(x) - \hat{\sigma}^0(x))]}{\sigma^0(x)(1 - F_\varepsilon^0(E_{ix}^{0T}))}(m^0(x) + \sigma^0(x)T)L(F_\varepsilon^0(T))f_\varepsilon^0(T) \right\} \\ &\quad + o_P((na_n)^{-1/2}). \end{aligned}$$

B_{4i} and B_{6i} are $|E_{ix}^{0T}|o_P((na_n)^{-1/2})$. For B_{5i}, $\hat{F}_\varepsilon^0(e)$ is replaced by $F_\varepsilon^0(e)$ and the remaining terms are $|E_{ix}^{0T}|o_P((na_n)^{-1/2})$ using integration by parts, the uniform consistency of $\hat{m}^0(\cdot)$, $\hat{\sigma}^0(\cdot)$ and $\hat{F}_\varepsilon^0(\cdot)$ and the fact that $\sup_e |ef_\varepsilon^0(e)| < \infty$. Then, use is made of the asymptotic representations of Propositions 4.8 and 4.9 of VKA (1999) such that B_{5i} is given by

$$\begin{aligned} &\frac{-(na_n)^{-1}f_X^{-1}(x)\sigma(x)}{1 - F_\varepsilon^0(E_{ix}^{0T})}\{\int_{E_{ix}^{0T}}^T L(F_\varepsilon^0(e))dF_\varepsilon^0(e)\sum_{j=1}^n K(\frac{x - X_j}{a_n})\eta(Z_j, \Delta_j|x) \\ &+ \int_{E_{ix}^{0T}}^T eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)\sum_{j=1}^n K(\frac{x - X_j}{a_n})\zeta(Z_j, \Delta_j|x)\} \\ &+ |E_{ix}^{0T}|o_P((na_n)^{-1/2}). \qquad (6.2.10) \end{aligned}$$

Finally, B_{7i} is $|E_{ix}^{0T}|O_P(n^{-1/2})$ using integration by parts.
From those developments, we can write

$$\begin{aligned}\Omega_{11}(x) \quad &= a_0 \sum_{i=1}^{n} W_i(x, a_n)\tilde{B}_1(Z_i, \Delta_i|x) \times \frac{1}{na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n})\eta(Z_j, \Delta_j|x) \\ &+a_0 \sum_{i=1}^{n} W_i(x, a_n)\tilde{B}_2(Z_i, \Delta_i|x) \times \frac{1}{na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n})\zeta(Z_j, \Delta_j|x) \\ &+o_P((na_n)^{-1/2}), \end{aligned} \tag{6.2.11}$$

where

$$\begin{aligned}\tilde{B}_1(Z_i, \Delta_i|x) = f_X^{-1}(x) \ \Big\{ &I(\Delta_i = 1)Z_i L'(F(Z_{ix}^T|x))f_\varepsilon^0(E_{ix}^{0T}) \\ &+I(\Delta_i = 0)\Bigg[f_\varepsilon^0(E_{ix}^{0T})(\frac{\int_{Z_{ix}^T}^{T_x} yL(F(y|x))dF(y|x)}{(1 - F(Z_{ix}^T|x))^2} \\ &-\frac{Z_{ix}^T L(F(Z_{ix}^T|x))}{1 - F(Z_{ix}^T|x)}) + f_\varepsilon^0(T)\frac{T_x L(F(T_x|x))}{1 - F(Z_{ix}^T|x)} \\ &-\sigma^0(x)\frac{\int_{E_{ix}^{0T}}^{T} L(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})}\Bigg]\Big\}, \end{aligned}$$

$$\begin{aligned}\tilde{B}_2(Z_i, \Delta_i|x) = f_X^{-1}(x) \ \Big\{ &I(\Delta_i = 1)Z_i L'(F(Z_{ix}^T|x))f_\varepsilon^0(E_{ix}^{0T})E_{ix}^{0T} \\ &+I(\Delta_i = 0)\Bigg[f_\varepsilon^0(E_{ix}^{0T})E_{ix}^{0T}(\frac{\int_{Z_{ix}^T}^{T_x} yL(F(y|x))dF(y|x)}{(1 - F(Z_{ix}^T|x))^2} \\ &-\frac{Z_{ix}^T L(F(Z_{ix}^T|x))}{1 - F(Z_{ix}^T|x)}) + f_\varepsilon^0(T)T\frac{T_x L(F(T_x|x))}{1 - F(Z_{ix}^T|x)} \\ &-\sigma^0(x)\frac{\int_{E_{ix}^{0T}}^{T} eL(F_\varepsilon^0(e))dF_\varepsilon^0(e)}{1 - F_\varepsilon^0(E_{ix}^{0T})}\Bigg]\Big\}, \end{aligned}$$

and $Z_{ix}^T = Z_i \wedge T_x$, $i = 1, \ldots, n$. Using Theorem 5.2.3 for the new data points $\tilde{B}_1(Z, \Delta|x)$, $\tilde{B}_2(Z, \Delta|x)$, $\eta(Z, \Delta|x)$ and $\zeta(Z, \Delta|x)$, the asymptotic representation of $\Omega_{11}(x)$ is

$$\frac{1}{na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n})\tilde{B}_3(Z_j, \Delta_j|x) + R_{n1}(x), \tag{6.2.12}$$

where

$$\tilde{B}_3(Z,\Delta|x) = a_0(E[\tilde{B}_1(Z,\Delta|x)|x]\eta(Z,\Delta|x) + E[\tilde{B}_2(Z,\Delta|x)|x]\zeta(Z,\Delta|x)),$$

and $\sup\{|R_{n1}(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$. Note that this rate can be obtained since $E[\eta(Z,\Delta|x)|x] = E[\zeta(Z,\Delta|x)|x] = 0$. For $\Omega_{12}(x)$, we readily obtain, using Theorem 5.2.3 with new data points equal to 1 and $\tilde{\phi}_1^*(Z_i,\Delta_i|x) - E[\tilde{\phi}_1^*(Z,\Delta|x)|x]$,

$$\frac{a_0}{na_n f_X(x)}\sum_{i=1}^{n} K(\frac{x-X_i}{a_n})(\tilde{\phi}_1^*(Z_i,\Delta_i|x) - E[\tilde{\phi}_1^*(Z,\Delta|x)|x]) + R_{n2}(x), \tag{6.2.13}$$

where $\sup\{|R_{n2}(x)|; x \in R_X\} = o_P((na_n)^{-1/2})$.
Next, rewrite the second term on the right hand side of (6.2.3) as

$$\sum_{j=1}^{k} a_j(D_{6j}(x) - D_{10j}(x))I(s_j \le F_{\phi 2}(T_x|x), s_j \le \hat{F}_{\phi 2}(T_x|x))$$
$$+\sum_{j=1}^{k} a_j D_{10j}(x)I(s_j \le F_{\phi 2}(T_x|x), s_j \le \hat{F}_{\phi 2}(T_x|x)) = \Omega_{21}(x) + \Omega_{22}(x),$$

where $D_{6j}(x) = \hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x - F_{\phi 2}^{-1}(s_j|x) \wedge T_x$ and

$$D_{10j}(x) = \frac{s_j \wedge F_{\phi 2}(T_x|x) - \hat{F}_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)}.$$

Using Theorem 6.2.7, $s_j \wedge F_{\phi 2}(T_x|x)$ can be replaced by $\hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)$ in $D_{10j}(x)$ of $\Omega_{21}(x)$. We next rewrite $\Omega_{21}(x)$ as

$$\sum_{j=1}^{k} \frac{a_j(D_{11j}(x) + D_{12j}(x))}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)} I(s_j \le F_{\phi 2}(T_x|x), s_j \le \hat{F}_{\phi 2}(T_x|x)), \tag{6.2.14}$$

where

$$D_{11j}(x) = f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)D_{6j}(x) - (F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)),$$

and

$$\begin{aligned} D_{12j}(x) \ &= F_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) \\ &\quad -\hat{F}_{\phi 2}(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) + \hat{F}_{\phi 2}(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x). \end{aligned}$$

Using a second order Taylor expansion, we get

$$
\begin{aligned}
&-[F(\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - F(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) - f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) D_{6j}(x)] \\
&= -\frac{f'(\theta_{jx}|x)}{2} D_{6j}(x)^2,
\end{aligned}
$$

where θ_{jx} is between $\hat{F}_{\phi 2}^{-1}(s_j|x) \wedge T_x$ and $F_{\phi 2}^{-1}(s_j|x) \wedge T_x$. Thus, using the proof of Theorem 6.2.1, the first term of (6.2.14) is $O_P((na_n)^{-1} \log a_n^{-1})$ since $\sup_{x,y} |f'(y|x)| < \infty$ and $\inf_x f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x) > 0$. Next, we treat the second term of (6.2.14). First, define $D_{13j}(x)$ as

$$
\frac{D_{12j}(x)}{f(F_{\phi 2}^{-1}(s_j|x) \wedge T_x|x)} I(s_j \leq F_{\phi 2}(T_x|x), s_j \leq \hat{F}_{\phi 2}(T_x|x)).
$$

The second term of (6.2.14) can then be rewritten as

$$
\begin{aligned}
&\sum_{j=1}^{k} a_j D_{13j}(x) I(|D_{6j}(x)| > d_n) \\
&+ \sum_{j=1}^{k} a_j D_{13j}(x) I(|D_{6j}(x)| \leq d_n),
\end{aligned}
$$

where $d_n \sim (na_n)^{-1/2} (\log a_n^{-1})^{1/2}$. The first term of this expression is negligible using Theorem 6.2.1 and the second one is $o_P((na_n)^{-1/2})$ using Theorem 6.2.7. Finally, $\Omega_{22}(x)$ can be written as

$$
\sum_{j=1}^{k} a_j D_{10j}(x) I(s_j \leq F_{\phi 2}(T_x|x)) + o_P((na_n)^{-1/2}),
$$

where use is made of Theorems 6.2.4 and 6.2.5.

Theorem 6.2.3 *Under the assumptions of Theorem 6.2.2,*

$$
(na_n)^{1/2} (\hat{m}_1^T(x) - m_1^T(x)) \xrightarrow{d} N(0, s^2(x)),
$$

where

$$
s^2(x) = f_X(x) \int K^2(u)\, du \sum_{\delta=0,1} \int B_1^2(z, \delta|x)\, dH_\delta(z|x).
$$

Proof. The result is obtained by using Lyapounov's Theorem. It's easy to check that the Lyapounov ratio is $O((na_n)^{-1/2})$ since $E[|Z|^\lambda] < \infty$.

6.2.2 Distribution results

Theorem 6.2.4 *Assume (A1) (i)-(iii), (A2) (i)-(ii), (A3) (i)-(ii), (A4) (i), (A6)' and (A8). Then,*

$$\sup_{x\in R_X}\sup_{-\infty<t<\infty}|\hat{F}_{\phi_2}(t|x)-F_{\phi_2}(t|x)|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).$$

Proof. Write

$$\begin{aligned}\hat{F}_{\phi 2}(t|x)-F_{\phi 2}(t|x) &=\sum_{i=1}^n W_i(x,a_n)\{I(\Delta_i=0)[\frac{\int_{\hat{E}_{ix}^{0T_t}}^{\hat{T}_t^x}d\hat{F}_\varepsilon^0(e)}{1-\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T})}-\frac{\int_{E_{ix}^{0T_t}}^{T_t^x}dF_\varepsilon^0(e)}{1-F_\varepsilon^0(E_{ix}^{0T})}]\}\\ &\quad+\sum_{i=1}^n W_i(x,a_n)\{\tilde{\phi}_{2t}^*(Z_i,\Delta_i|x)-E[\tilde{\phi}_{2t}^*(Z,\Delta|x)|x]\}\\ &=\sum_{i=1}^n W_i(x,a_n)\{I(\Delta_i=0)\Omega_{3t}(Z_i,\Delta_i|x)+\Omega_{4t}(Z_i,\Delta_i|x)\}.\end{aligned}$$

First, we treat $\Omega_{3t}(Z_i,\Delta_i|x)$. We have

$$\begin{aligned}\Omega_{3t}(Z_i,\Delta_i|x) &=(\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T})-F_\varepsilon^0(E_{ix}^{0T}))\frac{\int_{\hat{E}_{ix}^{0T_t}}^{\hat{T}_t^x}d\hat{F}_\varepsilon^0(e)}{(1-\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}))(1-F_\varepsilon^0(E_{ix}^{0T}))}\\ &\quad+\frac{\hat{F}_\varepsilon^0(\hat{T}_t^x)-F_\varepsilon^0(T_t^x)+F_\varepsilon^0(E_{ix}^{0T_t})-\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T_t})}{1-F_\varepsilon^0(E_{ix}^{0T})}\\ &=\Omega_{31t}(Z_i,\Delta_i|x)+\Omega_{32t}(Z_i,\Delta_i|x).\end{aligned}$$

Using the fact that

$$\sup_{x,t,z}|\hat{F}_\varepsilon^0(\frac{z\wedge T_x\wedge t-\hat{m}^0(x)}{\hat{\sigma}^0(x)})-F_\varepsilon^0(\frac{z\wedge T_x\wedge t-m^0(x)}{\sigma^0(x)})|=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}),$$

we have

$$\begin{aligned}&\sup_{x,t}|\sum_{i=1}^n W_i(x,a_n)\{I(\Delta_i=0)(\Omega_{31t}(Z_i,\Delta_i|x)+\Omega_{32t}(Z_i,\Delta_i|x))\}|\\ &=O_P((na_n)^{-1/2}(\log a_n^{-1})^{1/2}).\end{aligned}$$

For $\Omega_{4t}(Z_i,\Delta_i|x)$, we use Theorem 5.2.3 with new data points $\tilde{\phi}_{2t}^*(Z_i,\Delta_i|x)$ and we obtain the result.

Theorem 6.2.5 *Assume (A1) (i)-(iii), (A2) (i)-(ii), (A3) (i)-(ii), (A4) (i), (A6)' and (A8). Then, for any* $x \in R_X$,

$$\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x) = \frac{1}{na_n}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{a_n}\Big) A(t, Z_i, \Delta_i|x) + R_n(t|x),$$

where $\sup\{|R_n(t|x)|; x \in R_X\} = o_P((na_n)^{-1/2})$.

Proof. An asymptotic representation for the numerator of $\Omega_{32t}(Z_i, \Delta_i|x)$ is given by

$$\begin{aligned}\frac{1}{f_X(x)na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n}) \ \Big\{&[f_\varepsilon^0(T_t^x) + f_\varepsilon^0(E_{ix}^{0T_t})]\eta(Z_j, \Delta_j|x) \\ &[f_\varepsilon^0(T_t^x)T_t^x + f_\varepsilon^0(E_{ix}^{0T_t})E_{ix}^{0T_t}]\zeta(Z_j, \delta_j|x)\Big\} \\ &+o_P((na_n)^{-1/2}). \end{aligned} \tag{6.2.15}$$

For $\Omega_{31t}(Z_i, \Delta_i|x)$, it is straightforward that

$$\Omega_{31t}(Z_i, \Delta_i|x) = (\hat{F}_\varepsilon^0(\hat{E}_{ix}^{0T}) - F_\varepsilon^0(E_{ix}^{0T}))\frac{\int_{E_{ix}^{0T_t}}^{T_t^x} dF_\varepsilon^0(e)}{(1 - F_\varepsilon^0(E_{ix}^{0T}))^2} + o_P((na_n)^{-1/2}). \tag{6.2.16}$$

Therefore, with (6.2.15) and (6.2.16), we obtain

$$\begin{aligned}&\sum_{i=1}^{n} W_i(x, a_n)I(\Delta_i = 0)\Omega_{3t}(Z_i, \Delta_i|x) \\ &= \sum_{i=1}^{n} W_i(x, a_n)\tilde{B}_{4t}(Z_i, \Delta_i|x) \times \frac{1}{na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n})\eta(Z_j, \Delta_j|x) \\ &\quad + \sum_{i=1}^{n} W_i(x, a_n)\tilde{B}_{5t}(Z_i, \Delta_i|x) \times \frac{1}{na_n}\sum_{j=1}^{n} K(\frac{x - X_j}{a_n})\zeta(Z_j, \Delta_j|x) \\ &\quad + o_P((na_n)^{-1/2}), \end{aligned} \tag{6.2.17}$$

where

$$\begin{aligned}\tilde{B}_{4t} = f_X(x)^{-1}I(\Delta_i = 0) \ \Big\{&\frac{f_\varepsilon^0(T_t^x) + f_\varepsilon^0(E_{ix}^{0T_t})}{1 - F_\varepsilon^0(E_{ix}^{0T})} \\ &+ \frac{F_\varepsilon^0(T_t^x) - F_\varepsilon^0(E_{ix}^{0T_t})}{(1 - F_\varepsilon^0(E_{ix}^{0T}))^2} f_\varepsilon^0(E_{ix}^{0T})\Big\}, \end{aligned}$$

and

$$\tilde{B}_{5t} = f_X(x)^{-1} I(\Delta_i = 0) \left\{ \frac{f_\varepsilon^0(T_t^x) T_t^x + f_\varepsilon^0(E_{ix}^{0T_t}) E_{ix}^{0T_t}}{1 - F_\varepsilon^0(E_{ix}^{0T})} + \frac{F_\varepsilon^0(T_t^x) - F_\varepsilon^0(E_{ix}^{0T_t})}{(1 - F_\varepsilon^0(E_{ix}^{0T}))^2} E_{ix}^{0T} f_\varepsilon^0(E_{ix}^{0T}) \right\}.$$

Using Theorem 5.2.3 for the new data points $\tilde{B}_{4t}(Z, \Delta|x)$, $\tilde{B}_{5t}(Z, \Delta|x)$, $\eta(Z, \Delta|x)$ and $\zeta(Z, \Delta|x)$, the asymptotic representation for (6.2.17) is

$$\frac{1}{na_n} \sum_{j=1}^{n} K(\frac{x - X_j}{a_n}) \tilde{B}_{6t}(Z_j, \Delta_j|x) + R_{n1}(t|x), \tag{6.2.18}$$

where

$$\tilde{B}_{6t}(Z, \Delta|x) = E[\tilde{B}_{4t}(Z, \Delta|x)|x]\eta(Z, \Delta|x) + E[\tilde{B}_{5t}(Z, \Delta|x)|x]\zeta(Z, \Delta|x), \tag{6.2.19}$$

and $\sup\{|R_{n1}(t|x)|; x \in R_X\} = o_P((na_n)^{-1/2})$. For $\Omega_{4t}(Z_i, \Delta_i|x)$, we use Theorem 5.2.3 with new data points equal to 1 and $\tilde{\phi}_{2t}^*(Z_i, \Delta_i|x) - E[\tilde{\phi}_{2t}^*(Z, \Delta|x)|x]$ such that we obtain

$$\begin{aligned} &\sum_{i=1}^{n} W_i(x, a_n) \Omega_{4t}(Z_i, \Delta_i|x) \\ &= \frac{1}{na_n f_X(x)} \sum_{i=1}^{n} K(\frac{x - X_i}{a_n})(\tilde{\phi}_{2t}^*(Z_i, \Delta_i|x) - E[\tilde{\phi}_{2t}^*(Z, \Delta|x)|x]) + R_{n2}(t|x), \end{aligned} \tag{6.2.20}$$

where $\sup\{|R_{n2}(t|x)|; -\infty < t < \infty, x \in R_X\} = o_P((na_n)^{-1/2})$.

Theorem 6.2.6 *Under the assumptions of Theorem 6.2.5,*

$$(na_n)^{1/2}(\hat{F}_{\phi_2}(t|x) - F_{\phi_2}(t|x)) \xrightarrow{d} N(0, s^2(t|x)),$$

where

$$s^2(t|x) = f_X(x) \int K^2(u)\, du \sum_{\delta=0,1} \int A^2(t, z, \delta|x)\, dH_\delta(z|x).$$

Proof. The result is obtained by using Lyapounov's Theorem.

Theorem 6.2.7 *Assume (A1) (i)-(iii), (A2) (i)-(ii), (A3) (i)-(ii), (A4) (i), (A6)' and (A8). Then,*

$$\sup_{x\in R_X,\ |t-s|\leq d_n} |\hat{F}_{\phi 2}(t|x) - F_{\phi 2}(t|x) - \hat{F}_{\phi 2}(s|x) + F_{\phi 2}(s|x)| = o_P((na_n)^{-1/2}),$$

where $d_n \sim (na_n)^{-1/2}(\log a_n^{-1})^{1/2}$.

Proof. First, the expression in the theorem is written in terms of new data points :

$$\begin{aligned}
&\sup_{x\in R_X,\ |t-s|\leq d_n} |\sum_{i=1}^{n} W_i(x,a_n) \Big\{\hat{\phi}^*_{2t}(Z_i,\Delta_i|x) - \hat{\phi}^*_{2s}(Z_i,\Delta_i|x) \\
&\qquad\qquad -\tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) + \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x)\Big\}| \\
&\sup_{x\in R_X,\ |t-s|\leq d_n} |\sum_{i=1}^{n} W_i(x,a_n) \Big\{\tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) - \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x) \\
&\qquad\qquad -E[\tilde{\phi}^*_{2t}(Z,\Delta|x)|x] + E[\tilde{\phi}^*_{2s}(Z,\Delta|x)|x]\Big\}| \\
&= D_1 + D_2.
\end{aligned}$$

$D_2 = o_P((nan)^{-1/2})$ using Theorem 5.3.3. Using classical arguments, we then obtain

$$\begin{aligned}
&\hat{\phi}^*_{2t}(Z_i,\Delta_i|x) - \hat{\phi}^*_{2s}(Z_i,\Delta_i|x) - \tilde{\phi}^*_{2t}(Z_i,\Delta_i|x) + \tilde{\phi}^*_{2s}(Z_i,\Delta_i|x) \\
&= I(\Delta_i=0)\Big\{(\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}) - F^0_\varepsilon(E^{0T}_{ix})) \\
&\times \frac{F^0_\varepsilon(T^x_t) - F^0_\varepsilon(T^x_s) - F^0_\varepsilon(E^{0T_t}_{ix}) + F^0_\varepsilon(E^{0T_s}_{ix})}{(1-\hat{F}^0_\varepsilon(\hat{E}^{0T}_{ix}))(1-F^0_\varepsilon(E^{0T}_{ix}))} \\
&+ \frac{\hat{F}^0_\varepsilon(\hat{T}^x_t) - F^0_\varepsilon(T^x_t) - \hat{F}^0_\varepsilon(\hat{T}^x_s) + F^0_\varepsilon(T^x_s)}{1-F^0_\varepsilon(E^{0T}_{ix})} \\
&+ \frac{F^0_\varepsilon(E^{0T_t}_{ix}) - \hat{F}^0_\varepsilon(\hat{E}^{0T_t}_{ix}) + \hat{F}^0_\varepsilon(\hat{E}^{0T_s}_{ix}) - F^0_\varepsilon(E^{0T_s}_{ix})}{1-F^0_\varepsilon(E^{0T}_{ix})}\Big\} + O_P((na_n)^{-1}\log a_n^{-1}) \\
&= D_{21} + D_{22} + D_{23} + O_P((na_n)^{-1}\log a_n^{-1}).
\end{aligned}$$

$\sup_{x\in R_X,\ |t-s|\leq d_n} |D_{21}| = O_P((na_n)^{-1}\log a_n^{-1})$ using two Taylor developments and the fact that $\sup_e |f^0_\varepsilon(e)| < \infty$. Easy calculations show that

$$D_{22} = \frac{I(\Delta_i=0)}{1-F^0_\varepsilon(E^{0T}_{ix})}\left\{\frac{\hat{m}^0(x)-m^0(x)}{\hat{\sigma}^0(x)}(f^0_\varepsilon(T^x_s) - f^0_\varepsilon(T^x_t))\right.$$

$$+\frac{\hat{\sigma}^0(x)-\sigma^0(x)}{\hat{\sigma}^0(x)}((T_s^x-T_t^x)f_\varepsilon^0(T_s^x)$$
$$-T_t^x(f_\varepsilon^0(T_t^x)-f_\varepsilon^0(T_s^x)))\Big\}+o_P((na_n)^{-1/2}),$$

such that $\sup_{x\in R_X,\ |t-s|\leq d_n}|D_{22}|=o_P((na_n)^{-1/2})$. D_{23} is treated in a similar way and this finishes the proof.

6.3 Simulations

In this section, we compare the finite sample behaviour of the estimators $\hat{m}_1^T(x)$, $\tilde{m}^T(x)$ and $\hat{m}^T(x)$. We are interested in the behaviour of the integrated mean squared error, defined by $IMSE=\int E[(\hat{m}(x)-m(x))^2]\,dx$ for any estimator of $m(x)$. The simulations are carried out for samples of size $n=100$ and the results are obtained by using 250 simulations. We compare the three methods for the same four locations as in Section 4.3 : the conditional mean, the conditional truncated mean ($L(s)=(1/0.9)I(0.05<s\leq 0.95)$), the conditional median and conditional third quartile.

We use the same settings as in Section 4.3 (e.g. the same kernel, same bandwidth selection procedure, same choice of the score function J, etc.).

The first model we consider is

$$Y=\beta_0+\beta_1X+\beta_2X^2+\beta_3X^3+\sigma\varepsilon, \tag{6.3.1}$$

for various choices of $\beta_0,\beta_1,\beta_2,\beta_3$ and σ, where X has a uniform distribution on the unit interval or the interval $[0,3]$, and the error term ε is a normal random variable with zero mean and variance 1. The censoring variable C is generated as for the first setting (4.3.1) in Section 4.3 for several choices of α_0, α_1, α_2, α_3. We further assume that ε, ε^* and σ are independent of X and that ε is independent of ε^*.

Tables 6.1, 6.2 and 6.3 summarize the simulation results for different values of the parameters $\alpha_0,\alpha_1,\alpha_2,\alpha_3$, $\beta_0,\beta_1,\beta_2,\beta_3$ and σ (chosen according to the same characteristics as in Section 4.3). The proportion of censoring (in % and denoted by CP in the tables) is computed like in the previous chapters.

First, we compare $\hat{m}_1^T(x)$ with $\tilde{m}^T(x)$. The tables show that, in general, $\hat{m}_1^T(x)$ has smaller $IMSE$ than $\tilde{m}^T(x)$ for each of the four considered location functions. As for $\hat{m}^T(x)$, the higher the quantile, or the smaller the support of L, the worse the estimation but $\hat{m}_1^T(x)$ resists better than $\tilde{m}^T(x)$. Again, the locality of the Beran estimator and its inconsistency problems are highlighted. On the other hand, $\hat{m}_1^T(x)$ is a global estimator and

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
0	1	-1	1	30.1	0.089	0.090	0.098	0.110
4.1	-14	0	19.8	0.5	0.088	0.088	0.093	0.097
					0.084	0.086	0.090	0.098
0	1	0	0	35.3	0.124	0.126	0.141	0.168
3.6	-10.5	0	12	1	0.123	0.124	0.133	0.144
					0.115	0.120	0.126	0.142
0	1	-1	1	50.2	0.093	0.095	0.104	0.147
1.3	-6	4	3.2	0.5	0.091	0.092	0.098	0.110
					0.085	0.087	0.093	0.109
0	0.4	0	0	37.1	0.326	0.331	0.349	0.404
-0.4	1	-0.05	0	0.5	0.320	0.322	0.336	0.365
					0.322	0.325	0.341	0.377
0	0.4	0	0	58.8	0.390	0.396	0.454	0.569
0.24	0	0	0.02	0.5	0.390	0.388	0.408	0.507
					0.393	0.394	0.412	0.508
0	0.4	0	0	71.1	0.394	0.414	0.507	0.718
-0.3	0	0	0.05	0.5	0.384	0.390	0.445	0.586
					0.385	0.389	0.463	0.581

Table 6.1: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T(x)$ (second line) and $\hat{m}_1^T(x)$ (third line) for model (6.3.1) with large optimal bandwidth a_n. R_X is $[0,1]$ ($[0,3]$) for the three first (last) models.*

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
0	1	0	0	35.5	1.759	1.765	1.802	2.148
2	0	-0.2	0.09	0.5	1.749	1.747	1.762	1.772
					1.766	1.759	1.777	1.849
0	1	0	0	38.2	1.333	1.347	1.392	1.604
0.3	1	0	0	0.5	1.299	1.303	1.319	1.354
					1.305	1.318	1.351	1.438
0	1	0	0	58.0	1.631	1.681	1.862	1.926
0.5	0.13	0.2	0	0.5	1.517	1.525	1.547	1.676
					1.512	1.516	1.596	1.766
0	1	0	0	72.0	1.760	1.832	2.091	2.015
0	0.4	0.1	0	0.5	1.618	1.626	1.698	1.824
					1.616	1.623	1.745	1.853

Table 6.2: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T(x)$ (second line) and $\hat{m}_1^T(x)$ (third line) for model (6.3.1) with moderately large optimal bandwidth a_n. R_X is* $[0, 3]$.

its inconsistency problems are considerably less important than for the Beran estimator. Therefore, the observations of Section 4.3 concerning the shape of the function to estimate and the amount of censoring are also valid when comparing $\hat{m}_1^T(x)$ with $\tilde{m}^T(x)$: often, a more wiggly curve or an increase of the proportion of censoring affects more $\tilde{m}^T(x)$ than $\hat{m}_1^T(x)$.

We next compare the new estimator $\hat{m}_1^T(x)$ with its competitor $\hat{m}^T(x)$. As outlined in Section 1.8, the use of global information in the estimators of Chapter 4 sometimes penalizes the estimation procedure since the amount of local information is decreased by reducing the support of the score function J. These instability problems especially arise when the censoring probability curve contains some peaks in certain regions of the covariate space. Although both estimators $\hat{m}^T(x)$ and $\hat{m}_1^T(x)$ are based on $\hat{m}^0(\cdot)$ and $\hat{\sigma}^0(\cdot)$ (and can thus suffer from small supports of J), $\hat{m}_1^T(x)$ only uses them in the estimation of censored

synthetic data points and not in the construction of the estimator itself. Hence, it preserves uncensored local information and is less sensitive to these instability problems. This explains why $\hat{m}_1^T(x)$ often outperforms $\hat{m}^T(x)$ for the estimation of the mean and truncated mean. Note that the instability problem also depends on the shape of the curve to estimate and the amount of censoring. Therefore, Table 6.1 shows approximately the same results for the two new estimators since the chosen models are relatively flat. In Table 6.2 herebefore, $\hat{m}_1^T(x)$

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	σ^2	mean	trunc. mean	median	3^{rd} quartile
4	-7.5	6	-1.3	31.7	1.139	1.159	1.260	1.570
3.5	-7.45	7	-1.6	0.5	1.081	1.085	1.100	1.165
					1.059	1.067	1.125	1.276
4	-7.5	6	-1.3	38.2	1.047	1.066	1.161	1.513
4.3	-7.5	6	-1.3	0.5	1.030	1.034	1.043	1.111
					1.025	1.038	1.086	1.209
4	-7.5	6	-1.3	38.3	1.262	1.286	1.371	1.628
3.4	-7.45	7	-1.6	1	1.239	1.248	1.283	1.373
					1.231	1.248	1.313	1.450
4	-7.5	6	-1.3	51.3	1.251	1.314	1.508	1.559
3.2	-7.6	7	-1.6	0.5	1.142	1.158	1.188	1.315
					1.112	1.118	1.212	1.389
4	-7.5	6	-1.3	56.4	1.336	1.392	1.553	2.043
3	-7.6	7	-1.6	1	1.296	1.321	1.391	1.620
					1.283	1.308	1.423	1.665
4	-7.5	6	-1.3	74.7	1.493	1.590	2.412	2.119
3	-7.6	6	-1.3	1	1.512	1.576	2.005	2.176
					1.493	1.544	2.006	2.176

Table 6.3: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T(x)$ (second line) and $\hat{m}_1^T(x)$ (third line) for model (6.3.1) with small optimal bandwidth a_n. R_X is* $[0, 3]$.

outperforms $\hat{m}^T(x)$ for large proportions of censoring and more wiggly models while $\hat{m}_1^T(x)$ is the best one in Table 6.3 at all censoring levels. On the other hand, the estimator $\hat{m}^T(x)$ behaves better for quantile estimation. This is because $\hat{m}_1^T(x)$ is highly based on $\hat{F}_{\phi_2}(\cdot|x)$ which uses less homogeneous information (true data points mixed with data estimated by means of the general heteroscedastic model) than the global $\hat{F}_{\varepsilon}^0(\cdot)$ used by $\hat{m}^T(x)$. A comparative illustration of the behaviours of the two new methods is given in Figures 6.3.1 and 6.3.2. Those graphs are constructed in the same way as in Figures 3.3.1 and 3.3.2. The two new methods provide results very close to each other. Therefore, a comparative illustration between the second new method and the Beran estimator looks like Figures 4.3.1 and 4.3.2. For conditional mean and truncated mean, a small advantage is observed for the second method in Figure 6.3.1.

Next, we consider the case where model (1.7.1) is not satisfied. For this, we generate random response and censoring variables from Weibull distributions with the following parameters

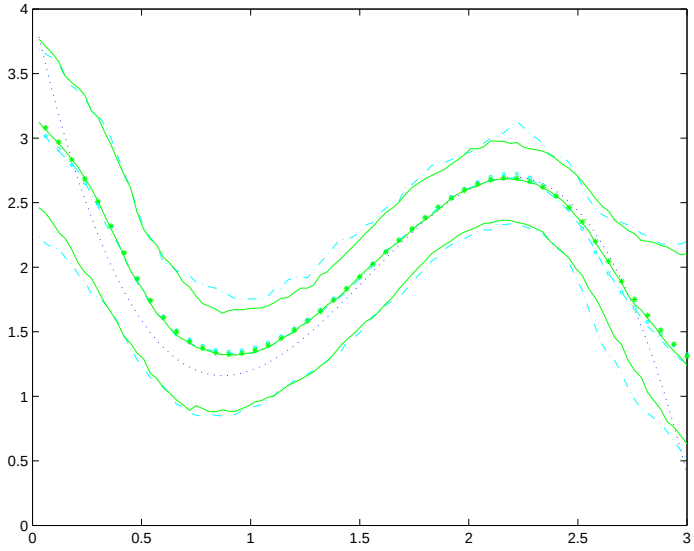
$$\begin{aligned} Y|X = x &\sim Weibull(x, d), \\ C|X = x &\sim Weibull((0.3 + x)/\xi, d), \end{aligned} \tag{6.3.2}$$

where X has a uniform distribution on the interval $[0, 3]$ and d and ξ are chosen to be positive for all $0 \leq x \leq 3$. From the conditional independence between Y and C for given X, it follows that the censoring probability curve is given by $P(\Delta = 0|X = x) = (0.3+x)/((\xi+1)x+0.3)$. Using conditional mean and standard deviation for Weibull distributions, we obtain

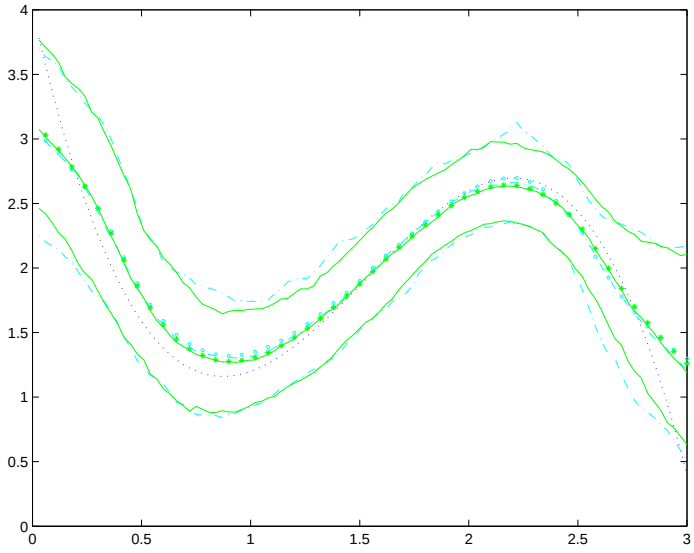
$$P(\varepsilon \leq t|x) = 1 - \exp(-\{t[\Gamma(1 + 2d^{-1}) - \Gamma^2(1 + d^{-1})]^{1/2} + \Gamma(1 + d^{-1})\}^d).$$

Therefore, choosing for instance $d = 2 + d_1 x$ enables to remove Y in (6.3.2) from model (1.7.1).

Table 6.4 shows the simulation results for model (6.3.2) with different values of d_1 and ξ. ξ is chosen such that the censoring probability curve is approximately the same for each d_1. The two new methods don't seem to be very sensitive to model assumption (1.7.1) since Beran's method obtains the largest IMSE for all values of d_1. When the value of d_1 increases, the second method seems to resist better than the first one for conditional mean or truncated mean estimation whereas the first method continues to outperform the second

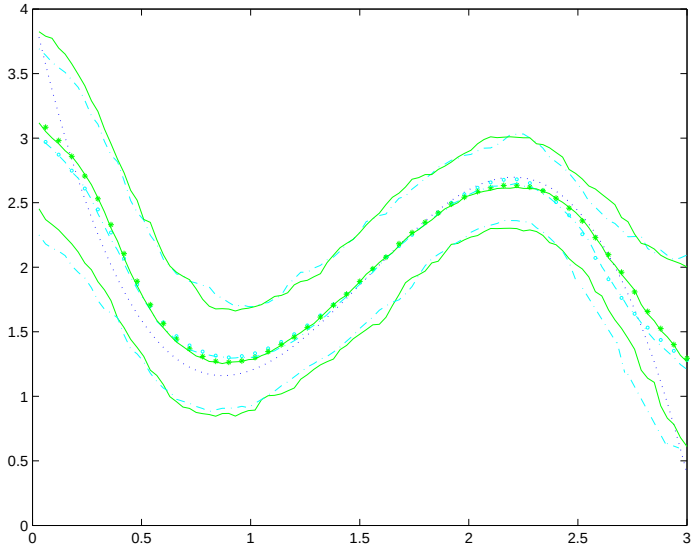


(a)

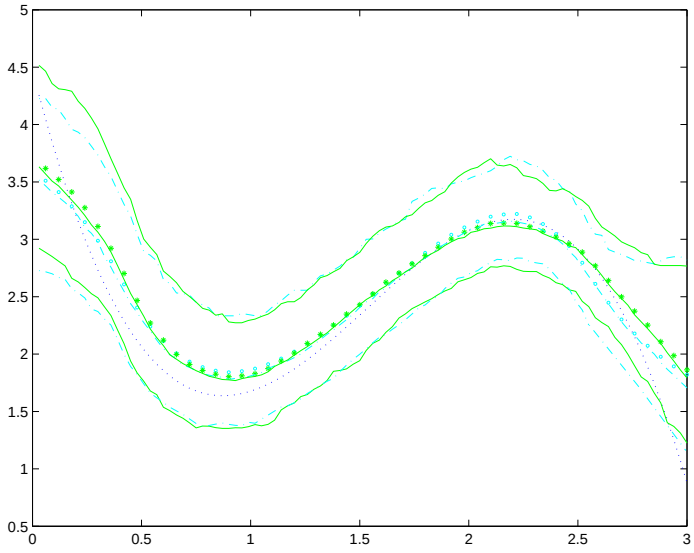


(b)

Figure 6.3.1: Comparison of pointwise biases and variabilities of the two new nonparametric methods. The dashed dotted, respectively, solid curves are obtained from the first, respectively, second new nonparametric method. Dots indicate the true curve and $\circ$, respectively, $*$ represent the mean curve for the first, respectively, second method. (a) Conditional mean for the last simulation of Table 6.2; (b) Conditional truncated mean for the last simulation of Table 6.2.



(a)



(b)

Figure 6.3.2: Comparison of pointwise biases and variabilities of the two new nonparametric methods. The dashed dotted, respectively, solid curves are obtained from the first, respectively, second new nonparametric method. Dots indicate the true curve and $\circ$, respectively, $*$ represent the mean curve for the first, respectively, second method. (a) Conditional first quartile for the last simulation of Table 6.2; (b) Conditional third quartile for the last simulation of Table 6.2.

one in quantile estimation. This can also be explained by the fact that the data set on which $\hat{F}_{\phi_2}(\cdot|x)$ is constructed becomes more and more heterogeneous as d_1 increases.

Finally, we consider in Table 6.5 the models (4.3.2) chosen in Section 4.3 to illustrate the heteroscedastic case. Even if the estimation of $\sigma^0(x)$ enters the estimation procedures of $\hat{m}^T(x)$ and $\hat{m}_1^T(x)$, the above conclusions seem to hold in the heteroscedastic case.

d_1	CP	$IMSE$			
ξ		mean	trunc. mean	median	3^{rd} quartile
0	36.58	1.488	1.672	1.507	3.882
1		1.479	0.999	1.412	3.271
		1.314	0.993	1.308	2.909
1	38.27	0.576	1.205	0.794	1.473
0.71		0.566	0.737	0.648	1.076
		0.517	0.605	0.622	1.022
2	38.03	0.362	1.069	0.515	0.769
0.55		0.348	0.754	0.388	0.602
		0.318	0.538	0.390	0.603
3	38.13	0.270	1.006	0.386	0.536
0.42		0.258	0.804	0.282	0.422
		0.241	0.541	0.288	0.448
4	38.28	0.235	0.953	0.319	0.377
0.32		0.218	0.843	0.221	0.346
		0.195	0.574	0.229	0.361

Table 6.4: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T(x)$ (second line) and $\hat{m}_1^T(x)$ (third line) for model (6.3.2)*

β_0	β_1	β_2	β_3	CP	$IMSE$			
α_0	α_1	α_2	α_3	γ^2	mean	trunc. mean	median	3^{rd} quartile
0	0.4	0	0	58.2	0.365	0.377	0.425	0.957
-0.1	0	0	0.1	0.1	0.338	0.347	0.335	0.943
					0.346	0.345	0.358	0.990
0	1	6	-4	48.9	0.621	0.631	0.638	0.950
0.5	1	-5	9	1	0.570	0.566	0.557	0.866
					0.582	0.563	0.566	0.922
0	1	6	-4	56.8	1.040	1.066	1.152	2.546
0.5	0.8	-6	8.5	5	1.032	1.032	1.069	2.161
					1.010	1.039	1.061	2.196

Table 6.5: *Results for $\tilde{m}^T(x)$ (first line), $\hat{m}^T(x)$ (second line) and $\hat{m}_1^T(x)$ (third line) for model (4.3.2). R_X is $[0,3]$ for the first model and $[0,1]$ for the two other ones.*

6.4 Data analysis

In this section, we add the new estimator $\hat{m}_1^T(x)$ to the analysis of the data set on spectral energy distributions of quasars. The choice of the bandwidth is achieved like in Section 4.4 where the bootstrap procedure of Remark 4.2.4 (adapted to $\hat{m}_1^T(x)$) is used. The selected bandwidth is approximately the same as for the two other methods. The results are given in Figures 6.4.3 to 6.4.6. The estimator $\hat{m}_1^T(x)$ also suggests to use linear functions for the four proposed locations. As expected, $\hat{m}_1^T(x)$ is less smooth than $\hat{m}^T(x)$, especially for the first quartile.

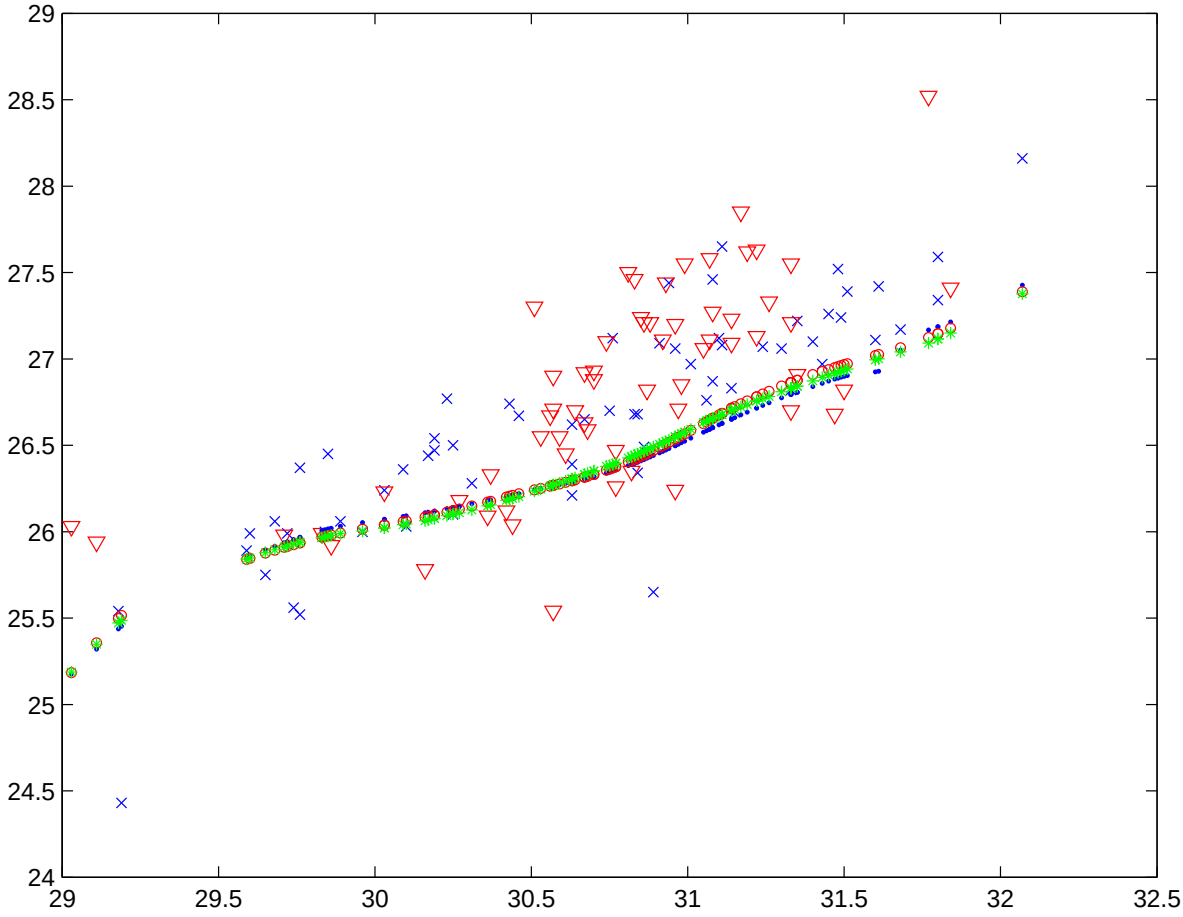


Figure 6.4.3: *Estimated conditional mean for the quasars data. The estimators* $\tilde{m}^T(x)$, $\hat{m}^T(x)$ *and* $\hat{m}_1^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

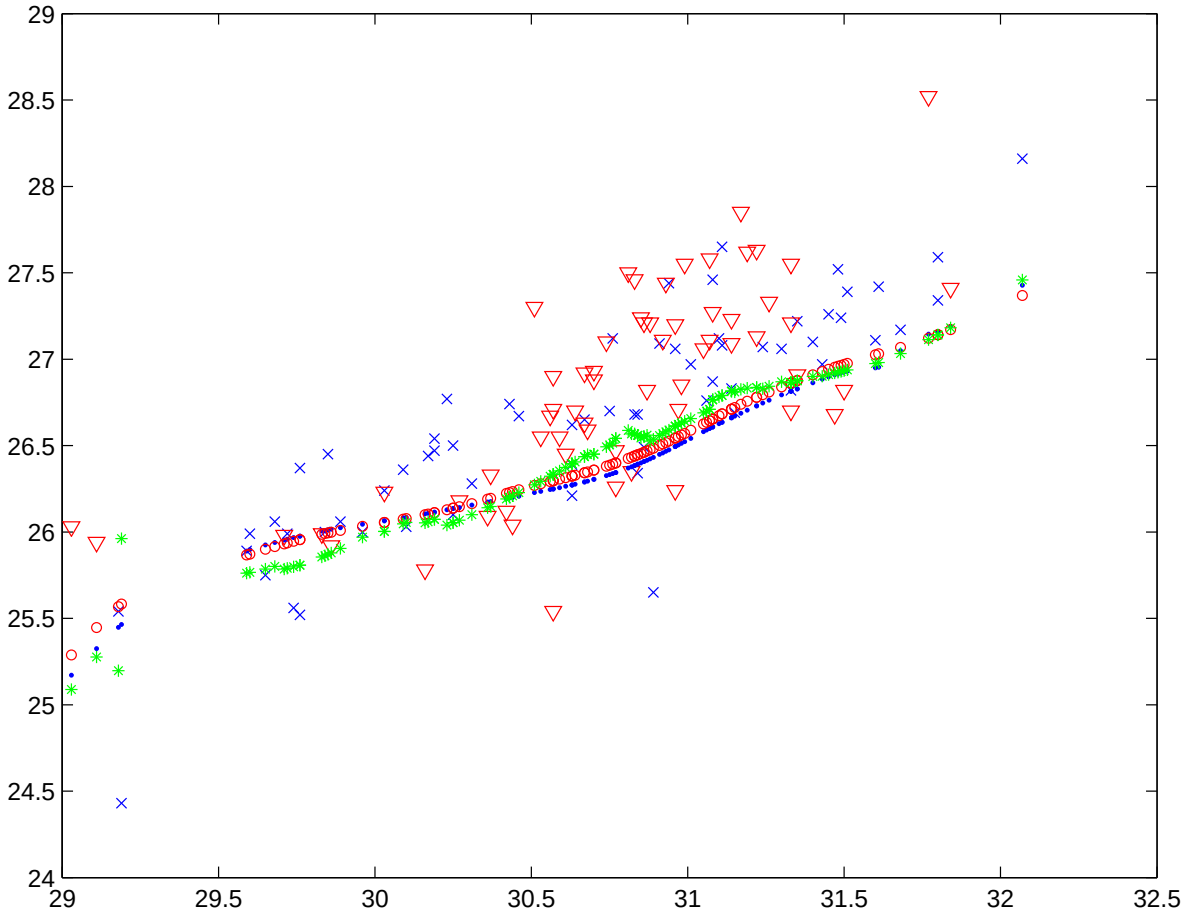


Figure 6.4.4: *Estimated conditional truncated mean for the quasars data (5 percent of truncation at both sides). The estimators* $\tilde{m}^T(x)$, $\hat{m}^T(x)$ *and* $\hat{m}_1^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

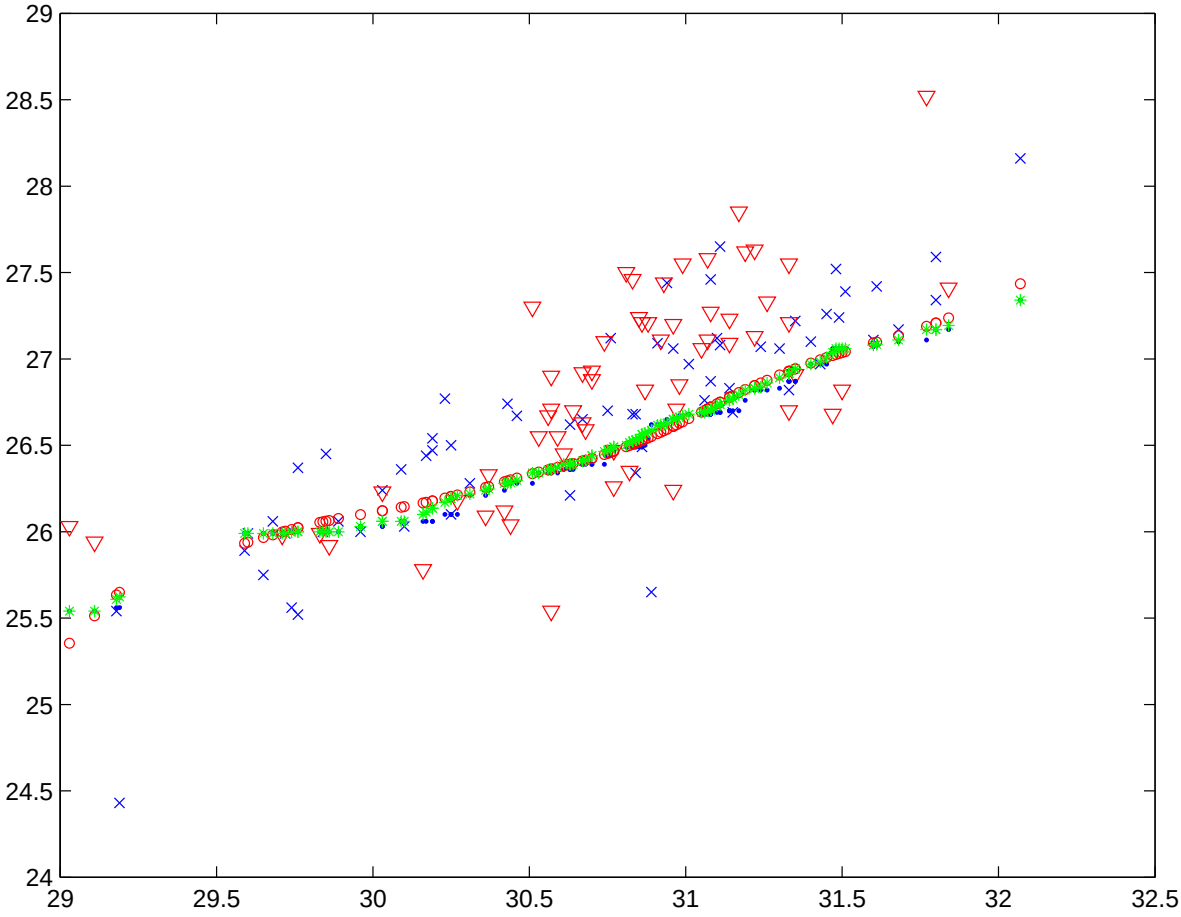


Figure 6.4.5: *Estimated conditional median for the quasars data. The estimators* $\tilde{m}^T(x)$, $\hat{m}^T(x)$ *and* $\hat{m}_1^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

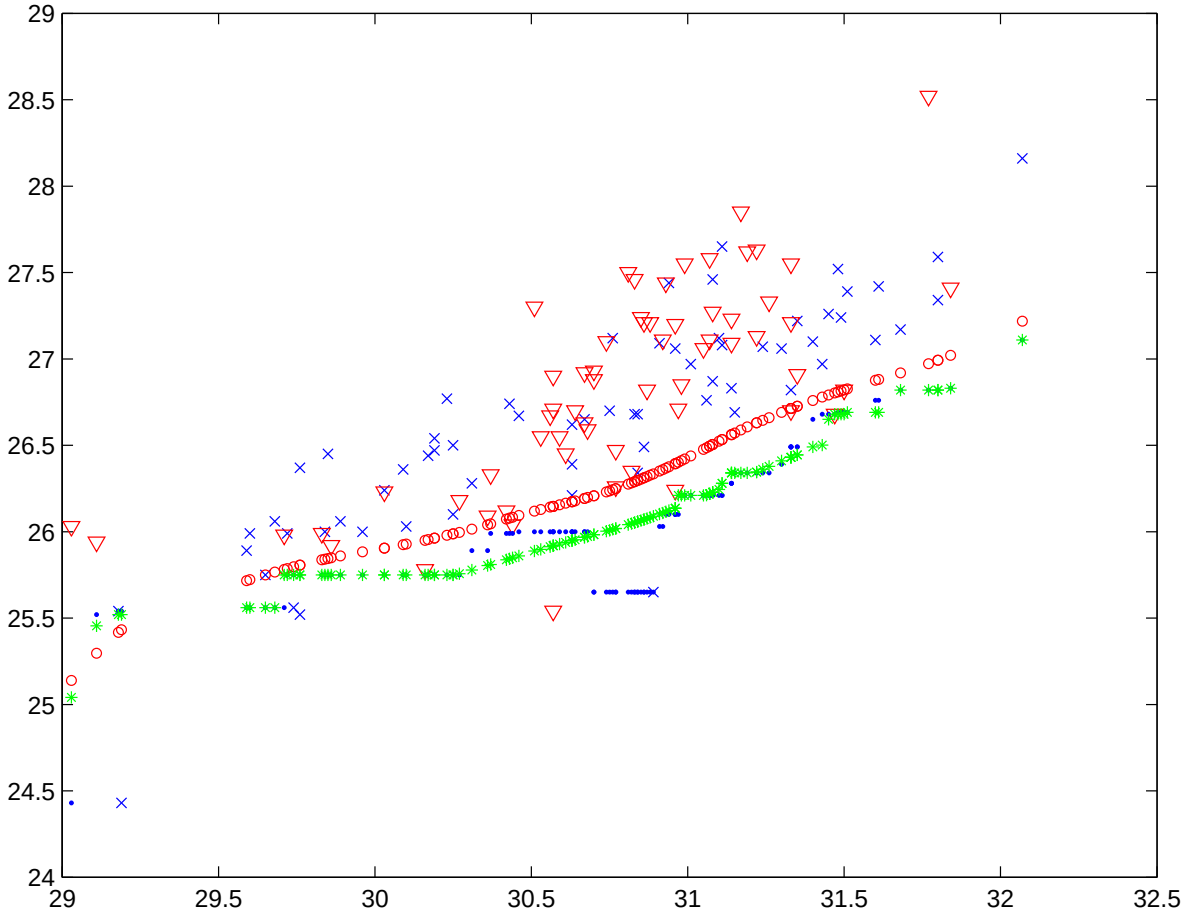


Figure 6.4.6: *Estimated conditional first quartile for the quasars data. The estimators* $\tilde{m}^T(x)$, $\hat{m}^T(x)$ *and* $\hat{m}_1^T(x)$ *are indicated by* $\cdot$, $\circ$ *and* $*$ *respectively. Uncensored data points are represented by* $\times$, *and (left) censored observations by* $\bigtriangledown$.

Further research

In the parametric case, the extension of the sum of squares to other loss functions can be considered. For instance, the parameters could be found by maximizing the likelihood function

$$L_n(\theta) = \prod_{i=1}^{n} \{\Delta_i \hat{f}_\varepsilon(\frac{Z_i - m_\theta(X_i)}{\hat{\sigma}_1^T(X_i)}) + (1-\Delta_i)(1 - \hat{F}_\varepsilon(\frac{Z_i - m_\theta(X_i)}{\hat{\sigma}_1^T(X_i)}))\},$$

where

$$\hat{f}_\varepsilon(\cdot) = \frac{1}{h_n} \int K(\frac{\cdot - z}{h_n}) d\hat{F}_\varepsilon(z),$$

$\hat{F}_\varepsilon(\cdot)$ is the Kaplan-Meier estimator based on the residuals $\frac{Z_i - m_\theta(X_i)}{\hat{\sigma}_1^T(X_i)}$ and $\hat{\sigma}_1^T(X_i)$ is a scale estimator obtained by the nonparametric technique of Chapter 6 (preserving the conditional second moment of the response). Furthermore, theory could be extended to general M-estimators (since general theory about M-estimation is already used in the proofs of Section 3.2).

Parametric conditional quantile functions could be obtained by minimizing

$$\sum_{i=1}^{n} (\hat{Y}_{Ti}^* - m_\theta(X_i))^2 \tag{6.4.1}$$

with respect to $\theta \in \Theta$, where $\hat{Y}_{Ti}^*$ is obtained by the nonparametric conditional quantile estimator $\hat{m}^T(X_i)$ proposed in Chapter 4 (in order to preserve the quantile regression function $m_\theta(X_i)$ at X_i). Note that, in the homoscedastic case, parametric conditional quantile functions can be immediately estimated from the method of Chapter 2 or 3 since those

methods provide an estimation of the conditional mean under the model $Y = E[Y|X] + \varepsilon$ with ε independent of X.

In the same way, those two techniques are also valid for the parametric estimation of a conditional location of the type $m_\theta(x) = \int yL(F(y|x))dF(y|x)$, where $\int_0^1 L(s)ds = 1$. In the expression (6.4.1), $\hat{Y}^*_{Ti}$ can therefore be replaced by $\hat{m}_1^T(X_i)$ proposed in Chapter 6 (for the location corresponding to $m_\theta(X_i)$) but also by

$$Y_i L(\hat{F}_1(Y_i \wedge T_{X_i}|X_i))\Delta_i + \frac{\int_{C_i \wedge T_{X_i}}^{T_{X_i}} yL(\hat{F}_1(y|X_i))d\hat{F}_1(y|X_i)}{1 - \hat{F}_1(C_i \wedge T_{X_i}|X_i)}(1 - \Delta_i).$$

In both cases, the location function $m_\theta(X_i)$ at X_i is preserved.

The resample scheme used all along this thesis doesn't use any model since the Y_i^* and C_i^* are taken from the Beran estimators $\tilde{F}(\cdot|X_i^*)$ and $\tilde{G}(\cdot|X_i^*)$ using the pilot bandwidth g_n. In the parametric case, the parameters $\hat{\theta}^T_{nj}$, $j = 1, \ldots, d$, obtained from the initial sample could be used to resample from $\hat{F}_\varepsilon(\frac{\cdot - m_{\hat{\theta}^T_n}(X_i^*)}{\hat{\sigma}_1^T(X_i^*)})$ instead of $\tilde{F}(\cdot|X_i^*)$, where $\hat{\sigma}_1^T(X_i^*)$ should use a pilot bandwidth. In the nonparametric case, the same idea would apply by resampling from the Kaplan-Meier estimator based on the residuals constructed with the location and scale estimators of Chapter 4 or 6 (using a pilot bandwidth). Bootstrap procedures based on those resampling schemes require lots of theoretical developments.

In order to improve the asymptotic behaviour of our estimators near the boundaries, we could use estimators of the local polynomial type. First, strong uniform consistency results of Chapter 5 could be shown with local polynomial weights instead of Nadaraya-Watson weights. Next, all the methods of this thesis could be adapted to the local polynomial case. Indeed, as shown b.e. in Kim and Truong (1998), the Beran estimator $\tilde{F}(y|x)$ can be written as

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \{1 - \hat{\Delta\Lambda}(Z_i|x)\},$$

where $\hat{\Delta\Lambda}(Z_i|x) = (\hat{\bar{H}}(Z_i - |x))^{-1}(\hat{\bar{H}}_1(Z_i - |x) - \hat{\bar{H}}_1(Z_i|x))$ and $\bar{H}(y|x) = P(Z > y|x)$ and $\bar{H}_1(y|x) = P(Z > y, \Delta = 1|x)$ can be approximated by estimators of the Stone type. Since Stone estimators are local constant estimators, they can be transformed into local polynomial estimators.

When the scale function $\sigma(\cdot)$ depends on the covariate, each square $(\hat{Y}^*_{Ti} - m_\theta(X_i))^2$ is usually weighted by the corresponding estimated conditional variance $\hat{\sigma}^2(X_i)$ such that the least squares problem becomes

$$\min_\theta \sum_{i=1}^{n} \frac{(\hat{Y}^*_{Ti} - m_\theta(X_i))^2}{\hat{\sigma}^2(X_i)}.$$

Therefore, a nonparametric estimation $\hat{\sigma}_1^T(\cdot)$ for $\sigma(\cdot)$ can be obtained from the method of Chapter 6.

Since we obtain parametric and nonparametric curves in this thesis, it is natural to develop goodness-of-fit tests by constructing Kolmogov-Smirnoff or Cramer-Von Mises statistics. Testing hypotheses is also possible by using empirical likelihood. Indeed, a likelihood ratio can be constructed to test the parameters of a model. The empirical likelihood function in the denominator could use the nonparametric residuals $(Z_i - \hat{m}_1^T(X_i))/\hat{\sigma}_1^T(X_i)$, $i = 1, \ldots, n$, whereas the numerator would use the residuals $(Z_i - m_{\hat{\theta}_n^T}(X_i))/\hat{\sigma}_1^T(X_i)$, $i = 1, \ldots, n$, under the constraint that their average is zero.

As explained in Remark 3.2.7, the methods proposed in this thesis cannot be generalized to the situation where more than one covariate are present. However, it is reasonable to think that some extensions to particular multiple regression models are possible. For instance, in a single index model of the type $Y = g(\theta_1 X_1 + \theta_2 X_2) + \varepsilon$ (ε independent of $X = (X_1 X_2)$ with zero mean, C independent of Y conditionally on X), a least squares minimization problem could be

$$\min_{\theta_1,\theta_2} \sum_{i=1}^{n} (\hat{Y}^*_{Ti}(\theta_1, \theta_2) - \hat{g}_{-i}(\theta_1 X_{1i} + \theta_2 X_{2i}))^2,$$

where $\hat{Y}^*_{Ti}(\theta_1, \theta_2)$, $i = 1, \ldots, n$, are new data points obtained by (2.1.5) with covariate X_i replaced by $\theta_1 X_{1i} + \theta_2 X_{2i}$ and $\hat{g}_{-i}(\theta_1 X_{1i} + \theta_2 X_{2i})$ is a nonparametric estimator of g which could be chosen as $\hat{m}^T_{1-i}(\theta_1 X_{1i} + \theta_2 X_{2i})$ (the second nonparametric estimator which doesn't use the data of index i). Another example of multiple regression model is the partial parametric model. For simplicity, consider the form $Y = m_\theta(X_1) + m(X_2) + \varepsilon$ (ε independent of X with zero mean, C independent of Y conditionally on X). In a first step, we could construct a backfitting algorithm based on the nonparametric procedures proposed in this thesis. And then, the vector of parameters θ should be estimated by a technique similar

to Chapter 3 replacing the new data points Y_{Ti}^* by new $Y_i - \hat{m}_2(X_{2i})$, where $\hat{m}_2(\cdot)$ is the estimator of $m(\cdot)$ obtained by the backfitting algorithm.

Bibliography

Aerts, M., Janssen, P. and Veraverbeke, N. (1994). Asymptotic theory for regression quantile estimators in the heteroscedastic regression model. In : *Asymptotic Statistics.*, (eds. P. Mandl and M. Hušková). Physica-Verlag, Heidelberg, 151–161.

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.*, **22**, 1299–1327.

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, **52**, 1342–1362.

Akritas, M. G. and La Valley, 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.

Andersen, P. K., Borgan, O., Gill, R. and Keiding, N. (1993). Statistical models based on counting processes. Springer-Verlag, New-York.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

Buchinski, M. and Hahn, J. (1998). An alternative estimator for censored quantile regression. *Econometrica*, **66**, 653–671.

Buckley, J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66**, 429–436.

Cai, Z. and Hong, Y. (2003). Weighted local linear approach to censored nonparametric regression. *Recent Advances and Trends in Nonparametric Statistics* (M. G. Akritas and D. M. Politis, eds.), 217-231.

Chen, S., Dahl, G. and Kahn, S. (2005). Nonparametric identification and estimation of a

censored location-scale regression model. *J. Amer. Statist. Assoc.*, **100**, 212–221.

Cox, D. R. (1972). Regression models and life-tables. *J. Royal Statist. Soc. Ser. B.*, **34**, 187–202.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181–197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.*, **17**, 1157–1167.

Dabrowska, D. M. (1992a). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Statist.*, **19**, 351–361.

Dabrowska, D. M. (1992b). Nonparametric quantile regression with censored data. *Sankhyā Ser. A.*, **54**, 252–259.

Doksum, K. and Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. In : *Festschrift for E. L. Lehmann* (eds. P. J. Bickel, K. A. Doksum and J. L. Hodges). Wadsworth, 140–156.

Doss, H. and Gill, D. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Amer. Statist. Assoc.*, **87**, 869–877.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312–319.

Einmahl, J. and Van Keilegom, I. (2004). Tests for independance in nonparametric regression (submitted).

Fan, J. and Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.*, **89**, 560–570.

Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete.*, **56**, 75–86.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.*, **22**, 732–746.

Fygenson, M. and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *Ann. Statist.*, **22**, 747–762.

González Manteiga, W. and Cadarso Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonparametic Statistics.*, **4**, 65–78.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157–178.

Härdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, 1428–1449.

Heuchenne, C. (2005). Strong uniform consistency of the weighted average of a conditional synthetic random variable. (submitted paper)

Heuchenne, C. and Van Keilegom, I. (2004). Polynomial regression with censored data based on preliminary nonparametric estimation (conditionally accepted by *Ann. Inst. Statist. Math.*).

Heuchenne, C. and Van Keilegom, I. (2005a). Nonlinear regression with censored data (submitted paper).

Heuchenne, C. and Van Keilegom, I. (2005b). Estimation in nonparametric location-scale regression models with censored data (submitted paper).

Heuchenne, C. and Van Keilegom, I. (2005c). Mean preservation in nonparametric regression with censored data (submitted paper).

Jennrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.*, **40**, 633–643.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients - a case study. *Statist. Neerlandica*, **37**, 103–125.

Kim, H. T. and Truong, Y. K. (1998). Nonparametric regression estimates with censored data : local linear smoothers and their applications. *Biometrics*, **54**, 1434–1444.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, **9**, 1276–1288.

Lai, T. and Ying, Z. (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.*, **19**, 1370–1402.

Lai, T. and Ying, Z. (1991b). Rank regression methods for left truncated and right-censored data. *Ann. Statist.*, **19**, 531-554.

Lai, T. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored or truncated data. *Ann. Statist.*, **22**, 1222–1255.

LeBlanc, M. and Crowley, J. (1995). Semiparametric regression functionals. *J. Amer. Statist. Assoc.*, **90**, 95–105.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301–309.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.

Li, G. and Datta, S. (2001). A bootstrap approach to non-parametric regression for right-censored data. *Ann. Inst. Statist. Math.*, **53**, 708–729.

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, **23**, 787–823.

Lin, D. Y. and Wei, L. J. (1992a). Linear regression analysis based on Buckley-James estimating equations. *Biometrics*, **48**, 679–681.

Lin, D. Y and Wei, L. J. (1992b). Linear regression analysis of multivariate failure time observations. *J. Amer. Statist. Assoc.*, **87**, 1091–1097.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Th. Rel. Fields*, **71**, 455–465.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

McKeague, I. W. and Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.*, **18**, 1172–1187.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical methods for reliability data*. Wiley, New-York.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449–464.

Müller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50**, 61–76.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.*, **9**, 141–142.

Nelson, W. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J. Testing Eval.*, **12**, 69–77.

Neumeyer, N., Dette, H. and Nagel, E.-R. (2004). Bootstrap tests for the error distribution in linear and nonparametric regression models (submitted).

Pascual, F. G. and Meeker, W. Q. (1997). Analysis of fatigue data with runouts based on

a model with nonconstant standard deviation and a fatigue limit parameter. *J. Testing Eval.*, **25**, 292–301.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.*, **98**, 1001–1012.

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics*, **32**, 143–155.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.*, **18**, 303–328.

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

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–645.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multiv. Analysis*, **45**, 89–103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *J. Multiv. Analysis*, **23**, 461–471.

Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, **9**, 1089–1102.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.*, **18**, 354–372.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

Van Keilegom, I. (2003). A note on the nonparametric estimation of the bivariate distribution under dependent censoring. *J. Nonpar. Statist.*, **16**, 659-670.

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.*, **27**, 1745–1784.

Van Keilegom, I. and Akritas, M. G. (2000). The least squares method in heteroscedastic censored regression models. In : *Asymptotics in Statistics and Probability*, Ed. M.L. Puri, VSP, 379–391.

Van Keilegom, I., Akritas, M.G. and Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data : a comparative study. *Comput. Statist. Data Anal.*, **35**, 487-500.

Van Keilegom, I. and Veraverbeke, N. (1996). Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. *Commun. Statist., - Theory Meth.*, **25**, 2251–2265.

Van Keilegom, I. and Veraverbeke, N. (1997a). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Commun. Statist., - Theory Meth.*,**26**, 853–869.

Van Keilegom, I. and Veraverbeke, N. (1997b). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467–491.

Van Keilegom, I. and Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *J. Statist. Planning Inf.*, **69**, 115-131.

Vignali, C., Brandt, W.N. and Schneider, D. P. (2003). X-ray emission from radio-quiet quasars in the SDSS early data release. The $\alpha_{\rm ox}$ dependence upon luminosity. *The Astronomical Journal*, **125**, 433–443.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A.*, **26**, 359–372.

Wei, L. J., Ying, Z. and Lin, D. (1990). Linear regression analysis of censored survival data based on linear rank tests. *Biometrika*, **77**, 845–851.

Yang, S. (1997). Extended weighted log-rank estimating functions in censored regression. *J. Amer. Statist. Assoc.*, **92**, 977–984.

Ying, Z. (1993). A large sample study of rank estimation for censored data. *Ann. Statist.*, **21**, 76–99.

Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.*, **90**, 178–184.

Zhou, M. (1992a). Asymptotic normality of the 'synthetic data' regression estimator for censored survival data. *Ann. Statist.*, **20**, 1002–1021.

Zhou, M. (1992b). M-estimation in censored linear models. *Biometrika.*, **79**, 837–841.