"Estimation in nonparametric regression under copula dependent censoring"

Sujica, Aleksandar

Abstract
In many studies in medicine, economics, demography, sociology, education, among others, one is often interested in the time until a certain event happens. This time can be the time until a patient dies or recovers from a disease (in a medical study), the time until an unemployed person finds a new job (in economics), the age at which a person marries (in demography), the time until a released prisoner gets re-arrested (in sociology), or the time taken to solve a problem (in education). The analysis of data of this kind is commonly called ‘survival analysis’ (or ‘duration analysis’ depending on the area of application). For this type of data it is common to be right censored. A typical assumption when working with randomly right censored data, is the independence between the variable of interest Y (the survival time) and the censoring variable C. This assumption, which is not testable, is however unrealistic in certain situations. In this thesis we assume that for a given cova...

Document type: Thèse (Dissertation)

Référence bibliographique
Sujica, Aleksandar. Estimation in nonparametric regression under copula dependent censoring. Prom. : Keilegom, Ingrid
Estimation in nonparametric regression under copula dependent censoring

Aleksandar Sujica

Thesis committee:
Ingrid Van Keilegom, supervisor (Université catholique de Louvain)
Johan Segers, President (Université catholique de Louvain)
Catherine Legrand (Université catholique de Louvain)
Noël Veraverbeke (Hasselt University)
Hohsuk Noh (Sookmyung Women’s University)
Acknowledgments

I dedicate this thesis to my exceptional mother.

This thesis was made possible by the persistent commitment, support, love for mathematics and patients, throughout hardship, of my mentor Ingrid Van Keilegom. Thank you Ingrid.
## Contents

1 Introduction ............................................. 7
   1.1 Introduction ......................................... 7
   1.2 Some basic concepts ................................... 9
      1.2.1 Independent censoring .............................. 10
      1.2.2 Dependent censoring .............................. 13
   1.3 Outline of the thesis ................................. 15

2 Estimation of location and scale functionals in nonparametric regression under copula dependent censoring ....... 17
   2.1 Introduction ......................................... 18
   2.2 Definitions and assumptions ........................... 21
   2.3 Asymptotic results .................................... 23
      2.3.1 Asymptotic results for the estimator of $F(y|x)$ .... 23
      2.3.2 Asymptotic results for the estimators of $m(x)$ and $\sigma^2(x)$ 25
   2.4 Simulations .......................................... 27
   2.5 Example ............................................. 33
   2.6 Appendix : Proofs .................................... 38
3 The copula-graphic estimator in censored nonparametric location-scale regression models 53

3.1 Introduction .................................................. 54
3.2 Estimation method ............................................ 56
3.3 Asymptotic results ............................................. 59
  3.3.1 Definitions and assumptions ........................... 59
  3.3.2 Asymptotic properties of the estimator \( \hat{F}_c(y) \) ...... 62
  3.3.3 Asymptotic properties of the estimator \( \hat{F}_{LS}(y|x) \) ... 63
3.4 Proofs of main results ........................................ 64
3.5 Simulations ................................................... 71
3.6 Appendix A ................................................... 78
3.7 Appendix B ................................................... 93

4 Conclusions and further research 109

Bibliography 113
Chapter 1

Introduction

1.1 Introduction

In many studies in medicine, economics, demography, sociology, education, among others, one is often interested in the time until a certain event happens. This time can be the time until a patient dies or recovers from a disease (in a medical study), the time until an unemployed person finds a new job (in economics), the age at which a person marries (in demography), the time until a released prisoner gets re-arrested (in sociology), or the time taken to solve a problem (in education). The analysis of data of this kind is commonly called ‘survival analysis’ (or ‘duration analysis’ depending on the area of application). It is common for this type of data that certain event times are not observed. Instead they can be subject to different types of incompleteness, like right censoring, left censoring, interval censoring, left or right truncation, and combinations of these. The most common type of incompleteness is right censoring, and we will therefore focus on this type in this thesis. Right censoring takes place when the occurrence of another event prevents the observation of the event of interest. For instance, in medical studies a patient might die due to another disease, of he or she might leave the study. In that case one observes a lower bound for the event time of interest (called the censoring time) and the patient is said to be right censored. It is necessary to take this censoring mechanism into account when one wishes to analyze this type of data, in order to avoid inconsistent estimators or tests. The main focus in survival analysis lies on developing methods and models that allow to do correct inference when the data are subject to right censoring.

A common assumption in survival analysis is the assumption of independence between the event (or survival) time and the censoring time. This assumption is in most contexts not testable, since we rarely have data for which
we observe both the survival and the censoring time. Hence, the relation between these two times can in these cases not be estimated from the data and an assumption needs to be imposed. The independence assumption is realistic, as long as the mechanism that causes data to be censored is unrelated to the actual survival time. We refer to this as the independent censoring mechanism. This is for instance the case for administrative censoring, which means that all subjects that are still at risk (i.e. who are still waiting for the event to happen) at the end of the study period are lost to follow up and are therefore censored. Clearly the censoring time is here unrelated to the actual survival time.

However, there are situations in which the assumption of independent censoring is not realistic. We give here four concrete examples of situations where this assumption is violated:

1. As an example of a medical study, one could look at data on male patients with prostate cancer who were diagnosed and treated in the early stage of the disease, that is, before the cancer spread to other parts of the body. In this case the time of interest is the time until death caused by prostate cancer, and censoring is caused by death due to other causes. The real data for this study can be found in the SEER database (and they were analyzed in more detail in Li et al. (2007)), where the cardio-vascular disease was the major source of censoring. Therefore, since prostate cancer and cardio-vascular disease share lots of risk factors, e.g. a high intake of fat, one would tend to assume positive dependence between time to death caused by prostate cancer and the censoring time.

2. Staying in a medical field we consider the PBC data set widely studied in Fleming and Harrington (1991). This study follows patients suffering from primary biliary cirrhosis, from the day of registration until death due to the illness. Right censoring is caused either by the end of the study or by liver transplantation. Now, while the former censoring is independent from the time until premature death, the reception of a new liver might influence the survival time of a patient in a positive way.

3. As a last specific example from the medical field, we consider the bone marrow transplantation data, which are described in Klein and Moeschberger (1997) (and are analyzed in Chapter 2). The data come from patients that are followed in their recovery from acute leukemia after bone marrow transplantation. One is interested in the disease-free survival time, i.e. the time until a patient has a relapse of leukemia. However, patients can be censored by two possible events: disease-free death or disease-free and alive at the end of the study. The censoring time is the time until the first
of these two events take place. Since the time until relapse and the time until disease-free death share common risk factors, i.e. overall health, it seems acceptable to assume that time until relapse and censoring are dependent.

4. Consider e.g. a more general situation where a patient decides to leave a medical study because he or she feels in very good shape and prefers therefore to stop the treatment. In such a case the censoring time will likely be negatively correlated with the survival time. On the other hand we might also have patients who decide to stop a certain treatment because they are not in good health and would e.g. prefer to change treatment or hospital. These are patients for which the survival time and the censoring time will tend to be positively correlated.

In this thesis we focus on the situation where the dependence structure between the survival and censoring time is given (both dependent and independent censoring are allowed) and we are interested in estimating certain quantities that are of common interest in survival analysis under the assumed dependence structure. We do this by considering that in addition to the survival time we also observe a continuous covariate. In the case of a medical study, which is our primary area of application, this explanatory variable can be the age, blood pressure, cholesterol level, or any other index that might have an impact on the survival time of a patient. We wish to take this covariate into account by considering a regression approach. We prefer to use so-called nonparametric location-scale regression models, that are very flexible and that are not based on heavy modeling assumptions.

Before we can explain what we plan to estimate in this nonparametric location-scale model and how we plan to take into account that the response in this model is subject to dependent right censoring, we have to introduce a number of definitions, estimators, concepts and references, that will be used throughout the thesis, and which serve as a basis for the remainder of this thesis.

1.2 Some basic concepts

Throughout the thesis we let \( Y \) be a (possible monotone transformation of a) survival time, and \( C \) denotes a censoring time. In the random right censoring model we observe \( (T, \Delta) \), where \( T = \min(Y, C) \) and \( \Delta = I(Y \leq C) \). The nonparametric estimation of the distribution of \( Y \) depends on whether \( Y \) and
Chapter 1. Introduction

C are independent or not, on the presence or not of a covariate \( X \), and (the case being) on the model imposed on \((X,Y)\). We therefore consider several cases and explain in each case how nonparametric estimators can be constructed.

1.2.1 Independent censoring

We start with the simplest case in which \( Y \) and \( C \) are independent and no covariates are present. We suppose that we have a sample at hand of i.i.d. replications \((T_i, \Delta_i)\) of \((T, \Delta)\). The nonparametric estimation of the distribution \( F(\cdot) = P(Y \leq \cdot) \) goes back to the seminal paper by Kaplan and Meier (1958), who proposed the following estimator (in the absence of ties):

\[
\hat{F}(y) = 1 - \left\{ \prod_{T(i) \leq y} \left( 1 - \frac{1}{n-i+1} \right)^{\Delta(i)} \right\},
\]

(1.2.1)

where \( T(1), \ldots, T(n) \) are the ordered \( T_i \), and \( \Delta(1), \ldots, \Delta(n) \) are the corresponding indicators \( \Delta_i \). This estimator is a step function which makes steps at the uncensored observations, and it reduces to the empirical distribution function when all data are uncensored. An asymptotic i.i.d. representation for this estimator has been obtained by Lo and Singh (1986), which is the starting point for many other asymptotic results.

Consider now the case where a one-dimensional covariate \( X \) is observed and it is assumed that \( Y \) and \( C \) are independent given \( X \). The data now consist of triplets \((X_i, T_i, \Delta_i)\), \( i = 1, \ldots, n \), which are i.i.d. and have the same distribution as \((X, T, \Delta)\). Then, a conditional Kaplan-Meier estimator has been proposed by Beran (1981) for the conditional distribution \( F(\cdot | x) = P(Y \leq \cdot | X = x) \).

When no ties are present, the estimator is defined by

\[
\tilde{F}(y|x) = 1 - \left\{ \prod_{T(i) \leq y} \left( 1 - \frac{W_{n(i)}(x, h_n)}{1 - \sum_{j=1}^{i-1} W_{n(j)}(x, h_n)} \right)^{\Delta(i)} \right\}.
\]

(1.2.2)

Here, \( W_{ni}(x, h_n) \) are Nadaraya-Watson weights, i.e.

\[
W_{ni}(x, h_n) = \frac{K \left( (x - X_i)/h_n \right)}{\sum_{j=1}^{n} K \left( (x - X_j)/h_n \right)},
\]

(1.2.3)

where \( K \) is a known density function, called the kernel function and \( h_n \) is a sequence of positive constants, converging to 0 as \( n \) tends to infinity, called the bandwidth sequence. Moreover, \( T(1), \ldots, T(n) \) are the ordered \( T_i \), and \( \Delta(1), \ldots, \Delta(n) \) and \( W_{n(1)}(x, h_n), \ldots, W_{n(n)}(x, h_n) \) are the corresponding indicators \( \Delta_i \) and weights \( W_{ni}(x, h_n) \). It is easily seen that the estimator \( \tilde{F}(y|x) \)
1.2. Some basic concepts

is a step function making jumps at the uncensored observations for which the corresponding covariate belongs to the window around \( x \). It reduces to the Kaplan-Meier estimator when \( W_n(x, h_n) = n^{-1} \) for all \( i \), and it reduces to the estimator proposed by Stone (1977) when \( \Delta_i = 1 \) for all \( i \). The latter estimator is a nonparametric kernel estimator of the conditional distribution \( F(y|x) \) in the absence of censoring. The asymptotic properties of the Beran-estimator have been studied by many authors, including Dabrowska (1989), González-Manteiga and Cadarso- Suárez (1994), Akritas (1994), Van Keilegom and Veraverbeke (1997) and Du and Akritas (2002).

We finally turn to the case where a nonparametric location-scale model is imposed to describe the relation between \( X \) and \( Y \):

\[
Y = m(X) + \sigma(X)\varepsilon, \tag{1.2.4}
\]

where we assume that \( \varepsilon \) and \( X \) are independent, and where the functions \( m(\cdot) \) and \( \sigma(\cdot) \) are smooth but unknown. We assume that \( m(\cdot) \) and \( \sigma(\cdot) \) are, respectively, a location and scale functional. This means that we can write \( m(x) = T(F_Y(\cdot|x)) \) and \( \sigma(x) = S(F_Y(\cdot|x)) \), for some functionals \( T \) and \( S \), such that

\[
T(F_{aY+b}(|x})) = aT(F_Y(\cdot|x)) + b
\]

and

\[
S(F_{aY+b}(|x})) = aS(F_Y(\cdot|x))
\]

for all \( a \geq 0 \) and all real \( b \), where \( F_Y(\cdot|x) \) denotes here the distribution of \( Y \) given \( X = x \) (see also Huber (1981), p. 59, 202).

Model (1.2.4) has been considered in Fan and Gijbels (1994) who studied estimation of \( m \) by local linear regression. When \( Y \) is the logarithm of the survival time, model (1.2.4) can be viewed as a nonparametric version of the accelerated failure time model.

Van Keilegom and Akritas (1999) studied the estimation of model (1.2.4). They first noticed that the conditional distribution of \( Y \) given \( X = x \) can be written as

\[
F(y|x) = F_e \left( \frac{y - m(x)}{\sigma(x)} \right), \tag{1.2.5}
\]

where \( F_e(\cdot) = P(\varepsilon \leq \cdot) \) is the error distribution. Hence, we can estimate \( F(y|x) \) by replacing \( F_e(\cdot), m(\cdot) \) and \( \sigma(\cdot) \) by suitable estimators. First, to estimate \( m \) and \( \sigma \) we work with the following location and scale functionals :

\[
m(x) = \int_{0}^{1} F^{-1}(s|x)J(s) \, ds,
\]
Chapter 1. Introduction

and

\[ \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s) \, ds - \bar m^2(x), \]

which are L-type functionals. Here, \( F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\} \) is the conditional quantile of order \( s \), and \( J(s) \) is a given score (weight) function satisfying \( \int_0^1 J(s) \, ds = 1 \). To estimate \( m \) and \( \sigma \) we replace the conditional quantile by the corresponding quantile of the Beran estimator \( \tilde F(\cdot|x) \) defined in (1.2.2). This yields the following estimators:

\[ \tilde m(x) = \int_0^1 \tilde F^{-1}(s|x) J(s) \, ds, \]

and

\[ \tilde \sigma^2(x) = \int_0^1 \tilde F^{-1}(s|x)^2 J(s) \, ds - \tilde m^2(x). \]

Define \( \tilde E_i = (Y_i - \tilde m(X_i))/\tilde \sigma(X_i) \), \( i = 1, \ldots, n \), and let

\[ \tilde F_{e}(y) = 1 - \left\{ \prod_{\tilde E_{(i)} \leq y} \left( 1 - \frac{1}{n - i + 1} \right)^{\Delta_{(i)}} \right\} \]  

(1.2.6)

be the Kaplan-Meier estimator of the error distribution, where \( \tilde E_{(1)} \leq \cdots \leq \tilde E_{(n)} \) are the ordered residuals, \( \Delta_{(1)}, \ldots, \Delta_{(n)} \) are the corresponding censoring indicators, and where as before we suppose that there are no ties. Finally, we obtain the following estimator of the conditional distribution function under model (1.2.4), inspired by relation (1.2.5):

\[ \tilde F_{LS}(y|x) = \tilde F_{e} \left( \frac{y - \tilde m(x)}{\tilde \sigma(x)} \right) \]  

(1.2.7)

(LS stands for location-scale). The asymptotic properties of the estimators \( \tilde F_{e}(y) \) and \( \tilde F_{LS}(y|x) \) can be found in Van Keilegom and Akritas (1999). They also explain what the advantages are of their estimator \( \tilde F_{LS}(y|x) \) compared to the Beran estimator \( \tilde F(y|x) \) in terms of the ability of both estimators to estimate well the right tail of the conditional distribution. As they explain, the location-scale model allows to estimate well the tails of the conditional distribution \( F(y|x) \), provided there is a region in the support of \( X \) where censoring is light. More precisely, thanks to relation (1.2.7), the quality of the estimation in the tail is determined by the quality of the estimator \( \tilde F_{e} \) in the tail. Since this estimator is a global estimator (as opposed to the Beran estimator \( \tilde F(\cdot|x) \))
which uses only local information), the estimator $\hat{F}_{LS}(\cdot|x)$ uses data coming from the whole support of $X$ and hence the quality in the tail is equally well for all values of $X$ and is determined by the region of $X$ where censoring is the lightest.

We finish this section with a digression from the development of our approach by mentioning two other more recent papers, in which the problem of estimating the conditional distribution $F(\cdot|x)$ in model (1.2.4) under censoring is also considered, and the same advantages in estimating the right tail are obtained. The papers by Lewbel and Linton (2002) and Chen et al. (2005) consider model (1.2.4), where they allow the covariate $X$ to be multidimensional. They focused on independent left censoring of type 1, that is, they work in them the model

$$ Y = m(X) + \sigma(X)\varepsilon $$
$$ T = \max(Y,c) $$

where $c$ is a known constant, $Y$ is the unobserved latent dependent variable and $T$ is the observed dependent variable, which is equal to the latent variable $Y$ when it exceeds the censoring point $c$. The error $\varepsilon$ is unknown and independent of $X$. Both papers derive estimators for $m$, $\sigma$ and $F(\cdot|x)$, and study their asymptotic properties. Additionally Lewbel and Linton (2002) develop an estimator for the partial derivatives $m_k(x) = \partial m(x)/\partial x_k$, for continuously distributed components $x_k$ of the covariate $x$, and they extend all the estimators to the truncated regression model, where $Y$ is observed only when it is not censored. Due to the technical nature of the construction of their estimators we refer to the original papers for further details.

### 1.2.2 Dependent censoring

We now turn to the case where the survival time $Y$ and the censoring time $C$ are not independent. As we have explained in Section 1, there are many situations where assuming that $Y$ and $C$ are independent is not realistic. We restrict attention here to the case where the dependence between $Y$ and $C$ is described by a copula model, and start with the case without covariates.

Zheng and Klein (1995) and Rivest and Wells (2001) were interested in estimating the distribution $F$ of $Y$ in the absence of covariates and under the assumption that the joint survival distribution of $Y$ and $C$ is given by

$$ P(Y > y, C > c) = C(1 - F(y), 1 - G(c)),$$
where $G$ is the distribution of the censoring time $C$, and $C$ is a known copula, i.e. a bivariate distribution function defined on $[0, 1] \times [0, 1]$ with uniform margins. Because of its nice properties and because of the broad range of different copula structures it covers, Rivest and Wells (2001) focused attention on the subclass of Archimedean copulas, i.e. they assumed that

$$P(Y > y, C > c) = \phi^{-1}\left[\phi(1 - F(y)) + \phi(1 - G(c))\right],$$

where $\phi$ is a generator function, i.e. a function from $(0, 1]$ to the positive real line, that is decreasing, convex and that satisfies $\phi(1) = 0$ and $\phi(0+) = \infty$. Under this framework, they proposed the following estimator of the distribution $F(\cdot)$, called the copula-graphic estimator:

$$\hat{F}(y) = 1 - \phi^{-1}\left\{ - \sum_{T_i \leq y, \Delta_i = 1} \left[\phi\left(1 - \hat{H}(T_i^-)\right) - \phi\left(1 - \hat{H}(T_i)\right)\right]\right\},$$

where $\hat{H}(y) = n^{-1} \sum_{i=1}^n I(T_i \leq y)$ is the empirical estimator of the distribution $H(y)$ of the observed survival time $T$, and $\hat{H}(y^-) = \lim_{t \downarrow y} \hat{H}(t)$. Simple calculus shows that the estimator $\hat{F}(y)$ reduces to the Kaplan-Meier estimator $\hat{F}(y)$ defined in (1.2.1) when the copula generator equals $\phi(t) = -\log(t)$, and so it is an extension of this estimator to the dependent data case.

Next, we consider an extension of the previous estimator to the case where a one-dimensional covariate $X$ is observed and it is assumed that $Y$ and $C$ are copula dependent given $X$. This means that

$$P(Y > y, C > c | X = x) = C_x(1 - F(y|x), 1 - G(c|x)),$$

where the copula $C_x$ is known and allowed to depend on $x$, and $G(c|x) = P(C \leq c | X = x)$ is the conditional distribution of $C$ given $X = x$. As in the unconditional case, we restrict attention to the case where $C_x$ is an Archimedean copula, and we denote its generator by $\phi_x$. Under this framework, Braekers and Veraverbeke (2005) proposed the following conditional copula-graphic estimator:

$$\hat{F}(y|x) = 1 - \phi_x^{-1}\left\{ - \sum_{T_i \leq y, \Delta_i = 1} \left[\phi_x\left(1 - \hat{H}(T_i^-|x)\right) - \phi_x\left(1 - \hat{H}(T_i|x)\right)\right]\right\},$$

where $\hat{H}(y|x) = n^{-1} \sum_{i=1}^n W_{ni}(x, h_n) I(T_i \leq y)$ is the Nadaraya-Watson estimator (Nadaraya, 1964, Watson, 1964), whose weights $W_{ni}(x, h_n)$ are defined in (1.2.3). It is easily seen that the estimator reduces to the Beran estimator $\hat{F}(y)$ defined in (1.2.2) if the copula $C_x$ is the copula that leads to independence, namely $C_x(u_1, u_2) = u_1 u_2$ or equivalently $\phi_x(t) = -\log(t)$. Also, when all weights $W_{ni}(x, h_n)$ are equal to $n^{-1}$ the estimator reduces to the estimator $\hat{F}(y)$ of Zheng and Klein (1995) and Rivest and Wells (2001).
1.3 Outline of the thesis

We are now ready to explain what the contribution of this thesis is. Table 1.3 summarizes the estimators that have been introduced in Section 2, and classifies these estimators depending on the type of censoring and on the assumed underlying model. It is clear from this table that there is a gap in the literature. This gap consists of the case where it is assumed that the relation between $Y$ and $X$ is given by a nonparametric location-scale model, and $Y$ and $C$ are dependent given $X$. The objective of this thesis is to fill this gap.

<table>
<thead>
<tr>
<th>Model</th>
<th>Type of censoring</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No covariates</td>
<td></td>
</tr>
<tr>
<td>No covariates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonparametric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location-scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No censoring</th>
<th>Type of censoring</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>Kaplan-Meier</td>
<td>(1958)</td>
</tr>
<tr>
<td>Dependent</td>
<td>Rivest-Wells</td>
<td>(2001)</td>
</tr>
<tr>
<td>Location-scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression</td>
<td>Stone</td>
<td>(1977)</td>
</tr>
<tr>
<td>Van Keilegom-</td>
<td>Beran</td>
<td>(1981)</td>
</tr>
<tr>
<td>regression</td>
<td>Braekers-Veraverbeke</td>
<td>(2005)</td>
</tr>
<tr>
<td>Location-scale</td>
<td>Akritas - Van Keilegom</td>
<td>(2001)</td>
</tr>
<tr>
<td>regression</td>
<td>Van Keilegom-</td>
<td>Akritas (1999)</td>
</tr>
<tr>
<td></td>
<td>Not yet studied</td>
<td>in the literature</td>
</tr>
</tbody>
</table>

Table 1.1: Schematic overview of the literature addressing models mentioned in this chapter.

More precisely, we will propose estimators of the error distribution and of the conditional distribution in this framework, and we will study their asymptotic properties. In order to do so, we first need to study estimators of the location functional $m(\cdot)$ and the scale functional $\sigma(\cdot)$ in the copula dependent case. This will be the topic of the next chapter, which is based on Sujica and Van Keilegom (2013). The estimators will be based on the conditional copula-graphic estimator $\hat{F}(\cdot|x)$ proposed by Braekers and Veraverbeke (2005). We will study these estimators by showing their asymptotic properties, and also via a simulation study which allows to verify the finite sample behavior of the estimators. As an illustration we will apply these estimators to leukemia data for which the independent censoring assumption is likely violated. In Chapter 3 we will study the estimators of the error distribution and of the conditional distribution. This chapter is an extended version of the paper by Sujica and Van Keilegom (2014). Again, we will carry out both an asymptotic study and a finite sample study of these estimators, and we will compare the proposed estimator of the conditional distribution with the conditional copula-graphic estimator $\hat{F}(\cdot|x)$ and also with the estimator $\hat{F}_{LS}(\cdot|x)$ proposed by Van Keilegom and Akritas (1999). Finally, in Chapter 4 we will summarize the main
results of this thesis and we will give some ideas for future research.
Chapter 2

Estimation of location and scale functionals in nonparametric regression under copula dependent censoring

**ABSTRACT:** Let \((X, Y)\) be a random vector, where \(Y\) denotes the variable of interest possibly subject to random right censoring, and \(X\) is a covariate. The variable \(Y\) is a (possible monotone transformation of a) survival time. The censoring time \(C\) and the survival time \(Y\) are allowed to be dependent, and the dependence is described via a known copula (this also includes the independent case). Under this setting we propose estimators of certain location and scale functionals of \(Y\) given \(X\). We derive their asymptotic properties, uniformly over the support of \(X\). In particular we derive an asymptotic representation and the uniform convergence rates for these estimators and their derivatives. We also prove asymptotic results for an estimator of the conditional distribution (the so-called conditional copula-graphic estimator), which generalizes previous results obtained by Braekers and Veraverbeke (2005). We also illustrate the results via simulations and the analysis of data on bone marrow transplantation.

**Key Words:** Asymptotic representation; convergence rates; copulas; dependent censoring; nonparametric regression, right censoring, survival analysis.

---

\(^1\)This chapter is a slightly edited version of Sujica and Van Keilegom (2013).
2.1 Introduction

Let \((X, Y)\) be a random vector where \(Y\) denotes a possible transformation of the variable of interest and \(X\) is a covariate. We assume that \(Y\) is subject to random right censoring, i.e. instead of observing \(Y\) we only observe \((T, \Delta)\), where \(T = \min(Y, C), \Delta = I(Y \leq C)\) and \(C\) represents the censoring time. Let \((T_i, X_i, \Delta_i), i = 1, \ldots, n\) be \(n\) independent vectors having the same distribution as \((T, X, \Delta)\).

In the statistical literature it is very common to assume that \(Y\) and \(C\) are independent given \(X\). Under this assumption a lot of work has been done on the nonparametric estimation of the conditional distribution \(F(\cdot | x) = P(Y \leq \cdot | X = x)\). We refer to Beran (1981), Dabrowska (1989), González-Manteiga and Cadarso-Suarez (1994), Akritas (1994), Van Keilegom and Veraverbeke (1997a,b), among many others. The nonparametric kernel estimator in this setting is often referred to as Beran’s estimator, and is a generalization of the Kaplan-Meier estimator to the inclusion of covariates.

However, in various situations it is unrealistic to assume that \(Y\) and \(C\) are conditionally independent given \(X\). Consider e.g. the situation where a patient decides to leave a medical study because he or she feels in very good shape and prefers therefore to stop the treatment. In such a case the censoring time \(C\) will likely be negatively correlated with the survival time \(Y\). On the other hand we might also have patients who decide to stop a certain treatment because they are not in good health and would e.g. prefer to change treatment or hospital. These are patients for which \(Y\) and \(C\) will tend to be positively correlated. In addition, in many situations the strength of the dependence between \(Y\) and \(C\) will depend on a certain number of external factors (or covariates). This motivates us to consider in this paper the situation where \(Y\) and \(C\) are dependent given \(X\).

However, in the context without covariates, Tsiatis (1975) showed in his seminal work that the joint distribution of \(Y\) and \(C\) cannot be identified by their minimum and the censoring indicator when the dependence between \(Y\) and \(C\) is unspecified. Crowder (1991) showed that even when the marginal distributions are known, the joint distribution function is not identifiable. Tsiatis’ observations have been the starting point for research on how to modify the model so as to identify the distribution of \((Y, C)\), and a variety of modeling approaches have been studied in the past. For example Basu and Ghosh (1978) show identifiability under the assumption that \((Y, C)\) is the bivariate normal, and the exponential distribution. Slud and Rubinstein (1983), instead of fixing the precise dependence structure of the joint distribution in advance, assume that
certain hazard ratio involving the survival and the censoring time is known. Another approach is proposed by Ebrahimi et al. (2003), where restrictions are imposed on partial derivative of the conditional distribution of the survival function given that the censoring time is larger than a given threshold. The approach we will use consists in modeling the dependence structure between \( Y \) and \( C \) by means of a copula function. The advantage of using copulas is that one only models the dependence structure without affecting the margins (see Sklar (1959)). Wang (2012) showed that even if we restrict attention to Archimedean copula functions, the model is in general not identifiable, whereas Schwarz, Jongbloed, and Van Keilegom (2013) showed that when the margins are known, the copula function can under certain conditions be identified. When the margins are unknown, Zheng and Klein (1995) and Rivest and Wells (2001) supposed that \( Y \) and \( C \) are dependent via some known copula and they showed that the marginal distribution of \( Y \) and \( C \) are identifiable under very weak conditions. They developed an estimator of the distribution of \( Y \), which they called the copula-graphic estimator. An extension of this copula-graphic estimator has been proposed by de Uña Álvarez and Veraverbeke (2013), when the full process is independently censored by some administrative censoring time.

In the presence of covariates, Braekers and Veraverbeke (2005) extended the work of Rivest and Wells (2001) to the case of a fixed design regression model. We follow their approach, except that we assume that \( X \) is random, and we model the conditional dependence between \( Y \) and \( C \) via a known copula \( C_x \) that is allowed to depend on the value of \( X \):

\[
P(Y > y, \, C > c | X = x) = C_x \left( 1 - F(y|x), \, 1 - G(c|x) \right),
\]

where \( 1 - G(c|x) = P(C > c | X = x) \) is the conditional survival function of the censoring time \( C \) given \( X = x \). Because of its nice properties and because of the broad range of different copula structures it covers, we focus attention in what follows on the subclass of Archimedean copulas, i.e. we assume that

\[
P(Y > y, \, C > c | X = x) = \phi_x^{-1} \left[ \phi_x \left( 1 - F(y|x) \right) + \phi_x \left( 1 - G(c|x) \right) \right], \tag{2.1.1}
\]

for an Archimedean copula generator \( \phi_x \), i.e. a function from \((0, 1)\) to \( \mathbb{R}^+ \) that is decreasing, convex and that satisfies \( \phi_x(1) = 0 \) and \( \phi_x(0^+) = +\infty \). Under these conditions Braekers and Veraverbeke (2005) proposed the conditional copula-graphic estimator, which generalizes the Beran estimator to the dependent setting (2.1.1). Their estimator reduces to Beran’s estimator when \( \phi_x(\cdot) = -\log(\cdot) \).

Under this setting of dependent censoring described by an Archimedean copula, we are interested in studying location and scale functionals of \( Y \) given
$X = x$, which we will denote by $m(x)$ and $\sigma^2(x)$. Because we are working in a completely nonparametric framework and because the response is subject to right censoring, we won’t be able to estimate the conditional mean and the conditional variance in a consistent way. Instead we will assume that $m(x)$ and $\sigma^2(x)$ take the following form:

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

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

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of $Y$ given $X = x$, and $J(s)$ is a given positive weight function such that $\int_0^1 J(s) ds = 1$. Note that the choice $J(s) \equiv 1$ would lead to the conditional mean and variance. But because of lack of data in the right tail of $F(\cdot|x)$, we will choose $J(s)$ so that no weight is given to the regions where the conditional distribution cannot be estimated consistently. This form of location and scale functionals (known as L-functionals) is very flexible and covers a broad range of common functionals (see e.g. Serfling (1980) to learn more about L-functionals).

The objective of this chapter is to propose appropriate estimators of $m(x)$ and $\sigma^2(x)$ and to study their asymptotic properties uniformly in $x$. These quantities have so far not been studied in the literature on nonparametric regression with copula dependent censoring. Indeed, attention has been focused on the estimation of the conditional distribution $F(y|x)$ for a fixed value of $x$ and under a fixed design setting. See Braekers and Veraverbeke (2005), Braekers and Veraverbeke (2008) and Gaddah and Braekers (2010a,b), where the latter three papers assume that the data satisfy a conditional Koziol-Green model. As a by-product we will also study an estimator of the conditional distribution $F(y|x)$. Compared to the aforementioned papers, the main difference is that we obtain results uniformly in $x$ and $y$, which are technically harder to prove.

The chapter is organized as follows. In Section 2.2 we define the estimators of $m(x)$ and $\sigma^2(x)$, and state the assumptions that will be needed for the asymptotic results. Section 2.3 contains the main asymptotic results of this chapter. We first study an estimator of the conditional distribution $F(y|x)$ and we next give an asymptotic representation and the uniform convergence rates for the estimators of $m(x)$ and $\sigma^2(x)$ and of their derivatives. In Section 2.4 we show the results of a small simulation study, and in Section 2.5 we illustrate our estimation method via the analysis of data on bone marrow transplantation. Finally, the appendix contains the proofs of the asymptotic results stated in Section 2.3.
2.2 Definitions and assumptions

We focus in this section on the estimation of the functions \( m(x) \) and \( \sigma^2(x) \) given in (2.1.2). These functions depend on the conditional distribution \( F(\cdot|x) \), which we need to estimate first. Braekers and Veraverbeke (2005) defined the so-called conditional copula-graphic estimator of \( F(\cdot|x) \):

\[
\hat{F}(y|x) = 1 - \phi_x^{-1} \left\{ - \sum_{T_i \leq y, \Delta_i = 1} \left[ \phi_x \left( \hat{H}(T_i|x) \right) - \phi_x \left( \hat{H}(T_i|x) \right) \right] \right\}, \quad (2.2.1)
\]

where \( \hat{H}(y|x) = \sum_{i=1}^{n} W_{ni}(x, h_n) I(T_i \leq y) \) is the Nadaraya-Watson estimator of the conditional distribution \( H(y|x) = P(T \leq y|X = x) \) of \( T \) given \( X = x \), \( \hat{H}(y|x) = 1 - \hat{H}(y|x) \), \( \hat{H}(y^-|x) = \lim_{t \uparrow y} \hat{H}(t|x) \),

\[
W_{ni}(x, h_n) = \frac{K \left( (x - X_i)/h_n \right)}{\sum_{j=1}^{n} K \left( (x - X_j)/h_n \right)}
\]

are Nadaraya-Watson weights, \( K \) is a probability density function (kernel), and \( h \equiv h_n \) is a bandwidth sequence tending to zero when \( n \) tends to infinity. The estimator \( \hat{F}(y|x) \) is an extension of the Beran estimator in the sense that it allows for dependent censoring, and it is also an extension of the estimator proposed by Zheng and Klein (1995), since it includes covariates.

This leads to the following estimators for \( m(x) \) and \( \sigma^2(x) \):

\[
\hat{m}(x) = \int_0^1 \hat{F}^{-1}(s|x) J(s) ds \quad \text{and} \quad \hat{\sigma}^2(x) = \int_0^1 \left[ \hat{F}^{-1}(s|x) - \hat{m}(x) \right]^2 J(s) ds.
\]

(2.2.2)

Note that these estimators are in the same spirit as the ones proposed by Van Keilegom and Akritas (1999), who worked under the assumption of conditional independence between \( Y \) and \( C \) given \( X \), and who estimated \( F(\cdot|x) \) by means of the Beran estimator instead of the estimator defined in (2.2.1). Therefore, if we take \( \phi_x(\cdot) = -\log(\cdot) \) in formula (2.2.1), the estimators \( \hat{m}(x) \) and \( \hat{\sigma}^2(x) \) reduce to the ones proposed in Van Keilegom and Akritas (1999).

The assumptions below are important for the establishment of the asymptotic results in Section 2.3. In addition to the distributions \( F(\cdot|x) \), \( G(\cdot|x) \) and \( H(\cdot|x) \) already defined above, they concern the subdistribution \( H^u(y|x) = P(T \leq y, \Delta = 1|X = x) \) of the uncensored observations and the distribution
Chapter 2. Estimation of location and scale functionals in nonparametric regression

under copula dependent censoring

\( F_X(x) = P(X \leq x) \) of the covariate. The probability density functions of the
distribution functions defined above will be denoted by the corresponding lower
case letters. Also, 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 an arbitrary (sub)distribution function \( L(y|x) \) we will use the notations
\( l(y|x) = L'(y|x) = \frac{\partial}{\partial y} L(y|x), \hat{L}(y|x) = \frac{\partial}{\partial y} L(y|x) \) and similar notations will
be used for higher order derivatives. (In the proofs, the function \( L(y|x) \) of
assumption (A5) below will be either \( H(y|x) \) or \( H^n(y|x) \)). We also use the
notation \( \mathcal{L}(y|x) = 1 - L(y|x) \) throughout the chapter.

(A1) (i) The sequence \( h_n \) satisfies \( nh_n^2 (\log n)^{-1} = O(1) \) and \( (nh_n)^{-1} \log n \to 0 \).

(ii) The support \( R_X \) of \( X \) is a bounded interval in \( \mathbb{R} \).

(iii) The probability density function \( K \) has compact support \([-M,M] \) for some \( M > 0 \), \( \int uK(u)du = 0 \), and \( K \) is twice continuously
differentiable.

(A2) (i) There exist \( 0 \leq s_0 \leq s_1 \leq 1 \) such that \( s_1 \leq \inf_x F(\tilde{T}_x|x), s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}, s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\} \) and
\( \inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0 \).

(ii) The function \( J \) is bounded and twice continuously differentiable on
the interval \((s_0,s_1), \int_{s_0}^{s_1} J(s)ds = 1 \) and \( J(s) \geq 0 \) for all \( 0 \leq s \leq 1 \).

(A3) (i) The distribution \( F_X \) is three times continuously differentiable on the
interior of \( R_X \) (and the right (left) limit of \( F_X'' \) at \( \min R_X \) (\( \max R_X \))
exists) and \( \inf_{x \in R_X} f_X(x) > 0 \).

(ii) The functions \( m \) and \( \sigma \) are twice continuously differentiable on \( R_X \)
and \( \inf_{x \in R_X} \sigma(x) > 0 \).

(A4) (i) The functions \( \phi_x(u) = \frac{u}{n^2} \phi_x(u), \phi_x''(u) \) and \( \phi_x^{(3)}(u) \) exist and are
continuous in \((x,u) \in R_X \times (0,1)\).

(ii) The functions \( \phi_x''(u) = \frac{u}{n^2} \phi_x(u), \phi_x^{(3)}(u) \) and \( \phi_x^{(4)}(u) \) exist and
are continuous in \((x,u) \in R_X \times (0,1)\).

(iii) The function \( \phi_x \) satisfies \( \phi_x'(1) < 0 \).

(A5) (i) \( L(y|x) \) is continuous in \((x,y)\).

(ii) \( L'(y|x) \) exists, is continuous in \((x,y)\) and \( \sup_{x,y} |yL'(y|x)| < \infty \).

(iii) \( L''(y|x) \) exists, is continuous in \((x,y)\) and \( \sup_{x,y} |y^2L''(y|x)| < \infty \).

(iv) \( \hat{L}(y|x) \) exists, is continuous in \((x,y)\) and \( \sup_{x,y} |y\hat{L}(y|x)| < \infty \).
2.3. Asymptotic results

(v) \( \tilde{L}(y|x) \) exists, is continuous in \((x, y)\) and \(\sup_{x,y} |y^2 \tilde{L}(y|x)| < \infty\).

\(\text{(A6)}\) There exist continuous and non-decreasing functions \(M_j\) with \(M_j(-\infty) = 0\) and \(M_j(\infty) < \infty\) \((j = 1, \ldots, 4)\) such that

\[
\begin{align*}
|H(y_2|x) - H(y_1|x)| &\leq |M_1(y_2) - M_1(y_1)|, \\
|H''(y_2|x) - H''(y_1|x)| &\leq |M_2(y_2) - M_2(y_1)|, \\
\left| \frac{\partial H(y_2|x)}{\partial x} - \frac{\partial H(y_1|x)}{\partial x} \right| &\leq |M_3(y_2) - M_3(y_1)|, \\
\left| \frac{\partial H''(y_2|x)}{\partial x} - \frac{\partial H''(y_1|x)}{\partial x} \right| &\leq |M_4(y_2) - M_4(y_1)|,
\end{align*}
\]

for all \(x \in R_X, -\infty < y_1, y_2 < +\infty\).

Note that assumption (A6) comes from Du and Akritas (2002), and is required to prove an i.i.d. representation for our estimator \(\hat{F}(y|x)\), whose remainder term is negligible uniformly in \(x\) and \(y\). The assumption (A6) is slightly stronger from the assumption (A5) (for details see Remarks 3.3.1 to 3.3.4.)

The following functions will also be needed in the sequel.

\[
g(T, \Delta, y|x) = \frac{-1}{\phi_x(F(y|x))} \left\{ \int_{-\infty}^{y} \phi_x' \left( \overline{H}(s|x) \right) \left[ I(T \leq s) - H(s|x) \right] dH''(s|x) \\
- \phi_x' \left( H(y|x) \right) \left[ I(T \leq y, \Delta = 1) - H''(y|x) \right] \\
- \int_{-\infty}^{y} \phi_x'' \left( \overline{H}(s|x) \right) \left[ I(T \leq s, \Delta = 1) - H''(s|x) \right] dH(s|x) \right\},
\]

\[
\eta(T, \Delta|x) = \int_{-\infty}^{+\infty} J(F(y|x))g(T, \Delta, y|x) dy,
\]

\[
\zeta(T, \Delta|x) = \int_{-\infty}^{+\infty} J(F(y|x))g(T, \Delta, y|x) \frac{y - m(x)}{\sigma(x)} dy.
\]

2.3 Asymptotic results

2.3.1 Asymptotic results for the estimator of \(F(y|x)\)

We start this section with some new results concerning the copula-graphic estimator \(\hat{F}(y|x)\). In particular, we will derive uniform consistency rates for \(\hat{F}(y|x)\), for its derivative \(\hat{F}'(y|x)\) and for the ‘derivative’ of order 1 + \(\delta\) defined by \((\hat{F}(y|x) - \hat{F}(y|x'))/(x - x')^{\delta}\). In addition we will also show an iid asymptotic representation for \(\hat{F}(y|x) - F(y|x)\), and establish the rate of convergence of the remainder term uniformly in \(x\) and \(y\).
These results are useful for establishing similar results for the estimators of \( m(x) \) and \( \sigma^2(x) \) in the next subsection. In addition, they generalize previous results obtained by Braekers and Veraverbeke (2005), who focused attention on the estimator \( \hat{F}(y|x) \) itself and who established the rate of convergence for a fixed value of \( x \) under a fixed design setting. We refer to Subsection 2.3.2 for additional motivation for studying these results.

**Proposition 2.3.1.** Assume (A1), (A3)(i), (A4)(i,iii), and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,

\[
\sup_{x \in \mathbb{R}X} \sup_{y \leq \hat{T}_x} |\hat{F}(y|x) - F(y|x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \quad \text{a.s.}
\]

**Proposition 2.3.2.** Assume (A1), (A3)(i), (A4)(i,iii), \( \hat{\phi}_x''(u) \) exists and is continuous in \((x,u)\), and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,

\[
\sup_{x \in \mathbb{R}X} \sup_{y \leq \hat{T}_x} |\hat{F}(y|x) - F(y|x)| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \quad \text{a.s.}
\]

**Proposition 2.3.3.** Assume (A1), (A3)(i), (A4), and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,

\[
\sup_{x,x' \in \mathbb{R}X} \sup_{y \leq \hat{T}_x \land \hat{T}_{x'}} |\hat{F}(y|x) - \hat{F}(y|x') + \hat{F}(y|x')| \leq |x - x'|^{\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \quad \text{a.s.},
\]

for any \( 0 < \delta < 1 \).

**Proposition 2.3.4.** Assume (A1), (A3)(i), (A4)(i,iii), (A6) and \( H \) and \( H^u \) satisfy (A5). Then,

\[
\hat{F}(y|x) - F(y|x) = \frac{1}{nh_n f_X(x)} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) g(T_i, \Delta_i, y|x) + R_n(y|x),
\]

where \( \sup_{x \in \mathbb{R}X} \sup_{y \leq \hat{T}_x} |R_n(y|x)| = O\left((nh_n)^{-3/4}(\log n)^{3/4}\right) \quad \text{a.s.}
\]

**Remark 2.3.1.** Without assumption (A4)(iii) all the results in this subsection remain valid if we replace \( \sup_{y \leq \hat{T}_x} \) by \( \sup_{\hat{T}_x \leq y \leq \hat{T}_x} \), where \( \hat{t}_x \) is chosen such that \( \sup_{x \in \mathbb{R}X} (1 - H(\hat{t}_x|x)) < 1 \). Also note that in the next subsection we do not need to assume (A4)(iii), since by assumption (A2)(i) the score function \( J(s) \) equals zero for \( s \) close to 0 or 1.
2.3. Asymptotic results

Remark 2.3.2. Note that the rates of convergence in Propositions 2.3.2 and 2.3.3 (and also Propositions 2.3.6 and 2.3.7 below) are \( o(1) \) a.s., if it is assumed that \( O((nh_n)^{-1/2}(\log n)^{1/2}) = o(1) \).

### 2.3.2 Asymptotic results for the estimators of \( m(x) \) and \( \sigma^2(x) \)

In this section we will derive some asymptotic properties of the location estimator \( \hat{m}(x) \) and the scale estimator \( \hat{\sigma}(x) \) defined in (2.2.2). In particular, we will show the uniform consistency rates of \( \hat{m}(x) \), of \( \hat{\sigma}(x) \) and of \( (\hat{m}(x) - \hat{m}(x'))/(x - x')^{\delta} \), and of their analogues for \( \hat{\sigma} \). We will also prove an asymptotic representation for \( \hat{m}(x) \) and \( \hat{\sigma}(x) \). All results will be obtained uniformly in \( x \in \mathbb{R}^X \).

These results are important, since they show that with probability tending to one, the estimators \( \hat{m} \) and \( \hat{\sigma} \) belong to the class \( C^{1+\delta}_K(R_X) \) of differentiable functions \( f : R_X \rightarrow \mathbb{R} \) for which

\[
\|f\|_{1+\delta} = \max \left\{ \sup_x |f(x)|, \sup_x |f'(x)| + \sup_{x,x'} \frac{|f'(x) - f'(x')|}{|x - x'|^{\delta}} \right\}
\]

is bounded by \( K < \infty \). This class plays a major role in empirical process theory, since its covering and bracketing number enjoy nice properties, which are needed to show that the class is Donsker (see Van der Vaart and Wellner (1996) for more details).

The results have additional value as they are necessary if one is interested in exploring the asymptotic properties of an estimator of the error distribution in a nonparametric location-scale regression model of the form \( Y = m(X) + \sigma(X)\epsilon \) with \( \epsilon \) and \( X \) independent. This is investigated in Chapter 3.

Proposition 2.3.5. Assume (A1)–(A3), (A4)(i), and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,

(a) \( \sup_{x \in R_X} |\hat{m}(x) - m(x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s.

(b) \( \sup_{x \in R_X} |\hat{\sigma}(x) - \sigma(x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s.

Proposition 2.3.6. Assume (A1)–(A3), (A4)(i), \( \phi''_x(u) \) exists and is continuous in \( (x,u) \), and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,
(a) \( \sup_{x \in \mathbb{R}} \left| \hat{m}'(x) - m'(x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \) a.s.

(b) \( \sup_{x \in \mathbb{R}} \left| \hat{\sigma}'(x) - \sigma'(x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \) a.s.

**Proposition 2.3.7.** Assume (A1)–(A3), (A4)(i,ii) and \( H \) and \( H^u \) satisfy (A5)(i,iv,v). Then,

(a) \[
\sup_{x, x' \in \mathbb{R}} \frac{\left| \hat{m}'(x) - m'(x) - \hat{m}'(x') + m'(x') \right|}{|x - x'|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \text{ a.s.}
\]

(b) \[
\sup_{x, x' \in \mathbb{R}} \frac{\left| \hat{\sigma}'(x) - \sigma'(x) - \hat{\sigma}'(x') + \sigma'(x') \right|}{|x - x'|^\delta} = O((nh_n^{3+2\delta})^{-1/2}(\log n)^{1/2}) \text{ a.s.,}
\]

for any \( 0 < \delta < 1 \).

**Proposition 2.3.8.** Assume (A1)–(A3), (A4)(i), (A6) and \( H \) and \( H^u \) satisfy (A5). Then,

(a) \[
\hat{m}(x) - m(x) = \frac{-1}{nh_n f(x)} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h_n} \right) \eta\left( T_i, \Delta_i x \right) + R_{n1}(x)
\]

(b) \[
\hat{\sigma}(x) - \sigma(x) = \frac{-1}{nh_n f(x)} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h_n} \right) \zeta\left( T_i, \Delta_i x \right) + R_{n2}(x),
\]

where \( \sup_{x \in \mathbb{R}} |R_{nj}(x)| = O\left( (nh_n)^{-3/4}(\log n)^{3/4} \right) \) a.s. \( (j = 1, 2) \).

The latter result is important for obtaining e.g. the asymptotic normality of \( (nh_n)^{1/2}(\hat{m}(x) - m(x)) \) and \( (nh_n)^{1/2}(\hat{\sigma}(x) - \sigma(x)) \) for a fixed \( x \) in \( \mathbb{R} \). More importantly, it can also be used as a first big step for constructing asymptotic confidence bands for the true unknown functions \( m(\cdot) \) and \( \sigma(\cdot) \), similarly as was done in Claeskens and Van Keilegom (2003) in the context of multiparameter local likelihood estimating equations, or for testing hypotheses concerning these functions.
2.4 Simulations

In this section we will illustrate the finite sample behavior of our estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ by means of Monte Carlo simulations. We will compare our estimators with the ones proposed by Van Keilegom and Akritas (1999). These estimators are based on the assumption that $Y$ and $C$ are conditionally independent, and are defined as follows:

\[
\hat{m}(x) = \int_0^1 \tilde{F}^{-1}(s|x)J(s)ds \quad \text{and} \quad \hat{\sigma}^2(x) = \int_0^1 [\tilde{F}^{-1}(s|x) - \hat{m}(x)]^2 J(s)ds,
\]

where $\tilde{F}(\cdot|x)$ is the Beran estimator:

\[
\tilde{F}(y|x) = 1 - \prod_{T_i \leq y, \Delta_i = 1} \left\{1 - \frac{W_{ni}(x,h_n)}{\sum_{j=1}^n I(T_j \geq T_i)W_{nj}(x,h_n)}\right\}.
\]

We expect that when the conditional dependence between $Y$ and $C$ is strong, the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ will behave poorly compared to the new estimators $\tilde{m}(x)$ and $\tilde{\sigma}(x)$, that take this dependence into account. The results will therefore illustrate the importance of correctly specifying the dependence between $Y$ and $C$. We measure the performance of the estimators by means of their mean squared error ($MSE$) and by means of 90% confidence intervals. The simulations are carried out for samples of size $n = 100, n = 200$ and $n = 400$, and the results are obtained by using 2000 Monte Carlo simulations.

We generate i.i.d. data from the following regression model:

\[
Y = 6(X - 0.5)^2 + 0.5\varepsilon_1, \quad (2.4.2)
\]

where $X$ has a uniform distribution on $[0,1]$ and the error $\varepsilon_1$ has a standard normal distribution and is independent of $X$. The censoring variable $C$ satisfies $C = \alpha_1(X - 0.5)^2 + \alpha_2 + 0.5\varepsilon_2$ for certain choices of $\alpha_1$ and $\alpha_2$, where $\varepsilon_2$ is standard normal and independent of $X$. The constants $\alpha_1$ and $\alpha_2$ are chosen so that the global censoring rate is 45% and the local censoring rate (for a fixed value of $x$) is between 42% and 48%. We further assume that the dependence between $Y$ and $C$ given $X = x$ (i.e. the dependence between $\varepsilon_1$ and $\varepsilon_2$ given $X = x$) is described via a Gumbel copula:

\[
C_x(u_1, u_2) = \exp \left\{ - [ - (\log u_1)^{\gamma(x)} - (\log u_2)^{\gamma(x)}]^{1/\gamma(x)} \right\},
\]

where $\gamma(x) = \max(5 - 5x, 1)$. This means that the corresponding Archimedean copula generator equals $\phi_x(u) = - (\log u)^{\gamma(x)}$. With this construction the conditional dependence between $Y$ and $C$ given $X = x$ decreases from strong...
positive dependence to complete independence as \( x \) goes from 0 to 0.8, and it continues to be independent on \([0.8, 1]\). Note that Kendall tau’s coefficient decreases from 0.8 to 0.

The functionals \( m(\cdot) \) and \( \sigma(\cdot) \) that can be estimated consistently have to have corresponding function \( J \) that satisfies \( J(s) = 0 \) for \( s > \inf_{x \in R_x} \tilde{F}(+\infty|x) \).

We will be estimating the functionals \( m(\cdot) \) and \( \sigma(\cdot) \) defined by the score function \( J(s) = b^{-1}I(0 \leq s \leq b) \), where the constant \( b = 0.8 \), was chosen to be smaller than the average of 1000 simulated infima \( \inf_{x \in R_x} \tilde{F}(+\infty|x) \). Note that the functionals that we are estimating are not the ones from (2.4.2).

Note that, in the setting above, an equivalent way of writing the estimator \( \hat{\sigma}^2(x) \) is

\[
\hat{\sigma}^2(x) = \sum_{i=1}^{n} [Y_i - \hat{m}(x)]^2 [\hat{F}_b(Y_i|x) - \hat{F}_b(Y_i^-|x)] J(\hat{F}_b(Y_i|x)),
\]

where \( \hat{F}_b(y|x) = \min\{\tilde{F}(y|x), b\} \) and \( b \) is any value larger than \( \sup\{s : J(s) \neq 0\} \). (In a general setting, we can show that the formula above is an asymptotic approximation of \( \hat{\sigma}^2(x) \), if we use \( \tilde{F} \) instead of \( \hat{F}_b \).) Furthermore, \( \hat{\sigma}^2(x) \) is close to

\[
\sum_{i=1}^{n} [Y_i - m(X_i)]^2 [\hat{F}_b(Y_i|x) - \hat{F}_b(Y_i^-|x)] J(\hat{F}_b(Y_i|x)).
\]

It can be easily seen that both estimators are asymptotically equivalent under certain conditions on the bandwidth. In the sequel we work with the latter estimator, since simulations showed that it outperforms somewhat the former (which is expected, since for a given covariate \( X_i \), the difference \( Y_i - m(X_i) \) is a real deviation of \( Y_i \) from \( m(X_i) \), whereas \( Y_i - m(x) \) is not). The same applies to the estimator \( \hat{\sigma}(x) \), for which we also work with the asymptotically equivalent variant.

For the weights that appear in our estimators \( \hat{m}(x) \) and \( \hat{\sigma}(x) \), and also in the estimators \( \tilde{m}(x) \) and \( \tilde{\sigma}(x) \), we use the kernel function \( K(u) = (15/16)(1 - u^2)^2I(|u| \leq 1) \), and we work with the score function \( J(s) = b^{-1}I(0 \leq s \leq b) \). In order to estimate the functionals \( m(\cdot) \) and \( \sigma(\cdot) \) consistently the constant \( b \) has to be smaller than or equal to \( \inf_{x \in R_x} \tilde{F}(+\infty|x) \). Therefore, we choose \( b = 0.8 \) which is smaller than the average of 1000 simulated infima.

For each of the four estimators \( \hat{m}, \hat{\sigma}, \tilde{m} \) and \( \tilde{\sigma} \) we use a different bandwidth \( h_n \). In the first step, to select the bandwidth for \( \hat{m} \), we minimize the integrated mean squared error \( IMSE(\hat{m}) = \int_{0.2}^{0.8} E[(\hat{m}(x) - m(x))^2]dFX(x) \) over a grid of 15 equidistant possible values of \( h_n \) between 0.05 and 0.40 (note that we do not take into account values of \( x \) close to the boundary of the support of \( X \).
to avoid boundary effects of the Nadaraya-Watson weights). To calculate this $IMSE(\hat{m})$, we use 2000 simulated data sets. For each simulated data set, we compute the integrated squared error $ISE(\hat{m}) = \int_{0.2}^{0.8} [\hat{m}(x) - m(x)]^2 dF_X(x)$, and we approximate $IMSE(\hat{m})$ by taking the average over these 2000 values of $ISE(\hat{m})$. In the second step, to select the bandwidth for $\hat{\sigma}$ we first write

$$\sum_{i=1}^{n} \left[ Y_i - \hat{m}_1(X_i) \right]^2 \left[ \hat{F}_{1b}(Y_i|x) - \hat{F}_{1b}(Y_i^-|x) \right] J(\hat{F}_{2b}(Y_i|X_i)),$$

where $\hat{m}_1$ and $\hat{F}_{1b}$ are the estimators based on the optimal bandwidth chosen in the first step, and $\hat{F}_{2b}$ is based on a second bandwidth. Now, we select this second bandwidth by minimizing the empirical $IMSE(\hat{\sigma})$ (which is estimated in the same way as in the first step) over the grid 0.05, 0.1, 0.2, ..., 0.9. The bandwidths for the estimators $\hat{m}$ and $\hat{\sigma}$ are chosen in an analogous way.

Note that we prefer to select the bandwidth for $\hat{m}$, $\hat{\sigma}$, $\tilde{m}$ and $\tilde{\sigma}$ by directly minimizing the $IMSE$ of these estimators, instead of minimizing the $IMSE$ of the estimators $\hat{F}(\cdot|x)$ and $\hat{F}(\cdot|x)$. This ensures that we control the quality of these estimators instead of controlling the quality of the intermediate estimators $\hat{F}(\cdot|x)$ and $\hat{F}(\cdot|x)$, which are of secondary importance in our estimation procedure.

The practical performance of the above bandwidth selection procedure is illustrated in Figures 2.1 and 2.2. The figures show the $IMSE$ as a function of the bandwidth for each considered estimator and each sample size. We see that the $IMSE$ for $\hat{m}$ and $\tilde{m}$ have a convexly shaped curve and that the optimal bandwidth decreases with the sample size. We also see that the bandwidths for the estimators of $\sigma$ are much larger than those for the corresponding estimators of $m$, which is natural since $\sigma$ is constant in our model. This highlights the importance of choosing different bandwidths when estimating $m$ and $\sigma$. Finally, we notice that the $IMSE$-curves for $\hat{m}$ and $\hat{\sigma}$ are quite a bit higher than the corresponding curves for $\tilde{m}$ and $\tilde{\sigma}$, and also the minimal values of these curves differ quite a lot, which suggests that $\hat{m}$ and $\hat{\sigma}$ are, globally speaking, behaving better than $\tilde{m}$ and $\tilde{\sigma}$.

Figures 2.3 and 2.4 show the $MSE$ of the estimators $\hat{m}(x)$, $\hat{\sigma}(x)$, $\tilde{m}(x)$ and $\tilde{\sigma}(x)$ for the bandwidths that minimize the corresponding $IMSE$ of these four estimators. The bandwidths that minimize the $IMSE$ corresponding to $n = 100, 200$ and $400$ are $0.200, 0.150$ and $0.125$ for $\hat{m}$, and $0.250, 0.225$ and $0.175$ for $\tilde{m}$, respectively, while the optimal bandwidths for both $\hat{\sigma}$ and $\tilde{\sigma}$ are 0.9 independently of $n$. As we expected, the new estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ outperform the estimators $\tilde{m}(x)$ and $\tilde{\sigma}(x)$, since the latter are incorrectly assuming that $Y$ and $C$ are independent given $X = x$ for all $x$, whereas this is
Chapter 2. Estimation of location and scale functionals in nonparametric regression under copula dependent censoring

Figure 2.1: $\text{IMSE}(\hat{m})$ (thick lines) and $\text{IMSE}(\tilde{m})$ (thin lines) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

Figure 2.2: $\text{IMSE}(\hat{\sigma})$ (thick lines) and $\text{IMSE}(\tilde{\sigma})$ (thin lines) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

only true for $0.8 \leq x \leq 1$. We see how the difference between the estimators becomes larger as $x$ decreases, i.e. as the conditional dependence between $Y$ and $C$ given $X = x$ becomes stronger. The ratio of the two MSE’s can be as high as 10 in case of the location function $m$ and up to 3 in case of the scale
2.4. Simulations

Figure 2.3: MSE of $\hat{m}(x)$ (left panel) and $\tilde{m}(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

Figure 2.4: MSE of $\hat{\sigma}(x)$ (left panel) and $\tilde{\sigma}(x)$ (right panel) for samples of size $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 400$ (solid line).

function $\sigma$. Note that for $x = 0.8$, the two estimators are not equal in Figures 2.3 and 2.4, although $\hat{m}(x)$ and $\tilde{m}(x)$ reduce to $\tilde{m}(x)$ and $\tilde{\sigma}(x)$, respectively, when $Y$ and $C$ are conditionally independent. This is because the bandwidths used to calculate the estimators are not the same, and are in fact determined
by the behavior of the estimators for all values of the covariate.

Figure 2.5 shows the quantiles of order 0.05 and 0.95 of the distribution of \( \hat{m}(x) - m(x) \) and \( \tilde{m}(x) - m(x) \) respectively. The reason why we consider the ‘standardized’ quantity \( \hat{m}(x) - m(x) \) instead of the more natural quantity \( \hat{m}(x) \), is that in the non-standardized graph the scale of the vertical axis is much wider, and the curves are therefore very close together and hard to distinguish. We see that as we are moving to the area of stronger dependence (small values of \( x \)) the new estimator \( \hat{m}(x) \) continues to behave well, while the estimator \( \tilde{m}(x) \) becomes increasingly biased and also slightly more variable. The same can be said for the estimators of \( \sigma(x) \) shown in Figure 2.6 (which we do not standardize since the true \( \sigma \)-curve is already constant). Again, the figure shows that the new estimator is only slightly biased, whereas the estimator \( \tilde{\sigma}(x) \) shows consistent biased behavior and slightly larger variance.

Figure 2.5: Quantiles of order 0.05 and 0.95 of the distribution of \( \hat{m}(x) - m(x) \) (left panel) and \( \tilde{m}(x) - m(x) \) (right panel) for samples of size \( n = 100 \) (dotted line), \( n = 200 \) (dashed line) and \( n = 400 \) (solid line).
2.5 Example

In this section we will illustrate our estimation method via the analysis of the bone marrow transplantation data, which are described in Klein and Moeschberger (1997). The data are collected during a study in which 137 patients are followed in their recovery from acute leukemia after bone marrow transplantation. We are interested in the disease-free survival time $Y$, i.e. the time until a patient has a relapse of leukemia. However, patients can be censored by two possible events: disease-free death or disease-free and alive at the end of the study. The censoring time $C$ is the time until the first of these two events takes place. It seems natural to believe that the time until relapse $Y$ depends on the age $X$ of the patient at transplantation, and on the time until disease-free death (for a given age). This indirectly implies that $Y$ and $C$ are dependent for a given covariate $X$. In Figure 2.7, we show a scatter plot of age versus disease-free survival time, where we distinguish between non-censored (relapsed) and censored patients. We note that the censoring rate is as high as 69%, caused by the rather short length of the study (less than 8 years).

First, from these data, for every fixed covariate $x$ ranging between 20 and 40 (the area containing most of the data) we will estimate the average of the lower 30% of relapse times. Also we will estimate the standard deviation of the
lower 30% of relapse times. We have to restrict to 30%, because of the rather high proportion of censoring. In fact, for some values of \( x \), the estimator of the conditional distribution function of \( Y \) given \( x \) is only consistent up to the 0.3-th quantile. The score function corresponding to this location and scale functional is \( J(s) = \frac{3}{4} I_{[0,0.3]}(s) \) (see (2.1.2)). To estimate these functionals we use the biweight kernel function \( K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1) \).

Second, it is important to note that our estimation procedure assumes that the dependence structure between the survival time \( Y \) and the censoring time \( C \) is completely known. Therefore, in a real data application we have to select the dependence structure based on external information. For instance, we can model the dependence based on an expert’s opinion or estimate it from additional data. To capture the expert’s opinion on the strength and direction of the dependence between \( Y \) and \( C \) we will use Kendall’s tau, which is defined as \( \tau(x) = 1 + 4 \int_0^1 \left( \frac{\phi_x(t)}{\phi'_x(t)} \right) dt \) (Nelsen (2006)) and has a range from \(-1\) to \(1\). The association gets stronger as \( \tau(x) \) gets further away from zero, while the concordance/discordance is determined by the sign.
In Figure 2.8 we show the estimator $\hat{m}(x)$ for $h = 8$ and $h = 15$, while Figure 2.9 shows the estimator $\hat{\sigma}(x)$ for $h = 8$ and $h = 15$. Each of the plots contains estimators constructed for different choices of the copula generator $\phi_x(t)$: the generator that assumes conditional independence between $Y$ and $C$ ($\phi_x(t) = -\log(t)$ and $\tau(x) = 0$), the Fréchet-Hoeffding lower bound ($\phi_x(t) = 1 - t$ and $\tau(x) = -1$), which assumes that $Y$ and $C$ are discordant, and the generators from the Frank family ($\phi_x(t) = -\log(\exp\{-\theta t\} - 1) + \log(\exp\{-\theta\} - 1)$) with $\theta$ taking values corresponding to $\tau(x) = -0.5, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.5$ and 0.9.

From Figure 2.8 we can see that, for all choices of the copula and the bandwidth, the average of the lower 30% of relapse times is decreasing as age grows from 20 to 40. This means that for the 30% of worst recipients, older people have a shorter time to relapse than younger people. From Figure 2.9 we see that, again in all cases, the standard deviation of the lower 30% of relapse times decreases as age grows from 20 to 40. This indicates that there is more uncertainty in how younger recipients respond to transplantation.
Figure 2.9: Estimator of the standard deviation of the lower 30% of relapse times for several choices of $\phi_x(t)$, not depending on $x$: independence (solid curve), Fréchet-Hoeffding lower bound (long dashed curve) and Frank family for $\phi_x$ corresponding to $\tau(x) = -0.5, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.5$ and $0.9$ (dotted curves going from highest to lowest, respectively).

Next, the figures also show that for the Frank copula, the value of Kendall’s tau is not influencing the overall pattern of the functions $\hat{m}$ and $\hat{\sigma}$ as long as the dependence between $Y$ and $C$ does not change in an extreme way with $X$. Hence, the overall shape of the curves is relatively robust to the value of Kendall’s tau. On the other hand, the value of the curves at a specific point is heavily influenced by the value of Kendall’s tau, which illustrates the importance of having accurate external information regarding the strength and direction of the dependence between $Y$ and $C$.

In practice, it might not be an easy task to select the best copula family based on external information. Therefore, we want to investigate the robustness of the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$ to the choice of the copula family under the same value of Kendall’s tau. In Figures 2.10 and 2.11 we show again the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$, respectively, for $h = 8$ and $h = 15$ and several choices of $\phi_x(t)$ (not depending on $x$): the Clayton family ($\phi_x(t) = \frac{1}{\theta}(t^{-\theta} - 1)$), the Frank family and the Gumbel family ($\phi_x(t) = -(\log t)^{\theta}$), for $\theta$ corresponding to $\tau = 0.1, 0.5$ and $0.8$ (representing small, significant and very strong positive
dependence, respectively). In this investigation we restrict ourselves to positive values of \( \tau \), since the Gumbel family is only able to produce positive correlation. Note that the dependence structure between \( Y \) and \( C \) differs a lot from one copula family to another: the Clayton family gives lots of weight to the left lower corner (or left tail) of the unit square, the Gumbel family to the left and the right tail, and the Frank family is distributed more homogeneously along all points close to the bisector.

Both Figures 2.10 and 2.11 show that, for a fixed \( \tau \), the difference between the estimators caused by different choices of the copula family is not exceeding 25%. All figures strongly indicate the general tendency of the curves, independently of the choice of the copula family, as long as the dependence between \( Y \) and \( C \) does not change in an extreme way with \( X \). (Note that even though the copula generators \( \phi_x(t) \) used in the figures do not depend on \( x \), one can easily see how the estimator would behave if \( \phi_x \) would change with \( x \).) Hence, the figures show that for a fixed value of Kendall’s tau, our estimators are quite robust to misspecification of the copula family.
Figure 2.11: Estimator of the standard deviation of the lower 30\% of relapse times for several choices of $\phi_x(t)$, not depending on $x$: Clayton family (long dashed curves), Frank family (dotted curves) and Gumbel family (solid curves), for $\phi_x$ corresponding to $\tau(x) = 0.1, 0.5$ and $0.8$ (curves going from highest to lowest, respectively).

2.6 Appendix : Proofs

Proof of Proposition 2.3.1. Let $x$ be an arbitrary value in $R_X$ and let $y \leq \tilde{T}_x$. Throughout this proof we will use that any random process $\alpha_n(x, y)$ that lies between $\tilde{H}(y|x)$ and $\hat{H}(y|x)$ for $n$ large enough, can be a.s. bounded from below for $n$ large enough:

$$\alpha_n(x, y) > \gamma \text{ on } \{(x, y) : x \in R_X, y \leq \tilde{T}_x\}$$

(2.6.1)

for some $\gamma > 0$. This follows from the uniform consistency of $\hat{H}(y|x)$ (see Proposition 1 in Akritas and Van Keilegom (2001)) and from the definition of $\tilde{T}_x$.

Using similar calculations as in the proof of Theorem 1 in Brakkers and
Veraverbeke (2005), we can write
\[ \hat{F}(y|x) - F(y|x) = \left[ \phi_x^{-1}(\tilde{H}(T_i^-|x) - \tilde{H}(T_i^-|x) - W_m(x, h_n)) \right] \]
\[ + \phi_x^{-1}(\tilde{H}(T_i^-|x))W_m(x, h_n) \]
\[ - \left[ \phi_x^{-1}(\tilde{H}(s^-|x))d\tilde{H}^u(s|x) \right] \]
\[ - \phi_x^{-1}(\tilde{H}(s^-|x))dH^u(s|x) \]
\[ := Q_1(x, y) + Q_2(x, y). \]

Next, we will calculate the order of convergence of each of these terms. Starting with the second term, we apply the mean value theorem and obtain
\[
Q_2(x, y) = O(1) \left[ - \int_{-\infty}^{y} \phi_x^{-1}(\alpha_2(x, s)) \tilde{H}^u(s|x) \right],
\]
where \( \alpha_2(x, s) \) lies between the terms \(- \int_{-\infty}^{y} \phi_x^{-1}(\tilde{H}(s^-|x))d\tilde{H}^u(s|x) \) and \(- \int_{-\infty}^{y} \phi_x^{-1}(\tilde{H}(s|x))dH^u(s|x) \). From (A4)(iii) we know that the first factor on the right hand side of (2.6.2) is uniformly bounded. Now, by adding and subtracting terms, we further obtain
\[
Q_2(x, y) = \left[ Q_2^{(1)}(x, y) + Q_2^{(2)}(x, y) \right],
\]
where
\[
Q_2^{(1)}(x, y) = - \int_{-\infty}^{y} \phi_x^{-1}(\tilde{H}(s^-|x)) - \phi_x^{-1}(\tilde{H}(s|x)) d\tilde{H}^u(s|x),
\]
\[
Q_2^{(2)}(x, y) = - \int_{-\infty}^{y} \phi_x^{-1}(\tilde{H}(s|x)) d\tilde{H}^u(s|x) - \int_{-\infty}^{y} \phi_x^{-1}(\tilde{H}(s|x)) dH^u(s|x).
\]
Applying the mean value theorem on the term \( Q_2^{(1)}(x, y) \) we get:
\[
Q_2^{(1)}(x, y) = - \int_{-\infty}^{y} \phi_x^{-1}(\alpha_2(x, s)) \tilde{H}(s^-|x) \tilde{H}^u(s|x) d\tilde{H}^u(s|x),
\]
where \( \alpha_2(x, s) \) lies between \( \mathcal{H}(s^{-} | x) \) and \( \mathcal{H}(s | x) \). This gives us that

\[
\sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| Q_{2}^{(1)}(x, y) \right| \leq \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \phi_{x}''(\alpha_2(x, y)) \right| \times \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \hat{H}(y^{-} | x) - H(y | x) \right|. \tag{2.6.3}
\]

From Proposition 1 in Akritas and Van Keilegom (2001), we have that

\[
\sup_{x, y} \left| \hat{H}(y^{-} | x) - H(y | x) \right| = O\left((nh_{n})^{-1/2}(\log n)^{1/2}\right) \text{ a.s.}
\]

Hence, applying (2.6.1) on the first supremum at the right hand side of (2.6.3) yields that

\[
\sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| Q_{2}^{(1)}(x, y) \right| = O\left((nh_{n})^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \tag{2.6.4}
\]

For \( Q_{2}^{(2)}(x, y) \) we integrate by parts, and using similar calculations as for \( Q_{2}^{(1)}(x, y) \), we easily obtain that

\[
\sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| Q_{2}^{(2)}(x, y) \right| \
\leq \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \phi_{x}'(\hat{H}(y | x)) \right| \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \hat{H}^{u}(y | x) - H^{u}(y | x) \right| \
+ \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \phi_{x}'(\hat{H}(y | x)) \right| \sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| \hat{H}^{u}(y | x) - H^{u}(y | x) \right| \
= O\left((nh_{n})^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \tag{2.6.5}
\]

Now, combining results (2.6.4) and (2.6.5), yields

\[
\sup_{x \in R_{x}} \sup_{y \leq T_{x}} \left| Q_{2}(x, y) \right| = O\left((nh_{n})^{-1/2}(\log n)^{1/2}\right) \text{ a.s.} \tag{2.6.6}
\]

For \( Q_{1}(x, y) \), we repeatedly apply the mean value theorem to get

\[
Q_{1}(x, y) = \frac{-1}{\phi_{x}'(\phi_{x}^{-1}(\alpha_3(x, y)))} \times \left\{ - \sum_{T_{i} \leq y, \Delta_{i} = 1} \left[ \phi_{x}(\hat{H}(T_{i}^{-} | x)) - \phi_{x}(\hat{H}(T_{i}^{-} | x) - W_{n_{1}}(x, h_{n})) \right] \right. \\
+ \sum_{T_{i} \leq y, \Delta_{i} = 1} \phi_{x}'(\hat{H}(T_{i}^{-} | x)) W_{n_{1}}(x, h_{n}) \left\} \\
= \frac{1}{2\phi_{x}'(\phi_{x}^{-1}(\alpha_3(x, y)))} \sum_{T_{i} \leq y, \Delta_{i} = 1} \phi_{x}''(\alpha_4(x, i)) W_{n_{1}}^{2}(x, h_{n}).
\]
where $\alpha_3(x,y)$ lies between $-\sum_{T_i \leq y, \Delta_i=1} [\phi_x(\overline{H}(T_i^- | x)) - \phi_x(\overline{H}(T_i^- | x) - W_{ni}(x,h_n))]$ and $-\sum_{T_i \leq y, \Delta_i=1} \phi_x'(\overline{H}(T_i^- | x)) W_{ni}(x,h_n)$, and $\alpha_4(x,i)$ lies between $\overline{H}(T_i^- | x)$ and $\overline{H}(T_i^- | x) - W_{ni}(x,h_n)$. This leads to

$$\sup_{x \in \mathbb{R}^n} \sup_{y \leq \overline{T}_n} \left| Q_1(x,y) \right| \leq \frac{1}{2} \left[ \inf_{x \in \mathbb{R}^n} \inf_{y \leq \overline{T}_n} \left| \phi_x' \left( \phi_x^{-1}(\alpha_3(x,y)) \right) \right| \right]^{-1} \times \sup_{x \in \mathbb{R}^n, i=1 \ldots n} \left| \phi_x''(\alpha_4(x,i)) \right| \sum_{i=1}^n W_{ni}^2(x,h_n).$$

The infimum is bounded from below because of (A4)(iii) and the supremum is bounded thanks to (A4) and equation (2.6.1). On the other hand, the term $\sum_{i=1}^n W_{ni}^2(x,h_n)$ is of order $O((nh_n)^{-1})$ a.s., since we know from standard kernel smoothing theory that $\sup_x \max_{i=1 \ldots n} W_{ni}(x,h_n) = O((nh_n)^{-1})$ a.s. It now follows that

$$\sup_{x \in \mathbb{R}^n} \sup_{y \leq \overline{T}_n} \left| Q_1(x,y) \right| = O \left( (nh_n)^{-1/2} (\log n)^{1/2} \right) \text{ a.s.} \quad (2.6.7)$$

The proof is finished by combining (2.6.7) and (2.6.6).

**Proof of Proposition 2.3.2.** Using the notation $U(x,y) := \phi_x(F(y|x))$ and $U_n(x,y) := \phi_x(\overline{F}(y|x))$, we calculate

$$\frac{\partial}{\partial x} \left[ U(x,y) - U_n(x,y) \right] = -\phi_x'(\overline{F}(y|x)) \hat{F}(y|x) + \phi_x'(\overline{F}(y|x))$$

$$+ \phi_x'(\overline{F}(y|x)) \hat{F}(y|x) - \phi_x'(\overline{F}(y|x)) = \phi_x'(\overline{F}(y|x)) [\hat{F}(y|x) - \hat{F}(y|x)]$$

$$+ \left[ \phi_x'(\overline{F}(y|x)) - \phi_x'(F(y|x)) \right] \hat{F}(y|x)$$

$$+ \phi_x''(\alpha(x,y)) [\hat{F}(y|x) - F(y|x)],$$

where $\alpha(x,y)$ is between $\overline{F}(y|x)$ and $F(y|x)$. From here we can write

$$\hat{F}(y|x) - \hat{F}(y|x) = \frac{1}{\phi_x'(\overline{F}(y|x))} \left\{ - \left[ \phi_x'(\overline{F}(y|x)) - \phi_x'(F(y|x)) \right] \hat{F}(y|x)$$

$$- \phi_x''(\alpha(x,y)) [\hat{F}(y|x) - F(y|x)] + \frac{\partial}{\partial x} [U(x,y) - U_n(x,y)] \right\}.$$
Hence,

$$
[\hat{F}(y|x) - F(y|x)] \left[ 1 + \frac{\phi'_x(\hat{F}(y|x)) - \phi'_x(F(y|x))}{\phi'_x(F(y|x))} \right]
= \frac{1}{\phi'_x(F(y|x))} \left\{ - [\phi'_x(\hat{F}(y|x)) - \phi'_x(F(y|x))] \hat{F}(y|x)
- \phi'_x(\alpha(x,y))[\hat{F}(y|x) - F(y|x)] + \frac{\partial}{\partial x}[U(x,y) - U_n(x,y)] \right\}.
$$

(2.6.8)

Hence, it follows from Proposition 2.3.1 and the conditions on $\phi_x$ that the statement of the proposition follows provided we can show that

$$
\sup_{x \in R} \sup_{y \leq T_x} \left| \frac{\partial}{\partial x}[U_n(x,y) - U(x,y)] \right| = O((nh^3)^{-1/2}(\log n)^{1/2}) \ a.s.
$$

Consider

$$
\frac{\partial}{\partial x}[U(x,y) - U_n(x,y)]
= \frac{\partial}{\partial x} \left[ - \sum_{T_i \leq y, \Delta_i = 1} \left\{ \phi_x(\hat{H}(T_i^-|x)) - \phi_x(\hat{H}(T_i^-|x) - W_i(x,h_n)) \right\}
+ \sum_{T_i \leq y, \Delta_i = 1} \phi'_x(\hat{H}(T_i|x))W_i(x,h_n) \right]
- \frac{\partial}{\partial x} \left[ \int_{-\infty}^y \phi'_x(\hat{H}(s|x))d\hat{H}^w(s|x) - \int_{-\infty}^y \phi'_x(\hat{H}(s|x))d\hat{H}^u(s|x) \right]
:= Q_1(x,y) + Q_2(x,y).
$$

The second term $Q_2(x,y)$ can be further decomposed in the sum of three terms:

$$
Q_2^{(1)}(x,y) = - \frac{\partial}{\partial x} \int_{-\infty}^y \left[ \phi'_x(\hat{H}(s|x)) - \phi'_x(\hat{H}(s|x)) \right]d\hat{H}^w(s|x),
Q_2^{(2)}(x,y) = \frac{\partial}{\partial x} \int_{-\infty}^y \phi'_x(\hat{H}(s|x))d\left[ H^w(s|x) - \hat{H}^w(s|x) \right],
Q_2^{(3)}(x,y) = - \frac{\partial}{\partial x} \int_{-\infty}^y \left[ \phi'_x(\hat{H}(s|x)) - \phi'_x(\hat{H}(s|x)) \right]d\left[ \hat{H}^u(s|x) - H^u(s|x) \right].
$$

We will show that $Q_2^{(1)}(x,y)$ is of the desired order. The terms $Q_2^{(2)}(x,y)$ and $Q_2^{(3)}(x,y)$ can be dealt with in a similar way with additional use of integration.
by parts. Write

\[
Q_2^{(1)}(x, y) = - \int_{-\infty}^{y} \left[ \phi_x''(\hat{H}(s|x)) \hat{H}(s|x) + \phi_x'(\hat{H}(s|x)) \right] ds
- \phi_x''(\hat{H}(s|x)) \hat{H}(s|x) - \hat{\phi}_x'(\hat{H}(s|x)) \right] h^n(s|x) ds
- \int_{-\infty}^{y} \left[ \phi_x'(\hat{H}(s|x)) - \phi_x'(\hat{H}(s|x)) \right] \frac{\partial}{\partial x} h^n(s|x) ds
= - \int_{-\infty}^{y} \left\{ \phi_x''(\hat{H}(s|x)) \left[ \hat{H}(s|x) - \hat{H}(s|x) \right] \\
+ \phi_x''(\alpha_1(x, s)) \left[ \hat{H}(s|x) - \hat{H}(s|x) \right] \hat{H}(s|x) \\
+ \phi_x''(\alpha_2(x, s)) \left[ \hat{H}(s|x) - \hat{H}(s|x) \right] h^n(s|x) ds
- \int_{-\infty}^{y} \phi_x''(\alpha_3(x, s)) \left[ \hat{H}(s|x) - \hat{H}(s|x) \right] \frac{\partial}{\partial x} h^n(s|x) ds,
\]

where \( \alpha_j(x, s) \) is between \( \hat{H}(s|x) \) and \( \hat{H}(s|x) \) (\( j = 1, 2, 3 \)). From assumption (A4) we know that \( \sup_{x \in R_x} \sup_{y \in T_x} |\phi_x''(\hat{H}(y|x))| < \infty \) and similarly for \( \phi_x' \) and \( \phi_x'' \). Hence, it follows from Proposition 1 in Akritas and Van Keilegom (2001) that \( \sup_{x \in R_x} \sup_{y \in T_x} |Q_2^{(1)}(x, y)| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \) a.s.

To finish the proof it remains to show that \( Q_1(x, y) \) is of the desired order. We use the abbreviated notation \( W_{ni} = W_{ni}(x, h_n) \), \( \hat{H}_{ix} = \hat{H}(T_i|x) \), \( \hat{H}_{ix} = \hat{H}(T_i^-|x) \) and similarly for \( \hat{H}_{ix} \) and \( \hat{H}_{ix} \). Now, we can write

\[
Q_1(x, y) = - \sum_{T_i \leq y, \Delta_i = 1} \left\{ \phi_x''(\hat{H}_{ix}) + \phi_x'(\hat{H}_{ix}) \hat{H}_{ix} - \hat{\phi}_x(\hat{H}_{ix}) \right\}
+ \sum_{T_i \leq y, \Delta_i = 1} \left\{ \phi_x'(\hat{H}_{ix}) W_{ni} + \phi_x''(\hat{H}_{ix}) W_{ni}(x, h_n) + \phi_x'''(\hat{H}_{ix}) W_{ni} \hat{H}_{ix} \right\}.
\]
After the second order Taylor expansion we have

\[ Q_1(x, y) = -\sum_{T_i \leq y, \Delta_i=1} \left\{ \frac{1}{2} \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) + \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) + \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) \right\} \]

\[ + \sum_{T_i \leq y, \Delta_i=1} \left\{ \frac{1}{2} \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) + \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) + \phi_x''(\hat{\mu}_{ix}) W_{ni}(x, h_n) \right\} \]

\[ = -\sum_{T_i \leq y, \Delta_i=1} \left\{ \frac{1}{2} \phi_x''(\alpha_1(x, i)) W_{ni}^2(x, h_n) + \phi_x''(\alpha_3(x, i)) W_{ni}(x, h_n) W_{ni} \right\} \]

\[ + \sum_{T_i \leq y, \Delta_i=1} \left\{ \frac{1}{2} \phi_x''(\alpha_1(x, i)) W_{ni}^2(x, h_n) + \phi_x''(\alpha_3(x, i)) W_{ni}(x, h_n) W_{ni} \right\} \]

\[ = \frac{1}{2} \phi_x''(\alpha_2(x, i)) W_{ni}^2(x, h_n) \hat{\mu}_{ix}, \]

(2.6.10)

where \( \alpha_j(x, i) \) is between \( \hat{\mu}_{ix} \) and \( \hat{\mu}_{ix}^{-1} (j = 1, 2, 3) \). From Proposition 1 in Akritas and Van Keilegom (2001) it follows that \( \inf_{x \in R_x} \min_{T_i \leq \bar{T}_x} \alpha_j(x, i) > 0 \) for \( j = 1, 2, 3 \) and for \( n \) large enough. This together with assumption (A4) implies that

\[ \sup_{x \in R_x, y \leq \bar{T}_x} |Q_1(x, y)| = C_1 \sum_{i=1}^{n} W_{ni}^2(x, h_n) + C_2 \sum_{i=1}^{n} |W_{ni}(x, h_n)| W_{ni} \]

for some \( 0 < C_1, C_2 < \infty \). The first term is of the order \( O((nh_n)^{-1}) \) a.s. because \( \max_i \sup_{x \in R_x} W_{ni}(x, h_n) = O((nh_n)^{-1}) \) a.s., whereas the second term is \( O((nh_n^3)^{-1}) = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \) a.s., since \( \max_i \sup_{x \in R_x} |W_{ni}(x, h_n)| = O((nh_n^3)^{-1}) \) a.s. \( \square \)

**Proof of Proposition 2.3.3.** When \( |x - x'| \geq h_n \), then Proposition 2.3.3 is trivially true. In what follows we consider the case when \( |x - x'| < h_n \). By using decomposition (2.6.8) we can easily see that

\[ \left| \hat{F}(y|x) - \hat{F}(y|x') \right| \left| x - x' \right|^{-\delta} \]

\[ \leq \left| Q_1(x, y) - Q_1(x', y) \right| \left| x - x' \right|^{-\delta} + \left| Q_2(x, y) - Q_2(x', y) \right| \left| x - x' \right|^{-\delta} \]  

(2.6.11)

\[ + O((nh_n^{1+2\delta})^{-1/2}(\log n)^{1/2}) \text{ a.s.}, \]

where \( Q_1(x, y) \) and \( Q_2(x, y) \) are defined in the proof of the previous proposition. For the first term on the right hand side of (2.6.11) we use (2.6.10), which leads
to
\[
Q_1(x, y) - Q_1(x', y) = - \sum_{T_i \leq y - \Delta_i = 1} \left\{ \frac{1}{2} \phi_x''(\alpha_1(x, i)) W_{n_i}^2(x, h_n) - \frac{1}{2} \phi_x''(\alpha_1(x', i)) W_{n_i}^2(x', h_n) + \phi_x''(x_3(i, x)) W_{n_i}^2(x, h_n) - \phi_x''(x_3(i, x')) W_{n_i}^2(x', h_n) + \frac{1}{2} \phi_x''(x_2(i, x)) W_{n_i}^2(x, h_n) - \frac{1}{2} \phi_x''(x_2(i, x')) W_{n_i}^2(x', h_n) \right\}. \tag{2.6.12}
\]

By adding and subtracting terms we can see that the sum of the first two terms of (2.6.12) equals
\[
- \frac{1}{2} \sum_{T_i \leq y - \Delta_i = 1} \left[ \phi_x''(\alpha_1(x, i)) - \phi_x''(\alpha_1(x', i)) \right] W_{n_i}^2(x, h_n)
- \frac{1}{2} \sum_{T_i \leq y - \Delta_i = 1} \phi_x''(\alpha_1(x', i)) [W_{n_i}^2(x, h_n) - W_{n_i}^2(x', h_n)],
\]
which under assumption (A4) when multiplied with \(|x - x'|^{-\delta}\) is of the order \(O((nh_n^{1+\delta})^{-1})\) a.s. In a similar way we can show the order of the other terms of (2.6.12).

By straightforward algebra and assumption (A4) the second term on the right hand side of (2.6.11) can be written in terms of \(|\hat{H}(y|x) - \tilde{H}(y|x) - \hat{H}(y|x') + \tilde{H}(y|x')| |x - x'|^{-\delta}\), which by Lemma 4.2 in Van Keilegom and Akritas (1999) is of the desired order.

\[\square\]

**Lemma 2.6.1.** Under the assumptions of Proposition 2.3.4, we have
\[
\sup_{x \in \mathcal{X}} \sup_{y \leq T_x} \left| \int_{-\infty}^{y} \left[ \phi_x'(H(t|x)) - \phi_x'(T(t|x)) \right] d\left[ H^u(t|x) - H^u(t|x) \right] \right| = O((nh_n)^{-3/4}(\log n)^{3/4}) \quad \text{a.s.}
\]

**Proof.** First note that it follows from Lemma 4.4 in Du and Akritas (2002) that
\[
\sup_{x \in \mathcal{X}} \sup_{|M(y_2) - M(y_1)| \leq a_n} \left| \frac{\hat{H}(y_2|x) - \hat{H}(y_1|x) - \tilde{H}(y_1|x) + \tilde{H}(y_1|x)}{\tilde{H}(y_1|x)} \right| = O((nh_n)^{-3/4}(\log n)^{3/4}) \quad \text{a.s.}, \quad \text{(2.6.13)}
\]
where \(a_n = O((nh_n)^{-1/2}(\log n)^{1/2})\) and \(M = M_1 + M_3\), and where the functions \(M_1\) and \(M_3\) come from assumption (A6). Equality (2.6.13) also holds for
Chapter 2. Estimation of location and scale functionals in nonparametric regression

under copula dependent censoring

$H^u$ with the function $M = M_2 + M_4$. Note that the result in Du and Akritas (2002) is in fact slightly less general, since they impose that $y_1$ and $y_2$ are bounded by some finite $T$, but it is easy to see that this is by no means necessary. Further note that after redefining the function $M$ as $M := M_1 + \ldots + M_4$, equality (2.6.13) holds for both $H$ and $H^u$.

Now, because of assumption (A6) we can partition $\mathbb{R}$ into $m = Ca_n^{-1}$ subintervals $[t_i, t_{i+1}]$ such that $|M(t_{i+1}) - M(t_i)| \leq a_n$. Consequently, we have

$$\sup_{x \in R_x} \sup_{y \leq \bar{T}_x} \left| \int_{-\infty}^y \left[ \phi'_x(\widehat{H}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d[H^u(t|x) - H^u(t|x)] \right|$$

$$\leq \sup_{x \in R_x} \sup_{y \leq \bar{T}_x} \left[ \sum_{i=2}^{k(y)} \int_{t_{i-1}}^{t_i} \left[ \phi'_x(\widehat{H}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d[H^u(t|x) - H^u(t|x)] \right]$$

$$+ \int_{t_{k(y)}}^{y} \left[ \phi'_x(\widehat{H}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d[H^u(t|x) - H^u(t|x)]$$

$$:= \sup_{x \in R_x} \sup_{y \leq \bar{T}_x} \left| \alpha_{n1}(x, y) + \alpha_{n2}(x, y) \right|,$$

where $t_{k(y)} \leq y < t_{k(y)+1}$. The term $\alpha_{n2}(x, y)$ can be treated in the same way as $\alpha_{n1}(x, y)$, so we will restrict ourselves to showing that $\sup_{x \in R_x} \sup_{y \leq \bar{T}_x} |\alpha_{n1}(x, y)| = O((nb_n)^{-3/4}(\log n)^{3/4})$ a.s. Write

$$\sup_{x \in R_x} \sup_{y \leq \bar{T}_x} |\alpha_{n1}(x, y)|$$

$$\leq \sup_{x \in R_x} \sum_{i=2}^{m(x)} \left| \int_{t_{i-1}}^{t_i} \left[ \phi'_x(\widehat{H}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] \cdot \left[ \phi'_x(\widehat{H}(t_{i+1}|x)) - \phi'_x(\overline{H}(t_{i+1}|x)) \right] \right|$$

$$+ \sup_{x \in R_x} \sum_{i=2}^{m(x)} \left[ \phi'_x(\widehat{H}(t_{i+1}|x)) - \phi'_x(\overline{H}(t_{i+1}|x)) \right] \int_{t_{i-1}}^{t_i} \left[ \phi'_x(\overline{H}(t|x)) - \phi'_x(\overline{H}(t|x)) \right] d[H^u(t|x) - H^u(t|x)],$$

where $m(x)$ is such that $t_{m(x)} \leq \bar{T}_x < t_{m(x)+1}$. Using the notation $J_{n_0}^x = \ldots$
where \( \xi \)

After applying a Taylor expansion we can bound the second term on the right hand side:

\[
\begin{align*}
|\alpha_{n2}| & \leq C_{\alpha_n}^{-1} \sup_{x \in R} \sup_{t \leq T} \phi_x''(\xi_{t,x}) \sup_{x \in R} \sup_{t \in T} \left| \hat{H}(t|x) - H(t|x) \right| \\
& \times \sup_{x \in R} \sup_{|M(y_2) - M(y_1)| \leq a_n} \left| \hat{H}''(t_1|x) - H''(t_1|x) - \hat{H}''(t_2|x) + H''(t_2|x) \right|,
\end{align*}
\]

where \( \xi_{t,x} \) is between \( H(t|x) \) and \( \hat{H}(t|x) \). By Proposition 1 in Akritas and Van Keilegom (2001) we know that for \( n \) large enough there is \( \gamma > 0 \) such that \( \inf_{x \in R} \sup_{t \leq T} |\xi_{t,x}| > \gamma \) a.s., from where we have that \( \sup_{x \in R} \sup_{t \leq T} \phi_x''(\xi_{t,x}) \) is a.s. finite. Now, by using Proposition 1 in Akritas and Van Keilegom (2001) and (2.6.13) to bound the second and third factor respectively, we get that \( \alpha_{n2} \) is of the desired order \( O((mh_n)^{-3/4}(\log n)^{3/4}) \) a.s. By again applying a Taylor expansion, we can bound \( \alpha_{n1} \):

\[
|\alpha_{n1}| \leq 2 \sup_{x \in R} \sup_{(t_1,t_2) \in J_{\alpha_n}} \left| \phi_x''(\hat{H}(t_1|x)) \right| \left| \hat{H}(t_1|x) - H(t_1|x) \right| \\
\cdot \left( H''(t_1|x) - \hat{H}(t_1|x) \right) \\
\cdot \sup_{x \in R} \left| \phi_x''(\hat{H}(t_2|x)) \right| \left| \hat{H}(t_2|x) - H(t_2|x) \right| \\
\cdot \left( H''(t_2|x) - \hat{H}(t_2|x) \right),
\]

where \( \xi_{t,x} \) is between \( H(t|x) \) and \( \hat{H}(t|x) \). Following the same argument as for \( \alpha_{n2} \), we have that \( \sup_{x \in R} \sup_{t \leq T} |\phi_x''(\xi_{t,x})| \) is a.s. finite. Now, we can conclude that the second term on the right hand side of (2.6.14) is of the order \( O((mh_n)^{-1} \log n) \) a.s. By Proposition 1 in Akritas and Van Keilegom (2001).
For the first term on the right hand side, we add and subtract terms followed by a Taylor expansion, and obtain the following bound:

\[ 2 \sup_{x \in \mathcal{X}} \sup_{(t_1, t_2) \in J^n} \left| \phi''(\mathcal{H}(t_1|x)) \left[ \mathcal{H}(t_1|x) - \mathcal{H}(t_1|x) - \mathcal{H}(t_2|x) + \mathcal{H}(t_2|x) \right] \right| + 2 \sup_{x \in \mathcal{X}} \sup_{(t_1, t_2) \in J^n} \left| \phi^{(3)}(\xi_{t_1, t_2,x}) \left[ \mathcal{H}(t_2|x) - \mathcal{H}(t_1|x) \right] \right|, \]

where \( \xi_{t_1, t_2,x} \) is between \( \mathcal{H}(t_1|x) \) and \( \mathcal{H}(t_2|x) \). Following an analogous reasoning as before we conclude that

\[ \sup_{x \in \mathcal{X}, t \leq \tilde{T}_x} |\phi''(\mathcal{H}(t|x))| \quad \text{and} \quad \sup_{x \in \mathcal{X}} \sup_{(t_1, t_2) \in J^n} |\phi^{(3)}(\xi_{t_1, t_2,x})| \]

are a.s. finite. The second term above is of the order \( O((nh_n)^{-1} \log n) \) a.s. by the definition of \( J^n \) and Proposition 1 in Akritas and Van Keilegom (2001). The first term is of the desired order \( O((nh_n)^{-3/4}(\log n)^{3/4}) \) a.s. thanks to (2.6.13).

**Proof of Proposition 2.3.4.** Using a similar decomposition as in the proof of Theorem 1 in Braekers and Veraverbeke (2005), we can write

\[ \hat{F}(y|x) - F(y|x) = \sum_{i=1}^{n} W_{ni}(x, h_n)g(T_i, \Delta_i, y|x) + R_n1(y|x) + R_n2(y|x) + R_n3(y|x) + R_n4(y|x), \]

where we refer to the aforementioned paper for the precise definition of the remainder terms \( R_{nj}(y|x) \) (\( j = 1, 2, 3, 4 \)). We can easily show that the terms \( R_n1(y|x), R_n2(y|x) \) and \( R_n3(y|x) \) are uniformly of the order \( O((nh_n)^{-1} \log n) \) a.s. by following the same arguments as in the proof of Theorem 1 in Braekers and Veraverbeke (2005). In fact, the most important difference with their proof is that we use Proposition 1 in Akritas and Van Keilegom (2001) (which gives the rate of convergence of \( \hat{H}(y|x) - H(y|x) \) uniformly in \( x \) and \( y \)) instead of Lemma A.4 in Van Keilegom and Veraverbeke (1997b) (which gives the same result but for Gasser-Müller weights and for a fixed value of \( x \)). The order of the remainder term \( R_{na}(y|x) \) is given by \( O((nh_n)^{-3/4}(\log n)^{3/4}) \) a.s. uniformly in \( x \) and \( y \leq \tilde{T}_x \) by Lemma 2.6.1.

This together with the uniform rate of convergence of \( \hat{F}(y|x) - F(y|x) \) given in Proposition 2.3.1, entails that

\[ \sup_{x \in \mathcal{X}} \sup_{y \leq \tilde{T}_x} \left| \sum_{i=1}^{n} W_{ni}(x, h_n)g(T_i, \Delta_i, y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \text{ a.s.}, \]
and hence
\[
\hat{F}(y|x) - F(y|x) = \frac{1}{nh_n f_X(x)} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) y(T_i, \Delta, y|x) + O((nh_n)^{-3/4}(\log n)^{3/4}),
\]

since \( \sup_n |f_X(x) - f_X(x)| = O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s. This finishes the proof.

\[\square\]

**Proof of Proposition 2.3.5.** For every \( x \in R_X \) we can write
\[
\hat{m}(x) - m(x)
= \int_{-\infty}^{1} \int_{-\infty}^{+\infty} I(\hat{F}^{-1}(s|x) \leq t \leq \hat{F}^{-1}(s|x)) I(\hat{F}^{-1}(s|x) > F^{-1}(s|x)) dt \, J(s) \, ds
\]
\[
- \int_{-\infty}^{1} \int_{-\infty}^{+\infty} I(\hat{F}^{-1}(s|x) \leq t \leq F^{-1}(s|x)) I(\hat{F}^{-1}(s|x) \leq F^{-1}(s|x)) dt \, J(s) \, ds
\]
\[
= \int_{-\infty}^{+\infty} \int_{F(t|x)}^{\hat{F}(t|x)} I(\hat{F}^{-1}(s|x) > F^{-1}(s|x)) J(s) \, ds \, dt
\]
\[
- \int_{-\infty}^{+\infty} \int_{F(t|x)}^{\hat{F}(t|x)} I(\hat{F}^{-1}(s|x) \leq F^{-1}(s|x)) J(s) \, ds \, dt
\]
\[
= \int_{-\infty}^{+\infty} \left[ L(F(t|x)) - L(\hat{F}(t|x)) \right] dt,
\]

where \( L(u) = \int_{0}^{u} J(s) ds \) for all \( 0 \leq u \leq 1 \). Using the substitution \( t = F^{-1}(s|x) \) we get
\[
\hat{m}(x) - m(x) = \int_{0}^{1} \left[ L(s) - L(\hat{F}(F^{-1}(s|x)|x)) \right] dF^{-1}(s|x)
\]
\[
= \int_{0}^{1} \left[ L(s) - L(\hat{F}(F^{-1}(s|x)|x)) \right] \frac{1}{f(F^{-1}(s|x)|x)} ds.
\]

Finally, by using a Taylor expansion, it follows that
\[
\sup_{x \in R_X} |\hat{m}(x) - m(x)| \leq \left[ \inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) \right]^{-1} \sup_{s_0 \leq s \leq s_1} \int_{s_0}^{s} \left( \sup_{x \in R_X} \sup_{F^{-1}(s_0|x) \leq y \leq F^{-1}(s_1|x)} |\hat{F}(y|x) - F(y|x)| \right) ds.
\]
Chapter 2. Estimation of location and scale functionals in nonparametric regression
under copula dependent censoring

Note that for all $x$, \([F^{-1}(s_0|x), F^{-1}(s_1|x)] \subset (-\infty, \tilde{T}_x]\), since $s_1 \leq \inf_x F(T_x|x)$ by assumption (A2). The result now follows by using again assumption (A2) together with Proposition 2.3.1. Part (b) can be shown in a similar way.

Proof of Proposition 2.3.6. The proof is very analogous to the proof of Proposition 4.6 in Van Keilegom and Akritas (1999). The only difference is that we use our Proposition 2.3.2 for the uniform rate of convergence of $\hat{F}(y|x) - \bar{F}(y|x)$, whereas they use the second statement of their Proposition 4.3.

Proof of Proposition 2.3.7. We follow exactly the same steps as in the proof of Proposition 4.7 in Van Keilegom and Akritas (1999), with the only exception that instead of using their Proposition 4.4, we use our Proposition 2.3.3.

Proof of Proposition 2.3.8. Using the notation $L(u) = \int_0^u J(s)ds$ we write

$$
\hat{m}(x) - m(x) = -\int_{-\infty}^{+\infty} \left[ L(\tilde{F}(y|x)) - L(F(y|x)) \right] dy \\
= -\int_{-\infty}^{+\infty} J(F(y|x)) \left[ \tilde{F}(y|x) - F(y|x) \right] dy \\
+ \frac{1}{2} \int_{-\infty}^{+\infty} J'(\beta(x,y)) \left[ \tilde{F}(y|x) - F(y|x) \right]^2 dy,
$$

with $\beta(x,y)$ between $\tilde{F}(y|x)$ and $F(y|x)$. By Proposition 2.3.1, the second term above is $O\left((nh_n)^{-1} \log n\right)$ a.s. Using Proposition 2.3.4, the first term can be written as

$$
-\frac{1}{nh_n f_X(x)} \sum_{i=1}^n K \left( \frac{x - X_i}{h_n} \right) \int_{-\infty}^{+\infty} J(F(y|x)) g(T_i, \Delta, y|x) dy \\
+ \int_{-\infty}^{+\infty} J(F(y|x)) R_n(y|x) dy \text{ a.s.}
$$

The rate of convergence of the last term above is $O\left((nh_n)^{-3/4} (\log n)^{3/4}\right)$ a.s., which completes the proof for part (a).

To prove the second assertion in Proposition 2.3.8, we mimic Van Keilegom and Akritas (1999) and write

$$
\hat{\sigma}(x) - \sigma(x) = \frac{\hat{\sigma}^2(x) - \sigma^2(x)}{2\sigma(x)} + \frac{\left[ \hat{\sigma}(x) - \sigma(x) \right]^2}{2\sigma(x)}.
$$
It follows from Proposition 2.3.5 that the second term above is $O\left((nh_n)^{-1} \log n\right)$ a.s. uniformly in $x \in R_X$. For the first term, we have similarly as in the proof of part (a),

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

$$= -2 \int_{-\infty}^{+\infty} \left[ L(\hat{F}(y|x)) - L(F(y|x)) \right] y \, dy - \left[ \hat{m}^2(x) - m^2(x) \right].$$

(2.6.15)

Now, by a Taylor expansion we have

$$\int_{-\infty}^{+\infty} \left[ L(\hat{F}(y|x)) - L(F(y|x)) \right] y \, dy$$

$$= \int_{-\infty}^{+\infty} J(y) \left[ \hat{F}(y|x) - F(y|x) \right] y \, dy$$

$$+ \frac{1}{2} \int_{-\infty}^{+\infty} J(x,y) \left[ L(\hat{F}(y|x)) - L(F(y|x)) \right]^2 \, y \, dy$$

$$= -\frac{1}{nh_n f(x)} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} J(F(y|x)) K\left(\frac{x - X_i}{h_n}\right) g(T_i, \Delta_i, y|x) \, dy$$

$$+ O\left((nh_n)^{-3/4} (\log n)^{3/4}\right) \text{ a.s.},$$

(2.6.16)

where $\gamma(x,y)$ is between $F(y|x)$ and $\hat{F}(y|x)$. In the above, the second equality follows from Proposition 2.3.1 and by using the same reasoning as at the end of part (a) of this proof. Next, we write

$$\hat{m}^2(x) - m^2(x)$$

$$= 2m(x) \left[ \hat{m}(x) - m(x) \right] + \left[ \hat{m}(x) - m(x) \right]^2$$

$$= -\frac{2m(x)}{nh_n f(x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \int_{-\infty}^{+\infty} J(F(y|x)) g(T_i, \Delta_i, y|x) \, dy$$

$$+ O\left((nh_n)^{-3/4} (\log n)^{3/4}\right) \text{ a.s.},$$

(2.6.17)

with the last equality following from part (a) and from Proposition 2.3.5(a). Combining (2.6.15), (2.6.16) and (2.6.17), we get the result.
Chapter 3

The copula-graphic estimator in censored nonparametric location-scale regression models

**ABSTRACT:** A common assumption when working with randomly right censored data, is the independence between the variable of interest $Y$ (the survival time) and the censoring variable $C$. This assumption, which is not testable, is however unrealistic in certain situations. In this chapter we assume that for a given covariate $X$, the dependence between the variables $Y$ and $C$ is described via a known copula. Additionally we assume that $Y$ is the response variable of a heteroscedastic regression model $Y = m(X) + \sigma(X)\epsilon$, where the error term $\epsilon$ is independent of the explanatory variable $X$, and the functions $m$ and $\sigma$ are ‘smooth’. We propose an estimator of the conditional distribution of $Y$ given $X$ under this model, and show the asymptotic normality of this estimator. We also study the small sample performance of the estimator, and discuss the advantages/drawbacks of this estimator with respect to competing estimators.

**Keywords and phrases:** Asymptotic normality, asymptotic representation, copula, dependent censoring, kernel estimator, nonparametric regression, right censoring.  

---

1This chapter is a slightly edited version of Sujica and Van Keilegom (2014).
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

3.1 Introduction

Consider the following nonparametric location-scale model

\[ Y = m(X) + \sigma(X) \varepsilon, \]  

(3.1.1)

where the error \( \varepsilon \) is assumed to be independent of a one dimensional covariate \( X \). The function \( m(\cdot) \) is a conditional location functional and \( \sigma(\cdot) \) is a conditional scale functional representing possible heteroscedasticity. We assume that \( Y \) is a possible (given) transformation of a survival time and is subject to random right censoring, i.e. instead of observing \( Y \) we only observe \( (T, \Delta) \), where \( T = \min(Y, C) \), \( \Delta = I(Y \leq C) \) and \( C \) represents the censoring time. Let \((T_i, X_i, \Delta_i), i = 1, \ldots, n\) be \( n \) independent vectors having the same distribution as \((T, X, \Delta)\).

The motivation for considering model (3.1.1) comes from the fact that the model offers important advantages with respect to the completely nonparametric model when one is interested in the estimation of the conditional distribution \( F(\cdot | x) = P(Y \leq \cdot | X = x) \) of \( Y \) given \( X = x \). Van Keilegom and Akritas (1999) showed how advantage can be taken from model (3.1.1) to estimate this conditional distribution. The advantages are especially apparent in the right tail of the distribution. In this region the completely nonparametric competitor proposed by Beran (1981) (see also Dabrowska (1989), González-Manteiga and Cadarso-Suárez (1994), Akritas (1994), Van Keilegom and Veraverbeke (1997a), Du and Akritas (2002), among others) suffers from inconsistency problems especially when censoring is heavy. This phenomenon is similar to what happens in the right tail of the Kaplan-Meier estimator in the absence of covariates. Under model (3.1.1), Van Keilegom and Akritas (1999) showed that the right tail of the distribution \( F(\cdot | x) \) can be well estimated for all values of \( X \), provided there is a region of \( X \) where censoring is light. This is because under model (3.1.1) the conditional distribution \( F(y|x) \) can be written as

\[ F(y|x) = F_\varepsilon\left(\frac{y - m(x)}{\sigma(x)}\right), \]  

(3.1.2)

where \( F_\varepsilon(\cdot) \) is the distribution of the error variable \( \varepsilon \), and this error distribution is the same for all \( x \).

Aside Van Keilegom and Akritas (1999) the nonparametric estimation of the above location-scale model with censored data has also been studied in other papers. See e.g. Lewbel and Linton (2002), Van Keilegom and Veraverbeke (2002), Chen, Dahl, and Khan (2005), Linton, Mammen, Nielsen, and Van Keilegom (2011) and Lambert (2013). Instead of studying the estimation of the model,
other authors have investigated testing procedures for several aspects of the model (again in the case of censored data). We refer to Pardo-Fernández and Van Keilegom (2006) and Dette and Heuchenne (2012) for goodness-of-fit tests for the location respectively scale function, and to Pardo-Fernández, Van Keilegom, and González-Manteiga (2007) for comparing regression curves under this model. Finally, semiparametric location scale models with censored data have been studied in Heuchenne and Van Keilegom (2007a,b) among others.

Whereas all the above papers restrict attention to the case where $Y$ and $C$ are independent given $X$, we will go one step further in this chapter, and consider the case where for a given value of $X$, the survival time and censoring time are related. The motivation for considering this situation is multifold. In many situations, the latent censoring mechanism is not of pure administrative or random nature, but is at the contrary linked (in a weak or strong way) to the survival time. This is e.g. the case when the medical condition of a patient (good or bad) makes him/her decide to leave the study or to change treatment. It is also the case in a study on the duration of unemployment, where a person might decide after long and unsuccessful attempts to find a job, to move to another region where the job market is more attractive, and hence this person will be lost to follow up. In addition, in many situations the strength of the dependence between $Y$ and $C$ will depend on the value of the covariate(s). Therefore, in this chapter we will allow the dependence between $Y$ and $C$ to depend on the value of $X$. We will model this dependence by means of a copula function, because copulas have the attractive feature to model the dependence structure without affecting the margins (see Sklar (1959)).

In the absence of covariates and leaving the marginal distributions of $Y$ and $C$ completely unspecified, Zheng and Klein (1995) and Rivest and Wells (2001) supposed that the dependence structure between $Y$ and $C$ is known and is described by a known copula, and they showed that the marginal distributions of $Y$ and $C$ are identifiable under very weak conditions. They developed an estimator of the distribution of $Y$, which they called the copula-graphic estimator, and which reduces to the Kaplan-Meier estimator when $Y$ and $C$ are independent. In the presence of covariates, Braekers and Veraverbeke (2005) extended the work of Rivest and Wells (2001) to the case of a fixed design regression model, and they proposed and studied an estimator of the conditional distribution $F(\cdot|X)$ for a given covariate $x$ without assuming any model restriction on $F(\cdot|x)$. Their estimator generalizes the Beran (1981) estimator, in the sense that it reduces to Beran’s estimator when the independence copula is chosen. In the case of a random design, Sujica and Van Keilegom (2013) (and Chapter 2) built further on the work of Braekers and Veraverbeke (2005). They proposed estimators of a location and scale function of $Y$ given $X$ and
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

studied their asymptotic properties.

In this chapter we will propose and study an estimator of the conditional distribution \( F(\cdot | X) \) assuming that \( Y \) and \( C \) are copula dependent given \( X \) (as in Braekers and Veraverbeke (2005)), and that \((X, Y)\) satisfy the nonparametric location-scale model (3.1.1). We will do this by first estimating the marginal error distribution \( F_e(\cdot) \) taking the dependence between \( Y \) and \( C \) into account. Next, the conditional distribution \( F(\cdot | x) \) will be estimated via relation (3.1.2), by plugging-in the obtained estimator of \( F_e \) and the estimators of \( m(\cdot) \) and \( \sigma(\cdot) \) studied by Sujica and Van Keilegom (2013) (and Chapter 2).

Before continuing we give an overview of the chapter. In the next section we introduce the precise definitions of the estimators of the error distribution and of the conditional distribution of \( Y \) given \( X \), and we state the assumptions needed for our asymptotic results. In Section 3.3 we give the main results. Proofs of the main results are in Section 3.4. Section 3.5 contains results of a small sample comparison between the suggested estimator and the estimators of Van Keilegom and Akritas (1999) and Braekers and Veraverbeke (2005). Additional lemmas needed for the main results are in Appendices A and B.

### 3.2 Estimation method

We start this section with a number of definitions. Let the random vector \((T, X, \Delta)\) be as defined in Section 3.1 and denote \( F(y|x) = P(Y \leq y | X = x) \), \( G(y|x) = P(C \leq y | X = x) \), \( H(y|x) = P(T \leq y | X = x) \), \( H_u(y|x) = P(T \leq y, \Delta = 1 | X = x) \) and \( F_X(x) = P(X \leq x) \). Further, denote \( F_e(y) = P(\varepsilon \leq y) = P(Y - m(X) \leq y) \), \( G_e(y|x) = P(C - m(X) \leq y | X = x) \), and for \( E = (T - m(X))/\sigma(X) \) denote \( H_e(y) = P(E \leq y) \), \( H^u_e(y) = P(E \leq y, \Delta = 1) \), \( H_e(y|x) = P(E \leq y | X = x) \) and \( H^u_e(y|x) = P(E \leq y, \Delta = 1 | X = x) \). The probability density functions of the distribution functions defined above will be denoted by the corresponding lower case letters, and for any distribution function \( F(\cdot) \), we denote the corresponding survival function by \( F(\cdot) = 1 - F(\cdot) \).

As explained in Section 3.1, we build further on the work of Zheng and Klein (1995), Rivest and Wells (2001), and Braekers and Veraverbeke (2005), and model the conditional dependence between \( Y \) and \( C \) via a known copula \( C_\alpha \) that is allowed to depend on the value of \( X \):

\[
P(Y > y, C > c | X = x) = C_\alpha \left( 1 - F(y|x), 1 - G(c|x) \right).
\]

Moreover, we will assume that the copula belongs to the family of Archimedean copulas, which have attractive properties and which cover a broad range of
3.2. Estimation method

different copula structures. This means that we suppose that
\[
P(Y > y, C > c \mid X = x) = \phi_x^{-1} \left[ \phi_x \left\{ 1 - F(y \mid x) \right\} + \phi_x \left\{ 1 - G(c \mid x) \right\} \right],
\] (3.2.1)
for an Archimedean copula generator \( \phi_x \), i.e. a function from \( [0,1] \) to \( \mathbb{R}^+ \) that is decreasing, convex and that satisfies \( \phi_x(1) = 0 \).

In order to construct an estimator of the conditional distribution \( F(y \mid x) \) given in (3.1.2), we start with focusing on the error distribution \( F_e(y) \). The assumption of an Archimedean copula allows to write \( F_e(y) \) in the following way:
\[
T_e(y) = \phi_{(y)}^{-1} \left\{ - \int_{B_y} \int_{-\infty}^y \phi_x' \left( \Pi_x(s \mid x) \right) dH_v(s \mid x) dF_X(x) \right\}
\] (3.2.2)
(see Lemma 3.6.1 in Appendix A), where \( \phi_{(y)}(u) = \int_{B_y} \phi_x(u) dF_X(x) \). Statement (3.2.2) holds for every nonempty set \( B_y \subset A_y := \{ x : \Pi_x(y \mid x) > \eta \} \) (\( \eta > 0 \)), which will be defined later. In order to derive an estimator of \( F_e(\cdot) \), we will replace the distribution functions \( H_v, H_v^n \) and \( F_X \) in (3.2.2) by corresponding estimators \( \hat{H}_v, \hat{H}_v^n \) and \( \hat{F}_X \).

We start with \( F_X \), which we estimate by the empirical distribution function \( \hat{F}_X(\cdot) = n^{-1} \sum_{i=1}^n I(X_i \leq \cdot) \). Next, to estimate \( H_v \) and \( H_v^n \), we first need to find appropriate estimators of the functions \( m \) and \( \sigma \), for which we use the following definitions:
\[
m(x) = \int_0^1 F^{-1}(s \mid x) J(s) ds \quad \text{and} \quad \sigma^2(x) = \int_0^1 F^{-1}(s \mid x)^2 J(s) ds - m(x)^2,
\] (3.2.3)
where \( F^{-1}(s \mid x) = \inf \{ y : F(y \mid x) \geq s \} \) and \( J(s) \) is a given score function satisfying \( \int_0^1 J(s) ds = 1 \).

To estimate the functions \( m(x) \) and \( \sigma(x) \), we replace the conditional distribution \( F(y \mid x) \) in (3.2.3) by the conditional copula-graphic estimator, introduced by Braekers and Veraverbeke (2005) and adopted to random design by Sujica and Van Keilegom (2013) (and Chapter 2):
\[
\hat{F}(y \mid x) = \phi_x^{-1} \left\{ - \sum_{T_i \leq y, \Delta_i = 1} \left[ \phi_x \left( \hat{\Pi}_i(\cdot \mid x) \right) - \phi_x \left( \hat{\Pi}(T_i \mid x) \right) \right] \right\}.
\] (3.2.4)
Here, \( \hat{H}(y \mid x) = \sum_{i=1}^n W_{ni}(x, h_n) I(T_i \leq y) \) is the Stone (1977) estimator of the distribution of \( T \) given \( X = x \), where
\[
W_{ni}(x, h_n) = \frac{K ((x - X_i)/h_n)}{\sum_{j=1}^n K ((x - X_j)/h_n)}
\]
are Nadaraya-Watson weights, $K$ is a kernel function, $h_n$ is a bandwidth sequence, and $\overline{H}(y \mid x) = \lim_{t \to y} \overline{H}(t \mid x)$. This leads to

$$\hat{m}(x) = \int_0^1 \hat{F}^{-1}(s|X)J(s)ds \quad \text{and} \quad \hat{\sigma}^2(x) = \int_0^1 \hat{F}^{-1}(s|X)^2J(s)ds - \hat{m}(x)^2,$$

(3.2.5)

where the score function $J(s)$ is chosen in such a way that $\hat{m}(x)$ and $\hat{\sigma}(x)$ are consistent. The estimators $\hat{m}(x)$ and $\hat{\sigma}^2(x)$ have been the object of study in Sujica and Van Keilegom (2013) (and Chapter 2), and are generalizations of the estimators proposed by Van Keilegom and Akritas (1999) to the case where $Y$ and $C$ are copula dependent given $X$.

Next, we estimate the (sub)distribution functions $H_c(y \mid x)$ and $H_c^u(y \mid x)$ by the following Stone (1977)-type estimators:

$$\hat{H}_c(y \mid x) = \sum_{i=1}^n W_{ni}(x, h_n)I(\hat{E}_i \leq y) \quad \text{and} \quad \hat{H}_c^u(y \mid x) = \sum_{i=1}^n W_{ni}(x, h_n)I(\Delta_i = 1)I(\hat{E}_i \leq y),$$

where $\hat{E}_i = (T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$, $i = 1, \ldots, n$.

Plugging-in the estimators $\hat{F}_X$, $\hat{H}_c$ and $\hat{H}_c^u$ in (3.2.2), we obtain the following estimator of the error distribution:

$$\hat{F}_e(y) = \hat{\phi}^{-1}_{(y)} \left\{ - \int_{B_y} \int_{-\infty}^y \phi_x'(\hat{H}_c(s \mid x))d\hat{H}_c^u(s \mid x)d\hat{F}_X(x) \right\},$$

(3.2.6)

where $\hat{\phi}_{(y)}(u) = \int_{B_y} \phi_x(u)d\hat{F}_X(x)$. We choose the set $B_y$ in (3.2.6) as a subset of the set $A_y$ that excludes all small segments, that is $B_y = \arg\max_{B \in \mathcal{B}, B \subseteq A_y} \lambda(B)$, where $\mathcal{B} = \{B \mid B = \bigcup_i B_i : B$ is nonempty, $B_i$ is convex, $\lambda(B_i) \geq \beta\}$, $\lambda$ is the Lebesgue measure and $\beta > 0$ is an arbitrary small constant. It is easy to show that $\{x \mapsto I_B(x) : B \in \mathcal{B}\}$ is a Donsker class of functions (where $I_B(x) = I(x \in B)$). We stress that the set $B_y$ could also be estimated, but proving the asymptotic results in Section 3.3 would fill the chapter with very technical details with no significant contribution (for more details see Remarks 3.3.5 and 3.3.6).

Finally, (3.2.6) together with (3.2.5) lead to our final estimator:

$$\hat{F}_{LS}(y \mid x) = \hat{F}_e \left( \frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right).$$

(3.2.7)
3.3 Asymptotic results

3.3.1 Definitions and assumptions

The primary objective of this section is to study the asymptotic distribution of the estimators \( \hat{F}_e(y) \) and \( \hat{F}_{e,S}(y|x) \), proposed in the previous section. For establishing the asymptotic representation of these estimators we will need the following functions:

\[
\xi_e(E, \Delta, y|X) = -\int_{E \land y}^y \phi''_X(\overline{F}_e(s|X))dH_n^u(s|X)
+ \int_{-\infty}^y \phi''_X(\overline{F}_e(s|X))H_e(s|X)dH_n^u(s|X)
+ \phi'_X(\overline{F}_e(s|X))I(E \leq y, \Delta = 1) - \int_{-\infty}^y \phi'_X(\overline{F}_e(s|X))dH_n^u(s|X),
\]

\[
\xi(T, \Delta, y|X) = \frac{-1}{\phi'_X(F(y|X))} \left\{ \int_{-\infty}^y \phi''_X(\overline{F}(s|X))[I(T \leq s) - H(s|X)]dH^u(s|X)
- \phi''_X(\overline{F}(y|X))[I(T \leq y, \Delta = 1) - H^u(y|X)]
- \int_{-\infty}^y \phi''_X(\overline{F}(s|X))[I(T \leq s, \Delta = 1) - H^u(s|X)]dH(s|X) \right\},
\]

\[
\eta(T, \Delta|X) = \int_{-\infty}^{+\infty} J(F(y|X)) \xi(T, \Delta, y|X) dy,
\]

\[
\zeta(T, \Delta|X) = \int_{-\infty}^y J(F(y|X))\xi(T, \Delta, y|X) \frac{y - m(X)}{\sigma(X)} dy,
\]

\[
\gamma_1(y|X) = -\int_{-\infty}^y \phi''_X(\overline{F}_e(s|X))h_e(s|X)dH_n^u(s|X)
+ \int_{-\infty}^y \phi'_X(\overline{F}_e(s|X))dh_n^u(s|X),
\]

\[
\gamma_2(y|X) = -\int_{-\infty}^y \phi''_X(\overline{F}_e(s|X))sh_e(s|X)dH_n^u(s|X)
+ \int_{-\infty}^y \phi'_X(\overline{F}_e(s|X))d(sh_n^u(s|X)).
\]

Finally, let \( \tau_{\eta} = \inf\{y : \overline{F}_e(y) > \eta\} \) for some small \( \eta > 0 \). The following assumptions are important for proving the asymptotic results in the next section.

(B1) \( (i) \) The sequence \( h_n \) satisfies \( nh_n^4 = o(1) \) and \( nh_n^{2+2\delta}(\log h_n^{-1})^{-1} \to \infty \) for some \( \delta > 0 \).

\( (ii) \) The support \( R_X \) of \( X \) is a bounded interval in \( \mathbb{R} \).
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

(iii) The probability density function $K$ has compact support $[-a, a]$ for some $a > 0$, $\int uK(u)du = 0$ and $K$ is twice continuously differentiable.

Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x \in R_x} (1 - H(\tilde{T}_x|x)) > 0$.

(B2) (i) There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 < \inf_x F(\tilde{T}_x|x)$, $s_0 < \inf\{s \in [0, 1] : J(s) \neq 0\}$, $s_1 > \sup\{s \in [0, 1] : J(s) \neq 0\}$ and $\inf_{x \in R_x} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$.

(ii) The function $J$ is bounded and twice continuously differentiable on the interval $(s_0, s_1)$, $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

(B3) (i) The distribution $F_X$ is three times continuously differentiable on the interior of $R_X$ (and the right (left) limit of $F''_X$ at min $R_X$ (max $R_X$) exists) and $\inf_{x \in R_X} F_X(x) > 0$.

(ii) The functions $m$ and $\sigma$ are twice continuously differentiable on $R_X$ and $\inf_{x \in R_X} \sigma(x) > 0$.

(B4) (i) The functions $\phi_x''(u) = \frac{\partial}{\partial u} \phi_x(u)$, $\phi_x'''(u)$ and $\phi_x^{(3)}(u)$ exist and are continuous in $(x, u) \in R_X \times (0, 1]$.

(ii) The functions $\tilde{\phi}_x''(u) = \frac{\partial^2}{\partial x^2 \partial u^2} \phi_x(u)$, $\tilde{\phi}_x^{(3)}(u)$ and $\phi_x^{(4)}(u)$ exist and are continuous in $(x, u) \in R_X \times (0, 1]$.

(iii) The function $\phi_x$ satisfies $\phi_x'(1) < 0$.

For a (sub)distribution function $L(y|x)$ we will use the notations $L'(y|x) = \frac{\partial}{\partial y} L(y|x)$, $\bar{L}(y|x) = \frac{\partial}{\partial y} L(y|x)$ and similar notations will be used for higher order derivatives. (In the proofs, the function $L(y|x)$ of assumption (B5) will be either $H(y|x)$, $H_c(y|x)$, $H''(y|x)$ or $H''_c(y|x)$.) Assumption (B5) is a standard set of assumptions used in the literature.

(B5) (i) $L(y|x)$ is continuous in $(x, y)$.

(ii) $L'(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L'(y|x)| < \infty$.

(iii) $L''(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L''(y|x)| < \infty$.

(iv) $\bar{L}(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L(y|x)| < \infty$.

(v) $\bar{L}(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L(y|x)| < \infty$.

(vi) $\bar{L}'(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L'(y|x)| < \infty$.

(vii) $\bar{L}'(y|x)$ exists, is continuous in $(x, y)$ and $\sup_{x,y} |y L'(y|x)| < \infty$. 


Remark 3.3.1. In the assumption (B1)(i) we assume that \( nh_n^k \to 0 \), which implies that the bias of estimators \( \hat{m}, \hat{\sigma}, \hat{F} \) or \( \hat{F}_{LS} \) is \( o(n^{-1/2}) \). By doing so we avoid deriving the explicit form of the asymptotic bias of these estimators, which would make this chapter even more technical than it is right now. However, this choice of bandwidth excludes the optimal ones, which minimize the mean square error of aforementioned estimators, and which are of order \( h_n = O(n^{-1/5}) \).

Remark 3.3.2. Assumption (B5) assures that the "wiggliness" (derivatives) of a distribution in the tails is not too extreme. An example of a distribution that does not satisfy (B5)(ii), (B5)(iii), (B5)(vi) and (B5)(vii) is a distribution whose density has a tail equal to \( L'(y|x) = g(y + 1/x) \) (for some density \( g \)), which allows that the expectation and the median go to infinity as \( x \) goes to zero.

Remark 3.3.3. Assumption (B5) also assures that the influence of the covariate \( x \) on the expectation and the median is not too extreme. An example of a distribution that does not satisfy (B5)(i, ii, iii) is a distribution whose density is equal to \( L'(y|x) = g(y + 1/x) \) (for some density \( g \)), which allows that the expectation and the median go to infinity as \( x \) goes to zero.

Remark 3.3.4. Note that assumption (B6) comes from Du and Akritas (2002), and is required to prove an i.i.d. representation for our estimator \( \hat{F}(y|x) \), whose remainder term is negligible uniformly in \( x \) and \( y \) (for details see Lemma 3.7.6). Assumption (B6) is slightly stronger than (B5)(ii, iii, vi): adding the factor \( y^{-\alpha} \) (for an arbitrary small \( \alpha > 0 \)) inside the absolute values in assumptions (B5)(ii, iii, vi), implies assumption (B6). An example of a distribution that does not satisfy this modification of (B5)(ii), (B5)(iii) and (B5)(iv) is a distribution whose density has a tail equal to \( L'_{\text{tail}}(y|x) = y^{-2}(1 - \sin(y^k x^2)) \) for \( k > 2 \), \( k > 1/2 \), \( k > 2 \) and \( k > 1/2 \), respectively, which is slightly bigger range of \( k \) than the one implied by the original assumptions (B5)(vi) and (B5)(vii)(see Remark 3.3.2). Additional restriction of adding factor \( y^{-\alpha} \) in (B5)(ii) excludes the border case possibility of expectation being infinity.

Throughout the rest of this chapter, we let \( C \) denote a generic positive constant, whose value may differ from line to line.
3.3.2 Asymptotic properties of the estimator $\hat{F}_e(y)$

We will extend the result in Van Keilegom and Akritas (1999) concerning the weak convergence of the estimator of the residual distribution function under independent censoring to the case where the dependence between censoring and survival time is described via a known copula. The weak convergence of the estimator will follow from its asymptotic representation, which we give first.

**Theorem 3.3.1.** [Asymptotic representation for $\hat{F}_e(y)$] Assume (B1)-(B4), and assume that (B5) and (B6) hold for $H(y|x)$ and $H^u(y|x)$. Let $y \leq \tau_\eta$. Then,

$$
\hat{F}_e(y) - F_e(y) = n^{-1} \sum_{i=1}^{k} k_y(T_i, \Delta_i, X_i) + R_n(y),
$$

where $\sup\{|R_n(y)| : -\infty < y \leq \tau_\eta\} = o_P\left(\frac{n^{-1/2}}{2}\right)$, and

$$
k_y(T, \Delta, X) = \frac{1}{\phi_y(F_e(y))} \left[ I(X \in B_y)\xi_e(E, \Delta, y|X) + I(X \in B_y)\eta(T, \Delta|X) \frac{\gamma_1(y|X)}{\sigma(X)} + I(X \in B_y)\zeta(T, \Delta|X) \frac{\gamma_2(y|X)}{\sigma(X)} + \left\{ I(X \in B_y) \int_{-\infty}^y \phi_x|\Pi_e(s|X))dH_e^u(s|X) \right\} 
\right.
\left. - \int_{B_y} \int_{-\infty}^y \phi_x|\Pi_e(s|x))dH_e^u(s|x)dF_X(x) \right] + \left\{ I(X \in B_y)\phi_x(F_e(y)) - \int_{B_y} \phi_x(F_e(y))dF_X(x) \right\}.
$$

Note that if we replace $B_y$ by $R_X$, $H_e(y|x)$ by $H_e(y)$, $H^u_e(y|x)$ by $H^u_e(y)$, and set $\phi_x(u) = -\log u$, the first term above corresponds to the i.i.d. representation of the usual Kaplan-Meier estimator due to Lo and Singh (1986). Furthermore, under the same changes, the first three terms give exactly the i.i.d. representation of the estimator studied in Van Keilegom and Akritas (1999) in the case of independent censoring. The second and third term come from the fact that in the estimating procedure we replaced $(T_i - m(X_i))/\sigma(X_i)$ by $(T_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$. Finally, the fourth and fifth terms above are caused by replacing $F_X(x)$ by $\hat{F}_X(x)$. 
3.3. Asymptotic results

**Corollary 3.3.1.** [Weak convergence of $\hat{F}_L(y)$] Under the assumptions of Theorem 3.3.1, the process $n^{1/2}(\hat{F}_L(y) - F(y))$, $-\infty < y \leq \tau_n$, converges weakly to a zero-mean Gaussian process $Z(y)$ with covariance function

$$
\text{Cov}(Z(y), Z(y')) = \text{Cov}(k_y(T, \Delta, X), k_y(T, \Delta, X)).
$$

**3.3.3 Asymptotic properties of the estimator $\hat{F}_{LS}(y|x)$**

Using the results from Theorem 3.3.1 and Corollary 3.3.1 we will show the asymptotic representation and the weak convergence of the estimator $\hat{F}_{LS}(y|x)$ of the conditional distribution under dependent censoring described via a known copula. This result will extend the results in Van Keilegom and Akritas (1999) which are obtained under independent censoring, to the case of dependent censoring described by a copula model.

**Theorem 3.3.2.** [Asymptotic representation for $\hat{F}_{LS}(y|x)$] Assume (B1)-(B4), and assume that (B5) and (B6) hold for $H(y|x)$ and $H^*(y|x)$. Let $(y - m(x))/\sigma(x) \leq \tau_n$. Then,

$$
\hat{F}_{LS}(y|x) - F(y|x) = \hat{F}_e \left( \frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right) - F_e \left( \frac{y - m(x)}{\sigma(x)} \right)
$$

$$
= (nh_n)^{-1} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) h_y(T, \Delta|x) + R_n(y|x),
$$

where $\sup \{ |R_n(y|x) : (x, y) \in \Omega \} = o_P((nh_n)^{-1/2})$, $\Omega = \{(x, y) : (y - m(x))/\sigma(x) \leq \tau_n, x \in R_X \}$ and

$$
h_y(T, \Delta|x) = \left[ \eta(T, \Delta|x) + \zeta(T, \Delta|x) \frac{y - m(x)}{\sigma(x)} \right] f_e \left( \frac{y - m(x)}{\sigma(x)} \right) \sigma^{-1}(x)f_X^{-1}(x).
$$

**Corollary 3.3.2.** [Weak convergence of $\hat{F}_{LS}(y|x)$] Under the assumptions of Theorem 3.3.2, the process $(nh_n)^{1/2}(\hat{F}_{LS}(y|x) - F(y|x))$, $x \in R_X$ fixed, $(y - m(x))/\sigma(x) \leq \tau_n$, converges weakly to a zero-mean Gaussian process $Z(y|x)$ with covariance function

$$
\text{Cov}(Z(y|x), Z(y'|x)) = f_X(x) \int K^2(u) du \text{Cov}(h_y(T, \Delta|x), h_{y'}(T, \Delta|x)|X = x).
$$

**Remark 3.3.5.** It can be shown that all the results from this section hold (with no extra terms in the i.i.d. representations in Theorems 3.3.1 and 3.3.2), if we replace $B_y$ by an estimator $\hat{B}_y$ that is converging a.s. to $B_y$ in Lebesgue.
measure for every fixed \(-\infty < y \leq \tau_n\), and for which the set \(\hat{B}_y\) belongs a.s. to \(B\) for all \(-\infty < y \leq \tau_n\). The required modifications of the proofs are very technical with a small contribution of novelty, and are therefore omitted.

**Remark 3.3.6.** Using the notation \(A_y(\eta) = \{ x : \mathcal{H}_c(y|x) > \eta \}\) and \(\hat{A}_y(\eta) = \{ x : \mathcal{H}_c(y|x) > \eta \}\), one potential estimator of \(B_y\) is \(\bar{B}_y = \text{argmax}\{ \lambda(B) : B \in \mathbb{B}, B \subset \hat{A}_y(\eta + a_n) \}\), where \(a_n = (nh_n)^{-1/2}(\log n)^{1/2+\alpha}\) and \(\alpha > 0\). This estimator satisfies the conditions of Remark 3.3.5 by the following reasoning. First, because of Lemma 3.7.4, we have \(P(\lim_{n \to \infty} I\{A_y(\eta + 2a_n) \subset \hat{A}_y(\eta + a_n) \subset A_y(\eta)\}) = 1\). Additionally, simple calculus shows that the function \(\eta \mapsto \lambda[A_y(\eta)]\) is a right continuous function, that is \(\lambda[A_y(\eta + 2a_n)] \to \lambda[A_y(\eta)]\) a.s. Hence, we have a.s. convergence in Lebesgue measure of \(\lambda[\hat{A}_y(\eta + a_n)]\) to \(\lambda[A_y(\eta)]\), which implies that \(\lambda[\hat{B}_y]\) converges to \(\lambda[B_y]\). On the other hand, by definition, the set \(\hat{B}_y\) belongs to \(B\) for all \(-\infty < y \leq \tau_n\).

### 3.4 Proofs of main results

In this section we give the proofs of the main results from Section 3.3. The proofs are based on a number of technical lemmas and propositions, which can be found in Sujica and Van Keilegom (2013) (and Chapter 2) and Appendices A and B.

**Proof of Theorem 3.3.1.** [Asymptotic representation for \(\hat{F}_c(y)\)]

First, we will break down \(\hat{F}_c(y) - F_c(y)\) into several terms, in such a way that each term depends on a single plug-in estimator. This decomposition will end at (3.4.6). Then, in a second step we will deal with each term in this decomposition. We start by using Lemma 3.6.1:

\[
\hat{F}_c(y) - F_c(y) = \phi_y^{-1} \left\{ \int_{B_y} \int_{-\infty}^y \phi_y'(s|x)d\mathcal{H}_c(s|x)dF_X(x) \right\} \\
- \hat{\phi}_y^{-1} \left\{ \int_{B_y} \int_{-\infty}^y \phi_y'(s|x)d\hat{\mathcal{H}}_c(s|x)d\hat{F}_X(x) \right\}
\]

\[
= \left[ \phi_y^{-1}\{U(y)\} - \hat{\phi}_y^{-1}\{U_n(y)\} \right] + \left[ \phi_y^{-1}\{U_n(y)\} - \phi_y^{-1}\{\hat{U}_n(y)\} \right]
\]

\[
+ \hat{\phi}_y^{-1}\{\hat{U}_n(y)\} - \hat{\phi}_y^{-1}\{\hat{U}_n(y)\}
\]

\[
:= \text{I} + \text{II} + \text{III},
\]  

(3.4.1)
3.4. Proofs of main results

where

\[ U(y) = - \int_{B_y} \int_{-\infty}^{y} \phi_x'(\mathcal{H}_c(s|x))dH_u^w(s|x)dF_X(x), \]

\[ U_n(y) = - \int_{B_y} \int_{-\infty}^{y} \phi_x'(\hat{\mathcal{H}}_c(s|x))d\hat{H}_u^w(s|x)dF_X(x) \]

and

\[ \tilde{U}_n(y) = - \int_{B_y} \int_{-\infty}^{y} \phi_x'(\hat{\mathcal{H}}_c(s|x))d\hat{H}_u^w(s|x)d\hat{F}_X(x). \]

(3.4.2)

Next, we examine each of the three terms. Starting with the first one, we have by a second order Taylor expansion:

\[ (I) = \frac{1}{\phi'_y \left( F_c(y) \right)} \left\{ U(y) - U_n(y) \right\} + R_{n1}^{(I)}(y), \]

where

\[ R_{n1}^{(I)}(y) = \frac{\phi''_y \left( \phi^{-1}_y(\varepsilon_1(y)) \right)}{2\phi'_y \left( \phi^{-1}_y(\varepsilon_1(y)) \right)^3} \left\{ U(y) - U_n(y) \right\}^2, \]

with \( \varepsilon_1(y) \) an intermediate value between \( U_n(y) \) and \( U(y) \). By adding and subtracting terms, we further have that

\[ (I) = \frac{1}{\phi'_y \left( F_c(y) \right)} \times \]

\[ \left\{ - \int_{B_y} \int_{-\infty}^{y} \phi_x''(\mathcal{H}_c(s|x)) \left[ \mathcal{H}_w(s|x) - H_w(s|x) \right] dH_u^w(s|x)dF_X(x) \right\} \]

\[ + R_{n1}^{(I)}(y) + R_{n2}^{(I)}(y) \]

\[ = \frac{1}{\phi'_y \left( F_c(y) \right)} \times \]

\[ \left\{ - \int_{B_y} \int_{-\infty}^{y} \phi_x''(\mathcal{H}_c(s|x)) \left[ \hat{H}_w(s|x) - \hat{H}_w(s|x) \right] d\hat{H}_u^w(s|x)d\hat{F}_X(x) \right\} \]

\[ + R_{n1}^{(I)}(y) + R_{n2}^{(I)}(y) + R_{n3}^{(I)}(y), \]

(3.4.3)
where
\[ R_{n2}^{(I)}(y) = \frac{1}{\phi'_y(F_y(y))} \times \int_{B_y} \int_{-\infty}^{y} \phi'_x(\tilde{H}_e(s|x)) - \phi'_x(\overline{H}_e(s|x)) d \left[ \tilde{H}_e^n(s|x) - H_e^n(s|x) \right] dF_X(x), \]
\[ R_{n3}^{(I)}(y) = \frac{1}{2\phi'_y(F_y(y))} \times \int_{B_y} \int_{-\infty}^{y} \phi''_\xi(\tilde{H}_e(s,x)) \left[ \tilde{H}_e(s|x) - H_e(s|x) \right]^2 dH_e^n(s|x) dF_X(x), \]
with \( \xi_1(s,x) \) between \( \tilde{H}_e(s|x) \) and \( \overline{H}_e(s|x) \).

Next, we examine (II):
\[ (II) = \frac{1}{\phi'_y(F_y(y))} \left\{ U_n(y) - \tilde{U}_n(y) \right\} + R_{n1}^{(II)}(y) + R_{n2}^{(II)}(y), \]
where
\[ R_{n1}^{(II)}(y) = \frac{\phi''_y \left( \phi^{-1}_y(\varepsilon_2(y)) \right)}{2\phi'_y \left( \phi^{-1}_y(\varepsilon_2(y)) \right)^3} \left\{ U_n(y) - \tilde{U}_n(y) \right\}^2, \]
\[ R_{n2}^{(II)}(y) = \left( \frac{1}{\phi'_y(\phi^{-1}_y(U_n(y)))} - \frac{1}{\phi'_y(F_y(y))} \right) \left\{ U_n(y) - \tilde{U}_n(y) \right\}, \]
with \( \varepsilon_2(y) \) between \( \tilde{U}_n(y) \) and \( U_n(y) \). Let us further decompose (II):
\[ (II) = \frac{1}{\phi'_y(F_y(y))} \int_{B_y} \int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) dH_e^n(s|x) d \left[ \hat{F}_X(x) - F_X(x) \right] \]
\[ + R_{n1}^{(II)}(y) + R_{n2}^{(II)}(y) + R_{n3}^{(II)}(y) + R_{n4}^{(II)}(y), \quad (3.4.4) \]
where
\[ R_{n3}^{(II)}(y) = \frac{1}{\phi'_y(F_y(y))} \times \int_{B_y} \int_{-\infty}^{y} \phi'_x(\tilde{H}_e(s|x)) - \phi'_x(\overline{H}_e(s|x)) d \tilde{H}_e^n(s|x) d \left[ \hat{F}_X(x) - F_X(x) \right], \]
\[ R_{n4}^{(II)}(y) = \frac{1}{\phi'_y(F_y(y))} \times \int_{B_y} \int_{-\infty}^{y} \phi'_x(\overline{H}_e(s|x)) d \left[ \tilde{H}_e^n(s|x) - H_e^n(s|x) \right] d \left[ \hat{F}_X(x) - F_X(x) \right]. \]
Next, we examine (III). By applying a second order Taylor expansion on \( \hat{\phi}_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \} - \hat{\phi}_n(y) \{ \tilde{U}_n(y) \} \), we obtain

\[
(III) = \phi_n^{-1}(\tilde{U}_n(y)) - \tilde{\phi}_n^{-1}(\tilde{U}_n(y)) \\
= \frac{1}{\hat{\phi}_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \}} \left[ \hat{\phi}_n(y) \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) - \tilde{U}_n(y) \right] + R_{n1}^{(III)}(y) \\
= \frac{1}{\phi_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \}} \left[ \hat{\phi}_n(y) \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) - \tilde{U}_n(y) \right] \\
+ R_{n1}^{(III)}(y) + R_{n2}^{(III)}(y) \\
= \frac{1}{\phi_n(y) \{ \phi_n^{-1}(U(y)) \}} \left[ \hat{\phi}_n(y) \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) - \tilde{U}_n(y) \right] \\
+ R_{n1}^{(III)}(y) + R_{n2}^{(III)}(y) + R_{n3}^{(III)}(y)
\]

where

\[
R_{n1}^{(III)}(y) = \frac{\hat{\phi}_n''(y) \{ \xi_2(y) \}}{2\hat{\phi}_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \} \left[ \phi_n^{-1}(\tilde{U}_n(y)) - \tilde{\phi}_n^{-1}(\tilde{U}_n(y)) \right]^2},
\]

\[
R_{n2}^{(III)}(y) = \left[ \frac{1}{\hat{\phi}_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \}} - \frac{1}{\phi_n'(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \}} \right] \\
\times \left[ \hat{\phi}_n(y) \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) - \tilde{U}_n(y) \right],
\]

\[
R_{n3}^{(III)}(y) = \left( \frac{1}{\phi_n(y) \{ \phi_n^{-1}(\tilde{U}_n(y)) \}} - \frac{1}{\phi_n'(y) \{ \phi_n^{-1}(U(y)) \}} \right) \\
\times \left[ \hat{\phi}_n(y) \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) - \tilde{U}_n(y) \right],
\]

with \( \xi_2(y) \) between \( \hat{\phi}_n^{-1}(\tilde{U}_n(y)) \) and \( \phi_n^{-1}(\tilde{U}_n(y)) \). Further, we continue with (III) by calculating

\[
(III) - R_{n1}^{(III)}(y) - R_{n2}^{(III)}(y) - R_{n3}^{(III)}(y) \\
= \frac{1}{\phi_n'(y) \{ \tilde{F}_c(y) \}} \int_{B_x} \phi_x \left( \phi_n^{-1}(\tilde{U}_n(y)) \right) d \left[ \tilde{F}_X(x) - F_X(x) \right] \\
= \frac{1}{\phi_n'(y) \{ \tilde{F}_c(y) \}} \int_{B_x} \phi_x \left( \phi_n^{-1}(U(y)) \right) d \left[ \tilde{F}_X(x) - F_X(x) \right] + R_{n4}^{(III)}(y) \\
= \frac{1}{\phi_n'(y) \{ \tilde{F}_c(y) \}} \int_{B_x} \phi_x \left( \tilde{F}_c(y) \right) d \left[ \tilde{F}_X(x) - F_X(x) \right] + R_{n4}^{(III)}(y), \quad (3.4.5)
\]
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

where

\[ R_{n4}^{(III)}(y) = \frac{1}{\phi_y'(F_e(y))} \times \int_{B_y} \phi_x \left( \phi_y^{-1}(U_n(y)) \right) - \phi_x \left( \phi_y^{-1}(U(y)) \right) d \left[ \hat{F}_X(x) - F_X(x) \right]. \]

Combining (3.4.1), (3.4.3), (3.4.4) and (3.4.5), we subsequently obtain

\[
\hat{F}_e(y) - F_e(y) = \frac{1}{\phi_y'(F_e(y))} \left\{ - \int_{B_y} \int_{-\infty}^{y} \phi_x''(H_e(s|x)) \left[ \hat{H}_e(s|x) - H_e(s|x) \right] dH_e^{(n)}(s|x) dF_X(x) \\
+ \int_{B_y} \int_{-\infty}^{y} \phi_x''(H_e(s|x)) d \left[ \hat{H}_e^{(n)}(s|x) - H_e^{(n)}(s|x) \right] dF_X(x) \\
+ \int_{B_y} \int_{-\infty}^{y} \phi_x'(H_e(s|x)) dH_e^{(n)}(s|x) d \left[ \hat{F}_X(x) - F_X(x) \right] \\
+ \int_{B_y} \phi_x \left( \hat{F}_e(y) \right) d \left[ \hat{F}_X(x) - F_X(x) \right] \right\} + R_n(y) \\
:= \frac{1}{\phi_y'(F_e(y))} \left\{ \hat{\Lambda}_1(y) + \hat{\Lambda}_2(y) + \hat{\Lambda}_3(y) + \hat{\Lambda}_4(y) \right\} + R_n(y). \tag{3.4.6}
\]

where \( R_n(y) = R_{n1}^{(I)} + R_{n2}^{(I)}(y) + R_{n3}^{(I)}(y) + R_{n4}^{(II)}(y) + R_{n5}^{(II)}(y) + R_{n6}^{(III)}(y) + R_{n7}^{(III)}(y) + R_{n8}^{(IV)}(y) + R_{n9}^{(IV)}(y) + R_{n10}^{(IV)}(y) + R_{n11}^{(IV)}(y) + R_{n12}^{(IV)}(y) + R_{n13}^{(IV)}(y) + R_{n14}^{(IV)}(y). \) Now, by applying Lemmas 3.6.2 and 3.6.3 for the main terms we get

\[
\hat{F}_e(y) - F_e(y) = \frac{1}{\phi_y'(F_e(y))} \frac{1}{n} \sum_{i=1}^{n} l_y(T_i, \Delta_i, X_i) + R_n(y) + o_P(n^{-1/2}),
\]
uniformly in \(-\infty < y \leq \tau_y\), where

\[
l_y(T_i, \Delta_i, X_i) = I(X_i \in B_y) \int_{-\infty}^y \phi_X'(\mathbf{F}(s|X_i)) \left[ - \frac{h_e(s|X_i)}{\sigma(X_i)} \right] \{s\zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i)\}
- I(E_i \leq s) + H_e(s|X_i) \right] dH_e^n(s|X_i)
+ I(X_i \in B_y) \int_{-\infty}^y \phi_X'(\mathbf{F}(s|X_i)) d \left[ \frac{h_e^n(s|X_i)}{\sigma(X_i)} \right] \{s\zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i)\}
+ I(E_i \leq s, \Delta_i = 1) - H_e^n(s|X_i)
+ I(X_i \in B_y) \int_{-\infty}^y \phi_X'(\mathbf{F}(s|X_i)) dH_e^n(s|X_i)
- \int_{B_y} \int_{-\infty}^y \phi_X'(\mathbf{F}(s|x)) dH_e^n(s|x) dF_X(x)
+ I(X_i \in B_y) \phi_X(\mathbf{F}(y)) - \int_{B_y} \phi_X(\mathbf{F}(y)) dF_X(x).
\]

From here we get the final form

\[
\hat{\mathbf{F}}_e(y) - \mathbf{F}_e(y) = \frac{1}{n} \sum_{i=1}^n k_y(T_i, \Delta_i, X_i).
\]

Remainder terms \(R^{(I)}_n(y)\) and \(R^{(I)}_n(y)\) are \(o(n^{-1/2})\) a.s. by Lemma 3.7.5 and Lemma 3.7.4, respectively. Now, by using a Taylor expansion, we can easily show that each of the other remainder terms in \(R_n(y)\) can be a.s. bounded by a product consisting of some of the following terms: \(U_n - U, \tilde{U}_n - U, \hat{\phi}(y) - \phi(y), \int_{B_y} G_n(y, x) d[\hat{F}_X(x) - F_X(x)]\) (where \(G_n\) is a function satisfying the assumptions of Lemma 3.7.2). From here it is easy to verify that all the remainder terms are \(o(n^{-1/2})\) a.s. by using the orders given in Lemmas 3.7.1 to 3.7.4.

\[\square\]

**Proof of Corollary 3.3.1.** We will prove the weak convergence by showing that the class of functions \(\mathcal{K} = \{k_y : -\infty < y \leq \tau_y\}\) from Theorem 3.3.1 is a Donsker class. The function \(x \mapsto [\phi_X'(\mathbf{F}(y))]^{-1}\) is a uniformly bounded, deterministic function because of assumptions (B3)(i), (B4)(iii) and because \(B_y\) is nonempty for all \(-\infty < y \leq \tau_y\). The term inside the square brackets in the definition of \(k_y\) belongs to a sum of Donsker classes (see (3.4.6) and Lemmas 3.6.2 and 3.6.3), which is also a Donsker class by Example 2.10.7 in Van der Vaart and Wellner (1996). Furthermore, it is easy to show that \(k_y\) is uniformly bounded. Hence, since multiplying uniformly bounded functions preserves the
Donsker property (by Example 2.10.8 in Van der Vaart and Wellner (1996)), the class of functions $\mathcal{K}$ is Donsker.

Proof of Theorem 3.3.2. [Asymptotic representation for $\hat{F}_{LS}(y|x)$] Write

$$
\hat{F}_{LS}(y|x) - F(y|x) = \left[ \hat{F}_e \left( \frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right) - F_e \left( \frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right) \right] \\
+ \left[ F_e \left( \frac{y - \hat{m}(x)}{\hat{\sigma}(x)} \right) - F_e \left( \frac{y - m(x)}{\sigma(x)} \right) \right] \\
+ \left[ F_e \left( \frac{y - m(x)}{\sigma(x)} \right) - F_e \left( \frac{y - m(x)}{\sigma(x)} \right) \right] \\
= \alpha_1^{(x,y)} + \alpha_2^{(x,y)} + \alpha_3^{(x,y)}.
$$

Note that $(nh_n)^{1/2} \alpha_1^{(x,y)} = o_P(1)$ uniformly in $(x,y) \in \Omega$, because of the weak convergence result established in Corollary 3.3.1. For $\alpha_2^{(x,y)}$ we have

$$
\alpha_2^{(x,y)} = -\frac{\hat{m}(x) - m(x)}{\hat{\sigma}(x)} \frac{y - m(x)}{\sigma(x)} + \frac{1}{2} \left( \frac{\hat{m}(x) - m(x)}{\sigma(x)} \right)^2 f'_e(\xi_x),
$$

for some $\xi_x$ between $(y - m(x))/\sigma(x)$ and $(y - \hat{m}(x))/\hat{\sigma}(x)$. The second term above is of order $O \left( \frac{(nh_n)^{-1}}{\log n} \right)$ a.s. by Proposition 2.3.5 in Chapter 2, together with the boundedness of $f'_e$ (which follows from assumption (B5)(ii)). For the first term, we first replace $\hat{\sigma}(x)$ by $\sigma(x)$ by using Proposition 3.5 and then apply Proposition 2.3.8 in Chapter 2 to obtain an asymptotic representation. For $\alpha_3^{(x,y)}$, we have

$$
\alpha_3^{(x,y)} = \frac{\sigma(x) - \hat{\sigma}(x)}{\hat{\sigma}(x)} \frac{y - m(x)}{\sigma(x)} \frac{y - m(x)}{\sigma(x)} + \frac{1}{2} \left( \frac{\sigma(x) - \hat{\sigma}(x)}{\hat{\sigma}(x)} \right)^2 \left( \frac{y - m(x)}{\sigma(x)} \right)^2 f'_e(\xi_{2x}),
$$

where $\xi_{2x}$ is between $(y - m(x))/\sigma(x)$ and $(y - m(x))/\hat{\sigma}(x)$. The second term above is $O \left( \frac{(nh_n)^{-1}}{\log n} \right)$ a.s. by Proposition 2.3.5 in Chapter 2, the fact that $\sup_y \left| y^2 f'_e(y) \right| < \infty$ and assumption (B3)(ii). After replacing $\hat{\sigma}$ with $\sigma$ by using again Proposition 2.3.5 in Chapter 2, the first term above has an asymptotic representation given by Proposition 2.3.8 in Chapter 2. This combined with the asymptotic representation for $\alpha_2^{(x,y)}$ completes the proof.

Proof of Corollary 3.3.2. The proof is similar to the proof of Corollary 3.4. in Van Keilegom and Akritas (1999). It suffices in fact to replace the function $\xi$ defined in Van Keilegom and Akritas (1999) by the function $\xi$ defined in Section 3.3.1. Apart from this, the proof is identical.
3.5 Simulations

In this section we illustrate the finite sample behavior of our estimator $\hat{F}_{LS}(y|x)$ by means of Monte Carlo simulations. We compare the estimator $\hat{F}_{LS}(y|x)$ proposed in this chapter with the estimator $\hat{F}(y|x)$ proposed by Braekers and Veraverbeke (2005) and defined in (3.2.4). We expect that under the assumption of the location-scale model (3.1.1), the estimator $\hat{F}_{LS}(y|x)$ outperforms $\hat{F}(y|x)$. Further, we compare the two estimators under misspecification of model (3.1.1), that is, when $\varepsilon$ and $X$ are dependent. Here we will explore the performance of $\hat{F}_{LS}(y|x)$ compared to $\hat{F}(y|x)$, as the dependence between $\varepsilon$ and $X$ increases.

Lastly, we compare $\hat{F}_{LS}(y|x)$ and $\tilde{F}_{LS}(y|x)$ (proposed by Van Keilegom and Akritas (1999)) under two settings: 1) dependence between $Y$ and $C$ given $X$ and 2) independence between $Y$ and $C$ given $X$. We expect that under 1) $\hat{F}_{LS}(y|x)$ outperforms $\tilde{F}_{LS}(y|x)$ and that under 2) they perform similarly.

To compare the performance of the estimators we use the mean squared error ($MSE$) and the integrated mean squared error ($IMSE$), to be defined further on. The simulations are carried out for samples of size $n = 100$, $n = 200$ and $n = 400$, and the results are obtained by using 1000 Monte Carlo simulations (and 200 Monte Carlo simulations in the case of Figures 3.3 and 3.4).

In the first setting, we generate i.i.d. data from the following regression model:

$$Y = 6(X - 0.5)^2 + 0.5\varepsilon,$$

where $X$ has a uniform distribution on $[0,1]$ and the error $\varepsilon$ has a standard normal distribution and is independent of $X$. The censoring variable $C$ satisfies $C = \alpha_1(X - 0.5)^2 + \alpha_2 + 0.5\tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is standard normal and independent of $X$, and the constants $\alpha_1$ and $\alpha_2$ are chosen so that the global censoring rate is 45% and the local censoring rate (for a fixed value of $x$) is between 42% and 48%. Finally, to model the dependence between $Y$ and $C$ given $X = x$ (i.e. the dependence between $\varepsilon$ and $\tilde{\varepsilon}$ given $X = x$) we use a Gumbel copula:

$$C_x(u_1, u_2) = \exp \left\{ -\left[ - (\log u_1)^{\gamma(x)} - (\log u_2)^{\gamma(x)} \right]^{1/\gamma(x)} \right\}, \quad (3.5.1)$$

where $\gamma(x) = \max(5 - 5x, 1)$. This means that the corresponding Archimedean copula generator equals $\phi_x(u) = -(\log u)^{\gamma(x)}$. Under this setting the conditional dependence between $Y$ and $C$ given $X = x$ decreases from strong positive dependence to complete independence as $x$ goes from 0 to 1 (Kendall’s tau coefficient decreases from 0.8 to 0). We work with the score function
Since, the function \(J(s) = b^{-1}I(0 \leq s \leq b)\). In order to estimate the functionals \(m(\cdot)\) and \(\sigma(\cdot)\) consistently the constant \(b\) has to be smaller than or equal to \(\inf_{x \in \mathbb{R}^n} \hat{F}_{LS}(+\infty|x)\). Therefore, we choose \(b = 0.8\) which is smaller than the average of 1000 simulated infima.

Note that, in the setting above, an equivalent way of writing the estimator \(\hat{\sigma}^2(x)\) is
\[
\hat{\sigma}^2(x) = \sum_{i=1}^{n} \left[ Y_i - \hat{m}(x) \right]^2 \left[ \hat{F}_b(Y_i|x) - \hat{F}_b(Y_i^-|x) \right] J(\hat{F}_b(Y_i|x)),
\]
where \(\hat{F}_b(y|x) := \min\{\hat{F}(y|x), b\}\). (In a general setting, we can show that the formula above is an asymptotic approximation of \(\hat{\sigma}^2(x)\), if we use \(\hat{F}\) instead \(\hat{F}_b\).) Furthermore, \(\hat{\sigma}^2(x)\) is close to
\[
\sum_{i=1}^{n} \left[ Y_i - \hat{m}(X_i) \right]^2 \left[ \hat{F}_b(Y_i|x) - \hat{F}_b(Y_i^-|x) \right] J(\hat{F}_b(Y_i|X_i)).
\]

In the sequel we work with the latter estimator, since simulations showed that it outperforms the former (which is expected, since for a given covariate \(X_i\), the difference \(Y_i - m(X_i)\) is a real deviation of \(Y_i\) from \(m(X_i)\), whereas \(Y_i - m(x)\) is not). Also, we can easily show that both estimators are asymptotically equivalent under certain assumptions on the bandwidth.

For the weights that appear in our estimators \(\hat{F}(y|x)\) and \(\hat{F}_{LS}(y|x)\) we use the kernel function \(K(u) = (15/16)(1 - u^2)^2I(|u| \leq 1)\). For a fixed \(x\), to select a bandwidth for \(\hat{F}(\cdot|x)\), we minimize the integrated mean squared error \(IMSE(\hat{F}(\cdot|x)) := E[\int \{\hat{F}(y|x) - F(y|x)\}^2 dF(y|x)]\) over a grid of 12 equidistant possible values of \(h_n\) between 0.050 and 0.400. The so-obtained estimator is denoted by \(\hat{h}_n(\hat{F}(\cdot|x))\). To calculate \(IMSE(\hat{F}(\cdot|x))\), we use 1000 simulated data sets. For each simulated data set, we compute the integrated squared error \(\int \{\hat{F}(y|x) - F(y|x)\}^2 dF(y|x)\), and we approximate \(IMSE(\hat{F}(\cdot|x))\) by taking the average over these 1000 values. On the other hand, to estimate \(\hat{F}_{LS}(\cdot|x)\) we need to choose 4 bandwidths corresponding to \(\hat{m}(\cdot), \hat{\sigma}(\cdot), \hat{\mu}_n(\cdot)\) and \(\hat{H}_n^*(\cdot)\). In the first step, to select the bandwidth for \(\hat{m}\), we minimize the integrated mean squared error \(IMSE(\hat{m}) = \int_0^\infty E[\hat{m}(x) - m(x)]^2 dF_X(x)\), in the same way as \(IMSE(\hat{\mu}(\cdot|x))\), over a grid of 15 equidistant possible values of \(h_n\) between 0.05 and 0.40. (Note that, in this case, we do not take into account values of \(x\) close to the boundary of the support of \(X\) to avoid addressing boundary effects of the Nadaraya-Watson weights.) In the second step, to select the bandwidth for
we first write
\[
\sum_{i=1}^{n} [Y_i - \hat{m}_1(X_i)]^2 [\hat{F}_{1b}(Y_i|x) - \hat{F}_{1b}(Y_i^-|x)] J(\hat{F}_{2b}(Y_i|X_i)),
\]
where \(\hat{m}_1\) and \(\hat{F}_{1b}\) are the estimators based on the optimal bandwidth chosen in the first step, and \(\hat{F}_{2b}\) is based on a second bandwidth. Now, we select this second bandwidth by minimizing the empirical IMSE(\(\hat{\sigma}\)) (which is estimated in the same way as in the first step) over the grid 0.05, 0.1, 0.2, ..., 0.9. In the next step we choose the same bandwidth for both \(\hat{H}_e(\cdot|x)\) and \(\hat{F}_{LS}(\cdot|x)\) so that it minimizes the corresponding approximated IMSE(\(\hat{F}_{LS}(\cdot|x)\)) (where \(\hat{F}_{LS}(\cdot|x)\) uses the bandwidths \(h_n(\hat{m})\) and \(h_n(\hat{\sigma})\) for \(\hat{m}\) and \(\hat{\sigma}\), obtained in the previous step). This bandwidth is denoted by \(h_n(\hat{F}_{LS}(\cdot|x))\).

![Figure 3.1: MSE of \(\hat{F}_{LS}(y|x)\) (solid line) and \(\hat{F}(y|x)\) (dashed line) for samples of size \(n = 100, n = 200\) and \(n = 400\) (row 1, row 2, row 3) and for covariate \(x = 0.2, x = 0.5\) and \(x = 0.8\) (column 1, column 2, column 3). The number under the curve represents the ratio of the approximated IMSE of \(\hat{F}(\cdot|x)\) and \(\hat{F}_{LS}(\cdot|x)\).](image)

Figure 3.1 shows the MSE of \(\hat{F}(y|x)\) and \(\hat{F}_{LS}(y|x)\) for bandwidths chosen by the above procedure. Each subgraph contains the ratio of the approximated IMSE of \(\hat{F}(\cdot|x)\) and \(\hat{F}_{LS}(\cdot|x)\), which shows that \(\hat{F}_{LS}(\cdot|x)\) outperforms \(\hat{F}(\cdot|x)\).
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

Figure 3.2: MSE of $\hat{F}_{LS}(y|x)$ (solid lines) and $\hat{F}(y|x)$ (dashed lines) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). Every subgraph contains the curves of $\hat{F}(|x)$ and $\hat{F}_{LS}(|x)$ for different values of the bandwidth ranging from 0.050 to 0.400 (with increments of 0.025). The bandwidths for estimating $m(\cdot)$ and $\sigma(\cdot)$ are taken equal to $h_n(\hat{m})$ and $h_n(\hat{\sigma})$ respectively.

for all sample sizes and all values of the covariate. We believe that this is a consequence of the fact that $\hat{F}_{LS}(y|x)$ uses the extra information given by the location-scale regression model (3.1.1). One could argue that in this comparison, having two optimization steps for choosing the bandwidths for $\hat{F}_{LS}(|x)$ gives an unfair advantage over a one step optimization for choosing the bandwidth for $\hat{F}(|x)$. To address this issue, in Figure 3.2, we plot the $MSE$ of $\hat{F}(y|x)$ and of $\hat{F}_{LS}(y|x)$ for several bandwidths ranging from 0.050 to 0.400. To estimate $m$ and $\sigma$, involved in estimating $\hat{F}_{LS}(y|x)$, we use the optimal bandwidths $h_n(\hat{m})$ and $h_n(\hat{\sigma})$ for all curves. Figure 3.2 shows that the second optimization step has little influence on the performance of the estimator $\hat{F}_{LS}(y|x)$, which once again shows the advantage of our estimator $\hat{F}_{LS}(y|x)$ over $\hat{F}(y|x)$.

To test the behavior of $\hat{F}_{LS}$ against $\hat{F}$ under misspecification of the location-scale model (3.1.1), we will use the same setting as above except that we allow
3.5. Simulations

Figure 3.3: MSE of $\hat{F}(y|x)$ (dashed line) and $\hat{F}_{LS}(y|x)$ (solid line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for parameter $c = 0$, $c = 0.5$ and $c = 1$ (column 1, column 2, column 3). The number under the curve represents the ratio of the approximated $IMSE$ of $\hat{F}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$.

$\varepsilon$ to depend on $X$:

$$\varepsilon|x = x \sim \sqrt{1 - (cx)^{1/4}t_{2/(cx)^{1/4}}},$$

where $0 \leq c \leq 1$ is the parameter controlling the kurtosis, which for a fixed $x$ goes to infinity as $c$ goes to 1. As $c$ goes to 0, the distribution of $\varepsilon$ converges to the normal distribution. We note that the percentage of censoring remains between 42% and 48%. To derive the estimators $\hat{F}_{LS}$ and $\hat{F}$ in this setting, we use the same estimation procedure as before.

Figure 3.3 shows the approximated $MSE$ and the ratio of the approximated $IMSE$ of the estimator $\hat{F}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$ obtained by the method above, for an arbitrary chosen covariate $x = 0.2$ and different values of the parameter $c$. As expected we see that the ratio of the $IMSE$ of $\hat{F}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$ decreases when $c$ increases. This feature is due to increase of the misspecification, that is, the increase of the dependence between $\varepsilon$ and $X$. We notice that the performance of the estimator $\hat{F}(\cdot|x)$ is also decreasing. We believe that this is
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

Figure 3.4: Ratio of the approximated IMSE of $\tilde{F}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$ (solid line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3).

Figure 3.5: MSE of $\tilde{F}_{LS}(y|x)$ (dashed line) and $\hat{F}_{LS}(y|x)$ (solid line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). The number in the plot represents the ratio of the approximated IMSE of $\tilde{F}_{LS}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$. 
3.5. Simulations

Figure 3.6: MSE of $\tilde{F}_{LS}(y|x)$ (dashed line) and $\hat{F}_{LS}(y|x)$ (solid line) for samples of size $n = 100$, $n = 200$ and $n = 400$ (row 1, row 2, row 3) and for covariate $x = 0.2$, $x = 0.5$ and $x = 0.8$ (column 1, column 2, column 3). The number under the curve represents the ratio of the approximated IMSE of $\tilde{F}_{LS}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$.

mainly caused by the fact that in general a conditional distribution function with higher kurtosis is harder to estimate. In Figure 3.4 we focus on the ratio of the approximated IMSE of $\hat{F}(\cdot|x)$ and $\hat{F}_{LS}(\cdot|x)$. We see that $\hat{F}_{LS}(\cdot|x)$ is constantly outperforming $\hat{F}(\cdot|x)$ except for the highest levels of misspecification.

Figure 3.5 shows the MSE of the estimator $\tilde{F}_{LS}(y|x)$ and the estimator $\hat{F}_{LS}(y|x)$. The estimator $\tilde{F}_{LS}(y|x)$ is defined by relation (1.2.7). The bandwidths for estimating $\tilde{m}$ and $\tilde{\sigma}$ are derived in exactly the same way as for $\hat{m}$, $\hat{\sigma}$. The only bandwidth selection for $\tilde{F}_{LS}(\cdot|x)$ is the one used for $\hat{m}$ and $\hat{\sigma}$. The second-step bandwidth for $\hat{F}_{LS}(\cdot|x)$ is chosen from the grid $[0.050,0.400]$ as before, so that it minimizes IMSE($\hat{F}_{LS}(\cdot|x)$). We see that in all the presented cases $\hat{F}_{LS}(y|x)$ outperforms $\tilde{F}_{LS}(y|x)$. The cases with smaller $x$ (bigger dependence between $Y$ and $C$ given $X$) result in bigger outperformance, which is expected since for small values of $x$ we are far from the independence assumption needed to use $\hat{F}_{LS}(y|x)$. Figure 3.6 shows again the MSE of the estimator $\tilde{F}_{LS}(y|x)$ and the estimator $\hat{F}_{LS}(y|x)$, but using the data from the modified first setting, where we assume that $Y$ and $C$ are independent given $X$ (i.e. in (3.5.1) we use the independent copula). As expected, we see that the MSE of the estimators $\tilde{F}_{LS}(y|x)$ and $\hat{F}_{LS}(y|x)$ are virtually identical, with
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

the ratio of the approximated IMSE of $\hat{F}_{LS}(\cdot|x)$ and $\tilde{F}_{LS}(\cdot|x)$ approaching 1, as the sample size increases.

3.6 Appendix A

Appendices A and B contain the results needed for the proof of Theorem 3.3.1. Lemma 3.6.1 is a small technical result that helps to break down $\hat{F}_c(y) - F_c(y)$ into a sum of terms each depending only on one plug-in estimator, and remainder terms (see (3.4.6)). The two main lemmas in this section are Lemmas 3.6.2 and 3.6.3, which derive asymptotic representations for the main terms in the representation of $\hat{F}_c(y) - F_c(y)$, while Appendix B contains the results needed for showing the negligibility of the remainder terms.

In the proofs of Lemmas 3.6.2 and 3.6.3 we use empirical process theory. The result that we use most frequently is Lemma 3.6.5, which is an adaptation of a result in Giné and Nickle (2008) to our framework. This result is allowing us to asymptotically approximate sums of the form $\sum_{i=1}^n \int \hat{g}_y(x, V_i) W_{ni}(x, h_n) \times dF_X(x) = \frac{1}{n} \sum_{i=1}^n \hat{g}_y(X_i, V_i) + o_P(n^{-1/2})$, where $V_i = (X_i, T_i)$. Lemma 3.6.4 is used to show the Donsker property of classes of functions appearing in the proofs of Lemmas 3.6.2 and 3.6.3.

Before proceeding, we define a certain set of functions that will be used throughout Appendix A:

$$
C_M^\alpha(R_X) := \{ f : R_X \mapsto R : \| f \|_\alpha < M \},
\tilde{C}_M^\alpha(R_X) := \{ f \in C_M^\alpha(R_X) : \inf_{x \in R_X} f(x) > b_{inf} \},
$$

where $0 < M < \infty$, $b_{inf} = \inf_{x \in R_X} \sigma(x)/2$ and

$$
\| f \|_\alpha := \max_{k \leq [\alpha]} \sup_{x \in R_X} |f^{(k)}(x)| + \sup_{x_1, x_2 \in R_X} |f^{[\alpha]}(x_1) - f^{[\alpha]}(x_2)| |x_1 - x_2|^{-\alpha - [\alpha]},
$$

where $[\alpha]$ is the greatest integer smaller then $\alpha$.

The following lemma gives a specific form for the error distribution $F_c$, which we use to construct the estimator $\hat{F}_c$.

Lemma 3.6.1. Assume that $H(y|x)$ and $H^u(y|x)$ satisfy (B5)(ii) and let $\phi'_x(u)$ exist for $(x, u) \in R_X \times (0,1)$ . Then for every $y \leq \tau_y$,

$$
\mathcal{F}_c(y) = \phi^{-1}_y \left\{ - \int_{B_y} \int_{-\infty}^y \phi'_x(u) (\mathcal{P}_c(s|x)) dH_c^u(s|x) dF_X(x) \right\}.
$$
Lemma 3.6.2. [Asymptotic representation for \( \Lambda_1(y) \)]

Assume (B1)-(B4), and assume that (B5) holds for \( \Lambda_1 \). Let \( \hat{\Lambda}_1 \) be defined as in (3.4.6). Then, using relation (3.2.1), we can easily show that \( \hat{\Lambda}_1 \) is a Donsker class of functions.

Proof. Define

\[
\hat{\Lambda}_1(y) = \frac{1}{\tau_\eta} \sum_{i=1}^{\tau_\eta} \left( X_i, T_i, \Delta_i \right) + R_n(y),
\]

where \( g^{(1)}_y(X, T, \Delta) = I(X \in B_y) \int_0^T \frac{\hat{g}(s)}{\hat{\sigma}(X)} ds \) and \( \hat{\Lambda}_1 \) is defined in (3.4.6). Furthermore, the class of functions \( \mathcal{G}^{(1)} = \{ (x, t, \delta) : -\infty < y \leq \tau_\eta \} \) is a Donsker class of functions.

Proof. By using the notation

\[
g^{[1]}_{y, m, \hat{\sigma}}(x_1, x_2, t) = I(x_1 \in B_y) \times \int_{-\infty}^T \frac{\hat{g}(s)}{\hat{\sigma}(x_1)} \left[ H_c(s|x_1) - I \left( t - \hat{\sigma}(x_2) \right) \right] h^2(s|x_1) ds,
\]
where $g$

Now, thanks to Lemma 3.6.5 we have

To deal with the term corresponding class of functions

The second term on the right hand side is of the desired form, and the corre-

By adding and subtracting terms we get,

we can write

$$
\hat{\Lambda}_1(y) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B_y) \int_{-\infty}^{y} \phi_{X_i}^{n}(\mathcal{P}_e(s|X_i)) \left[ H_e(s|X_i) - I(\hat{E}_i \leq s) \right] \Phi^{n}(s) ds.
$$

Now, thanks to Lemma 3.6.5 we have

$$
\hat{\Lambda}_1(y) = \alpha P(n^{-1/2})
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B_y) \int_{-\infty}^{y} \phi_{X_i}^{n}(\mathcal{P}_e(s|X_i)) \left[ I(E_i \leq s) - I(\hat{E}_i \leq s) \right] \Phi^{n}(s) ds.
$$

By adding and subtracting terms we get,

$$
\hat{\Lambda}_1(y) = \alpha P(n^{-1/2})
$$

The second term on the right hand side is of the desired form, and the corre-

The class of functions $\mathcal{G}^{[2]} := \left\{ (x,t) \mapsto g_{y,m_1,\sigma_1}^{[2]}(x,t) : y \leq \gamma_n, \ m_1 \in C^{1+\delta}_M, \ \sigma_1 \in \mathbb{C}^{1+\delta}_M \right\}$,

where $g_{y,m_1,\sigma_1}^{[2]}(x,t) = I(x \in B_y) \int_{-\infty}^{y} \phi_{X_i}^{n}(\mathcal{P}_e(s|X_i)) I(\frac{t-m_1(x)}{\sigma_1(x)} \leq s) - I(\frac{t-m_1(x)}{\sigma_1(x)} \leq s) \Phi^{n}(s) ds$. The class of functions $\mathcal{G}^{[2]}$ is Donsker by Remark 3.6.1. Therefore, by Corollary 2.3.12 in the book of Van der Vaart and Wellner (1996) we have

$$
\lim_{\alpha \downarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{g^{[2]} \in \mathcal{G}^{[2]} \atop \mathbb{V} \mathbb{a} r(g^{[2]}) < \alpha} \left| \sum_{i=1}^{n} g^{[2]}(X_i, T_i) - E[g^{[2]}(X, T)] \right| > \epsilon \right) = 0,
$$

we have

$$
\lim_{\alpha \downarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{g^{[2]} \in \mathcal{G}^{[2]} \atop \mathbb{V} \mathbb{a} r(g^{[2]}) < \alpha} \left| \sum_{i=1}^{n} g^{[2]}(X_i, T_i) - E[g^{[2]}(X, T)] \right| > \epsilon \right) = 0,
$$

where $g^{[2]}(x,t) = I(x \in B_y) \int_{-\infty}^{y} \phi_{X_i}^{n}(\mathcal{P}_e(s|X_i)) I(\frac{t-m_1(x)}{\sigma_1(x)} \leq s) - I(\frac{t-m_1(x)}{\sigma_1(x)} \leq s) \Phi^{n}(s) ds$. The class of functions $\mathcal{G}^{[2]}$ is Donsker by Remark 3.6.1. Therefore, by Corollary 2.3.12 in the book of Van der Vaart and Wellner (1996) we have
for every \( \varepsilon > 0 \). Since it follows from Propositions 2.3.5, 2.3.6 and 2.3.7 in Chapter 2 that \( \lim_{n \to \infty} P(g_{y, \hat{\sigma}, \hat{\theta}}^{[2]} \in \mathcal{G}^{[2]}) = 1 \), and since it can be easily verified that \( \text{Var}(g_{y, \hat{\sigma}, \hat{\theta}}^{[2]}(X, T)) = o(1) \) a.s., we can approximate \( g_{y, \hat{\sigma}, \hat{\theta}}^{[2]} \) in \( \hat{\Lambda}_{11}(y) \) by its expectation:

\[
\hat{\Lambda}_{11}(y) - o_P \left( n^{-1/2} \right) = E \left\{ I(\hat{X} \in B_y) \int_{-\infty}^{y} \phi_X^{n}(\Pi_c(s|X)) \left[ I(\hat{E} \leq s) - I(\hat{E} \leq s) \right] h_{\xi}(s|X)ds \right\}.
\]

Further, we calculate

\[
\hat{\Lambda}_{11}(y) - o_P \left( n^{-1/2} \right)
= \int_{-\infty}^{y} \int_{B_y} \phi_X^{n}(\Pi_c(s|x)) h_{\xi}(s|x) \times \left[ P(E \leq s|X = x) - P \left( \hat{E} \leq s|\hat{X}_n, X = x \right) \right] dF_X(x)ds
= \int_{-\infty}^{y} \int_{B_y} \phi_X^{n}(\Pi_c(s|x)) h_{\xi}(s|x) \times \left[ H_{\xi}(s|x) - H_{\xi} \left( \frac{s\hat{\sigma}(x) + \hat{m}(x) - m(x)}{\sigma(x)} \right) \right] dF_X(x)ds
= - \frac{1}{2} \int_{-\infty}^{y} \int_{B_y} \phi_X^{n}(\Pi_c(s|x)) h_{\xi}(s|x) \times \frac{h'_x(s|x)}{\sigma(x)^2} \left\{ s[\hat{\sigma}(x) - \sigma(x)] + \hat{m}(x) - m(x) \right\} dF_X(x)ds
\]

where \( s_1 \) is between \( s \) and \( \sigma(x)^{-1}[s\hat{\sigma}(x) + \hat{m}(x) - m(x)] \), and \( \mathcal{X}_n = \{ (T_i, \Delta_i, X_i) : i = 1, \ldots, n \} \). Because of assumption (B5)(iii) we have that \( \sup_{y, x} |y^2 h_x'(y|x)| < \infty \), so the second term above is of order \( O((nh_n)^{-1} \log n) \) a.s. by Proposition 2.3.5 in Chapter 2. By using the asymptotic representation for \( \hat{m} \) and \( \hat{\sigma} \) given
in Proposition 2.3.8 in Chapter 2, we get
\[
\hat{\Lambda}_{11}(y) - o_P \left( n^{-1/2} \right) = \frac{1}{n} \sum_{i=1}^{n} \int_{B_y} \int_{-\infty}^{y} \phi''_x(\overline{T}_e(s|x))
\]
\[
\times \frac{h_x(s|x)}{\sigma(x)} \left[ s\zeta(T_i, \Delta_i|x) + \eta(T_i, \Delta_i|x) \right] \frac{K \left( \frac{x - X_i}{h_n} \right)}{h_n} d\overline{H}_n^o(s|x) dx
d\]
\[
:= \frac{1}{n} \sum_{i=1}^{n} \int_{B_y} g_{y}^{[3]}(X_i + uh_n, T_i, \Delta_i)K(u) du.
\]

We can write
\[
g_{y}^{[3]}(x, t, \delta) = -I(x \in B_y) \int_{-\infty}^{y} g_s(x, t, \delta) d\overline{H}_n^o(s|x),
\]
where \( g_s(x, t, \delta) = \phi''_x(\overline{T}_e(s|x)) \frac{h_x(s|x)}{\sigma(x)} \left[ s\zeta(t, \delta|x) + \eta(t, \delta|x) \right] \). By assumptions (B4) and (B5)(ii,iii) we have that for every \( s \leq y \leq \tau_\eta \) the function \( x \mapsto \phi''_x(\overline{T}_e(s|x)) \frac{h_x(s|x)}{\sigma(x)} \) is uniformly bounded (by a constant not depending on \( s \) and \( y \)) in \( x \in B_y \), as well as the corresponding first derivative. Therefore, using the notation \( \Omega = \{ x \mapsto \phi''_x(\overline{T}_e(s|x)) \frac{h_x(s|x)}{\sigma(x)} : s \leq \tau_\eta \} \), we have that the bracketing number \( N_{[1]}(\Omega, L_2(P)) \) equals \( \exp(-C\varepsilon) \), for some constant \( C \) (see Corollary 2.7.2 in Van der Vaart and Wellner (1996)). By a similar reasoning we have that the bracketing number of the class \( \Omega' = \{ x \mapsto \phi''_x(\overline{T}_e(s|x)) \frac{h_x(s|x)}{\sigma(x)} : s \leq \tau_\eta \} \) equals \( \exp(-C\varepsilon) \). Now, \( \Omega \) and \( \Omega' \) are Donsker by Theorem 2.5.6 in Van der Vaart and Wellner (1996). Hence, Remark 3.6.1 entails that \( G^{[3]} = \{ (x, t, \delta) \mapsto g_{y}^{[3]}(x, t, \delta) : -\infty < y \leq \tau_\eta \} \) is a Donsker class of functions, while Lemma 3.6.5 entails that
\[
\hat{\Lambda}_{11}(y) = \frac{1}{n} \sum_{i=1}^{n} g_{y}^{[3]}(X_i, T_i, \Delta_i) + o_P \left( n^{-1/2} \right).
\]

Now, by using (3.6.2) we get
\[
\hat{\Lambda}_1(y) = \frac{1}{n} \sum_{i=1}^{n} g_{y}^{[1]}(X_i, T_i, \Delta_i) + o_P \left( n^{-1/2} \right).
\]

The class of functions \( G^{[1]} \) is a Donsker class as it is a sum of two Donsker classes (see Example 2.10.7 in Van der Vaart and Wellner (1996)), which completes the proof.

\[\Box\]

**Lemma 3.6.3.** [Asymptotic representation for \( \hat{\Lambda}_2(y), \hat{\Lambda}_3(y) \) and \( \hat{\Lambda}_4(y) \)]

Assume (B1)-(B4), and assume that (B5) holds for \( H(y|x) \) and \( H^o(y|x) \). Then
for $y \leq \tau_\eta$,

$$\hat{\Lambda}_2(y) = \frac{1}{n} \sum_{i=1}^{n} g^{(2)}(X_i, T_i, \Delta_i) + R_n(y),$$

$$\hat{\Lambda}_3(y) = \frac{1}{n} \sum_{i=1}^{n} g^{(3)}(X_i, T_i, \Delta_i),$$

$$\hat{\Lambda}_4(y) = \frac{1}{n} \sum_{i=1}^{n} g^{(4)}(X_i, T_i, \Delta_i),$$

where

$$g^{(2)}(X, T, \Delta) = I(X \in B_y) \int_{-\infty}^{y} \phi_X(T_e(s|X))d\left[ \frac{h^w(s|X)}{\sigma(x)} [s\zeta(T, \Delta|X) + \eta(T, \Delta|X)] + I(E \leq s, \Delta = 1) - H^w_c(s|X) \right],$$

$$g^{(3)}(X, T, \Delta) = I(X \in B_y) \int_{-\infty}^{y} \phi_X(T_e(s|X))dH^w_c(s|X)$$

$$- \int_{B_y} \int_{-\infty}^{y} \phi'_x(T_e(s|x))dH^w_c(s|x)dF_X(x),$$

$$g^{(4)}(X, T, \Delta) = I(X \in B_y)\phi_X(F_e(y)) - \int_{B_y} \phi_x(F_e(y))dF_X(x),$$

$$\sup_{y \leq \tau_\eta} |R_n(y)| = o_P(n^{-1/2}),$$

and the terms $\hat{\Lambda}_2(y)$, $\hat{\Lambda}_3(y)$ and $\hat{\Lambda}_4(y)$ are defined in (3.4.6). Furthermore, the class of functions $G^{(i)} = \{(x, t, \delta) \mapsto g^{(i)}_y(x, t, \delta) : -\infty < y \leq \tau_\eta \}$ is a Donsker class of functions for every $i = 2, 3, 4$.

**Proof.** The expressions of $\hat{\Lambda}_3(y)$ and $\hat{\Lambda}_4(y)$ are obtained from simple algebra. Showing the Donsker property of $G^{(3)}$ and $G^{(4)}$ is very similar, and therefore we will only show it for $G^{(3)}$. The first factor of the first term in $g^{(3)}_y(x, t, \delta)$ belongs to a class of functions that is Donsker by construction (see Section 3.2). The second factor of the first term in $g^{(3)}_y(x, t, \delta)$ is monotone in the parameter $y$ and uniformly bounded in $y \leq \tau_\eta$. Therefore, its $\varepsilon$-bracketing number is bounded by $C\varepsilon^{-1}$ (see Theorem 2.7.5 in Van der Vaart and Wellner (1996)), which by Theorem 2.5.6 in Van der Vaart and Wellner (1996) implies the Donsker property. The second term of $g^{(3)}_y(x, t, \delta)$ is a deterministic, uniformly bounded function in $y \leq \tau_\eta$ by (B4), and therefore trivially Donsker. Since multiplying and adding uniformly bounded functions preserves the Donsker property (by Examples 2.10.7 and 2.10.8 in Van der Vaart and Wellner (1996)), the class of
functions \( G^{(i)} \) is also Donsker. To deal with the expression of \( \hat{\Lambda}_{21}(y) \) we start by using integration by parts:

\[
\hat{\Lambda}_{21}(y) = \int_{B_y} \phi'_x(\overline{H}_e(y|x)) \left[ \hat{H}_e^n(y|x) - H_e^n(y|x) \right] dF_X(x)
\]

\[
+ \int_{B_y} \int_y \phi'_x(\overline{H}_e(s|x)) h_e(s|x) \left[ \hat{H}_e^n(s|x) - H_e^n(s|x) \right] ds dF_X(x)
\]

\[
= \hat{\Lambda}_{21}(y) + \hat{\Lambda}_{22}(y).
\]

The first term \( \hat{\Lambda}_{21}(y) \) can be written as:

\[
\hat{\Lambda}_{21}(y) = \sum_{i=1}^n \int_{B_y} \phi'_x(\overline{H}_e(y|x)) I(\Delta_i = 1) \left[ I(\hat{E}_i \leq y) - I(E_i \leq y) \right] W_{ni}(x, h_n) dF_X(x)
\]

\[
+ \sum_{i=1}^n \int_{B_y} \phi'_x(\overline{H}_e(y|x)) [I(E_i \leq y)I(\Delta_i = 1) - H^n_e(y|x)] W_{ni}(x, h_n) dF_X(x).
\]

As in the proof of Lemma 3.6.2, we can apply Lemma 3.6.5 to get

\[
\hat{\Lambda}_{21}(y) - o_P \left( n^{-1/2} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^n I(X_i \in B_y) \phi'_x(\overline{H}_e(y|X_i)) I(\Delta_i = 1) \left[ I(\hat{E}_i \leq y) - I(E_i \leq y) \right]
\]

\[
+ \frac{1}{n} \sum_{i=1}^n I(X_i \in B_y) \phi'_x(\overline{H}_e(y|X_i)) [I(\Delta_i = 1)I(E_i \leq y) - H^n_e(y|X_i)]
\]

\[
: = \hat{\Lambda}_{211}(y) + \hat{\Lambda}_{212}(y). \tag{3.6.3}
\]

The second term on the right hand side is of the desired form. Now, to deal with the first term on the right hand side we repeat similar calculations as in the proof of Lemma 3.6.2 for \( \hat{\Lambda}_{11} \) involving Corollary 2.3.12 in Van der Vaart and Wellner (1996) and the i.i.d representation of \( \hat{m} \) and \( \hat{s} \) to get

\[
\hat{\Lambda}_{211}(y) - o_P \left( n^{-1/2} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int_{B_y} \phi'_x(\overline{H}_e(y|x)) \frac{h^n_e(y|x)}{\sigma(x)} \left[ y\zeta(T_i, \Delta_i|x) + \eta(T_i, \Delta_i|x) \right] K \left( \frac{x-X_i}{h_n} \right) dx.
\]

Again we can apply Lemma 3.6.5 to replace integrals with indicators:

\[
\hat{\Lambda}_{211}(y) - o_P \left( n^{-1/2} \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^n \phi'_x(\overline{H}_e(y|X_i)) \frac{h^n_e(y|X_i)}{\sigma(X_i)} I(X_i \in B_y) [y\zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i)]. \tag{3.6.4}
\]
3.6. Appendix A

To deal with the term $\hat{\Lambda}_{22}(y)$ we calculate

$$\hat{\Lambda}_{22}(y) = \int_{-\infty}^{y} \int_{B_y} \phi_{\epsilon}^{\mu}(\mathcal{H}_c(s|x))h_c(s|x) \left[ \hat{H}_c^{\mu}(s|x) - H_c^{\mu}(s|x) \right] dF_X(x)ds$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{y} \int_{B_y} \phi_{\epsilon}^{\mu}(\mathcal{H}_c(s|x))h_c(s|x)I(\Delta_i = 1)$$

$$\times \left[ I(\hat{E}_i \leq s) - I(E_i \leq s) \right] W_{ni}(x,h_n)dF_X(x)ds$$

$$+ \sum_{i=1}^{n} \int_{B_y} \int_{-\infty}^{y} \phi_{\epsilon}^{\mu}(\mathcal{H}_c(s|x))h_c(s|x)$$

$$\times \left[ I(\Delta_i = 1)I(E_i \leq s) - H_c^{\mu}(s|x) \right] W_{ni}(x,h_n)dF_X(x)ds.$$

Again we can apply Lemma 3.6.5 to get

$$\hat{\Lambda}_{22}(y) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{y} I(X_i \in B_y)\phi_{X_i}^{\mu}(\mathcal{H}_c(s|X_i))h_c(s|X_i)$$

$$\times I(\Delta_i = 1) \left[ I(\hat{E}_i \leq s) - I(E_i \leq s) \right] ds$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \int_{B_y} \int_{-\infty}^{y} I(X_i \in B_y)\phi_{X_i}^{\mu}(\mathcal{H}_c(s|X_i))h_c(s|X_i)$$

$$\times \left[ I(\Delta_i = 1)I(E_i \leq s) - H_c^{\mu}(s|X_i) \right] ds$$

$$:= \hat{\Lambda}_{221}(y) + \hat{\Lambda}_{222}(y). \quad (3.6.5)$$

The second term on the right hand side is of the desired form. Now to deal with the first term on the right hand side we repeat similar calculations as in the proof of Lemma 3.6.2 for $\hat{\Lambda}_{11}$ which give us

$$\hat{\Lambda}_{221}(y) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{y} \int_{B_y} \phi_{X_i}^{\mu}(\mathcal{H}_c(s|x))h_c(s|x)$$

$$\times \left[ s \zeta(T_i, \Delta_i|x) + \eta(T_i, \Delta_i|x) \right] \frac{1}{h_n} \frac{h_c^{\mu}(s|x)}{\sigma(x)} dx + o_P \left( n^{-1/2} \right).$$

Again we can apply Lemma 3.6.5 to get

$$\hat{\Lambda}_{221}(y) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{y} I(X_i \in B_y)\phi_{X_i}^{\mu}(\mathcal{H}_c(s|X_i))h_c(s|X_i)$$

$$\times \frac{h_c^{\mu}(s|X_i)}{\sigma(X_i)} \left[ s \zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i) \right] ds + o_P \left( n^{-1/2} \right). \quad (3.6.6)$$
By combining (3.6.3), (3.6.4), (3.6.5) and (3.6.6) we have

\[
\hat{\Lambda}_2(y) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B_y) \phi'_{X_i}(\overline{H}_{e_i}(y|X_i)) \times \left[ \frac{h_{e}(y|X_i)}{\sigma(X_i)} \right] [y \zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i)] + I(E_i \leq y) I(\Delta = 1) - H_{e_i}(y|X_i) \\
+ \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B_y) \int_{-\infty}^{y} \phi''_{X_i}(\overline{H}_{e_i}(s|X_i)) \\
\times \left[ \frac{h_{e}(s|X_i)}{\sigma(X_i)} \right] [s \zeta(T_i, \Delta_i|X_i) + \eta(T_i, \Delta_i|X_i)] + I(\Delta = 1) I(E_i \leq s) - H_{e_i}(s|X_i)] ds + o_P\left(n^{-1/2}\right)
\]

The last equality follows by applying of integration by parts. This proofs that \( \hat{\Lambda}_2(y) \) is of the desired form. Now, by looking at the form of \( g^{(2)} \) in the first equality, the Donsker property of \( g^{(2)} \) follows from applying Remark 3.6.1 below.

**Remark 3.6.1.** The class of functions that is obtained by adding the classes of functions, in Lemma 3.6.4, is again Donsker from the fact that adding classes preserves the Donsker property (see Example 2.10.7 in Van der Vaart and Wellner (1996)).

The following lemma shows the Donsker property for the classes of functions that are showing up in the proofs of Lemmas 3.6.2, 3.6.3 and 3.6.5.

**Lemma 3.6.4.** Assume (B1)-(B4), and assume that (B5) holds for \( H(y|x) \)
and $H^u(y|x)$. Then, the following classes of functions are Donsker:

\[ \mathcal{G}_1 = \left\{ (x_1, x_2, t) \mapsto g_{y, m_1, \sigma_1}(x_1, x_2, t) : 
\quad y \leq \tau_y, \ m_1 \in C_1^{1+\delta}(R_X), \ \sigma_1 \in \tilde{C}_1^{1+\delta}(R_X) \right\}, \]

\[ \mathcal{G}_2 = \left\{ (x_1, x_2, t) \mapsto g_{y, m_1, \sigma_1, 2}(x_1, x_2, t) : 
\quad y \leq \tau_y, \ m_1 \in C_1^{1+\delta}(R_X), \ \sigma_1 \in \tilde{C}_1^{1+\delta}(R_X) \right\}, \]

\[ \mathcal{G}_3 = \left\{ (x, t, \delta) \mapsto g_{y, \psi, 3}(x, t, \delta) : y \leq \tau_y, \psi \in \{\zeta, \eta\} \right\}, \]

\[ \mathcal{G}_4 = \left\{ (x, t, \delta) \mapsto g_{y, \psi, 4}(x, t, \delta) : y \leq \tau_y, \psi \in \{\zeta, \eta\} \right\}, \]

where

\[ g_{y, m_1, \sigma_1, 1}(x_1, x_2, t) = I(x_1 \in B_y) \int_{-\infty}^{y} q_s(x_1) \left[ I \left( \frac{t - m_1(x_2)}{\sigma_1(x_2)} \leq s \right) - H_c(s|x_1) \right] h^u_s(s|x_1) ds, \]

\[ g_{y, m_1, \sigma_1, 2}(x_1, x_2, t) = I(x_1 \in B_y) q_y(x_1) \left[ I \left( \frac{t - m_1(x_2)}{\sigma_1(x_2)} \leq y \right) - H_c(y|x_1) \right] h^u(y|x_1), \]

\[ g_{y, \psi, 3}(x, t, \delta) = I(x \in B_y) \int_{-\infty}^{y} q_s(x) \psi(t, \delta|x) h^u(s|x) ds, \]

\[ g_{y, \psi, 4}(x, t, \delta) = I(x \in B_y) q_y(x) \psi(t, \delta|x) h^u(y|x), \]

the sets $C_1^{1+\delta}(R_X)$ and $\tilde{C}_1^{1+\delta}(R_X)$ are defined in (3.6.1), and the class \( \{x \mapsto q_y(x) : y \leq \tau_y\} \) is a Donsker class of non-negative (non-positive) functions, such that $\sup_{y \leq \tau_y} \sup_{x \in B_y} (|q_y(x)|, \frac{\partial}{\partial x} q_y(x)) < \infty$.

**Proof.** First, we will prove that the class of functions $\mathcal{G}_2$ is Donsker. Note that the class of functions \( \{(x_1, x_2, t) \mapsto I(\frac{t - m_1(x_2)}{\sigma_1(x_2)} \leq y) - H_c(y|x_1) : y \leq \tau_y, \ m_1 \in C_1^{1+\delta}(R_X), \ \sigma_1 \in \tilde{C}_1^{1+\delta}(R_X)\} \) is Donsker by calculations in Lemma A.1 in Van Keilegom and Akritas (1999). The class of functions \( \{x \mapsto I_B(x) : B \in \mathfrak{B}\} \) is Donsker (see Section 3.2). Hence, the class of functions $\mathcal{G}_2$ is Donsker since it is a product of uniformly bounded, Donsker classes of functions (see Example 2.10.8 in the book of Van der Vaart and Wellner (1996)). Since it is easy to see that the functions $\zeta$ and $\eta$ are uniformly bounded, $\mathcal{G}_4$ is a product of
uniformly bounded, Donsker classes of functions, and therefore Donsker itself (see Example 2.10.8 in Van der Vaart and Wellner (1996)).

Proving that $\mathcal{G}_1$ and $\mathcal{G}_3$ are Donsker is similar, therefore we will only prove that $\mathcal{G}_1$ is Donsker, which is the hardest of the two. We will use results from the book of Van der Vaart and Wellner (1996). By Theorem 2.5.6 in their book it is sufficient to show that

$$\int_0^\infty \sqrt{\log N_\varepsilon} (\varepsilon, g, \| \cdot \|) d\varepsilon < \infty. \quad (3.6.7)$$

We will restrict ourselves to showing (3.6.7) for the class of functions $\mathcal{G}_1^1 = \{(x_1, x_2, t) \mapsto g_{y, m_1, \sigma_1}(x_1, x_2, t) : y \leq \tau_0, m_1 \in C_\lambda^1 R_X, \sigma_1 \in \tilde{C}_\lambda^1 R_X \}$, where $g_{y, m_1, \sigma_1}(x_1, x_2, t) = \int_0^\infty q_\varepsilon(x_1) I\left( \frac{t - m_1(x_2)}{\sigma_1(x_2)} \leq s \right) h^\varepsilon(s|x_1) ds$. By Theorem 2.7.1 in the aforementioned book we know that $C_\lambda^1 R_X$ and $\tilde{C}_\lambda^1 R_X$ can be covered by $M_1 = \exp(C_1 \varepsilon^{-1/(1+\delta)})$ and $M_2 = \exp(C_2 \varepsilon^{-1/(1+\delta)})$ $\varepsilon$-brackets with respect to the supremum norm, respectively. Let $\{[m_j^1, m_j^2] : j = 1, \ldots, M_1 \}$ and $\{[\sigma_k^1, \sigma_k^2] : k = 1, \ldots, M_2 \}$ be those $\varepsilon$-brackets. Let $x_1, \ldots, x_{M_3}$ be a grid of $R_X$ such that $x_{r+1} - x_r \leq \varepsilon$, $r = 1, \ldots, M_3 - 1$ and let $\{y_{ri} : r = 1, \ldots, M_3, i = 1, \ldots, M_4 \}$ be such that $H^\varepsilon_k(y_{ri+1}|x_j) - H^\varepsilon_k(y_{ri}|x_j) \leq \varepsilon$. Let $\{y_i : i = 1, \ldots, M_5 \}$ be the union of all $y_{ri}$ in ascending order. There are $M_5 = C \varepsilon^{-2}$ of them. Now, we define brackets $\{[y_i, m_j^1, \sigma_k^1, \sigma_k^2] \in i,j,k$. There are at most $C \varepsilon^{-2} \exp\{C \varepsilon^{-1/(1+\delta)}\}$ of them. Hence, they satisfy condition (3.6.7) and they cover $\mathcal{G}_1^1$, because $g_{y, m, \sigma}(\cdot, \cdot)$ is a monotone function of its parameters. Now, to show that $\mathcal{G}_1^1$ is Donsker, we only need to show that the defined brackets are $\varepsilon$-brackets:

$$\|g_{y, m_1^1, \sigma_1^1} - g_{y, m_1^2, \sigma_1^2}\|_\infty \leq \|g_{y, m_1^1, \sigma_1^2} - g_{y, m_1^1, \sigma_1^1}\|_\infty + \|g_{y, m_1^2, \sigma_1^1} - g_{y, m_1^2, \sigma_1^2}\|_\infty. \quad (3.6.8)$$

We start with the second term:

$$\|g_{y, m_1^1, \sigma_1^1} - g_{y, m_1^2, \sigma_1^2}\|_\infty \leq C \sup_{t \leq \tau_0, x_1, x_2 \in R_X} \left| I \left( \frac{t - m_1^1(x_2)}{\sigma_1^1(x_2)} \leq s \right) - I \left( \frac{t - m_1^2(x_2)}{\sigma_1^2(x_2)} \leq s \right) \right| h^\varepsilon(s|x_1) ds \leq C \sup_{t \leq \tau_0, x_1, x_2 \in R_X} \left| H^\varepsilon_k \left( \frac{t - m_1^1(x_2)}{\sigma_1^1(x_2)} \right) - H^\varepsilon_k \left( \frac{t - m_1^2(x_2)}{\sigma_1^2(x_2)} \right) \right| x_1 \leq \|H^\varepsilon_k\|_\infty \sup_{x_2 \in R_X} \frac{|m_1^1(x_2) - m_1^2(x_2)|}{\sigma_1^2(x_2)} \leq C \|m_1^1 - m_1^2\|_\infty \leq C \varepsilon.$$

Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models
The fourth inequality follows from (B5)(ii) for $H^u_{c}$, and the fact that $\sigma_k$ is bounded away from zero as a function belonging to $\tilde{C}_{1+\delta}^1(R_X)$. Similarly, we can bound the first term:

$$
\|g_{y,m_j^i,\sigma_k^i} - g_{y,m_j^i,\sigma_k^i}\|_{\infty} \leq \|h^u_{c}\|_{\infty} \sup_{y \leq \tau_n, x_2 \in R_X} \left| \frac{(y - m_j^i(x_2))(\sigma_k^i(x_2) - \sigma_k^i(x_2))}{\sigma_k^i(x_2)} \right| \\
\leq C \|\sigma_k^i - \sigma_k^i\|_{\infty} \leq C\varepsilon.
$$

The second inequality follows from (B5)(ii) for $H^u_{c}$, and the fact that $\sigma_k$ and $\sigma_1$ are bounded away from zero as functions belonging to $\tilde{C}_{1+\delta}^1(R_X)$. For the third term we calculate

$$
\|g_{y,m_j^i,\sigma_k^i} - g_{y+1,m_j^i,\sigma_k^i}\|_{\infty} \\
\leq C \sup_{x \in R_X} \int_{y}^{y+1} h^u_{c}(s|x)ds \\
\leq C \sup_{x \in R_X} |H^u_{c}(y+1|x) - H^u_{c}(y|x)| \\
\leq C \max_{j} |H^u_{c}(y_j+1|x_j) - H^u_{c}(y_j|x_j)| + 2\sup_{y,x} |h^u_{c}(y|x)|\varepsilon \leq C\varepsilon.
$$

To get the last inequality we used the uniform boundedness of $h^u_{c}$. Now, we have that expression (3.6.8) is bounded by $C\varepsilon$. Because $C$ is independent of the bracket selection (i.e. of $\varepsilon$), we have that the brackets $\{[g_{y,m_j^i,\sigma_k^i}, g_{y+1,m_j^i,\sigma_k^i}] : i, j, k\}$ are $\varepsilon$-brackets, therefore by the previous reasoning $\mathcal{G}_1^2$ (and hence $\mathcal{G}_1^2$) is Donsker.

The next lemma is an adaptation of the result on convergence of smoothed processes in Giné and Nickle (2008) to our framework. We will use this result in the proofs of Lemmas 3.6.2 and 3.6.3 to asymptotically replace expressions of the form $\sum_{i=1}^{n} \int \tilde{g}_y(x, V_i) W_m(x, h_n) dF_X(x)$ by the simpler expressions $\frac{1}{n} \sum_{i=1}^{n} \tilde{g}_y(X_i, V_i)$, where $V_i = (X_i, T_i)$.

**Lemma 3.6.5.** Assume the conditions and notations of Lemma 3.6.4. Then, using the notation $\tilde{g}_y,r := g_{y,\bar{m},\bar{\sigma},r}$, for $r = 1, 2$, and $\tilde{g}_y,r := g_{y,\psi,r}$, for $r = 3, 4$, we have

$$
\sup_{y \leq \tau_n} \left| \frac{1}{n} \sum_{i=1}^{n} \int \tilde{g}_y,r(X_i + uh_n, V_i) K(u) du - \tilde{g}_y,r(X_i, V_i) \right| = o_P \left( n^{-1/2} \right),
$$

(3.6.9)

$$
\sup_{y \leq \tau_n} \left| \frac{1}{n} \sum_{i=1}^{n} \int \tilde{g}_y,r(t, V_i) W_m(t, h_n) f_X(t) dt - \frac{1}{n} \tilde{g}_y,r(X_i, V_i) \right| = o_P \left( n^{-1/2} \right).
$$

(3.6.10)
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

Proof. Before we start, we give the results that will be used throughout the proof and which will be shown at the end of this proof. For \( r = 1, 2, 3, 4 \) we have

\[
\sup_{g \in \mathcal{G}} \|g\|_{\infty} < \infty, \tag{3.6.11}
\]

\[
\lim_{n \to \infty} P \left( \left\{ \hat{g}_{y,r} : -\infty < y \leq \tau_n \right\} \subset \mathcal{G}_r \right) = 1, \tag{3.6.12}
\]

\[
\sup_{g \in \mathcal{G}_r} E \left[ \int g(X + uh_n, V) K(u) \, du - g(X, V) \right]^2 = o(1), \tag{3.6.13}
\]

\[
\sup_{y \leq \tau_n} \left| E \left[ \int \hat{g}_{y,r}(X + uh_n, V) K(u) \, du - \hat{g}_{y,r}(X, V) \mid X_n \right] \right| = o(n^{-1/2}) \quad \text{a.s.}, \tag{3.6.14}
\]

where \( V = (X, T) \) and \( X_n = \{(X_i, T_i, \Delta_i) : i = 1, \ldots, n\} \). The aforementioned results also hold when we replace \( \hat{g}_{y,r} \) and \( \mathcal{G}_r \) by \( \hat{g}'_{y,r} = \hat{g}_{y,r} \frac{f_X - \bar{f}_X}{\bar{f}_X} \) and \( \mathcal{G}'_r := \mathcal{G}_r \times C^{1+\delta}(R_X) \), respectively, where \( \bar{f}_X(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \).

In the following calculations we will use the notation \( \hat{g}_{y,r} \) to represent \( \hat{g}_{y,r} \), for all \( r = 1, 2, 3, 4 \), since the proof is the same in all cases. Similarly, for all \( r = 1, 2, 3, 4 \), we will use the notation \( g_r \) for \( g_r \in \mathcal{G}'_r \), and \( \mathcal{G}'_r \). Conditions (3.6.11) to (3.6.14) allow us to apply Theorem 2 (a) in Giné and Nickle (2008) to the term in (3.6.9), which gives

\[
\sup_{y \leq \tau_n} \left| \frac{1}{n} \sum_{i=1}^n \left\{ \int \hat{g}_{y}(t, V_i) \frac{1}{h_n} K\left(\frac{X_i - t}{h_n}\right) \, dt - \hat{g}_{y}(X_i, V_i) \right\} \right| = \sup_{y \leq \tau_n} \left| E \left[ \int \hat{g}_{y}(t, V) \frac{1}{h_n} K\left(\frac{X - t}{h_n}\right) \, dt - \hat{g}_{y}(X, V) \mid X_n \right] \right| + o_P(n^{-1/2}).
\]

The first term on the right hand side is \( o(n^{-1/2}) \) a.s. by (3.6.14), which implies (3.6.9). Next, to prove statement (3.6.10), we write

\[
Q(\hat{g}_y) := \sum_{i=1}^n \int \hat{g}_y(t, V_i) W_{ni}(t, h_n) f_X(t) \, dt
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int \hat{g}_y(t, V_i) \frac{f_X(t)}{f_X(t)} \frac{1}{h_n} K\left(\frac{X_i - t}{h_n}\right) \, dt
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int \hat{g}_y(t, V_i) \frac{1}{h_n} K\left(\frac{X_i - t}{h_n}\right) \, dt
\]

\[
+ \frac{1}{n} \sum_{i=1}^n \int \hat{g}_y(t, V_i) \frac{f_X(t)}{f_X(t)} \frac{1}{h_n} K\left(\frac{X_i - t}{h_n}\right) \, dt + o(n^{-1/2}) \quad \text{a.s.}
\]

\[
:= Q_1(\hat{g}_y) + Q_2(\hat{g}_y) + o(n^{-1/2}) \quad \text{a.s.}
\]
3.6. Appendix A

Because of conditions (3.6.11)-(3.6.14) for $\hat{g}'_y$ and $G'$, we can use (3.6.9) to write $Q_2(\hat{g}_y)$ as

$$Q_2(\hat{g}_y) = \frac{1}{n} \sum_{i=1}^{n} \hat{g}_y(X_i, V_i) f_X(X_i) - \hat{f}_X(X_i) f_X(X_i) + o_P \left( n^{-1/2} \right).$$

By Corollary 2.3.12 in the book of Van der Vaart and Wellner (1996) we have

$$\lim_{\alpha \downarrow 0} \lim \sup_{n \to \infty} P \left( \sup_{g' \in G', \var{g'} < \alpha} \left| \sum_{i=1}^{n} g'(X_i, V_i) - E (g'(X, V)) \right| > \varepsilon \right) = 0,$$

for every $\varepsilon > 0$. Since, $\lim_{n \to \infty} P(\hat{g}_y f_X - \hat{f}_X \in G') = 1$ and $\var{\hat{g}_y f_X - \hat{f}_X} = o(1)$, we can use (3.6.15) to approximate $Q_2(\hat{g}_y)$ with the corresponding expectation:

$$Q_2(\hat{g}_y) = E \left[ \hat{g}_y(X, V) f_X(X) - \hat{f}_X(X) f_X(X) \right | x_n] + o_P \left( n^{-1/2} \right)$$

$$= E \left[ E[\hat{g}_y(X, V)|X, x_n] f_X(X) - \hat{f}_X(X) f_X(X) \right | x_n] + o_P \left( n^{-1/2} \right)$$

$$= o_P \left( n^{-1/2} \right).$$

To obtain the last equality we used that $\|f_X - \hat{f}_X\|_\infty = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. and $\sup_{x \in R_x} |E[\hat{g}_y(X, V)|X = x, x_n]| = O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. (see Proposition 2.3.5 in Chapter 2). Now, applying (3.6.9) on $Q_1(\hat{g}_y)$ we get

$$Q(\hat{g}_y) = \frac{1}{n} \sum_{i=1}^{n} \hat{g}_y(X_i, V_i) + o_P \left( n^{-1/2} \right),$$

which concludes the proof.

Now, it remains to show that conditions (3.6.11), (3.6.12), (3.6.13) and (3.6.14) are satisfied (proving the corresponding conditions for $\hat{g}'_y$ and $G'$ will be omitted since it uses the same techniques and reasoning). Condition (3.6.11) is easily verified, while condition (3.6.12) follows from Propositions 2.3.5, 2.3.6 and 2.3.7 in Chapter 2. Because of condition (3.6.11) we can bound the expectation in condition (3.6.13) by

$$C \sup_{u \in [-a, a]} E[|g(X + uh_n, V) - g(X, V)|].$$
Chapter 3. The copula-graphic estimator in censored nonparametric location-scale regression models

Now, we decompose the function $g$ as $g(X, V) = I(X \in B_y)z(X, V)$ and write

\[ Q_y(uh_n) = E[E[g(X + uh_n, V)|X] - E[g(X, V)|X]] = E[I(X + uh_n \in B_y)E[z(X + uh_n, V)|X] - I(X \in B_y)E[z(X, V)|X]]. \]

By using assumptions (B4) and (B5)(iv,vi), it is easy to see that the function $(x, e) \mapsto E[z(x + e, V)|X = x]$ has uniformly bounded partial derivative over $e$, for $e$ small enough. This allows us to use a Taylor expansion and to obtain

\[ Q_y(uh_n) = E \{[I(X + uh_n \in B_y) - I(X \in B_y)]|E[z(X, V)|X]]\} + O(uh_n) \]

where $\Delta$ is the symmetric difference, $B_y = \bigcup_{i=1}^{k} B_{yi}$, the sets $B_{yi}$ are convex and $k \leq \lambda(R_X)/\beta < \infty$ (for the definition of $\beta$ see Section 3.2). This implies condition (3.6.13). To prove condition (3.6.14), we use the notation $E[\cdot|\cdot] = E[\cdot|X_n]$ and write the expectation in condition (3.6.14) as

\[ \int E\left[\tilde{g}_y(X + uh_n, V) - \tilde{g}_y(X, V)\right]K(u)du. \]

Using the decomposition $\tilde{g}_y(x, v) = I(x \in B_y)\tilde{z}_y(x, v)$, we can write the expectation in the integral above as

\[ Q_{\tilde{g}_y}(uh_n) = E \left\{ I(X + uh_n \in B_y) \left[ E[\tilde{z}_y(X + uh_n, V)|X] - E[\tilde{z}_y(X, V)|X] \right] \right\} 
   + E \left\{ I(X + uh_n \in B_y) - I(X \in B_y) \left[ E[\tilde{z}_y(X, V)|X] \right] \right\} 
   := Q_{\tilde{g}_y,1}(uh_n) + Q_{\tilde{g}_y,2}(uh_n). \]

In order to bound the second term on the right hand side, we use similar
calculations as when dealing with $Q_g(uh_n)$ to get

$$
|Q_{g,2}(uh_n)| \leq C \sum_{i=1}^{k} \int_{\sup B_{u_i} - uh_n}^{\sup B_{u_i} + uh_n} |E[\tilde{z}_g(x, V) | X = x]| dx \\
+ C \sum_{i=1}^{k} \int_{\inf B_{u_i} - uh_n}^{\inf B_{u_i} + uh_n} |E[\tilde{z}_g(x, V) | X = x]| dx \\
\leq 2Ck \sup_{y \leq \tau_n} \sup_{d(x, B_y) \leq uh_n} |E[\tilde{z}_g(x, V) | X = x]| uh_n \\
= O \left( (nh_n)^{-1/2}(\log n)^{1/2}h_n \right) = o(n^{-1/2}) \text{ a.s.,}
$$

uniformly in $y \leq \tau_n$ and $u \in [-a, a]$, where $d(x, B_y) = \inf \{|x - x_1| : x_1 \in B_y\}$. To get the first equality we used Proposition 2.3.5 in Chapter 2, and assumptions (B4) and (B5)(iv,vi). The term $Q_{g,1}(uh_n)$ can be bounded as follows:

$$
|Q_{g,1}(uh_n)| \leq \sup_{y \leq \tau_n} \sup_{d(x, B_y) \leq uh_n} \left| E \left[ \frac{\partial}{\partial x} E[\tilde{z}_g(x, V) | X = x] \right] uh_n \right| \\
= O \left( (nh_n)^{-1/2}(\log n)^{1/2}h_n \right) = o(n^{-1/2}) \text{ a.s.}
$$

uniformly in $y \leq \tau_n$ and $u \in [-a, a]$. To get the first equality we used again Proposition 2.3.5 in Chapter 2, and assumptions (B4) and (B5)(iv,vi). This implies (3.6.14). \qed

### 3.7 Appendix B

This section contains results on uniform convergence rates of quantities needed for showing that the remainder terms in (3.4.6) of Theorem 3.3.1 are $o(n^{-1/2})$ a.s.

**Lemma 3.7.1.** Assume (B1)(i,ii) and (B4). Then, for $\alpha \in (0, 1]$ and $0 < M < \infty$,

$$
\sup_{v \in [a, 1], \; y \leq \tau_0} |\hat{\phi}_y(v) - \phi_y(v)| = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s.,} \quad (3.7.1)
$$

$$
\sup_{v \in [a, 1], \; y \leq \tau_0} |\hat{\phi}_y'(v) - \phi_y'(v)| = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s.,} \quad (3.7.2)
$$

$$
\sup_{v \in [a, 1], \; y \leq \tau_0} |\hat{\phi}_y''(v) - \phi_y''(v)| = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s.,} \quad (3.7.3)
$$

$$
\sup_{u \in [0, M], \; y \leq \tau_0} |\hat{\phi}_y^{-1}(u) - \phi_y^{-1}(u)| = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s.} \quad (3.7.4)
$$
Proof. Statements (3.7.1), (3.7.2) and (3.7.3) can be proven in an analogous way. Therefore, we will prove only statement (3.7.3). We can write
\[
\tilde{\phi}_x^n(v) - \phi_x^n(v) = \int_{B_y} \phi_x^n(v) d[F_X(x) - F_X(x)].
\]
Now, since \(\phi_x^n(v)\) and \(\phi_x^{''n}(v)\) are uniformly bounded in \((v, x) \in \alpha, b\times R_X\), we can easily show by following calculations done for proving (3.7.5) in Lemma 3.7.2 below, that statement (3.7.3) is true. To prove statement (3.7.4) we use a first order Taylor expansion to get
\[
\tilde{\phi}_x^{-1}(u) - \phi_x^{-1}(u) = \frac{-1}{\phi_x'(u)(\xi(u, y))} \left[ \phi_x(u) \left( \phi_x^{-1}(u) \right) - \tilde{\phi}_x(u) \left( \phi_x^{-1}(u) \right) \right],
\]
where \(\xi(u, y)\) is between \(\tilde{\phi}_x^{-1}(u)\) and \(\phi_x^{-1}(u)\). The second factor on the right hand side is \(O(n^{-1/3}(\log n)^{1/2})\) a.s. by (3.7.1). The first factor on the right hand side is a.s. uniformly bounded (in \(u \in [0, M]\) and \(y \leq \eta_n\)) because of assumption (B4)(iii).

When showing the negligibility of the remainder terms in the proof of Theorem 3.3.1 via Lemma 3.7.2 below, we need to verify certain assumptions regarding the rate of \(G_n(y, x)\) and \(\partial\phi_nG_n(y, x)\), where \(G_n(y, x)\) is a stochastic process. This will either be trivial or will reduce to verifying that \(\sup_{y \in X} |\tilde{L}(y|x) - L(y|x)| = O((nh_n)^{-1/2}(\log n)^{1/2})\) a.s., \(\sup_{y \in X} \tilde{L}(y|x) - L(y|x)| = O(1)\) a.s. or \(\sup_{y \in X} |\tilde{L}(y|x) - L(y|x)| = O((nh_n^3)^{-1/2}(\log n)^{1/2})\) a.s., for \(L \in \{H, H_e\}\), which is true by Lemma 3.7.4.

Lemma 3.7.2. Assume (B1)(i,ii). Let \(G_n(y, x)\) be a stochastic process that is satisfying \(\sup_{y \leq \eta_n} \sup_{x \in B_n} \{ |G_n(y, x)|, |\partial \phi_n G_n(y, x)| \} = O(1)\) a.s. Then,
\[
\sup_{y \leq \eta_n} \int_{B_y} G_n(y, x) d \left[ \tilde{F}_X(x) - F_X(x) \right] = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s.} \quad (3.7.5)
\]
If additionally \(\sup_{y \leq \eta_n} \sup_{x \in B_n} |G_n(y, x)| = O((nh_n)^{-1/2}(\log n)^{1/2})\) a.s. and \(\sup_{y \leq \eta_n} \sup_{x \in B_n} |\partial \phi_n G_n(y, x)| = O((nh_n^3)^{-1/2}(\log n)^{1/2})\) a.s., then
\[
\sup_{y \leq \eta_n} \int_{B_y} G_n(y, x) d \left[ \tilde{F}_X(x) - F_X(x) \right] = o(n^{-1/2}) \text{ a.s.} \quad (3.7.6)
\]
Proof. We start by partitioning \(R_X = [a, b]\) using a grid \(a = x_1 < x_2 < \ldots < x_m = b\) such that \(x_{i+1} - x_i < a_n, i = 1, \ldots, m - 1\), where \(m = C a_n^{-1}\) and \(a_n\) is...
a sequence of constants to be specified further on. We can then write

\[
\left| \int_{B_y} G_n(y, x) d \left( \hat{F}_X(x) - F_X(x) \right) \right|
\]

\[
= \left| \sum_{i=1}^{m-1} G_n(y, x_i) \int_{B_y \cap [x_i, x_{i+1}]} d \left( \hat{F}_X(x) - F_X(x) \right) \right|
\]

\[
+ \left| \sum_{i=1}^{m-1} \int_{B_y \cap [x_i, x_{i+1}]} [G_n(y, x) - G_n(y, x_i)] d \left( \hat{F}_X(x) - F_X(x) \right) \right|
\]

\[
\leq C a_n^{-1} \sup_{g \leq r_\alpha} \sup_{x \in B_y} |G_n(y, x)| \times
\]

\[
\sup_{|x-x'| < a_n} \left| \hat{F}_X(x') - F_X(x') - \hat{F}_X(x) + F_X(x) \right|
\]

\[
+ 2k \sup_{g \leq r_\alpha} \sup_{x \in B_y} \left| \frac{\partial}{\partial x} G_n(y, x) \right| a_n,
\]

where the last equality holds for \( n \) big enough, and \( k \leq \lfloor \lambda(R_X)/\beta \rfloor \). Under the assumption that \( \frac{\partial}{\partial x} G_n(y, x) \) is a.s. uniformly bounded, by defining \( a_n = n^{-1/3} \), the second term on the right hand side is uniformly of the order \( O(a_n) = O(n^{-1/3}) \) a.s. By using Theorem 0.2 in Stute (1982) we can bound the first term by \( a_n^{-1} O(n^{-1/2} a_n^{1/2} (\log n)^{1/2}) = O(n^{-1/3} (\log n)^{1/2}) \) a.s. This proves (3.7.5).

By taking now \( a_n = n^{-1/2} \), statement (3.7.6) is true, because under the additional assumptions for (3.7.6), the second term above is uniformly of the order \( a_n O(n^{-1/2} a_n^{-3/2} (\log n)^{1/2}) = o(n^{-1/2}) \) a.s. Again by using Theorem 0.2 in Stute (1982) we can bound the first term above by

\[
a_n^{-1} O((n a_n)^{-1/2} (\log n)^{1/2}) \times O(n^{-1/2} a_n^{1/2} (\log n)^{1/2}) = o(n^{-1/2}) a.s.
\]

Lemma 3.7.3. Under the conditions of Theorem 3.3.1 we have

\[
\sup_{y \leq r_\alpha} \left| U_n(y) - U(y) \right| = O \left( (n a_n)^{-1/2} (\log n)^{1/2} \right) \text{ a.s.,} \quad (3.7.7)
\]

\[
\sup_{y \leq r_\alpha} \left| \tilde{U}_n(y) - U(y) \right| = O \left( n^{-1/3} (\log n)^{1/2} \right) \text{ a.s.,} \quad (3.7.8)
\]

where \( U(y), U_n(y) \) and \( \tilde{U}_n(y) \) are defined in (3.4.2)
Proof. To deal with (3.7.7) we calculate

\[ U_n(y) - U(y) = \int_{B_y} \int_{-\infty}^{y} \left( \frac{\widehat{H}_e(s|x)}{\widehat{H}_c(s|x)} \right) d \left[ \widehat{H}_n^u(s|x) - H_n^u(s|x) \right] dF_X(x) \]

\[ + \int_{B_y} \int_{-\infty}^{y} \left( \frac{\widehat{H}_e(s|x)}{\widehat{H}_c(s|x)} - 1 \right) d\widehat{H}_c(s|x) dF_X(x). \]

Now, by using integration by parts for the first term, and a first order Taylor expansion for the second term we get

\[ U_n(y) - U(y) \]

\[ = \int_{B_y} \phi_y^\prime \left( \frac{\widehat{H}_e(y|x)}{\widehat{H}_c(y|x)} \right) \left[ \widehat{H}_n^u(y|x) - H_n^u(y|x) \right] dF_X(x) \]

\[ + \int_{B_y} \int_{-\infty}^{y} \phi_y^\prime \left( \frac{\widehat{H}_e(s|x)}{\widehat{H}_c(s|x)} \right) \left[ \widehat{H}_n^u(s|x) - H_n^u(s|x) \right] d\widehat{H}_c(s|x) dF_X(x) \]

\[ - \int_{B_y} \int_{-\infty}^{y} \phi_y^\prime \left( \frac{\widehat{H}_e(s|x)}{\widehat{H}_c(s|x)} \right) \left[ \widehat{H}_c(s|x) - \widehat{H}_e(s|x) \right] dH_n^u(s|x) dF_X(x) \]

\[ + \frac{1}{2} \int_{B_y} \int_{-\infty}^{y} \phi_y^\prime \left( \frac{\widehat{H}_e(s|x)}{\widehat{H}_c(s|x)} \right) \left( \widehat{H}_c(s|x) - \widehat{H}_e(s|x) \right)^2 \right] dH_n^u(s|x) dF_X(x), \]

where \( \xi(s, x) \) is between \( \widehat{H}_c(s|x) \) and \( \widehat{H}_e(s|x) \). Now, by using assumption (B4) and Lemma 3.7.4 below we get the desired order.

Statement (3.7.8) can be bounded by \( \sup_{y \in \tau_n} |U_n(y) - U(y)| = O(n^{-1/3}(\log n)^{1/2}) \) a.s. by Lemma 3.7.2, while the second term is \( O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s. by (3.7.7). This concludes the proof. \( \square \)

The following lemma is a generalization of Lemma 4.1 in Van Keilegom and Akritas (1999) regarding the uniform (in \( y \in R \) and \( x \in R_X \)) rate of convergence of the difference \( \widehat{H}(y|x) - H(y|x) \) to the uniform rate of convergence of \( \widehat{H}_e(y|x) - H_e(y|x) \). The former difference is a sum of i.i.d. random variables, while the latter difference is a sum of non-independent random variables.

Lemma 3.7.4. Assume (B1)-(B3), and assume that (B5)(i,ii,iv,v) holds for \( H_e(y|x) \) and \( H_e^u(y|x) \). Then,

(i) \( \sup_{x \in R_X} \sup_{y \in R} \left| \widehat{H}_e(y|x) - H_e(y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s.

(ii) \( \sup_{x \in R_X} \sup_{y \in R} \left| \widehat{H}_e(y|x) - H_e(y|x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \) a.s.
(iii) \[ \sup_{x \in R_x} \sup_{y \in \mathbb{R}} \left| \hat{H}_e^*(y|x) - H_e^*(y|x) \right| = O((nh_n)^{-1/2}(\log n)^{1/2}) \text{ a.s.} \]

(iv) \[ \sup_{x \in R_x} \sup_{y \in \mathbb{R}} \left| \hat{H}_e^*(y|x) - \hat{H}_e^*(y|x) \right| = O((nh_n^3)^{-1/2}(\log n)^{1/2}) \text{ a.s.} \]

**Proof.** The proofs of statements (i), (ii), (iii) and (iv) use the same idea so we will only show (i). For that we define a new estimator

\[ \hat{H}_e^*(y|x) = \sum_{i=1}^{n} W_{ni}(x, h_n) I(E_i \leq y), \]

for which statement (i) is true by Lemma 4.1 in Van Keilegom and Akritas (1999). To finish the proof we will show that the difference between \( \hat{H}_e \) and \( \hat{H}_e^* \) is uniformly of the order \( O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s. Consider

\[
\left| \hat{H}_e(y|x) - \hat{H}_e^*(y|x) \right| \\
= \left| \sum_{i=1}^{n} W_{ni}(x, h_n) \left( I(\hat{E}_i \leq y) - I(E_i \leq y) \right) \right| \\
\leq \sum_{i=1}^{n} W_{ni}(x, h_n) \left| I(\hat{E}_i \leq |y|\beta_n + \alpha_n) - I(E_i \leq |y|\beta_n - \alpha_n) \right| \\
= \hat{H}_e^*(y + |y|\beta_n + \alpha_n|x) - \hat{H}_e^*(y - |y|\beta_n - \alpha_n|x),
\]

where \( \alpha_n = \sup_{x \in R_x} \frac{\hat{m}(x) - m(x)}{\sigma(x)} \) and \( \beta_n = \sup_{x \in R_x} \frac{\beta(x) - \sigma(x)}{\sigma(x)} \). Now, by adding and subtracting \( H_e(y + |y|\beta_n + \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x) \), we get

\[
\sup_{y \in \mathbb{R}, x \in R_x} \left| \hat{H}_e(y|x) - \hat{H}_e^*(y|x) \right| \\
\leq \sup_{y \in \mathbb{R}, x \in R_x} \left| \hat{H}_e(y + |y|\beta_n + \alpha_n|x) - H_e(y + |y|\beta_n + \alpha_n|x) \right| \\
+ \sup_{y \in \mathbb{R}, x \in R_x} \left| \hat{H}_e(y - |y|\beta_n - \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x) \right| \\
+ \sup_{y \in \mathbb{R}, x \in R_x} \left| H_e(y + |y|\beta_n + \alpha_n|x) - H_e(y - |y|\beta_n - \alpha_n|x) \right| \\
= 2 \sup_{y \in \mathbb{R}, x \in R_x} \left| \hat{H}_e(y|x) - H_e(y|x) \right| \\
+ \sup_{y \in \mathbb{R}, x \in R_x} \left| h_e(\xi_{x,y}|x) [2\beta_n|y| + 2\alpha_n] \right|. 
\]
where \( \xi_{x,y} \) is between \( y-|y|\beta_n-\alpha_n \) and \( y+|y|\beta_n+\alpha_n \). As explained in the beginning the first term on the right hand side is of the order \( O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s. The second term is of the same order because \( \sup_{y \in \mathbb{R}} \sup_{x \in \mathbb{R}_x} |h_c(y|x)| < \infty \) by assumption \((B5)(ii)\) and because \( \alpha_n \) and \( \beta_n \) are \( O((nh_n)^{-1/2}(\log n)^{1/2}) \) a.s. by Proposition 2.3.5 in Chapter 2.

\[ \text{Proof} \]

The proof is very analogous to the proof of Lemma 2.6.1 in Chapter 2. The only difference is that we use Lemma 3.7.6 below for the uniform rate of convergence of the modulus of continuity \( \hat{H}_c(y_1|x) - \hat{H}_c(y_2|x) - H_c(y_1|x) + H_c(y_2|x) \), whereas they use Lemma 4.4 in Du and Akritas (2002) for \( \hat{H}(y_1|x) - \hat{H}(y_2|x) - H(y_1|x) + H(y_2|x) \). Details of the proof are omitted and can be found in Lemma 2.6.1 in Chapter 2.

\[ \text{Lemma 3.7.5.} \quad \text{Under the assumptions of Theorem 3.3.1 we have} \]
\[
\sup_{y \leq \tau_n} \sup_{x \in B_y} \int_{-\infty}^{y} \left[ \phi_x'(\tilde{T}_c(s|x)) - \phi_x'(T_c(s|x)) \right] d \left[ \hat{H}_c^n(s|x) - H_c^n(s|x) \right] = O((nh_n)^{-3/4}(\log n)^{3/4+\alpha}) \text{ a.s.,} \tag{3.7.9}
\]

where \( \alpha > 0 \) is an arbitrarily small constant.

\[ \text{Proof.} \]

The following lemma is a uniform modulus of continuity result for the Nadaraya-Watson type estimator \( \hat{H}_c(y|x) \). It is an adaptation of Lemma 4.4 in Du and Akritas (2002). A major difficulty in this adaptation is that \( \hat{H}_c(y|x) \) is not a sum of independent random variables.

\[ \text{Lemma 3.7.6.} \quad \text{Assume \((B1)-(B4)\), and assume that \((B5)\) and \((B6)\) hold for \( \hat{H}_c(y|x) \). Let \( a_n = O((nh_n)^{-1/2}(\log n)^{1/2}) \). Then,} \]
\[
\sup_{x \in \mathbb{R}_x} \sup_{(y_1,y_2) \in J_{a_n}} |\hat{H}_c(y_1|x) - \hat{H}_c(y_2|x) - H_c(y_1|x) + H_c(y_2|x)| = O \left( a_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2+\alpha} + a_n^2 + a_n h_n \right) = o \left( n^{-1/2} \right) \text{ a.s.,} \tag{3.7.10}
\]

for an arbitrarily small constant \( \alpha > 0 \), \( J_{a_n} = \{(y_1,y_2) : |M(y_1) - M(y_2)| \leq a_n \} \) and \( M(y) = \sum_{i=1}^{n} M_i(y) \) (see \((B6)\) for \( L = H_c \)).

\[ \text{Proof.} \]

In this proof we will use an index \((-r)\) to denote estimators that leave
out the random variables $E_r$ and $\Delta_r$:

\[
\bar{H}_{(-r)}(y|x) := \sum_{i=1, i \neq r}^{n} W_{ni}(x, h_n) I (T_i \leq y),
\]

\[
\bar{F}_{(-r)}(y|x) := \phi^{-1}_x \left\{ - \sum_{\Delta_i = \bar{F}_{(-r)}(T_i | x), \bar{F}_{(-r)}(T_i | x)} \phi_x \left( \bar{F}_{(-r)}(T_i | x) \right) \right\},
\]

\[
\tilde{m}_{(-r)}(x) := \sum_{i=1, i \neq r}^{n} \Delta_{\bar{F}_{(-r)}(T_i | x)} T_j (\bar{F}_{(-r)}(T_i | x)),
\]

\[
\tilde{\sigma}_{(-r)}^2(x) := \sum_{i=1, i \neq r}^{n} \Delta_{\bar{F}_{(-r)}(T_i | x)} T_j^2 (\bar{F}_{(-r)}(T_i | x)) - \tilde{m}_{(-r)}^2(x).
\]

We will also use the notation $\Delta(y, x) = y \Delta_2(x) + \Delta_1(x)$, where $\Delta_2(x) = \bar{\sigma}_x - \bar{m}_x$ and $\Delta_1(x) = \bar{m}_{(x) - m}(x)$. Further we will denote $\Delta_{(-r)}$, $\Delta_{1(-r)}$ and $\Delta_{2(-r)}$ when replacing $\tilde{m}$ and $\tilde{\sigma}$ by $\tilde{m}_{(-r)}$ and $\tilde{\sigma}_{(-r)}$ in the functions $\Delta$, $\Delta_1$ and $\Delta_2$, respectively.

We will prove the main statement of the Lemma by showing the following two statements:

\[
\sup_{x \in \mathcal{R}_X} \sup_{(y_1, y_2) \in J_{an}} | \bar{H}_c(y_1 | x) - \bar{H}_c(y_2 | x) - \bar{H}_c(y_1 | x) + \bar{H}_c(y_2 | x) | = O(a_n^{1/2} (nh_n)^{-1/2} (\log n)^{1/2 + \alpha}) \text{ a.s.},
\]

(3.7.11)

\[
\sup_{x \in \mathcal{R}_X} \sup_{(y_1, y_2) \in J_{an}} | \bar{H}_c(y_1 | x) - \bar{H}_c(y_2 | x) - H_c(y_1 | x) + H_c(y_2 | x) | = O(a_n^{2} + a_n h_n) \text{ a.s.},
\]

(3.7.12)

where $\bar{H}_c(y | x) = \sum_{r=1}^{n} W_{nr}(x, h_n) H_c(y + \Delta_{(-r)}(y, X_r) | X_r)$. To show (3.7.12) we calculate, for all $y_1$ and $y_2$ such that $|(M_2 + M_3 + M_4(y_1) - (M_2 + M_3 + \ldots$
where $x'_r$ is between $x$ and $X_r$, $\xi_{ir}$ is between $y_i$ and $y_i + \Delta_{(-r)}(y_i, X_r)$ for $i = 1, 2$ and $r = 1, \ldots, n$. In the last equality we used assumptions (B5)(ii,iii), Proposition 2.3.5 in Chapter 2, and the fact that $\sup_{x} \max_{y = 1, \ldots, n} |W_{nr}(x, h_n)| = O((nh_n)^{-1})$ a.s., which is easy to show.

To prove (3.7.11) we start by partitioning $\mathbb{R}$ into $m_n = \lfloor M(+\infty)/a_n \rfloor$ subintervals $-\infty = y_0 < y_1 < \ldots < y_{m_n} = \infty$, such that $M(y_{i+1}) - M(y_i) = a_n := M(+\infty)/m_n$. For each $i = 1, \ldots, m_n - 1$ define $I_{ni} = [y_i, y_{i+1})$. Further partition each interval $I_{ni}$ into $2b_n$ smaller intervals $[y_{ij}, y_{ij+1})$ for $j = -b_n, -b_n + 1, \ldots, b_n$, where $b_n = O(a_n/n^{1/2}(\log n)^{-1/2})$, such that $M(y_{ij+1}) - M(y_{ij}) = \frac{a_n}{b_n}$. It can be easily verified that $a_n \leq a_n \leq 2a_n$ for $n$ large enough, and for any $y_1, y_2 \in \mathbb{R}$ with $|M(y_1) - M(y_2)| < a_n$, there exists an interval $I_{ni}$ such that $y_1, y_2 \in I_{ni}$. Hence, by using the monotonicity of $\hat{H}_e(\cdot | x)$, it can be seen that (3.7.11) is bounded by

\[
\sup_{x \in \mathbb{R}} \max_{1 \leq i \leq m_n - 1} \max_{-b_n \leq j \leq b_n} | \hat{H}_e(y_{ik} | x) - \hat{H}_e(y_{ij} | x) - \hat{H}_e(y_{ik} | x) + \hat{H}_e(y_{ij} | x) | \tag{3.7.13}
\]

\[
+ 2 \sup_{x \in \mathbb{R}} | \hat{H}_e(y_1 | x) - \hat{H}_e(y_2 | x) | : |M(y_1) - M(y_2)| \leq a_n/b_n. \tag{3.7.14}
\]

We can write the term between absolute values in (3.7.14) as

\[
\sum_{r=1}^{n} W_{nr}(x, h_n) \left\{ H_e(y_1 + \Delta_{(-r)}(y_1, X_r) | X_r) - H_e(y_2 + \Delta_{(-r)}(y_2, X_r) | X_r) \right\}
\]
\[
\sum_{r=1}^{n} W_{nr}(x, h_n) \left\{ H_c(y_1|X_r) - H_c(y_2|X_r) \\
+ H_c'(y_1|X_r) \Delta_{(-r)}(y_1, X_r) - H_c'(y_2|X_r) \Delta_{(-r)}(y_2, X_r) \\
+ \frac{1}{2} H_c''(\xi_1|X_r) \Delta_{(-r)}^2(y_1, X_r) - \frac{1}{2} H_c''(\xi_2|X_r) \Delta_{(-r)}^2(y_2, X_r) \right\},
\]

where \( \xi_{ir} \) is between \( y_i \) and \( y_i + \Delta_{(-r)}(y_i, X_r) \) for \( i = 1, 2 \). The first difference on the right hand side is of the order \( a_n = O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \). The third difference is of the order \( O\left((nh_n)^{-1}\log n\right) \) a.s., because of assumption (B5)(iii), and because the terms \( \Delta_{1(-r)}(y_i, X_r) \) and \( \Delta_{2(-r)}(y_i, X_r) \) are uniformly \( O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \) a.s. (by Proposition 2.3.5 in Chapter 2, and the relation \( \sup_{x \in R_X} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| = O((nh_n)^{-1}) \) a.s.). To show that the second difference is negligible we calculate

\[
\left| H_c'(y_1|X_r) y_1 - H_c'(y_2|X_r) y_2 \Delta_{2(-r)}(X_r) \\
+ H_c'(y_1|X_r) - H_c'(y_2|X_r) \right| \Delta_{1(-r)}(X_r) \right| \\
\leq |M(y_1) - M(y_2)| \left| \Delta_{2(-r)}(X_r) \right| + |M(y_1) - M(y_2)| \left| \Delta_{1(-r)}(X_r) \right| \\
= O\left(\frac{a_n}{b_n} (nh_n)^{-1/2}(\log n)^{1/2}\right) \text{ a.s.}
\]

Hence, we showed that (3.7.14) is \( O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \) a.s.

To deal with (3.7.13) we define the grid \( a = x_0 < x_1 < \ldots < x_k = b \) of \( R_X = [a, b] \) such that \( x_i - x_{i-1} \leq a_n^{1/2} (nh_n)^{-1/2}(\log n)^{1/2} b_n^3, \ i = 1, \ldots, k_n \).

By similar calculations as in Lemma 4.2 in Du and Akritas (2002), we have uniformly, up to a remainder term of order \( O\left((nh_n)^{-1/2}(\log n)^{1/2}\right) \) a.s., that (3.7.13) is bounded by

\[
A_n := \max_{l=1,\ldots,k_n} \max_{1 \leq j \leq m_n, 1 \leq b_n, k \leq b_n} \left| \tilde{H}_c(y_{kj}|x_l) - \tilde{H}_c(y_{lj}|x_l) \\
- \frac{1}{2} \tilde{H}_c(y_{kj}|x_l) + \tilde{H}_c(y_{lj}|x_l) \right|.
\]

Before continuing we define \( D_n \) as a set where for a fixed constant \( 0 < C' < \infty \) the following conditions are satisfied:

\[
\sup_{x \in R_X} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| \leq C'(nh_n)^{-1}(\log n)^{\alpha}, \quad (3.7.15)
\]
\[
\sup_{x \in R_X} \max_{r=1,\ldots,n} \left| \Delta_{1(-r)}(x) \right| \leq C'(nh_n)^{-1/2}(\log n)^{1/2+\alpha}, \quad (3.7.16)
\]
\[
\sup_{x \in R_X} \max_{r=1,\ldots,n} \left| \Delta_{2(-r)}(x) - \Delta_{1}(x) \right| \leq C'(nh_n)^{-1}(\log n)^{\alpha}. \quad (3.7.17)
\]
We can show that \( P(\cup_{m=1}^{\infty} \cap_{n=1}^{\infty} D_n) = 1 \). Indeed, as mentioned before we have that \( \sup_{x \in \mathcal{R}_x} \max_{r=1,\ldots,n} |W_{nr}(x, h_n)| = O((nh_n)^{-1}) \) a.s. By additionally using Proposition 2.3.5 in Chapter 2, we have that the term on the left hand side in (3.7.16) is \( O((nh_n)^{-1}) \) a.s. The term on the left hand side in (3.7.17) is \( O((nh_n)^{-1}\log n) \) a.s. by direct use of Lemma 3.7.7. On the set \( D_n \), we can write \( \hat{H}_e(y_{ik}|x) - \hat{H}_e(y_{ij}|x) - \hat{H}_e(y_{ik}|x_t) + \hat{H}_e(y_{ij}|x_t) = \sum_{r=1}^{n} X_{rijkl} \), where

\[
X_{rijkl} = \tilde{W}_{nr}(x_t; h_n)
\times \left\{ \begin{array}{l}
I(E_r \leq y_{ik} + \Delta_r(y_{ik}, X_r)) - I(E_r \leq y_{ij} + \Delta_r(y_{ij}, X_r)) + Z_{rijkl} \\
- H_e(y_{ik} + \Delta_r(y_{ik}, X_r)|X_r) + H_e(y_{ij} + \Delta_r(y_{ij}, X_r)|X_r) \end{array} \right\},
\]

\[
Z_r(y) := I(E_r \leq y + \Delta(y, X_r)) - I(E_r \leq y + \Delta_r(y, X_r)) - Z_r(y_{ij}|I_{D_{1\ldots n}}, \tilde{W}_{nr}(x_t; h_n) = W_{nr}(x_t; h_n) I_{D_{1\ldots n}}, \tilde{\Delta}_r(y, x) = y\Delta_2(-r)(x) + \Delta_1(-r)(x) + \tilde{\Delta}_i(-r)(x) = \Delta_i(-r)(x) I_{D_{1\ldots n}}, i = 1, 2 \).
\]

We define a (centered) version of the random variable \( X_{rijkl} \) by \( \tilde{X}_{rijkl} = X_{rijkl} - E[\tilde{X}_{rijkl}|X_{(-r)}, X_r] \), where \( X_{(-r)} = \{ X_t, T_t, \Delta_t \}_{t \neq r} \). Now in order to prove (3.7.13) we will use a modification of Bernstein’s inequality (see Theorem 1.2A in de la Peña (1999)). Using the notation \( \tilde{X}_{r-1,ijkl} = \{ \tilde{X}_{1,ijkl}, \ldots, \tilde{X}_{r-1,ijkl} \} \), if the following is satisfied

\[
E[\tilde{X}_{rijkl}|\tilde{X}_{r-1,ijkl}] = 0, \quad \text{ (3.7.18)}
\]

\[
\sum_{r=1}^{n} E[\tilde{X}_{rijkl}^2|\tilde{X}_{r-1,ijkl}] \leq v_n, \quad \text{ (3.7.19)}
\]

\[
E[\tilde{X}_{rijkl}^p|\tilde{X}_{r-1,ijkl}] \leq \frac{1}{2} E[\tilde{X}_{rijkl}^2|\tilde{X}_{r-1,ijkl}] E[p-2, p!], \quad \text{ (3.7.20)}
\]

where \( L_n > 0 \) and \( v_n > 0 \) are constants, we have

\[
P \left( \sum_{r=1}^{n} \tilde{X}_{rijkl} > \lambda_n \right) \leq \exp \left\{ -\frac{\lambda_n^2}{2 [v_n + L_n\lambda_n]} \right\}.
\]

We will use now, and show later that the conditions (3.7.18), (3.7.19) and (3.7.20) can be verified for constants \( v_n = C a_n(nh_n)^{-1}(\log n)^a \) and \( L_n = C (nh_n)^{-1}(\log n)^a \). Now, for \( \lambda_n = c_1 a_n^{1/2} (nh_n)^{-1/2}(\log n)^{1/2+a} \), where \( c_1 \) is a
positive constant to be specified further on, we have

\[ P \left( \{ A_n > 2\lambda_n \} \cap D_n \right) \]

\[ \leq P \left( \max_{t=1,\ldots,k_n} \max_{1 \leq i \leq m_n-1} \max_{-b_n \leq j,k \leq b_n} \sum_{r=1}^n X_{rijkl} > 2\lambda_n \right) \]

\[ \leq P \left( \max_{t=1,\ldots,k_n} \max_{1 \leq i \leq m_n-1} \max_{-b_n \leq j,k \leq b_n} \sum_{r=1}^n \tilde{X}_{rijkl} > \lambda_n \right) \]

\[ \leq 2k_n(m_n - 1)(2b_n + 1)^2 \exp \left\{ - C \lambda_n^2 / \left( \frac{a_n (\log n)^\alpha}{nh_n} + \frac{\lambda_n (\log n)^\alpha}{nh_n} \right) \right\} \]

\[ \leq 2k_n(m_n - 1)(2b_n + 1)^2 n^{-c_1 C}, \]

for some constant \( C > 0 \). For the second inequality we used that \( \lambda_n > \sum_{r=1}^n E[\tilde{Z}_{rijkl}|X_{(-r)},X_r] \) by (3.7.23) below, while for the third inequality we used the modified Bernstein's inequality. Since, by proper choice of \( c_1 \), this can be made summable, using the Borel-Cantelli lemma we get \( P(\bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \{ A_n > 2\lambda_n \} \cap D_n) = 0 \). From here we can calculate

\[ P \left( \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \{ A_n > 2\lambda_n \} \right) \]

\[ \leq P \left( \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \left\{ \{ A_n > 2\lambda_n \} \cap D_n \right\} \cup D_n^C \right) \]

\[ \leq P \left( \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \{ A_n > 2\lambda_n \} \right) + P \left( \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty D_n^C \right). \]

As we have just shown, the first term on the right hand side is 0 by the Borel-Cantelli lemma, while the second term is 0, since \( P(\bigcup_{m=1}^\infty \bigcap_{n=m}^\infty D_n) = 1 \). Therefore, this proves (3.7.13) and subsequently (3.7.11). Finally, to show the conditions that allowed us to use the modified Bernstein’s inequality, we start with condition (3.7.18), that is satisfied since by definition \( E[\tilde{X}_{rijkl}|X_{(-r)},X_r] = 0 \). Condition (3.7.20) follows from

\[ E[\tilde{X}_{rijkl}^p|\tilde{X}_{r-1,ijkl}] \leq E[\tilde{X}_{rijkl}^2 \tilde{W}_{nr}^{-2}(x_1,h_n)|\tilde{X}_{r-1,ijkl}] \]

\[ \leq E[\tilde{X}_{rijkl}^2|\tilde{X}_{r-1,ijkl}][(C'(nh_n)^{-1}(\log n)^\alpha)^{p-2}], \]

where the last inequality is uniform in \( r, i, j, k \) and \( l \). Before continuing we
calculate:

\[
E[\tilde{X}_{rijkl}^2|X_{(-r)}, X_r] = \overline{W}_{nr}(x_t, h_n) \left\{ \left[ H_c(y_{ik} + \tilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_c(y_{ij} + \tilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r) \right] \\
- \left[ H_c(y_{ik} + \tilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_c(y_{ij} + \tilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r) \right]^2 \\
+ E \left[ (\tilde{Z}_{ij} - E[\tilde{Z}_{ij}])|X_{(-r)}, X_r, X_r^2 \right] \right\} \tag{3.7.21}
\]

\[
= \overline{W}_{nr}(x_t, h_n) [a_n + R_n],
\]

where \( |R_n| \leq C(a_n^2 \log n)^{2\alpha} + (nh_n)^{-1}(\log n)\alpha \) uniformly in \( r, i, j, k \) and \( l \). To bound the first and the second difference above we use the following calculation:

\[
H_c(y_{ik} + \tilde{\Delta}_{(-r)}(y_{ik}, X_r)|X_r) - H_c(y_{ij} + \tilde{\Delta}_{(-r)}(y_{ij}, X_r)|X_r)
= [H_c(y_{ik}|X_r) - H_c(y_{ij}|X_r)]
+ [H'_c(y_{ik}|x)|X_r] - H'_c(y_{ij}|x)|X_r) \tilde{\Delta}_{(-r)}(y_{ij}, X_r)]}
+ \frac{1}{2} H''_c(\xi_{ik}|x)|X_r) \tilde{\Delta}_{(-r)}^2(y_{ij}, X_r) \tag{3.7.22}
\]

uniformly in \( r, i, j \) and \( k \), where \( |R'_n| \leq C(a_n^2 \log n)^{\alpha} + a_n^2 \log n)^{2\alpha} \) and \( \xi_{it} \) is between \( y_{it} \) and \( y_{it} + \tilde{\Delta}_{(-r)}(y_{it}, X_r) \), \( t = k, j \). For the second and the third term of (3.7.22) we used (3.7.16) and assumptions (B5)(ii) and (B5)(iii). To bound the third term in (3.7.21) we use the following reasoning. We first define \( c_n = c'(nh_n)^{-1}(\log n)^{\alpha} \). Now, on the set \( D_n \) we can conclude that the variable \( I(E_r \leq s + \tilde{\Delta}(s, X_r)) \) lies in between \( I(E_r \leq s + \tilde{\Delta}_{(-r)}(s, X_r) - sc_n - c_n) \) and \( I(E_r \leq s + \tilde{\Delta}_{(-r)}(s, X_r) + sc_n + c_n) \). Therefore, we have that

\[
E[Z_r(s)I_{D_n}|X_{(-r)}, X_r] \leq E \left[ I(E_r \leq s + \tilde{\Delta}_{(-r)}(s, X_r) + sc_n + c_n) \right. \\
- I(E_r \leq s + \tilde{\Delta}_{(-r)}(s, X_r) - sc_n - c_n) \right] |X_{(-r)}, X_r]
\leq H_c \left( s + \tilde{\Delta}_{(-r)}(s, X_r) + sc_n + c_n |X_r \right)
- H_c \left( s + \tilde{\Delta}_{(-r)}(s, X_r) - sc_n - c_n |X_r \right)
\leq 2 \sup_{s,x} H'_c(s|x) |s|sc_n + c_n\]
\leq C (nh_n)^{-1}(\log n)^{\alpha}, \tag{3.7.23}
\]

for some constant \( C > 0 \), uniformly in \(-\infty < s < \infty \) and for all \( r \leq n \). The last inequality follows from assumption (B5)(ii). This proves (3.7.21). Condition
(3.7.19) is now satisfied, since \( \sum_{r=1}^{n} E[\hat{X}_{rijkl}^2 | \hat{X}_{r-1,ijkl}] \leq C a_n (nh_n)^{-1} (\log n)^\alpha \) uniformly, because of (3.7.21). This completes the proof. \( \square \)

**Lemma 3.7.7.** Under the assumptions of Lemma 3.7.6 we have,

\[
\sup_{x \in R_X} \max_{r=1,\ldots,l} |\hat{m}(x) - \hat{m}_{(-r)}(x)| = O \left((nh_n)^{-1}\right) \quad a.s., \tag{3.7.24}
\]

\[
\sup_{x \in R_X} \max_{r=1,\ldots,l} |\hat{\sigma}(x) - \hat{\sigma}_{(-r)}(x)| = O \left((nh_n)^{-1}\right) \quad a.s. \tag{3.7.25}
\]

**Proof.** Before we start, we give the results that will be used throughout the proof and which will be shown at the end of this proof.

\[
\sup_{x \in R_X} \max_{r=1,\ldots,n} W_{ni}(x, h_n) = O \left((nh_n)^{-1}\right) \quad a.s., \tag{3.7.26}
\]

\[
\sup_{x \in R_X \ y \leq \hat{T}_x} \max_{r=1,\ldots,n} |\hat{F}(y|x) - \hat{F}_{(-r)}(y|x)| = O \left((nh_n)^{-1}\right) \quad a.s., \tag{3.7.27}
\]

\[
\sup_{x \in R_X} \max_{r=1,\ldots,n} \Delta \hat{F}(T_i|x) = O \left((nh_n)^{-1}\right) \quad a.s., \tag{3.7.28}
\]

\[
\sup_{x \in R_X \ r,i=1,\ldots,n} \Delta \hat{F}_{(-r)}(T_i|x) = O \left((nh_n)^{-1}\right) \quad a.s., \tag{3.7.29}
\]

\[
\Delta \hat{F}(T_i|x) - \Delta \hat{F}_{(-r)}(T_i|x) = C_{irx} W_{ni}(x, h_n), \tag{3.7.30}
\]

where \( \sup_{x \in R_X, r,i=1,\ldots,n} |C_{irx}| = O((nh_n)^{-1}) \) a.s. Additionally, we will assume that indices result in ascendingly ordered \( T_i \). Showing (3.7.24) is very similar to showing (3.7.25). Therefore, we will show only the latter and will do so by demonstrating that \( \hat{\sigma}^2(x) - \hat{\sigma}_{(-r)}^2(x) \) is uniformly of order \( O((nh_n)^{-1}) \) a.s.

\[
\hat{\sigma}^2(x) - \hat{\sigma}_{(-r)}^2(x)
\]

\[
= \sum_{i \neq r} \hat{F}(T_i|x) [T_i - \hat{m}(x)]^2 \left[ J(\hat{F}(T_i|x)) - J(\hat{F}_{(-r)}(T_i|x)) \right]
\]

\[
+ \sum_{i \neq r} \Delta \hat{F}(T_i|x) \{ [T_i - \hat{m}(x)]^2 - [T_i - \hat{m}_{(-r)}(x)]^2 \} J \left( \hat{F}_{(-r)}(T_i|x) \right)
\]

\[
+ \sum_{i \neq r} [\Delta \hat{F}(T_i|x) - \Delta \hat{F}_{(-r)}(T_i|x)] [T_i - \hat{m}_{(-r)}(x)]^2 J \left( \hat{F}_{(-r)}(T_i|x) \right)
\]

\[
+ \Delta \hat{F}(T_i|x)[T_i - \hat{m}(x)]^2 J(\hat{F}(T_i|x)).
\]

By combining assumptions (B2)(i) and (B5)(ii) we have \( \inf_{x \in R_X} \inf_{s=s_0} \inf_{s \leq s_1} (F^{-1}(s|x))^{-1} > 0 \), that is \( \sup_{x \in R_X} \sup_{s_0 \leq s \leq s_1} (F^{-1}(s|x)) < M < \infty \). From here for \( t > M \) we have \( \hat{F}(t|x) \geq \hat{F}(M|x) \geq F(M|x) + o(1) \geq s_1 + o(1) \) a.s. (for the second inequality see Proposition 2.3.1 in Chapter 2). Now, because of
the assumption $s_1 > \sup\{s \in [0, 1] : J(s) > 0\}$, we have a.s., for $n$ big enough, $J(\hat{F}(t|x)) = 0$, uniformly $\forall t > M$. Similarly for $t < -M$ we have a.s., for $n$ big enough $J(\hat{F}(t|x)) = 0$ and therefore for $|t| > M$ we have $J(\hat{F}(t|x)) = 0$ a.s., for $n$ big enough. We can also show that for $|t| > M$ we have a.s., for $n$ big enough that $J(\hat{F}_{(-r)}(t|x)) = 0$, by using (3.7.27). From here we know that all the $T_i$ that are occurring in the decomposition of $\hat{\sigma}^2(x) - \hat{\sigma}^2_{(-r)}(x)$ are uniformly $O(M)$ a.s. Combining this with the fact that $\sup_{x \in \mathbb{R}^k} |\hat{m}(x)| = O(1)$ a.s. (which follows from the Proposition 2.3.5, equality (3.7.24) and the uniform boundedness of $m$), we have

$$\hat{\sigma}^2(x) - \hat{\sigma}^2_{(-r)}(x)$$

$$= O(1) \sum_{i \neq r} \Delta \hat{F}(T_i|\hat{x}) \left[J(\hat{F}(T_i|x)) - J(\hat{F}_{(-r)}(T_i|x))\right]$$

$$+ O(1) \sum_{i \neq r} \Delta \hat{F}(T_i|x) \left\{|\hat{m}(x)^2 - \hat{m}_{(-r)}(x)^2| + |\hat{m}(x) - \hat{m}_{(-r)}(x)|\right\}$$

$$+ O(1) \sum_{i \neq r} \left[\Delta \hat{F}(T_i|x) - \Delta \hat{F}_{(-r)}(T_i|x)\right]$$

$$+ O(1) \Delta \hat{F}(T_r|x) \text{ a.s.}$$

$$:= \alpha_1(x) + \alpha_2(x) + \alpha_3(x) + \alpha_4(x) \text{ a.s.}$$

The term $\alpha_r(x)$ is uniformly $O((nh_n)^{-1})$ a.s. by applying a first order Taylor expansion on $J$ and using (3.7.27). Terms $\alpha_2(x)$, $\alpha_3(x)$ and $\alpha_4(x)$ are uniformly $O((nh_n)^{-1})$ a.s. by using (3.7.24), (3.7.30) and (3.7.30), respectively.

To finish the proof it remains to show (3.7.26), (3.7.27), (3.7.28), (3.7.29) and (3.7.30). Equality (3.7.26) can be easily proven by using standard result $\sup_{x \in \mathbb{R}^k} |\hat{F} - F_t| = o(1)$ a.s. and the assumption $\inf_{x \in \mathbb{R}^k} f_X(x) > 0$. Next, we deal with (3.7.27). By using a first order Taylor expansion, for $y \leq \hat{F}_r$, we have

$$\hat{F}(y|x) - \hat{F}_{(-r)}(y|x)$$

$$= \frac{1}{\phi_x'(\gamma_r(y, x))} \left\{- \sum_{T_i \leq y, \Delta_i = 1} \phi_x \left(\hat{H}(T_i^+|x)\right) - \phi_x \left(\hat{H}(T_i|x)\right)\right\}$$

$$+ \sum_{T_i \leq y, \Delta_i = 1, \neq r} \phi_x \left(\hat{H}_{(-r)}(T_i^+|x)\right) - \phi_x \left(\hat{H}_{(-r)}(T_i|x)\right),$$

where $\gamma_r(y, x)$ is between $\hat{F}(y|x)$ and $\hat{F}_{(-r)}(y|x)$. Again by using Taylor ex-
3.7. Appendix B

We have

\[ \bar{F}(y|x) - \bar{F}_{-1}(y|x) \]

\[ = \frac{1}{\phi_x'(\gamma_r(y,x))} \left\{ - \sum_{T_i \leq y, \Delta_i = 1} \phi_x'(\xi_{i1}(x)) W_{ni}(x, h_n) \right. \]

\[ + \sum_{T_i \leq y, \Delta_i = 1, i \neq r} \phi_x'(\xi_{i2(-r)}(x)) W_{ni}(x, h_n) \right\} \]

\[ = \frac{1}{\phi_x'(\gamma_r(y,x))} \left\{ \sum_{T_i \leq y, \Delta_i = 1, i \neq r} \phi_x'(\xi_{i3(-r)}(x)) \left[ \xi_{i2(-r)}(x) - \xi_{i1}(x) \right] W_{ni}(x, h_n) \right. \]

\[ + \phi_x'(\xi_{r1}(x)) W_{nr}(x, h_n) \left. \right\} \]

\[ : = O(1) \left\{ \sum_{i \neq r} \alpha_1(x, i, r) + \alpha_2(x, r) \right\} \]

where the term \( \xi_{i1}(x) \) is between \( \overline{H}(T_i|x) \) and \( \overline{H}(T_{i-1}|x) \), \( \xi_{i2(-r)}(x) \) is between \( \overline{H}_{-1}(T_i|x) \) and \( \overline{H}_{-1}(T_{i-1}|x) \), and \( \xi_{i3(-r)}(x) \) is between \( \xi_{i1}(x) \) and \( \xi_{i2(-r)}(x) \). For the last equality we used that by (B4) we have for some \( \inf_x |\phi_x'(t)| > 0 \). From here we conclude that \( \sup_{x \in \mathbb{R}, y \leq T_r, r = 1, \ldots, n} |\bar{F}(y|x) - \bar{F}_{-1}(y|x)| = O((nh_n)^{-1}) \) a.s., by the following reasoning. First we will show a technical result, that \( \xi_{i1}(x) \), \( \xi_{i2(-r)}(x) \) and \( \xi_{i3(-r)}(x) \) are uniformly bounded from below by \( \alpha + o(1) \) a.s., for some \( \alpha > 0 \). We do the calculations only for \( \xi_{i3(-r)}(x) \), since the proofs are almost identical:

\[ \inf_{x \in \mathbb{R}} \min_{T_i \leq y, r = 1, \ldots, n} \xi_{i3(-r)}(x) \]

\[ \geq \inf_{x \in \mathbb{R}} \inf_{y \leq T_r} \min_{r = 1, \ldots, n} \left\{ \overline{H}_{-1}(y|x), \overline{H}(y|x) \right\} \]

\[ \geq \inf_{x \in \mathbb{R}} \inf_{y \leq T_r} \overline{H}(y|x) - \sup_{x \in \mathbb{R}}, y \in \mathbb{R} \min_{r = 1, \ldots, n} \left| \overline{H}_{-1}(y|x) - \overline{H}(y|x) \right| \]

\[ \geq \inf_{x \in \mathbb{R}} \inf_{y \leq T_r} \overline{H}(y|x) - \sup_{x \in \mathbb{R}}, y \in \mathbb{R} \left| \overline{H}(y|x) - \overline{H}(y|x) \right| + O((nh_n)^{-1}) \text{ a.s.} \]

\[ \geq \alpha + O((nh_n)^{-1/2}|\log h_n|^{1/2}) + O((nh_n)^{-1}) \text{ a.s.} \]

\[ \geq \alpha + o(1) \text{ a.s.} \quad (3.7.31) \]

For the third inequality we used (3.7.26) and for the fourth we used Lemma 4.1 in Van Keilegom and Akritas (1999). The first factor in \( \alpha_2(x, r) \) is uniformly of order \( M + o(1) \) a.s. by use of (3.7.31) and (B4). Hence, by (3.7.26) we have that \( \alpha_2(x, r) \) is uniformly \( O((nh_n)^{-1}) \) a.s. The term \( \alpha_1(x, i, r) \) is uniformly \( O((nh_n)^{-1}) \) a.s., since the term \( \phi_x'(\xi_{i3(-r)}(x)) \) is again uniformly of order \( M + \)
To prove (3.7.28), similarly as before, we apply twice a first order Taylor expansion to get

$$
\Delta \hat{F}(T_i|x) = \frac{1}{\phi_x'(\gamma_i(x))} \phi_x'(\xi_i(x)) W_{ni}(x, h_n),
$$

where \( \gamma_i(x) \) is between \( \hat{F}(T_i|x) \) and \( \hat{F}(T_{i-1}|x) \), and \( \xi_i(x) \) is between \( \hat{H}(T_i|x) \) and \( \hat{H}(T_{i-1}|x) \). From here, statement (3.7.28) follows by use of assumption (B4), (3.7.31) and (3.7.26). To prove (3.7.29), we repeat the same calculations as above to get

$$
\Delta \hat{F}_{(-r)}(T_i|x) = \frac{1}{\phi_x'(\gamma_{(-r)i}(x))} \phi_x'(\xi_{(-r)i}(x)) W_{ni}(x, h_n),
$$

where \( \gamma_{(-r)i}(x) \) is between \( \hat{F}_{(-r)}(T_i|x) \) and \( \hat{F}_{(-r)}(T_{i-1}|x) \), and \( \xi_{(-r)i}(x) \) is between \( \hat{H}_{(-r)}(T_i|x) \) and \( \hat{H}_{(-r)}(T_{i-1}|x) \). From here (3.7.29) follows by the same reasoning as in (3.7.28). Finally, to show (3.7.30), we calculate

$$
\Delta \hat{F}(T_i|x) - \Delta \hat{F}_{(-r)}(T_i|x)
$$

\[=
\left[\frac{1}{\phi_x'(\gamma_i(x))} - \frac{1}{\phi_x'(\gamma_{(-r)i}(x))}\right] \phi_x'(\xi_i(x)) W_{ni}(x, h_n)
+ \frac{1}{\phi_x'(\gamma_{(-r)i}(x))} \left[\phi_x'(\xi_{(-r)i}(x)) - \phi_x'(\xi_i(x))\right] W_{ni}(x, h_n)
\]

\[=
\frac{\phi_x'(\gamma_{(-r)i}(x)) [\gamma_{(-r)i}(x) - \gamma_i(x)]}{\phi_x'(\gamma_i(x)) \phi_x'(\gamma_{(-r)i}(x))} \phi_x'(\xi_i(x)) W_{ni}(x, h_n)
+ \frac{1}{\phi_x'(\gamma_{(-r)i}(x))} \phi_x'(\xi_{(-r)i}(x)) [\xi_i(x) - \xi_{(-r)i}(x)] W_{ni}(x, h_n)
\]

\[= O((nh_n)^{-1}) W_{ni}(x, h_n) a.s.,
\]

where \( \xi_i(x) \) is between \( \xi_i(x) \) and \( \xi_{(-r)i}(x) \), and \( \gamma_{(-r)i}(x) \) is between \( \gamma_{(-r)i}(x) \) and \( \gamma_i(x) \). In the third equality we used that the denominators are uniformly bounded by assumption (B4)(iii), and we used the uniform boundedness of \( \phi_x'(\cdot) \) and \( \phi_x''(\cdot) \), which follows from similar reasoning as for proving (3.7.27). In the last equality above we used that \( \gamma_{(-r)i}(x) - \gamma_i(x) \) is uniformly \( O((nh_n)^{-1}) \) a.s. by use of (3.7.26) and (3.7.27), and we used that \( \xi_i(x) - \xi_{(-r)i}(x) \) is uniformly \( O((nh_n)^{-1}) \) a.s. by applying relation (3.7.26).
Conclusions and further research

In this thesis we studied nonparametric location-scale regression models of the form

\[ Y = m(X) + \sigma(X)\varepsilon, \]  

(4.0.1)

where \( \varepsilon \) is assumed to be independent of \( X \), and \( m(\cdot) \) and \( \sigma(\cdot) \) are unknown but smooth. The distribution of \( \varepsilon \) is also unknown. We supposed that the response in this regression model is subject to random right censoring and that the covariate is completely observed. This means that instead of observing \( Y \), we observe \( T = \min(Y, C) \) and \( \Delta = I(Y \leq C) \), where \( C \) represents the censoring time. In order to identify the distribution of \( Y \) given \( X \), an assumption needs to be made regarding the dependence between \( Y \) and \( C \) given \( X \). Instead of assuming that \( Y \) and \( C \) are independent given \( X \), which is common in the statistics literature, we assume that the dependence between \( Y \) and \( C \) for a given value of \( X \) is characterized by a copula function, which includes the conditional independence as a special case. This copula is supposed to be known and belonging to the Archimedean family, and it is allowed to depend on the value of \( X \), which allows us to model different dependence between \( Y \) and \( C \) for different values of covariate \( X \).

Under these model assumptions, we studied in Chapter 2 the nonparametric estimation of the location functional \( m(\cdot) \) and the scale functional \( \sigma(\cdot) \). The proposed estimators are nonparametric kernel estimators that are based on the estimator of the conditional distribution function proposed by Braekers and Veraverbeke (2005). We showed that the estimators satisfy an asymptotic representation and we proved the uniform convergence rates of the estimators and of their derivatives. We also studied their small sample performance through a simulation study and we applied the estimators to data on acute leukemia.
Chapter 4. Conclusions and further research

The object of Chapter 3 was to propose and study estimators of the error distribution and of the conditional distribution of $Y$ given $X$ under model (4.0.1). The estimator of the error distribution makes use of the assumed copula dependence and has a rather complicated structure. Once this estimator was found, the estimator of the distribution of $Y$ given $X$ was easily derived by using the independence between $\varepsilon$ and $X$. For both estimators, we proved an i.i.d. asymptotic representation, from which the weak convergence of the estimators followed easily. The finite sample behavior of the estimators was investigated in a small simulation study, in which both models that satisfy the location-scale structure (4.0.1) and that do not satisfy this structure were considered.

Although the thesis has tried to give answers to a number of questions related to the estimation of model (4.0.1) under the assumed censoring mechanism, a number of questions remain open and are worth studying in the future:

1. **Bandwidth selection**
   In the simulations in Chapters 2 and 3, we selected the bandwidth for estimating $m(\cdot)$, $\sigma(\cdot)$ and the conditional distribution of $Y$ given $X$ by minimizing the integrated mean squared error (IMSE) of these estimators, obtained from Monte Carlo simulation. This IMSE is however unknown in practice. It would be very useful to develop a data driven procedure to estimate this IMSE, by either using a plug-in procedure or a bootstrap procedure. This was however out of the scope of this thesis.

2. **Multivariate covariates**
   We assumed throughout the thesis that $X$ is univariate. However, in practice, one often has several covariates. There are basically two possible extensions of the results in this thesis to multiple covariates. First, one could extend the estimation of $m(\cdot)$ and $\sigma(\cdot)$ from one-dimensional smoothing to multi-dimensional smoothing. However, this might create curse-of-dimensionality problems, and moreover the results of Braeckers and Veraverbeke (2005) on which our results are heavily based, ought to be extended first to multi-dimensional covariates. A second possibility is to use a semiparametric model for $m(\cdot)$ and/or $\sigma(\cdot)$, like a single-index, partial linear, additive or varying coefficient model (or combinations of these) and to estimate $m(\cdot)$ and $\sigma(\cdot)$ under these model assumptions. Whereas this second option has the advantage of not suffering from curse-of-dimensionality problems, it does not seem an easy task to estimate $m(\cdot)$ and $\sigma(\cdot)$ in a semiparametric way.
3. Confidence bands

Regarding the functionals $m(x)$ and $\sigma(x)$, the theoretical focus of this thesis has been mainly on obtaining consistency rates and asymptotic representation of the local estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$, thereby providing the main ingredients to construct pointwise confidence intervals for $m(x)$ and $\sigma(x)$, the functionals of interest evaluated at $x$. However, to get an idea about the variability of the estimator of the whole curve, or to correctly answer questions about the shape of the functionals, we have to go a step further and derive confidence bands for $m(x)$ and $\sigma(x)$. The asymptotic results developed in the thesis can be used for that purpose.

4. Hypothesis testing

In the literature there is a variety of tests applicable to the specific case of the model presented in this thesis, when assuming the independence between $Y$ and $C$ given $X$. As discussed through the thesis, in practice the assumption of independency is often unjustified. Here we mention some tests that would be useful and, in our opinion, possible to generalize to the dependent case.

(a) Goodness-of-fit for parametric regression

Throughout the thesis we assume that $m$ is nonparametric. However, in many practical situations, one would benefit from using parametric regression models, since they describe the relationship between the response and the covariate in a simple way and usually allow for interpretability of the parameters (for instance, in linear regression). Nevertheless, if the parametric model is wrongly assumed, then the conclusions would be erroneous. Any parametric analysis should be accompanied by a test to check its validity and to avoid misspecification and wrong conclusions. Therefore, it would be very useful to develop specific goodness-of-fit tests for parametric models in regression. Under the assumption of independence between $Y$ and $C$ given $X$ such tests have been developed by Pardo-Fernández et al. (2007). These tests compare an estimator of the error distribution based on parametric residuals to another estimator relying on non-parametric residuals. It is also worth mentioning that these tests allow us (under some mild conditions) to test for characteristics of location functions, which cannot even be estimated consistently.

(b) Comparison of regression curves

If we can distinguish two or more groups (gender, treated patients and non-treated patients) in the population of interest, one may be interested in testing the equality of the corresponding regression curves. This allows to verify if the effect of the covariate on
the variable of interest is the same across all the groups. Assuming independence between $Y$ and $C$ given $X$, Pardo-Fernández and Van Keilegom (2006) have developed such tests based on the difference between estimators of the error distribution obtained under assumptions of parametric and nonparametric regression.

(c) Goodness-of-fit for parametric error distribution
Knowing that the distribution of the error $\varepsilon$ has a certain parametric form offers important advantages when doing inferences for the functionals $m$ and $\sigma$. Without assuming any parametric form for $m$ and $\sigma$, Heuchenne and Van Keilegom (2010) proposed tests for goodness-of-fit for parametric error distribution based on the difference between a nonparametric estimator of the error distribution and a parametric one.

5. Extension to cure models
In this thesis we implicitly assumed that the distribution of $Y$ given $X$ is a proper distribution, i.e. it reaches one at infinity. However, in certain contexts this assumption is not realistic. Consider e.g. Example 1 from Chapter 2, where the survival time of interest is the time from prostate cancer treatment to death caused by prostate cancer. In the corresponding data lots of patients were deemed cured, in a sense that an individual will have little or no risk of death due to prostate cancer. As a consequence some survival times are equal to infinity. Furthermore, survival data on any curable disease will share this feature, and given the staggering advancements in medical treatments there is an increasing number of curable diseases, resulting in need for proper statistical tools. It is therefore interesting to verify how the results in this thesis can be extended to the case where some survival times are equal to infinity. Models that take this feature into account are called cure models (see e.g. Maller and Zhou (1996), for an introduction). The literature on cure models with dependent censoring is very scarce. To the best of our knowledge, the only two papers that combine these two features are Li, Tiwari, and Guha (2007) (who consider the case without covariates), and Othus, Li, and Tiwari (2009)(who consider the case with covariates). Before studying the extension of the model in this thesis to the presence of a cure fraction, we first need to study the nonparametric location-scale model with independent censoring in the context of cure models. The extension to dependent censoring would then be a natural next step.


