Université catholique de Louvain
Institute of Statistics, Biostatistics and Actuarial Sciences

Quantile Regression with Censored Data

—

An investigation of new estimation procedures

Mickaël De Backer

A thesis submitted to the
Université catholique de Louvain
in partial fulfillment of the
requirements for the degree of
Doctor of Sciences

Thesis Committee:

Prof. **Anouar El Ghouch**	*Université catholique de Louvain*	Advisor
Prof. **Ingrid Van Keilegom**	*KU Leuven, Université catholique de Louvain*	Advisor
Prof. **Irène Gijbels**	*KU Leuven*	
Prof. **Olivier Lopez**	*Université Pierre et Marie Curie Paris VI*	
Prof. **Davy Paindaveine**	*Université Libre de Bruxelles*	
Prof. **Johan Segers**	*Université catholique de Louvain*	Chairman

June 2018

Fall in love with some activity and do it! Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.

Richard Feynman

Acknowledgements

En cette fin d'aventure doctorale, il est une célèbre phrase que je voudrais brièvement reprendre afin de m'aider à exprimer toute la gratitude ressentie lors de la rédaction de ces quelques lignes. Cette phrase est communément attribuée, lors d'une correspondance avec son rival Robert Hooke, à l'un des plus grands esprits scientifiques que l'humanité ait engendré, Sir Isaac Newton:

"If I have seen further, it is only by standing upon the shoulders of giants."

Que l'on se rassure, contrairement à ce géant parmi les géants, mon intention n'est ici nullement de reprendre cette image à des fins scientifiques par rapport au travail repris dans ce manuscrit. Simplement, j'aime à penser que l'opportunité même de tenter de voir plus loin lors de mon parcours de recherche, à une échelle tellement plus modeste, est avant tout le fruit d'un environnement et d'un soutien exceptionnels que je tiens ici à mettre en avant. Ainsi, au-delà d'un accomplissement académique personnel, la rédaction de ce manuscrit est à mes yeux une opportunité de prendre le temps d'apprécier les géants de mon quotidien qui ont contribué à la fois à mon épanouissement personnel et intellectuel repris dans ce texte. C'est donc dans cette optique que je voudrais débuter par ces quelques lignes qui témoignent de ma plus sincère gratitude envers toutes ces personnes m'ayant, de près ou de loin, porté sur leurs épaules tout au long de mon parcours.

Pour commencer, mes plus sincères et chaleureux remerciements vont aux deux formidables instigateurs de l'aventure qu'a été cette thèse, Ingrid et Anouar. Le début de notre collaboration est un fantastique fruit du hasard qui a résulté en quatre années d'apprentissage où votre bienveillance, bonne humeur, et confiance ont rythmé nos réunions et mon évolution quotidienne. Votre encadrement et vos qualités humaines n'ont cessé de m'encourager à travers les années, et je me considère particulièrement chanceux d'avoir pu profiter du formidable duo que vous formez. Pour tout cela, je vous remercie très sincèrement.

Je voudrais ensuite également exprimer ma profonde gratitude et reconnaissance envers tous les autres membres du Jury : Madame et Messieurs les Professeurs Irène Gijbels, Olivier Lopez, Davy Paindaveine et, enfin, Johan Segers que je remercie également pour avoir accepté d'endosser le rôle de président de ce Jury. Votre intérêt pour nos travaux a été un réel honneur et une grande source de fierté, et vos diverses questions et remarques lors de la défense privée ont sans conteste contribué à améliorer la qualité de ce manuscrit.

Je souhaite également consacrer quelques lignes en ce début de manuscrit à une autre Professeure, car mon parcours académique aurait sans aucun doute pris une tournure bien

différente si je n'avais pas eu la chance de pouvoir profiter de ses remarquables qualités de pédagogue durant mes trois années d'études aux Facultés Universitaires Saint-Louis. Il s'agit de la Professeure Dominique Deprins qui, avec un enthousiasme à toute épreuve, tente de convaincre son public d'étudiants que le monde d'aujourd'hui est plus statistique que jamais, en mettant compréhension et intuition au cœur de son enseignement au détriment des 'recettes' que d'aucuns souhaiteraient pourtant. Il va de soi que cette thèse n'aurait vu le jour sans cet enseignement qui m'a profondément inspiré pour la suite de mon parcours.

Je me considère ensuite particulièrement chanceux d'avoir pu vivre mon parcours doctoral dans ce chaleureux cadre que constitue l'ISBA, mettant bien à mal le cliché selon lequel un doctorant évoluerait reclus dans un environnement des plus solitaires. Mes remerciements envers toutes les personnes le constituant se dirigent ici en premier lieu vers son formidable staff administratif, composé de Nadja, Nancy, Maguy, Sophie et Tatiana. Au risque d'oublier quelqu'un, je voudrais ensuite également remercier pour l'excellente ambiance au travail l'ensemble des collègues doctorants et post-doctorants que j'ai pu avoir durant ces dernières années: Adrien, Alexandre, Anna, Antoine B., Antoine S., Aurélie, Benjamin, Cédric, Cheikh, Edward, Florian, François, Gildas, Gilles, Hélène, Hugues, John-John, Joris, Kassu, Maïlis, Manon, Michał, Nathalie Lu., Nathan, Nicolas, Oswaldo, Rebecca, Samuel, Sophie, Stefka, Sylvie et Vincent. Entre concours de pronostics, rosés en terrasse devant l'Euro et journées d'études/billard à Fréjus, ce parcours doctoral aura su trouver un joli équilibre entre sérieux et moins sérieux à vos côtés.

Je tiens ensuite également à remercier l'ensemble de l'équipe du SMCS, qui participe plus qu'activement à l'ambiance de nos couloirs tout en nous permettant de nous 'salir' les mains avec de vraies données lors de nos fameuses premières consultances. Je tiens donc à remercier en particulier Alain, Aurélie, Catherine, Céline, Nathalie Le., Lieven, Séverine et (presque) Vincent, pour l'opportunité unique et la confiance attribuée lors de nos premières joies de consultant. Merci également à Alain pour l'aide et l'usage plus que raisonnable des serveurs 'SMCS'.

Je me tourne ensuite vers mes proches, amis et famille qui m'ont toujours soutenu[1] durant l'ensemble de mon parcours et que je remercie vivement. En particulier, je remercie du fond du cœur mes parents, qui ont toujours tout mis en œuvre pour que les cinq frères que nous sommes puissions nous épanouir dans la vie. Avec les années qui passent (et la maturité !) on se rend compte que, ce qui semble acquis dans les yeux d'un enfant, traduit en réalité une chance inouïe de vous avoir comme parents.

Enfin, je remercie également du fond du cœur Marine qui partage le quotidien de cette thèse depuis son origine. Son inestimable soutien de géant consiste à croire en moi bien plus que je ne l'ose moi-même, me portant au quotidien. Cette fin de thèse marque la fin d'un chapitre important à tes côtés, et maintenant que celui-ci touche à sa fin, je me réjouis de l'avenir qui nous tend les bras.

Mickaël De Backer, 11 juin 2018

[1] "Cette thèse, cette put*** de thèse"

Preface

One of the most fundamental purposes of applied statistics is to provide an appropriate description of the relationship between certain stochastic covariate variables and a response of interest. While classical statistical theory historically commits for this task to the most popular mean regression modelling, a growing interest has emerged these last decades for the analysis of conditional quantiles instead, as the latter allow, among other conveniences, for a more complete representation of the relationship of interest.

Providing practitioners convenient tools for an appropriate estimation of a quantile regression model, be it parametrically, semiparametrically or nonparametrically specified, then constitutes a major challenge for statisticians. This statement is further emphasized in situations where the data at hand slightly diverges from the ideal scenario where all observations are completely recorded. One example of such nuisance in the data that is often encountered in practice concerns the occurrence of right-censoring of the responses, where one no longer completely observes the true response variable of interest, but instead only the minimum of it and an interfering censoring variable.

In this context, this thesis aims at developing new estimation procedures for quantile regression models where right-censoring of the response variable may occur, although part of our work will be observed to encompass some novelties for situations with strictly complete observations as well. In particular, this manuscript is organized as follows:

Chapter 1: Overview.

This first chapter offers an overview of the notations and main material required in order to comfortably grasp the content of this thesis. In particular, the chapter starts by covering a brief introduction on the notions of quantiles and quantile regression, by highlighting among other things a certain duality between two distinct approaches when it comes to the estimation of conditional quantiles in practice. The chapter then proceeds to introduce some fundamental notions related to copulas and copula-based regression modelling that will

come in handy for some of our contributions. Then, motivated by the main context of this thesis, a brief introduction to survival analysis and censored quantile regression modelling is considered, before closing the chapter by highlighting clearly where the contributions of this thesis fit with respect to the literature on censored quantile regression estimation.

Chapter 2: Semiparametric Copula Quantile Regression for Complete or Censored Data.

In this chapter, two main contributions are developed for the convenient combination of copulas with quantile regression modelling. Building on the work of Noh et al. (2015), the first contribution concerns the estimation of a quantile regression for strictly completely observed data, where an alternative estimation strategy is here considered in order to bypass the shortcomings of the approach of Noh et al. illustrated in the literature. In a second contribution, the proposed estimation procedure is then adapted in order to accommodate for the potential presence of right-censored data in the response, hereby preventing one from completely observing the latter. For the specific handling of censoring, a judicious use of the so-called inverse-censoring-probability technique is considered here, resulting in a competitive semiparametric estimation procedure for quantile regressions with censored data. However, although extensively used in the literature on censored quantile regressions, a major flaw of the inverse-censoring-probability technique is that the latter may in fact be traced back to some notorious work in mean regression modelling with censored responses, hereby suggesting that the latter is not specifically designed to account for the specificities of quantiles in its handling of censoring. This then motivates the development of the next chapter.

Chapter 3: An Adapted Loss Function for Censored Quantile Regression.

This third chapter introduces a reconsideration of the usual handling of censored data in the literature on quantile regression estimation. Following the insights of the previous chapters, the objective is here to provide a general framework for the consideration of censored data in quantile regressions. To that end, and mimicking the pioneering work of Koenker and Bassett (1978) expressing conditional quantiles as the result of minimizing a conditional expected loss, the idea is here to define an alternative loss function that already, and automatically, accounts for censoring in the responses. This results in a very simple accommodation of the loss function routinely employed for complete observations, and hence could allow for a simple transcription to the world of survival data of numerous models that have been initially developed in the literature on complete observations. An illustrative application of the general methodology to the linear regression framework is then considered and further studied in detail, both from theoretical and practical points of view.

Chapter 4: Linear Censored Quantile Regression: A Novel Minimum-Distance Approach.

In this chapter, a new procedure is investigated for the estimation of a linear quantile regression with censored data, motivated by several insights following our previous work. To that end, rather than building the procedure on the conventional technique of minimizing a conditional expected loss, the estimation of the linear coefficients is here carried out by minimizing an alternative measure of distance, taken as the distance between the quantile level of interest and the conditional distribution of the response given the covariates, the latter being evaluated at the value of the linear function for each observation. In essence, this estimation approach is thus similar to the well-known technique consisting in estimating the conditional distribution of the response given the covariates in a first step, and inverting it in a second step in order to recover the conditional quantiles. The main novelty then comes here from the somewhat atypical consideration of the latter technique in the linear framework, however motivated in this context by several arguments following our work and experience with quantile regressions for censored data.

Chapter 5: Discussion and Extensions.

This last chapter closes the thesis, for which a detailed conclusion following our journey is first provided, before considering the early, and illustrative developments of three suggestions that could be further investigated for future contributions in the domain of censored quantile regression estimation.

Contents

Chapter 1

Overview

In this thesis, we aim at extending statistical methodologies for the estimation of general quantile regression models with both complete and censored response variables, although a stronger emphasize will be placed on the latter situation. Regression analysis is arguably the most common and most powerful statistical tool when it comes to investigating the relationship between certain covariate variables and a response of interest. While conditional mean models historically dominated the regression landscape, the last decades have witnessed the emergence of a wide variety of regression models focusing on conditional quantiles instead, stemming from the seminal work of Koenker and Bassett (1978) who introduced conditional quantiles as a natural extension of ordinary quantiles when considering additional covariate information. Quantile regression has since then received a notable interest in the literature on both theoretical and applied statistics, with several conveniences over mean regression models acting as driving forces for the latter research, as will be made explicit in the following sections.

This introductory chapter is devoted to the establishment of all the material required in order to easily apprehend the suggestions of our work. To that end, a natural opening reported in Section 1.1 is a brief overview of the basic notions of ordinary and conditional quantiles. Next, a quick necessary detour to the concept of copulas will be undertaken and motivated in Section 1.2. Then, as our primary motivation comes from the world of survival analysis, Section 1.3 reports the main notions and singularities related to the latter domain. Section 1.4 next aims at bridging Sections 1.1 and 1.3 by resuming earlier work in the literature on censored quantile regression techniques, to which our proposed procedures will be brought into contrast. Finally, Section 1.5 concludes this chapter by providing a succinct overview of the outline and contributions of the thesis.

1.1 Quantiles and quantile regression for complete observations

This first section is devoted to the necessary introduction of some basic notations and concepts related, in first instance, to ordinary quantiles. The extension to conditional quantiles will then be considered next, while a few motivational words on the relevance of the quantile regression methodology for real data applications will complete the section. All the concepts and methodologies related to quantile regression are detailed in numerous textbooks, of which the work of Koenker (2005) is arguably the most popular.

1.1.1 Quantiles without covariates

For notational purposes, let us introduce here a real-valued (continuous) random variable T with cumulative distribution function (c.d.f.) $F_T(t) = \mathbb{P}(T \leq t)$. In the context of this thesis, T will be mainly to be read as a time-to-event variable, or some monotone transformation of the latter, although this is inconsequential for this introductory section. Now, for any $\tau \in (0, 1)$, we denote by m_τ the τ-th *ordinary quantile* of the variable T, formally defined as

$$m_\tau = \inf\{t \in \mathbb{R} : F_T(t) \geq \tau\} \equiv F_T^{-1}(\tau). \tag{1.1.1}$$

A textbook approach to motivating the role of quantiles is to bring the latter notion into contrast with the expectation of T, denoted by $\mathbb{E}(T)$, on two specific features. First, a crucial element of the quantile function $F_T^{-1}(\tau)$ is its inherent capacity of fully characterizing the distribution of T, whereas the expectation allows for identifying only one characteristic of the latter. In that sense, quantiles provide by nature a much broader picture than the expectation. A second appealing point of comparison is that both the expectation and quantiles allow in fact for an alternative formulation expressed as the minimization of an expected loss. For the case of $\mathbb{E}(T)$, it is indeed well-known that the latter is simply recovered from minimizing an expected quadratic loss, that is, $\mathbb{E}(T) = \arg\min_{a \in \mathbb{R}} \mathbb{E}[(T - a)^2]$. Similarly, one may easily show by derivation that any quantile m_τ of T is recovered by an analogous minimization problem given by

$$m_\tau = \arg\min_{a \in \mathbb{R}} \mathbb{E}\left[\rho_\tau(T - a)\right], \tag{1.1.2}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u \leq 0))$ is the convex "check" loss function, illustrated in Figure 1.1 for different values of τ, and $\mathbb{1}(\cdot)$ is the indicator function. Although long known, the latter formulation has long "languished in the status of curiosum, appearing for example as an exercise in Ferguson (1967, p. 51)", as stated in the seminal paper of Koenker and Bassett (1978). From a modelling point of view however, formulation (1.1.2) strikes as being very appealing as it allows one to

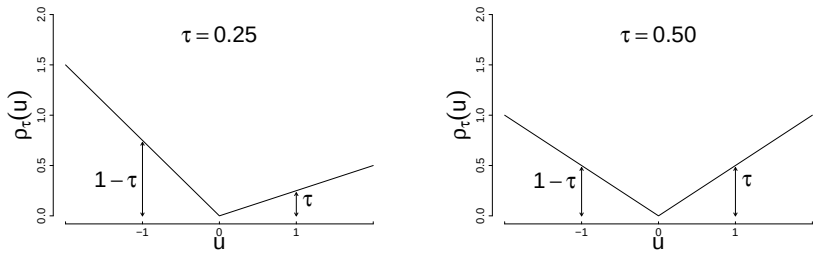

Figure 1.1: Check loss function for quantile levels $\tau \in \{0.25, 0.5\}$.

retrieve a similar, and hence familiar, framework as for the determination of the expectation. The latter observation is of particular convenience when considering the inclusion of covariates, as will be exposed in the next section.

Now, in anticipation of Chapter 3, it is at this point interesting to digress slightly and discuss one of the links between the definition of quantiles and their check-based formulation. To that end, with the definition of m_τ for a continuous random variable at hand, our trivial starting point is that $\mathbb{P}(T \leq m_\tau) = \tau$. Hence, finding m_τ may be written equivalently as finding the root in a of the equation

$$\mathbb{E}[\mathbb{1}(T > a) - (1 - \tau)] = 0. \tag{1.1.3}$$

Then, instead of using this estimation equation in practice, the idea is here to define a loss function such that, when taken in expectation and minimized, one eventually retrieves expression (1.1.3). Giving ourselves a realization t of the random variable T, this is easily achieved by defining the function $a \mapsto \int_a^t (\mathbb{1}(t > s) - (1 - \tau))\,\mathrm{d}s$. Simple calculations then reveal that the latter loss function is nothing more than the check function $\rho_\tau(t - a)$, hereby linking the latter with the definition of m_τ as $\partial/\partial a\, \mathbb{E}[\rho_\tau(T - a)] = (1 - \tau) - \mathbb{P}(T > a)$. This very simple observation of how one may come to define the check function will in fact be seen to lie at the heart of the methodology of Chapter 3.

Next, if instead of knowing the true distribution F_T, one only disposes of an independent and identically distributed (i.i.d.) sample $T_1, \ldots, T_n$ drawn from F_T, formulations (1.1.1) and (1.1.2) both offer decent starting points for the practical estimation of m_τ. That is, one may for instance embrace formulation (1.1.1) to estimate m_τ by

$$\widehat{m}_\tau = \inf\{t \in \mathbb{R} : \widehat{F}_T(t) \geq \tau\}, \tag{1.1.4}$$

where $\widehat{F}_T(t)$ is an appropriate estimator of $F_T(t)$. An obvious candidate for the latter estimator is the empirical distribution function $n^{-1} \sum_{i=1}^n \mathbb{1}(T_i \leq t)$, although

more refined candidates may be considered (for instance by smoothing, see e.g. Hubert et al. (2013)), as will be evoked in Chapter 4 as well. If instead one wishes to consider formulation (1.1.2) for the estimation of m_τ, then this may be carried out numerically by

$$\widehat{m}_\tau = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(T_i - a). \tag{1.1.5}$$

A crucial lesson stemming from the latter formulation is that we may replace the notion of sorting, which seemed inextricably tied to the definition of sample quantiles, by an elementary optimization problem. Similarly to existing literature on the subject, we will refer to the two described estimation strategies in the following of this thesis as "inverse-c.d.f." and "check-based", for rather obvious reasons. This simple duality between inverse-c.d.f. and check-based modelling is in fact at the heart of the motivation for Chapter 4, given it naturally reappears when considering covariate information as will be reported next.

1.1.2 Conditional quantiles

In practice, one often wishes to include the information of an additional covariate vector $\boldsymbol{X}$ of dimension $d \geq 1$ into a regression model in order to try and capture the relationship between the latter and the response variable T. For this objective, and in an attempt primarily motivated by the need for a robust (to outliers in the response) alternative to the usually employed mean regression model, Koenker and Bassett pioneered the development of quantile regression. Their rationale, initially concentrated solely on linear regression models, is in fact a direct generalization of the previous section. To introduce the latter, let us denote here by $m_\tau(\boldsymbol{x})$, for any $\tau \in (0,1)$, the τ-th *conditional quantile function* of T, or some monotone transformation of the latter, given $\boldsymbol{X} = \boldsymbol{x}$. Specifically,

$$m_\tau(\boldsymbol{x}) = \inf\{t \in \mathbb{R} : F_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \geq \tau\}, \tag{1.1.6}$$

where $F_{T|\boldsymbol{X}}$ denotes the conditional c.d.f. of T given $\boldsymbol{X}$. Similarly to ordinary quantiles, Koenker and Bassett then noted that a general and equivalent check-based formulation for conditional quantiles may be given by

$$m_\tau(\boldsymbol{x}) = \arg\min_{a\in\mathbb{R}} \mathbb{E}\Big[\rho_\tau(T-a)|\boldsymbol{X} = \boldsymbol{x}\Big]. \tag{1.1.7}$$

Clearly, a major appealing feature of this check-based formulation exploited by Koenker and Bassett and many following authors is its comparability with the well-studied mean regression framework where one minimizes a conditional expected quadratic loss. For illustration of this statement, let us consider as in Koenker and Bassett the example of a simple linear regression model, and hence, based on an i.i.d. sample $(T_i, \boldsymbol{X}_i), i = 1, \ldots, n$, from $(T, \boldsymbol{X})$, let us assume that the regression

model is of the form $m_\tau(\boldsymbol{X}_i) = \beta_\tau^\mathsf{T} \boldsymbol{X}_i$, where $\boldsymbol{X}$ is here $(d+1)$-dimensional given it includes a component relative to the intercept, and β_τ is the $(d+1)$-dimensional unknown quantile coefficient vector. Then, based on (1.1.7) and by similarity with mean regression models, Koenker and Bassett simply suggested to estimate β_τ by

$$\widehat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^{d+1}} \sum_{i=1}^{n} \rho_\tau(T_i - \beta^\mathsf{T} \boldsymbol{X}_i). \tag{1.1.8}$$

Hence, by expressing quantile regression as the minimization of a conditional expected loss, Koenker and Bassett opened the door to numerous creative regression modelling techniques that were typically initially suggested for mean regressions, and then further applied to the quantile context, see e.g. the review of Yu et al. (2003). In fact, the literature on check-based quantile regression blossomed so broadly that the celebrated book of Koenker (2005) exclusively documents check-based models, with one copula-based exception, leaving out any room for regression models based on formulation (1.1.6) instead.

Apart from this resemblance with mean regressions, a second main ingredient supporting this historical success of check-based modelling may undoubtedly be attributed to the powerful numerical techniques developed for the practical minimization of the latter formulation when confronted to finite samples. Taking again the example of a linear model, the effectiveness of some linear programming algorithms applied to the latter framework is such that, for sufficiently large samples, they may succeed in providing numerical solutions even faster than their mean regression analogue relying for their part on closed-form solutions, as exposed in Portnoy and Koenker (1997). These two features stemming from formulation (1.1.7) are undeniably at the cornerstone of the check-based triumph for practical quantile regression modelling.

On the other hand, and to a lesser extent, in certain particular situations the inverse-c.d.f. formulation (1.1.6) also emerged in the literature as a valuable starting point to estimate conditional quantiles in practice. One such particular situation is the domain of nonparametric regression, where a straightforward extension of (1.1.4) is to estimate $m_\tau(\boldsymbol{x})$ by

$$\widehat{m}_\tau(\boldsymbol{x}) = \inf\{t \in \mathbb{R} : \widehat{F}_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \geq \tau\}, \tag{1.1.9}$$

where $\widehat{F}_{T|\boldsymbol{X}}$ is some appropriate estimator of $F_{T|\boldsymbol{X}}$. This is the starting point of several papers in the literature, see e.g. Yu and Jones (1998), Li and Racine (2008) and Li et al. (2013), where the exposed numerical results globally suggest that for nonparametric regression, the inverse-c.d.f. method tends to outperform its check-based counterpart, especially for 'central' quantile levels where one is not required to estimate $F_{T|\boldsymbol{X}}$ in the tails of the response. For instance, Yu and Jones explicitly recommend to adopt the inverse-c.d.f. version of their local linear estimator rather than its check-based counterpart. This observation, along with further arguments that will be developed when necessary, is at the origin of Chapter 4.

Conclusively, the extension from ordinary quantiles to conditional quantiles gives again rise to a certain duality between inverse-c.d.f. and check-based modelling, although the latter offers here a clear modelling advantage in most situations as will be exploited in Chapters 2 and 3. For both approaches however, taking one step back, it is worth mentioning how the study of conditional quantiles may allow an analyst to complement, or simply improve, his regression analysis in practice. This is the topic the next section.

1.1.3 Motivational features of quantile regression

We provide in this section, through two examples, a brief overview on the usually praised features of quantile regressions over mean regressions. The first example is taken from data going back to the study of Engel (1857) on households' food expenditure dependence on their income. Figure 1.2 depicts the original data along with superimposed linear mean and quantile regression estimators. Clearly, a first motivation for applying here a quantile regression analysis is the evident increase in variability of food expenditures when the level of income grows. While standard mean regressions do not allow for detecting such heteroscedasticity, quantile regressions are perfectly able to detect the latter by exploring the conditional distribution with a set of different values of τ. An additional information highlighted by this exploration is the left-skewed feature of the conditional distribution, given the

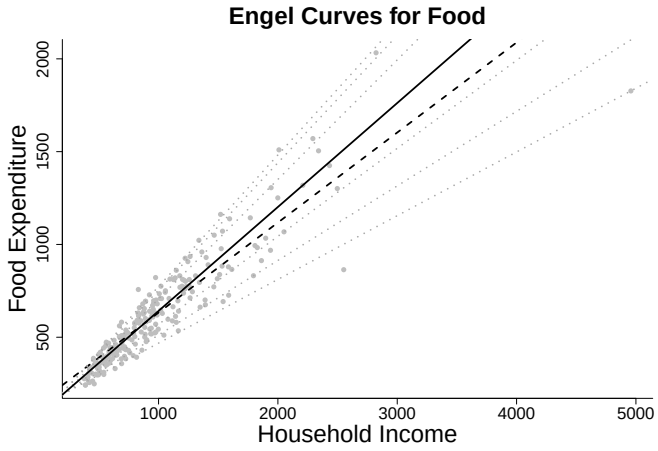

Figure 1.2: Household food expenditure data (available in the R library `quantreg`). The dashed black line represents the least squares estimate of the conditional mean while the plain line depicts the estimated median linear regression. Six other quantile regression estimates for $\tau \in \{0.05, 0.1, 0.25, 0.75, 0.9, 0.95\}$ are depicted in 'increasing' order in dotted grey lines.

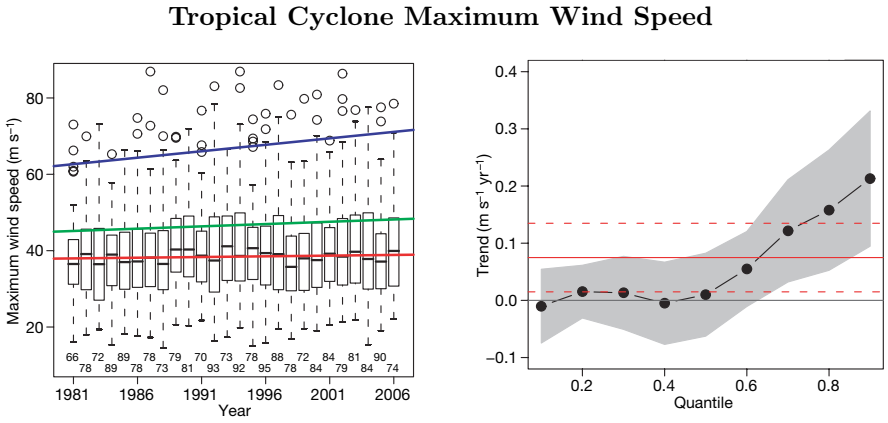

Figure 1.3: Tropical cyclone maximum wind speed graphs of Elsner et al. (2008). The left plot reports boxplots of the data by year, with trend lines for $\tau = 0.5$, $\tau = 0.75$ and 1.5 times the interquartile range. The right plot illustrates a typical graph in quantile regression analysis: the slope coefficient of the quantile regression estimation with confidence bands (90%) depicted as a function of τ. The red lines represent here the mean regression slope estimation along with its 90% point-wise confidence band.

regression lines tend to be closer to each other for higher values of τ. Now, comparing specifically the mean and median regression estimations, the chosen example also highlights that the former may in some cases not even be most appropriate to explore the central behavior of the conditional distribution. Indeed, as one can observe, the least square estimates provides here a very poor fit for the conditional mean of households with lowest incomes, as these points are all below the represented dashed line. This finding may here be attributed to both the presence of a few unusual data points and the aforementioned asymmetry of the conditional distribution. In opposition to this, the median regression does not suffer as much from these combined effects in its estimation of the central tendency of the conditional distribution, hereby illustrating the robustness of quantile regressions to such undesirable effects in practice.

In some situations however, the mean regression estimation, while truthfully reflecting the central behavior of the data, simply does not report the regression information of interest in opposition to tails of the conditional distribution. This is namely the situation in many scientific fields where extremes are of primary interest; for instance, it might be socially desirable to implement an educational policy that has a significant reducing impact on the proportion of students in great difficulty, even though its effect on the average level of all students is negligible. The second example of this section graphically further illustrates this statement in the domain of environmental studies. Figure 1.3 reports here data coming from a study

of Elsner et al. (2008) on maximum wind speeds of the strongest tropical cyclones over the years. In this situation where both the mean and median regressions depict little trend over time, the authors proposed a quantile regression study for multiple values of τ in order to identify that upper quantiles clearly report an increasing trend over the years, a finding that is believed to be in coherence with the global warming of oceans. Hence, a simple application of the quantile regression procedure allows here the analysts to provide clear evidence of the increasing intensity of upper extremes, a verdict much more crucial to climatologists than the central behavior of the data cloud.

Altogether, these simple illustrations exhibit the main reasons for which quantile regressions are usually praised in the literature on both applied and theoretical statistics, as they provide a broader picture than mean regressions, are robust to outliers in the response and flexible with respect to the error distribution. This then broadly motivates the study of such models in this thesis. However, before covering in detail the latter study, it will now be necessary to take one step back from quantile regressions in the following section, and introduce a few auxiliary notions related to copulas that are key to some of our work.

1.2 Copulas and copula-based quantile regression

In anticipation of the methodology of Chapter 2 proposing a copula-based quantile regression modelling, this section here offers a brief review of the basic notions related to copulas along with their link to conditional quantiles. To that end, still considering the situation where one disposes of a complete response variable T and a d-variate covariate vector $\boldsymbol{X}$, we first introduce the definitions and characteristics that are at the cornerstone of copula theory and its recent success in statistical methodologies. We then briefly discuss different approaches promoted in the literature for the estimation of copulas when confronted to an i.i.d. sample $(T_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, and complete the section by discussing the intrinsically close relationship between copulas and both inverse-c.d.f. and check-based quantile regression modelling. Of course, this section is not intended to be an exhaustive review of the copula literature at all, but rather focuses on the notions reappearing in later chapters.

1.2.1 Definitions and generalities

We start this section by introducing the definition of a copula function, and then recall the fundamental theorem of Sklar (1959) lying at the core of copula theory. A book length treatment of copulas may further be found in Nelsen (2006) and Joe (2014).

Now, let us suppose that all X_j, $j = 1, \ldots, d$, are continuous random variables with c.d.f. denoted by $F_1, \ldots, F_d$, respectively. Throughout this thesis, we

further denote by f_j and f_T the density of $X_j, j = 1, \dots, d$, and T, respectively. The *copula* function $C_{T\boldsymbol{X}}$ of $(T, \boldsymbol{X})$ is then defined as the joint c.d.f. of $(F_T(T), F_1(X_1), \dots, F_d(X_d))$, that is,

$$C_{T\boldsymbol{X}}(u_0, u_1, \dots, u_d) := \mathbb{P}(U_0 \leq u_0, U_1 \leq u_1, \dots, U_d \leq u_d),\ (u_0, \dots, u_d) \in [0,1]^{d+1},$$

with $U_0 = F_T(T)$ and $U_j = F_j(X_j)$, $j = 1, \dots, d$. While copulas were initially introduced and studied in probability theory, their thriving success across the last decades in the statistical literature is easily to be attributed to the pioneering theorem of Sklar (1959). The latter states that, for every $\boldsymbol{x} = (x_1, \dots, x_d)^\mathsf{T}$, the multivariate c.d.f. $F_{T\boldsymbol{X}}$ of $(T, \boldsymbol{X})$ evaluated at $(t, \boldsymbol{x})$ may be decomposed into marginal distributions and the unique copula $C_{T\boldsymbol{X}}$ as follows:

$$F_{T\boldsymbol{X}}(t, \boldsymbol{x}) = C_{T\boldsymbol{X}}(F_T(t), \boldsymbol{F}(\boldsymbol{x})), \quad \forall (t, \boldsymbol{x}) \in \mathbb{R}^{d+1}, \tag{1.2.1}$$

where $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1), \dots, F_d(x_d))^\mathsf{T}$. Equation (1.2.1) is usually interpreted in the way that the copula $C_{T\boldsymbol{X}}$ disjoints the marginal behaviors of T and $\boldsymbol{X}$ from their stochastic dependence structure, hereby allowing a great modelling flexibility by separating the effects coming from one or the other part. This particular nature of copulas shaped them into an indispensable tool in many scientific fields relying on flexible multivariate modelling, such as insurance, finance or even astronomy. For practical implementation, one then needs to resort to appropriate estimation procedures for copulas, which is the topic of the next section.

1.2.2 Copula estimation

As reported above, in multivariate frameworks copulas allow to abstract from the complications of marginal behaviors and focus solely on the form of the dependence. This motivates the necessity in the statistical literature to propose practical procedures for the estimation of copulas, often implemented via the *copula density*. Taking the example of the copula $C_{T\boldsymbol{X}}$, the latter density associated to $C_{T\boldsymbol{X}}$ is defined as $c_{T\boldsymbol{X}}(u_0, u_1, \dots, u_d) = \partial^{d+1} C_{T\boldsymbol{X}}(u_0, u_1, \dots, u_d)/\partial u_0 \partial u_1 \dots \partial u_d$, for $(u_0, \dots, u_d) \in [0,1]^{d+1}$.

Now, a major specificity of such copula density estimation, be it for parametric or nonparametric approaches as will be distinguished later, is that one typically only observes a sample $(T_i, \boldsymbol{X}_i)$, $i = 1, \dots, n$, whereas $c_{T\boldsymbol{X}}$ is in fact the density of $(U_0, \dots, U_d) = (F_T(T), F_1(X_1), \dots, F_d(X_d))$, where F_T and $F_j, j = 1, \dots, d$, are usually unknown. As a consequence, general estimation of a copula density is unfeasible as such and requires to work in two sequential steps[1]: first, compute 'pseudo-observations' $(\widehat{U}_{0i}, \dots, \widehat{U}_{di}) = (\widehat{F}_T(T_i), \widehat{F}_1(X_{1i}), \dots, \widehat{F}_d(X_{di})), i = 1, \dots, n$, where $\widehat{F}_T$ and $\widehat{F}_j$ are some appropriate estimators of F_T and F_j, $j = 1, \dots, d$,

[1] A parametrization of the marginals as well may in fact allow for a one-step estimation only, but this will not be treated here and is seldom recommended in the literature given the strong exposure to possible misspecification.

and then, in a second step, treat the obtained sample as if it were coming from C_{TX} in order to estimate the copula density. In the literature, in order to avoid undesirable misspecifications issues at this point, it is common practice to estimate the marginal distributions using some rescaled version of the empirical c.d.f. in order to further keep the constructed observations in the interior of $[0,1]$, that is, $\widehat{U}_{0i} = (n+1)^{-1}\sum_{k=1}^{n} \mathbb{1}(T_k \leq T_i)$ and $\widehat{U}_{ji} = (n+1)^{-1}\sum_{k=1}^{n} \mathbb{1}(X_{jk} \leq X_{ji})$, for $i = 1, \ldots, n$, and $j = 1, \ldots, d$. Throughout this thesis, we will adopt this standard practice as well.

Now, disposing of a sample of pseudo-observations at hand, as for any density function estimation, typically two school of thoughts may be opposed for the estimation of the copula density, depending on a possible parametrization of the latter. As both approaches will be considered in our work, we distinguish and detail next some relevant estimation procedures for the two of them.

Semiparametric estimation

First, a wide literature is devoted to the parametric modelling of the copula, that is, assuming that the latter belongs to a certain parametric family $\mathcal{C} = \{c(u_0, \ldots, u_d; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$, for some $p \in \mathbb{N}$. Among these families, perhaps the most popular and hence worth mentioning are the families of Archimedean copulas (e.g. Clayton, Frank or Gumbel) and elliptical copulas (e.g. Gaussian, Student t). In this situation, apart from the particularity of disposing only of pseudo-observations, estimating the copula density basically amounts to fitting a parametric multivariate density, for which likelihood methods are undoubtedly most popular. Specifically, estimation of the copula relies here entirely on the estimation of $\boldsymbol{\theta}$ through

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \log c_{T\boldsymbol{X}}(\widehat{F}_T(T_i), \widehat{F}_1(X_{1i}), \ldots, \widehat{F}_d(X_{di}); \boldsymbol{\theta}). \tag{1.2.2}$$

This estimation procedure, spearheaded by Genest et al. (1995) and Shih and Louis (1995), is qualified as being overall semiparametric as it relies on nonparametric estimators in a first stage for the determination of the pseudo-sample. Parametric copula modelling and subsequent estimation has been proved to be very useful in a large set of domains such as finance (e.g. Genest et al. (2009a)), biology (e.g. Kim et al. (2008)), hydrology (e.g. Salvadori and De Michele (2007)) and so on.

Of course, as for any parametric procedure, this first approach to copula estimation is based on the somewhat speculative thought[2] that the true copula $C_{T\boldsymbol{X}}$ indeed belongs to the chosen family $\mathcal{C}$, or is at the least sufficiently and conveniently close to the latter. In practice however, there are inevitable situations where parametric specifications fail to provide the flexibility needed to accurately specify the

[2]Of course, a lot of interesting research is dedicated to a more formal study on how well the parametric assumption conforms to the data. This is here however left out for sake of brevity.

underlying natural law, and the copula literature, although granting an extensive range of parametric copula families, is naturally not immune to this drawback. The motivational example of Section 2.2.2 offers a precise and obvious illustration of such a pitfall. In these situations, data analysis must rely on the nonparametric front of statistical modelling, where no rigid assumptions are made on the underlying distribution. Estimation of such nonparametric tools in this context is then the topic of the next section.

Nonparametric estimation

For ease of presentation, this nonparametric approach to copula estimation will here be reviewed solely for the particular bivariate situation, that is, assuming $d = 1$ and hence concentrating on the copula C_{TX_1} of (T, X_1). This is of course common practice in the literature given that, while theoretical extensions to higher dimensions are often claimed to be straightforward, practical implementation is on the other hand only rational within the restrictions laid by the *curse of dimensionality*.

Now, the probably earliest step in the direction of nonparametric copula estimation may be traced back to the work of Deheuvels (1979) with the study of the *empirical copula*

$$\widehat{C}_{TX_1}(u_0, u_1) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(\widehat{F}_T(T_i) \leq u_0, \widehat{F}_1(X_{1i}) \leq u_1), \tag{1.2.3}$$

for $(u_0, u_1) \in [0, 1]^2$ and where $\widehat{F}_T$ and $\widehat{F}_1$ are again rescaled versions of the empirical c.d.f. (where the rescaling was later suggested by Omelka et al. (2009)). Numerous subsequent work may be related to this first approach, see e.g. the study of the associated empirical process in Fermanian et al. (2004), Tsukahara (2005) or Segers (2012), or the work related to some smoothed version of (1.2.3) in Fermanian et al. (2004), Chen and Huang (2007) and Omelka et al. (2009).

On the other hand, rather than focusing on the copula c.d.f., an extensive literature has concentrated on the nonparametric estimation of its density function instead, driven by the main argument that the latter is often more readable for interpretation. Among these techniques, our attention is here drawn in particular to kernel methods, mainly for their simplicity of construction and interpretation, although alternative strategies are certainly worth mentioning, see e.g. the spline smoothing methods of Shen et al. (2008) and Kauermann et al. (2013), the Bernstein polynomial methods of Sancetta and Satchell (2004), or the wavelet methods of Genest et al. (2009b).

Now, in addition to observing only a pseudo-sample, two characteristics of copula densities are central to the development of appropriate kernel estimation procedures, as bivariate copula densities are, first, only supported on the unit square and may, second, typically be unbounded in some regions of the latter. Ignoring momentarily these features in order to motivate the construction of appropriate estimators in this situation, suppose first one wishes to consider the straightforward,

yet naive, approach of estimating the density c_{TX_1} at $(u_0, u_1) \in [0,1]^2$ by

$$\frac{1}{n\,|\boldsymbol{H}|^{1/2}} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{1/2}\begin{pmatrix} u_0 - \widehat{U}_{0i} \\ u_1 - \widehat{U}_{1i} \end{pmatrix}\right),$$

where $\boldsymbol{H}$ is a symmetric positive-definite *bandwidth* matrix and $\boldsymbol{K} : \mathbb{R}^2 \to \mathbb{R}$ is a bivariate *kernel* function. This construction, although tempting, clearly completely ignores the boundedness of the support on the unit square, and hence places positive mass outside the latter support through the kernel $\boldsymbol{K}$. Additionally, this basic kernel approach is further known not to be consistent for unbounded densities. Hence, to remedy the first, and arguably most inconvenient, of these two issues, several distinct schemes have been designed in the literature to correct for well-known bias issues at the boundaries, such as the mirror reflection method of Gijbels and Mielniczuk (1990) or the boundary kernel method of Chen and Huang (2007). An additional procedure was further proposed by Charpentier et al. (2006) and Geenens et al. (2017), and exploits the key invariance property of copulas with respect to increasing transformations of their margins. Here, in order to avoid having to estimate copulas on a bounded support, the idea is then to work in three steps; first, the initial data is appropriately projected on an unbounded support, then, the bivariate density of the projected data is estimated in a second step by means of standard techniques, and, lastly, estimation of the copula density is implemented in the third step by simple back-transformation on the unit square. For instance, using the example of a probit transformation, estimation of the copula density c_{TX_1} at $(u_0, u_1) \in (0,1)^2$ is carried out by

$$\frac{\widehat{f}_{01}(\Phi^{-1}(u_0), \Phi^{-1}(u_1))}{\phi(\Phi^{-1}(u_0))\phi(\Phi^{-1}(u_1))}, \tag{1.2.4}$$

where ϕ and Φ stand for the standard normal density and c.d.f., respectively, and where $\widehat{f}_{01}$ is an appropriate bivariate density estimator of the projected pseudo-sample $(\Phi^{-1}(\widehat{U}_{0i}), \Phi^{-1}(\widehat{U}_{1i})), i = 1, \ldots, n$. In order to cope with the last above-mentioned peculiarity of copula densities, i.e. their possible unboundedness, Geenens et al. (2017) subsequently suggested to couple (1.2.4) with polynomial local-likelihood estimation for f_{01}. This results in a bivariate copula density estimator illustrated to outperform its competitors in most simulation scenarios, as further emphasized in Nagler (2014) whose work reports a more exhaustive comparison of existing methodologies for bivariate copula density estimation. Hence, when necessary in the following of this thesis, our preference will naturally go to the latter described technique.

We now finish this section on copula estimation by including a few words on higher-dimensional copula modelling, be it for parametric or nonparametric copulas. Clearly, multivariate settings (i.e. $d > 1$) offer a significant challenge to copula modelling; parametric approaches may lack even more flexibility than for bivariate modelling, and nonparametric procedures are on the other hand quickly exposed to

the curse of dimensionality. A unified and natural suggestion in the literature then seems to consider breaking down the multivariate copula into lower-dimensional, and hence more tractable and flexible, components. Under this general strategy, several methodologies in the literature then include, among others, nested Archimedean copulas (see e.g. Hofert and Pham (2013) and Joe (2014)), factor copulas (see e.g. Oh and Patton (2012)) and, arguably the most popular, vine copulas where only bivariate blocks are used in an iterative scheme (see e.g. Czado (2010), Joe (2014) and references therein). Without entering into too much further details, while these techniques inevitably induce some approximations that one hopes to be controlled for (see e.g. the work of Hobæk Haff et al. (2010) and Stöber et al. (2013) for the approximations related to vine copulas, or the work of Gijbels et al. (2017) and references therein for the related issue of testing for covariate effects in conditional copulas), their interest clearly resides in the need for only estimating parametrically or nonparametrically low-dimensional copulas, for which the above-mentioned techniques are perfectly suited. In Chapter 2, when confronted to higher-dimensional copulas, we will then embrace such a line of thought as well.

Now, to introduce how we actually come to require the estimation of copulas in a quantile regression model, we devote the next section to the intrinsically close connection between copulas and conditional quantiles.

1.2.3 Copula-based quantile regression

Given the common interpretation of copulas as dependence modelling objects, it certainly comes as little surprise that copulas may closely be linked to quantile regression modelling. As a matter of fact, copula-based quantile regression modelling is observed to be considered in the literature, in line with the vocabulary of Section 1.1, both in terms of inverse-c.d.f. and check-based approaches. We thus review and discuss in this section both strategies. A first common observation that is worth mentioning is that a general copula-based approach allows the analyst to straightforwardly avoid common regression issues such as the consideration of transformed covariates or the possible inclusion of interactions among the latter. This forms a first considerable convenience of such copula-based techniques.

Inverse-c.d.f. approach

The first approach to copula-based quantile regression comes from the somewhat expected observation that conditional quantiles may appropriately and straightforwardly be expressed in terms of a conditional copula and univariate margins. Taking the example of univariate conditional quantiles for ease of reading, an early relation in this direction reported for instance in Joe (1997) is that

$$F_{T|X_1}\left(F_T^{-1}(u_0)|F_1^{-1}(u_1)\right) = \frac{\partial C_{TX_1}(u_0,u_1)}{\partial u_1}, \quad \forall (u_0,u_1) \in [0,1]^2,$$

where $F_{T|X_1}$ is the conditional c.d.f. of T given X_1. A direct implication of the latter relation, which is in fact the cornerstone of the sequential nature of vine copula modelling, was exploited in (an earlier version of) Bouyé and Salmon (2009), and then further applied in the time-series works of Chen and Fan (2006) and Chen et al. (2009). Their starting point is thus that for $\tau \in (0,1)$,

$$F^{-1}_{T|X_1}(\tau|x_1) \equiv m_\tau(x_1) = F^{-1}_T\left(C^{-1}_{T|X_1}(\tau|F_1(x_1))\right),$$

where $C^{-1}_{T|X_1}$ is the τ-th conditional copula quantile function of $F_T(T)$ given $F_1(X_1)$, also to be seen as the inverse of $\partial C_{TX_1}(u_0,u_1)/\partial u_1$ with respect to u_0. Hence, estimation of conditional quantiles simply relies here on estimation and appropriate inversion of marginal distributions and a bivariate copula, that is,

$$\widehat{m}_\tau(x_1) = \widehat{F}^{-1}_T\left(\widehat{C}^{-1}_{T|X_1}(\tau|\widehat{F}_1(x_1))\right),$$

for some appropriate estimators $\widehat{F}_T$, $\widehat{F}_1$ and $\widehat{C}^{-1}_{T|X_1}$. While marginal c.d.f. estimation is again typically carried out nonparametrically, the subsequent suggestion of Bouyé and Salmon is to consider a parametrization of the copula. By means of several examples exposed in their work, this is then observed to allow for an explicit expression of $C^{-1}_{T|X_1}$ and results in a semiparametric model that may, depending on the chosen copula family and underlying marginal distributions, be non-linear in the parameters.

Now, in addition to the aforementioned possible lack of flexibility of parametric copula modelling, the extension of this technique to multivariate covariates is argued in the literature to be less straightforward than for univariate covariates, given the required inversion of the conditional copula $C_{T|\boldsymbol{X}}$ of T given $\boldsymbol{X}$. In some situations however, assuming a particular structure for the multivariate copula may result in significant reduction of the complexity of this task. This is for instance exploited in the recent work of Kraus and Czado (2017) and Schallhorn et al. (2017), where a subclass of vine copulas, the D-vine structure ('Drawable vine', see e.g. Aas et al. (2009)), is shown to lead, under the usual vine copula simplifying assumption, to an analytical rewriting of the conditional copula quantile function in terms of recursive partial derivatives of bivariate copulas and their inverses. As for any vine decomposition, the essential feature of this rewriting is therefore the need to rely here only on bivariate building blocks, for which numerous modelling and estimation methodologies exist, as previously-described.

Additionally, while the work of Kraus and Czado only considers parametric copula modelling, and is hence not immune to the shortcomings of the latter approach in the regression context as illustrated in Section 2.2.2, Schallhorn et al. provide an extension of the methodology by namely including nonparametric estimation of some bivariate copulas forming the bulding blocks of the vine structure. In some sense, and anticipating at this point a few arguments detailed later in the text, the latter work has thus in its essence parallels with arguments of Chapter 2, although posterior to the publication of our work. However, in addition to

tackling the regression modelling with a different approach, a major contrast with our work worth mentioning here is that the D-vine structure does not allow for a direct modelling, and consequent estimation, of the bivariate copulas of T with every included covariate. To a certain extent, this arguably exposes the inability of the D-vine structure to fully take account of the regression context where one is mostly interested in the dependence between the response and every covariate at hand. At this point, these arguments are most probably to be put in perspective with the developments and numerical results of Section 2.2.2 and Section 2.5.1 for a better understanding, along with a more detailed discussion on D-vine copula modelling left here out for sake of brevity.

Conclusively, this first approach to copula quantile regression modelling offers what one could categorize as an inverse-c.d.f. approach to quantile regression by using either low-dimensional covariates in order to exploit the parametric bivariate copula literature or, more recently, by using particular vine copula structures for higher-dimensional problems. A check-based counterpart to this will then be exposed next, with computational and theoretical convenience acting as main incentives for the latter approach.

Check-based approach

The check-based approach to copula-based regression modelling may in fact be initially related to the work of Noh et al. (2013) on mean regression models, and is again a direct implication of Sklar's theorem. The starting point here is indeed to observe that the conditional density of T given $\boldsymbol{X}$, denoted by $f_{T|\boldsymbol{X}}$, may simply be rewritten using the latter theorem (expressed in terms of densities) as

$$f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = \frac{f_{T\boldsymbol{X}}(t,\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} = f_T(t)\,\frac{c_{T\boldsymbol{X}}(F_T(t),\boldsymbol{F}(\boldsymbol{x}))}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))}, \quad \forall (t,\boldsymbol{x}) \in \mathbb{R}^{d+1}, \tag{1.2.5}$$

where $f_{T\boldsymbol{X}}$ is the joint density of $(T,\boldsymbol{X})$, $f_{\boldsymbol{X}}$ is the joint density of $\boldsymbol{X}$ and $c_{\boldsymbol{X}}$ is the corresponding copula density. Exploiting this relation in a context where one constructs upon a conditional expectation then yields the following copula- and check-based approach to quantile regression:

$$\begin{aligned} m_\tau(\boldsymbol{x}) &= \arg\min_{a\in\mathbb{R}} \mathbb{E}\Big[\rho_\tau(T-a)|\boldsymbol{X}=\boldsymbol{x}\Big] \\ &= \arg\min_{a\in\mathbb{R}} \mathbb{E}\Big[\rho_\tau(T-a)\,c_{T\boldsymbol{X}}(F_T(T),\boldsymbol{F}(\boldsymbol{x}))\Big]. \end{aligned} \tag{1.2.6}$$

This appealing formulation as a weighted quantile of the response suggests, as described in Noh et al. (2015), to estimate the quantile regression by taking the empirical analogue of (1.2.6) where all unknown quantities are replaced by appropriate estimators. That is,

$$\widehat{m}_\tau(\boldsymbol{x}) = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(T_i-a)\,\widehat{c}_{T\boldsymbol{X}}(\widehat{F}_T(T_i),\widehat{\boldsymbol{F}}(\boldsymbol{x})),$$

with $\widehat{\boldsymbol{F}}(\boldsymbol{x}) = (\widehat{F}_1(x_1), \ldots, \widehat{F}_d(x_d))^\intercal$. Quite naturally in first instance, Noh et al. suggested subsequently to leave the marginal distributions unspecified while assuming a parametric model for the copula. Their resulting semiparametric estimator then leads to easily obtained asymptotics for both i.i.d. and time-series settings, with a notable parametric rate of convergence independently of the dimension d of the covariate. Furthermore, the estimator has the additional appealing property of being automatically monotonic across quantile levels, hereby naturally avoiding typical, and worrisome in practice, quantile crossing issues.

However, similarly to the inverse-c.d.f. copula-based approach, the parametric assumption for the required copula estimation has since the publication of Noh et al. been exposed to somewhat vexing issues in the regression context, as highlighted specifically by Dette et al. (2014). In particular, their work uncovered possible shortcomings that are induced by the misspecification of the parametric copula, by taking the example of a univariate quantile regression function that is non-monotonic in the covariate. In this situation, it is unpleasantly found that most of the common parametric copula families still yield a monotone estimation of the regression function, thereby providing a rather poor fit of the latter. This example is further detailed in Section 2.5.1. Hence, while offering some advantages over its inverse-c.d.f. counterpart, the check-based approach to copula quantile regression summarized here still inherits from a fundamental shortcoming restraining its practical use. This is the starting point of Chapter 2, where the latter comment will naturally need to be addressed.

Conclusively, this section briefly reviewed the major notions and the global context related to copula-based regression prior to our work, with an accent on the issues already appearing here for complete observations. Now, as exposed earlier, an additional motivation for the following chapters is the inclusion in this context of possible censoring of the responses. Hence, the next section is devoted to the description of the main statistical domain accommodating for the issues related to censoring.

1.3 Survival analysis

Many interesting real data applications, of which a few will be mentioned next, require an appropriate data analysis while only disposing of possibly incomplete or incorrect data at hand. In this thesis, we propose quantile regression methodologies designed to handle one such particular incompleteness, known as *censoring*. The latter phenomenon, characterized by disturbances preventing the observer to acquire all information about the data, is usually integrated in the literature under the domain of survival analysis, the branch of statistics aiming at modelling time-to-event data. As the term suggests, the latter data describe the time between a defined origin and an event of interest. Examples of the study of such times, also referred to as *survival times* or *failure times*, include medical applications (e.g. the time until full recovery after a disease), actuarial applications (e.g. the time

between subscription to an insurance and the first car accident) or engineering applications (e.g. the study of the lifetime of a certain machine) to name a few. Clearly, the association of these data with possible censoring comes as a natural consequence of the very structure of time-to-event data, as the latter are by definition not measured instantly, and hence exposed to various types of interferences during the lapse of time required for completely collecting the data.

The aim of this section is here to provide a brief presentation of the peculiarities of such survival data for statistical analysis. To that end, we start by reviewing the usually employed quantities of interests for such data, and describe more formally the notion of censoring. We then discuss estimation of prevailing survival quantities, and end the section by examining different particular regression models in this context when considering covariate information. Numerous textbooks are devoted to a more detailed description of survival analysis, including Therneau and Grambsch (2000), Kalbfleisch and Prentice (2002) and Moore (2016) to name a few.

1.3.1 Characteristics of survival data

As already evoked earlier, survival data may be characterized by two fundamental features, as they usually describe a time and are thus positive, and, more fundamentally, they are typically subject to nuisances in their measures that need to be accounted for in any statistical analysis. From a methodological point of view, this mainly implies that the paradigm of concentrating primarily on the density and c.d.f. of T will here shift towards other quantities, although of course related. These statistical quantities will be described next, after a quick review of the notion of censoring.

General censoring occurs when any nuisance prevents the observer from measuring the exact survival time of the objects under study. Quite naturally, different types of censoring may be distinguished, depending on the nature of the interfering nuisance. For instance, in the event of only observing an upper bound for the survival time, we refer to the phenomenon as *right censoring*. A textbook example is here the situation of a clinical trial where some patients may not have experienced the event of interest by the end of the study, hereby providing only the partial information that their survival time is at least greater than a certain value. Conversely, observing only a lower bound is referred in the literature to as *left censoring*, for which the statistical tools may of course be related to the context of right censoring, as will be namely illustrated in the application of Chapter 4. Lastly, if the observer may only report that the event of interest took place somewhere between two recorded times, then this is denoted as *interval censoring*.

Now, our attention in this thesis is drawn in particular to *random right censoring*[3], that is, situations in which the censoring time, i.e. the lower bound for the

[3]Note that the term 'random' here will be considered implicit in the following sections and chapters when referring to 'right censoring'.

survival time, is a random variable itself. Building on the aforementioned clinical trial example, the latter characteristic allows for modelling situations where patients may enter and leave the study at any time, be it for experiencing the event of interest, or any other reason leading to censoring (e.g. lost of follow up or end of the trial). In contrast, other types of right censoring not covered here include particular mechanisms implying that all censored observations share a common censoring time. This is the situation in *Type I right censoring*, where all individuals enter the trial at the same time and the latter ends at a fixed point in time, and in *Type II right censoring*, where all individuals again enter the trial at the same time but the study is ended after a predefined number of event of interests have occurred. Lastly, taking one step back, while a clear motivation exposed here for modelling right censoring comes from the world of time-to-event data, it is worth emphasizing that censoring does not exclusively affect the latter type of data, as numerous other applications may be subject to this incompleteness of the information, such as in insurance (where the actual cost of a claim may take years before being fully known) or in astronomy (where detection limits may prevent full observation of astronomical objects).

Now, to introduce the formal context without covariates, and denoting by C the random censoring variable, right censoring implies that the observed data consists here only in (Y, Δ) instead of T, where

$$Y = \min(T, C),$$

and

$$\Delta = \mathbb{1}(T \leq C)$$

is called the *censoring indicator*, allowing for a distinction between complete and censored observations. For identifiability concerns, we further assume throughout the thesis that T and C are independent, or conditionally independent given the covariates if any. This namely implies that the following relation between the distributions of Y, T and C holds for every t:

$$1 - F_Y(t) = (1 - F_T(t))(1 - G_C(t)),$$

where F_Y is the c.d.f. of Y and G_C is the c.d.f. of C. Now, observing only (Y, Δ) implies that classical statistical methodologies have here to be adapted, as it is easily seen that treating all responses as if they were complete, or even discarding the censored responses, will both lead to inadequate statistical inference given that the largest survival times will be under-represented. The latter remark comes from the natural observation that these largest survival times are the ones more likely to be censored given their longer exposure to the latter phenomenon. The next section will be devoted to the accommodation of censored data for the estimation of classical statistical quantities.

Before turning to the latter subject, we now close this section with a few words on the statistical quantities usually employed in survival analysis and thereby sometimes reappearing in the following sections and chapters. In opposition to the

situation with only complete observations, where the main objects of interest are the density and c.d.f. of T, the literature on survival analysis mostly builds models on the survival function, the hazard function and the cumulative hazard function, all of which may of course be related to the distribution function. For the non-negative time-to-event variable T, these are defined as follows: the *survival function* $S_T(t)$ of T represents the probability that an individual, or object, is still at risk at time t for experiencing the event of interest, that is,

$$S_T(t) = 1 - F_T(t) = \mathbb{P}(T > t), \quad t \in \mathbb{R}^+.$$

The *hazard function* on the other hand is defined as

$$h(t) = \lim_{\eta \to 0} \frac{\mathbb{P}(t \leq T \leq T + \eta | T \geq t)}{\eta} = \frac{f_T(t)}{S_T(t)}, \quad t \in \mathbb{R}^+.$$

Hence, for a small $\eta > 0$, $\eta \times h(t)$ represents approximately the probability of experiencing the event of interest in the interval $[t, t+\eta]$, conditionally on the knowledge that the latter did not take place before the time t. By letting η shrink towards 0, $h(t)$ is thus generally interpreted as the instantaneous risk of experiencing the event at time t for items that are surviving till that time. With this interpretation in mind, lastly, the *cumulative hazard function* represents the accumulated instantaneous risk over time, that is,

$$H(t) = \int_0^t h(s) \, \mathrm{d}s, \quad t \in \mathbb{R}^+.$$

With these notions and characteristics laying the grounds for survival analysis, we may now undertake in the next section the first steps towards estimation of statistical quantities based on finite samples exposed to censoring.

1.3.2 Nonparametric estimation with censored data

In this section, we are interested in the estimation of several quantities based, first, on an i.i.d. sample $(Y_i, \Delta_i), i = 1, \ldots, n$, of (Y, Δ), and second, on an i.i.d. sample including covariate information as well. In particular, we are here mainly interested in the estimation of the c.d.f. and the conditional c.d.f. of the unobservable variable T, given the covariates $\boldsymbol{X}$ if present. As a preliminary remark, we are here solely interested in nonparametric estimation of the latter, as parametric methodologies such as survival likelihood will not appear in the remainder of this thesis. Quite naturally, we first discuss estimation of the unconditional c.d.f. before turning to the estimation of the conditional c.d.f.

Nonparametric c.d.f. estimation with censored data

The certainly most widely used procedure to estimate the c.d.f. F_T of the unobservable variable T goes back to Kaplan and Meier (1958), with[4]:

$$\widehat{F}_T(t) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{1}{\sum_{j=1}^{n} \mathbb{1}(Y_i \leq Y_j)} \right\}^{\mathbb{1}(Y_i \leq t,\, \Delta_i = 1)}. \tag{1.3.1}$$

The Kaplan-Meier estimator is observed to define a step function jumping only at complete observations. Furthermore, as one might expect, in case of no censoring (i.e. $Y_i = T_i, i = 1, \ldots, n$), the estimator simply boils down to the empirical c.d.f. with jumps n^{-1} at every observation. An alternative rewriting of the estimator, reported for instance in Miller (1981), allows for a further connection between situations with and without censoring, as the estimator is shown to take the following form:

$$\widehat{F}_T(t) = \frac{1}{n} \sum_{i=1}^{n} W_{in} \mathbb{1}(Y_i \leq t), \qquad W_{in} = \frac{\Delta_i}{1 - \widehat{G}_C(Y_i-)}, \tag{1.3.2}$$

where $\widehat{G}_C$ is the Kaplan-Meier estimator of the c.d.f. of C (obtained by replacing Δ_i with $1-\Delta_i$ in (1.3.1)), and $G_C(y-)$ denotes the left limit of G_C at y. Note that (1.3.2) clearly exhibits that the jumps of the estimator increase for higher values in the sample, which is in line with compensation of the previously-mentioned underrepresentation of the largest survival times due to censoring. This formulation further emphasizes that, contrary to the situation with no censoring, the estimator is as such no longer an average of i.i.d. terms, as explicitly highlighted by the subscript n in the weights W_{in}. This, of course, substantially complicates the study of its asymptotic properties, for which much work has been provided in the literature, see e.g. Gill (1983) and Lo and Singh (1986) to name a few.

Nonparametric conditional c.d.f. estimation with censored data

When studying the inclusion of covariate information, for identifiability concerns we first recall that we consider throughout the thesis that the censoring and survival times are independent conditionally on the covariates, leading namely for every t to

$$1 - F_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = (1 - F_{T|\boldsymbol{X}}(t|\boldsymbol{x}))(1 - G_C(t|\boldsymbol{x})),$$

where $F_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = \mathbb{P}(Y \leq t|\boldsymbol{X} = \boldsymbol{x})$ and $G_C(t|\boldsymbol{x}) = \mathbb{P}(C \leq t|\boldsymbol{X} = \boldsymbol{x})$. In the objective here of estimating now the conditional c.d.f. $F_{T|\boldsymbol{X}}$, a classical approach for both complete and censored data initiates from the idea that, provided $F_{T|\boldsymbol{X}}(t|\boldsymbol{x})$

[4]In order not to complicate the notations, and since this will not induce any confusion in the following chapters, we do not specifically add here a sub- or superscript to stress that $\widehat{F}_T$ is here indeed the Kaplan-Meier estimator of F_T.

is continuous in $\boldsymbol{x}$, the information contained in observations whose $\boldsymbol{X}_i$ is, in some sense, close to $\boldsymbol{x}$ should be prioritized over observations laying further away from $\boldsymbol{x}$. For complete observations, this line of thought leads to the estimator of Stone (1977), representing a simple extension of the empirical c.d.f. by replacing the uniform weights n^{-1} with weights depending on the relative distance of observations $\boldsymbol{X}_i, i = 1, \ldots, n$, with respect to the point of interest $\boldsymbol{x}$. For censored observations, a similar extension of the Kaplan-Meier estimator was carried out by Beran (1981), leading to the following:

$$\widehat{F}_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{B_{ni}(\boldsymbol{x})}{\sum_{j=1}^{n} \mathbb{1}(Y_i \leq Y_j) B_{nj}(\boldsymbol{x})} \right\}^{\mathbb{1}(Y_i \leq t,\, \Delta_i = 1)}, \qquad (1.3.3)$$

for a sequence of weights $B_{nj}(\boldsymbol{x})$ adding up to 1. A popular choice for the latter weights that we will embrace in this thesis is the particular case of Nadaraya-Watson type of weights given by

$$B_{nj}(\boldsymbol{x}) = \frac{K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h}\right)}{\sum_{k=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_k}{h}\right)},\ j = 1, \ldots, n,$$

where K_d is a multivariate kernel density function, $K_d(\boldsymbol{u}/h) = K_d(u_1/h, \ldots, u_d/h)$, and h is a positive bandwidth parameter converging to zero as $n \to \infty$. Equivalently to the situation with no covariates, Beran's estimator, sometimes called the *conditional Kaplan-Meier estimator*, may also be written in a similar form as (1.3.2) with weights $W_{ni}, i = 1, \ldots, n$, now depending on the relative distance of $\boldsymbol{X}_i, i = 1, \ldots, n$, with respect to the point of interest $\boldsymbol{x}$. Again, in case of no censoring, (1.3.3) is observed to reduce to Stone's estimator. A more detailed study of the properties of Beran's estimator may be found in various settings in Dabrowska (1989), Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Veraverbeke (1997a) among others.

Lastly, to close this section and given the context of this thesis, it is worth mentioning here a few elementary words on the estimation of quantile regression with censored data based on (1.3.3). Given such an estimator of $F_{T|\boldsymbol{X}}$, one may indeed readily suggest to adopt an inverse-c.d.f. approach as in (1.1.9) in order to estimate the conditional quantile function. This was for instance studied in Van Keilegom and Veraverbeke (1996) and Van Keilegom and Veraverbeke (1997b), leading to a fully nonparametric quantile regression estimator. In that sense, we note that this section on Beran's estimator slightly encompasses the following section on regression modelling with censored data. Finally, note that in Chapter 4, when considering a similar inverse-c.d.f. approach but in a somewhat different context, instead of using Beran's estimator as such we will rely on a "double-kernel" version of the latter. The motivations underlying this particular choice will be exposed later when necessary. For now, we may concentrate in the next section on different seminal regression techniques that are abundantly used in practice for an analysis

with censored data, and compare the latter with our studied quantile regression framework.

1.3.3 Regression modelling with censored data

We propose in this section a succinct review of the commonly employed regression models in survival analysis when confronted to censored data. As a preliminary remark, similarly to the simpler situation without censoring, the latter models quite naturally include fully parametric, semiparametric and nonparametric approaches. We propose here to concentrate on two specific families of models that are preeminent for the analysis of covariate effects on survival data, namely the *Cox proportional hazards* models (Cox (1972)) and the *accelerated failure time* models (AFT, mentioned in Cox (1972) and further considered in Prentice (1978) for instance). For ease of comparison with quantile regression modelling in the present context, we further choose to adopt here a unified approach for the presentation of the latter models, and assume a linear effect of the covariates in the latter. Then, the starting point highlighted by Doksum and Gasko (1990) is to observe that most survival models may be written in a unified form as

$$g(T_i) = \beta^\mathsf{T} \boldsymbol{X}_i + \epsilon_i,\ i = 1, \ldots, n, \tag{1.3.4}$$

where g is some appropriate monotone transformation depending on the particular model of interest, $\boldsymbol{X}$ again includes a component relative to the intercept and is thus of dimension $d+1$, β is the $(d+1)$-dimensional unknown regression coefficient vector and $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. random variables with some c.d.f. F_ϵ. We now detail how (1.3.4) encompasses both the Cox and the AFT models, and then conclude the section by a concise comparison with quantile regression models. Further details on the latter comparison may be found in Koenker and Bilias (2001), Koenker and Geling (2001) and Portnoy (2003).

Cox proportional hazards model

The Cox regression, along with its numerous variations, is historically undoubtedly the predominant technique for analyzing the relationship between covariates and a possibly censored response variable. As the denomination suggests, the latter approach initially aims at modelling the influence of covariates on the conditional hazard function rather than on the survival times directly. The formulation, with an unknown coefficient vector denoted here by γ, then goes as follows:

$$h(t|\boldsymbol{X}_i) = h_0(t) \exp(\gamma^\mathsf{T} \boldsymbol{X}_i),\ i = 1, \ldots, n,$$

where the coefficient in γ relative to the intercept is set to 0 in order to avoid identifiability issues, and where $h_0(t)$ is the so-called baseline hazard function, common to all observations and usually left unspecified. The term 'proportional hazards' comes from the simple consideration that the conditional hazard ratio of

two observations is constant over values of t, hereby introducing a strong global assumption through the model specification. Now, the link with formulation (1.3.4) is obtained by simple rearrangements of the previously-described survival analysis quantities: starting with $\log h(t|\boldsymbol{X}_i) = \log h_0(t) + \gamma^\mathsf{T}\boldsymbol{X}_i$, $i = 1, \ldots, n$, and recalling that $h(t) = -\mathrm{d}S_T(t)/S_T(t)$ and similarly for a conditional version, we have

$$\log\big(-\log S_{T|\boldsymbol{X}}(t|\boldsymbol{X}_i)\big) = \log H_0(t) - \gamma^\mathsf{T}\boldsymbol{X}_i,\ i = 1, \ldots, n,$$

where $S_{T|\boldsymbol{X}}$ is the conditional survival function of T given $\boldsymbol{X}$ and $H_0(t)$ is the cumulative baseline hazard function. From the latter expression, we may reformulate the Cox model as

$$\log H_0(T_i) = \beta^\mathsf{T}\boldsymbol{X}_i + \epsilon_i,\ i = 1, \ldots, n, \tag{1.3.5}$$

where $\beta = -\gamma$ and $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. random variables with (extreme value) distribution $F_\epsilon(s) = 1 - \exp(-\exp(s))$. Hence, the latter formulation of the Cox regression clearly exhibits that the proportional hazards model implicitly designs a linear regression with a transformation of the form $g(T_i) = \log H_0(T_i)$, and where the covariates only affect the latter through a simple location-shift effect. An additional discussion on the subject is deferred to the section devoted to the comparison with quantile regression models.

Accelerated failure time model

The AFT model is usually introduced in the literature as a natural extension to the Cox regression allowing the analyst to relax the stringent proportional hazards assumption. This stems from the observation that the model may be introduced as follows:

$$h(t|\boldsymbol{X}_i) = h_0\{\exp(\gamma^\mathsf{T}\boldsymbol{X}_i)t\}\exp(\gamma^\mathsf{T}\boldsymbol{X}_i),\ i = 1, \ldots, n,$$

hereby clearly preventing the ratio of conditional hazards of two observations to be automatically constant over values of t. The term 'accelerated failure time' may here be seen as originating from the multiplicative impact of the covariates on t in the baseline hazard, or baseline survival depending on the chosen formulation, as this is usually interpreted in the sense that the covariates will change the time scale of observations with respect to a baseline time (i.e. when $\boldsymbol{X} = 0$). Now, although the latter formulation clearly exhibits the alleviation of the proportional hazards assumption, the AFT model is usually presented in a form similar to (1.3.4) as this allows for a direct and valuable interpretation of the coefficients on the (log) survival time:

$$\log T_i = \beta^\mathsf{T}\boldsymbol{X}_i + \epsilon_i,\ i = 1, \ldots, n, \tag{1.3.6}$$

where $\beta = -\gamma$ and where the distribution F_ϵ has here the convenience of being left unspecified. A similar observation as for the Cox model however holds at this point, as covariates are again observed to solely exert a location shift impact of the

(log) survival times, hereby again drastically restraining the permitted influence on the conditional distribution. This will then be put into contrast with quantile regression models in the next section.

On the use of quantile regressions for censored data

The starting point of our comparison between the above-described procedures and (linear) quantile regression models for the analysis of censored survival data emanates from the acknowledgment that the former both rely on an i.i.d. error assumption in (1.3.4), hereby considering only location shift effects of the covariates. In contrast, linear quantile regression models are usually formulated for their part as

$$m_\tau(\boldsymbol{X}_i) = \beta_\tau^\mathsf{T} \boldsymbol{X}_i,\ i = 1, \ldots, n, \tag{1.3.7}$$

where all coefficients in β_τ are allowed to vary with the value of τ. This is to be put into perspective with formulation (1.3.4), where one would have had the following if adopting a similar assumption for quantile regressions:

$$m_\tau(\boldsymbol{X}_i) = \beta^\mathsf{T} \boldsymbol{X}_i + F_\epsilon^{-1}(\tau),\ i = 1, \ldots, n.$$

In this situation, the regression coefficient depends on τ exclusively through the intercept term, hereby constituting in fact a special case of (1.3.7). Similarly to Section 1.1.3, this thus again reports that quantile regression models may accommodate for a larger variety of forms of heterogeneity in the conditional distribution of the response variable than the historically dominating methodologies described above. Additionally, requiring assumptions on the conditional distribution in (1.3.7) that are only local with respect to τ prevents the analyst from worrying that global assumptions may undesirably, and unnecessarily, prejudice the results in the region of interest of his study. This is for instance nicely illustrated in the survival context in the application of Koenker and Geling (2001) on the extreme longevity of some medflies, where the focus lies thus solely on the upper tail of the conditional survival distribution.

Now, to dig further into the comparison between quantile regressions and the particular Cox model, we may determine explicitly how conditional quantile functions are in fact implied by the formulation of the latter. In this objective, simple rearrangement of (1.3.5) leads to

$$m_\tau(\boldsymbol{X}_i) = H_0^{-1}(-\log(1-\tau)\, \exp(-\gamma^\mathsf{T} \boldsymbol{X}_i)),\ i = 1, \ldots, n.$$

As already noted above, the very structure of the Cox regression is thus observed to greatly restrict the effect of covariates on quantiles, as these are all necessarily monotone in $-\log(1-\tau)$. Specifically, the proportional hazards assumption also implies that the conditional quantile functions, or equivalently the conditional survival functions, of two distinct observations may never cross. That is, for

two different quantile levels τ_1 and τ_2, $m_{\tau_1}(\boldsymbol{X}_i) \leq m_{\tau_1}(\boldsymbol{X}_j)$ automatically implies that $m_{\tau_2}(\boldsymbol{X}_i) \leq m_{\tau_2}(\boldsymbol{X}_j)$ for $i, j \in \{1, \ldots, n\}, i \neq j$. Note that the terminology 'crossing of conditional quantiles' employed here should not be confused with the usual notion of 'quantile crossing' where one refers to the possible issue that $\widehat{m}_{\tau_1}(\boldsymbol{X}_i) \leq \widehat{m}_{\tau_2}(\boldsymbol{X}_i)$ for $\tau_1 > \tau_2$ and for some $\boldsymbol{X}_i, i = 1, \ldots, n$, whereas we refer here to the possibility that the ranking of conditional quantiles between several observations may change with the quantile levels. This inability to account for the latter phenomenon is of course a highly undesirable implication for conditional quantiles, and is often violated in practice as one may for instance simply think of the example of the mortality rates between men and women that are known to reverse at advanced ages in comparison with earlier ages.

Conclusively, this section briefly summarized some of the main features of quantile regressions that are usually put forward when considering a regression analysis with censored data. In particular, quantile regressions have again been argued to provide more flexibility than their main competing procedures, as illustrated here for the easily interpretable and particularly much exploited linear paradigm. The last remaining step before entering the main body of this thesis is then to review the main estimation procedures prior to our work for quantile regression models with censored data.

1.4 Censored quantile regression modelling

We provide in this section a summary of the main techniques proposed in the literature in order to account for possible censoring in a quantile regression model. An appealing characteristic of censored quantile regression is that the latter ideas that were developed prior to our work are in fact easily illustrated using no covariates at all. Hence, after identifying the specific impact of censored data in the present context, and although the title of this section includes the term 'regression', we may allow ourselves to consider for a major part of the section the simpler situation devoted to the estimation of unconditional quantiles m_τ of a possibly censored variable T. As a preliminary remark, we concentrate here in first instance exclusively on check-based approaches to censored quantile regression, given that inverse-c.d.f. techniques in the literature prior to our work are simple inversions of the (conditional) Kaplan-Meier estimator as described in Section 1.3.2. Specifically, we detail here the two main check-based approaches that are widely exploited in the literature, and introduce them without covariates in order to disregard any model assumption besides the quantile context. For completeness, we then include a few words on an alternative estimation strategy that will however not be exploited in this thesis for reasons exposed when necessary. We then conclude the section by considering at last the inclusion of covariates, for which some insights are of crucial help to describe the outline of our work in the next section.

Now, our starting point is to first consider the trivial impact of censored data on the estimation of m_τ. To that end, we consider a toy example that will be helpful

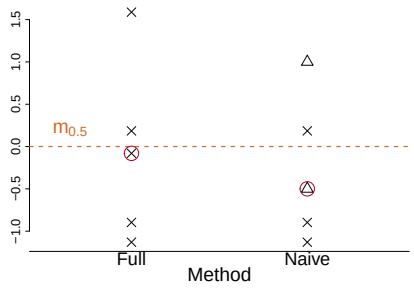

Figure 1.4: Toy example to illustrate the impact of censored data in the estimation of the median. Uncensored data points are given by ×, while censored observations are represented by △. The estimators of the median are circled in red.

to illustrate some techniques described below, and study the problem of estimating the median of a (normal) distribution for which five observations are available, among which two are censored. The resulting data is reported in Figure 1.4, where we consider estimation of the median by simple sorting of the observations, both for the complete and censored samples. For the latter case, the term 'naive' in the figure stands here for considering all observations as if they were complete. Note that the other naive strategy of simply disregarding the censored observations from the sample would have resulted in the same conclusion we bring here into spotlight.

Clearly, if needed be illustrated at this point, the impact of censored data reported in Figure 1.4 is to stimulate underestimation of the quantile of interest. The chosen example also highlights an additional, and more interesting, insight that is more specific to the estimation of quantiles with censored data. As one may observe, the example indeed illustrates that not all censored observations have the same impact here on the underestimation of the median. In fact, by the very sorting nature of quantiles, it is noted that the largest censored observation has no underestimating impact on the estimation of the median at all, in the sense that the latter would still have been appropriately estimated if all other observations had been fully observed. This thus provides a major distinction with the impact of censored data on the estimation of, for instance, the expectation rather than quantiles, and suggests that all censored data are in this context not be treated similarly if possible. The latter comment is in fact to be seen as the starting point of the *redistribution-of-mass* technique described below. General techniques that are specifically designed for the handling of censored data in the quantile context will further conveniently be qualified as 'quantile-specific'.

Now, as will be observed in the next section, not all techniques designed in the literature are built upon distinctive treatments of censored observations in the sample, and hence do not automatically fall into the quantile-specific denomination. This may arguably partly be attributed to the influence of mean regression mo-

delling, but also comes as a natural consequence of the difficulties of building, and subsequently estimating, appropriate censoring techniques that are not restricted to particular regression models. This remark will be emphasized further in the text but motivates at this point the study of the following strategy that, although not being quantile-specific in its primary formulation, will be denoted as 'model-free' in opposition to the redistribution-of-mass technique restricted by nature to a certain particular class of regression models.

1.4.1 Inverse-censoring-probability without covariates

The class of techniques described in this section is built on a check-based formulation of quantiles, and is undoubtedly stemming from the work of Koul et al. (1981), and in fact more precisely Zhou (1992), on mean regression modelling with censored data. The starting point is to observe after some simple calculations based on the tower property of conditional expectations that, for any measurable function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\left[\varphi(Y)\frac{\Delta}{1-G_C(Y-)}\right] &= \mathbb{E}\left[\varphi(T)\frac{\mathbb{E}(\Delta|T)}{1-G_C(T-)}\right] \\ &= \mathbb{E}[\varphi(T)], \end{aligned} \tag{1.4.1}$$

provided T and C are independent, and where the denomination *inverse-censoring-probability* (ICP) of the technique appears as self-explanatory. Hence, interpreting φ now as a loss function, the latter result illustrates that the expected loss of the unobservable response can easily be recovered from the actually observed response, by only keeping the loss of uncensored observations in the procedure and correctly adjusting through the censoring distribution. In the context of quantile estimation, given the check-based formulation (1.1.2), this then suggests to propose

$$m_\tau = \arg\min_{a\in\mathbb{R}} \mathbb{E}\left[\rho_\tau(Y-a)\frac{\Delta}{1-G_C(Y-)}\right], \tag{1.4.2}$$

from which, based on an i.i.d. sample (Y_i, Δ_i), $i = 1, \ldots, n$, and an appropriate estimator $\widehat{G}_C$ of G_C (usually the Kaplan-Meier estimator (1.3.1)), we have

$$\widehat{m}_\tau^{ICP} = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(Y_i-a)\frac{\Delta_i}{1-\widehat{G}_C(Y_i-)}. \tag{1.4.3}$$

Note that the minimization in the latter formulation involves a convex function alone, and is hence easily carried out in practice. Without entering into more details here, further note that this technique may of course be closely linked to (1.3.2) and the so-called *Kaplan-Meier integrals*.

Now, as previously evoked, this strategy to handle censored data is characterized here as being model-free in the sense that, when considering covariate information,

application of the latter technique is observed to be straightforward for numerous regression models. The only adjustment will indeed be to include a conditional censoring c.d.f. in (1.4.1) rather than the unconditional one. Examples of papers profiting from this almost universal adaptability of the technique will be given later. For now, similarly to the mean regression context, it is worth noting that the suggested weighting scheme however suffers from two notable drawbacks. First, an efficiency loss in the estimation is usually put forward in the literature given that roughly only uncensored observations are preserved in the procedure due to the presence of the censoring indicator in the numerator, and second, (1.4.3) highlights the need to evaluate an estimator of G_C in the right tail as well, for which classical estimators are known to be typically unstable due to sparsity of the data.

Two valuable and distinctive attempts to bypass these issues have been proposed in Zhou (2006) and El Ghouch and Van Keilegom (2009). For completeness, and given the ideas may be rallied under what has already been stated, we briefly review the latter even though they will not reappear in the following of this thesis given the complexity of implementation or the inability to easily extend to conditional quantiles. A common observation for both ideas is that their proposal consists in an attempt to take the quantile context into account, driving these procedures in some sense to be more quantile-specific in their handling of censored data than formulation (1.4.2). Starting with the suggestion of Zhou (2006), the latter attempts at avoiding computation of the estimator $\widehat{G}_C$ in the right-tail by introducing an artificial censoring that emanates from the previously-mentioned observation that alteration of data beyond the quantile of interest should have no impact on the estimation of the latter. Specifically, Zhou suggests to introduce an appropriate constant $M > m_\tau$, and estimate the quantile by

$$\widehat{m}_\tau^Z = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(Y_i^M - a) \frac{\Delta_i^M}{1 - \widehat{G}_C(Y_i^M -)},$$

where $Y_i^M = \min(Y_i, M)$, $\Delta_i^M = \Delta_i + (1-\Delta_i)\mathbb{1}(M \leq Y_i)$, $i = 1, \ldots, n$, and $\widehat{G}_C$ is of course still estimated with the original data. While providing an interesting attempt to take the robustness of quantiles into account, the procedure relies on an appropriate, and data dependent, choice of M that will be particularly delicate when including covariates. Furthermore, it does not succeed in avoiding the efficiency loss of handling the censoring indicator in the numerator of the procedure.

The second attempt, first developed in El Ghouch and Van Keilegom (2009), emanates here from the investigation that the check function in (1.4.2) may in fact be split up in order to introduce the ICP weights inside of it. Writing $W \equiv \Delta/(1 - G_C(Y-))$, the development is as follows:

$$\begin{aligned}\mathbb{E}\left[\rho_\tau(Y-a)W\right] &= \mathbb{E}\left[(Y-a)(\tau - \mathbb{1}(Y \leq a))W\right] \\ &= \mathbb{E}\left[\tau(Y-a)(W-1)\right] + \mathbb{E}\left[(Y-a)(\tau - W\mathbb{1}(Y \leq a))\right].\end{aligned}$$

Using similar developments as in (1.4.1), the first term is then observed to be equal to $\mathbb{E}[\tau(T-Y)]$, and hence does not depend on a. In the context of estimating

quantiles, we may then write

$$\widehat{m}_{\tau}^{EGVK} = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{n} (Y_i - a) \left[\tau - \mathbb{1}(Y_i \leq a) \frac{\Delta_i}{1 - \widehat{G}_C(Y_i-)} \right]. \tag{1.4.4}$$

Here, inserting the ICP weight into the check function allows, by the presence of $\mathbb{1}(Y_i \leq a)$, for treating both drawbacks at once: first, it prevents from evaluating $\widehat{G}_C$ in the right tail (provided the quantile level τ is not too high) and, second, it does not automatically force the loss of all censored observations to be set to 0. This is somewhat in line with our above-described intuition that the only censored observations that should be affected by a weighting scheme are those lying below the quantile of interest. Inconveniences of this approach however include the need to adapt numerical tools for the practical minimization of (1.4.4), and the failure to be systemically equivalent to a check-based formulation when considering covariates, in the sense that not all regression models allow for such a customization of the check function. Our copula-based approach of Chapter 2 is, as a matter of fact, an illustration of the latter remark.

Overall, returning to the basic ICP technique, the latter provides satisfactorily a general framework for estimating quantiles with censored data, but is suggested to be somewhat victim of this extensive flexibility. The literature indeed often illustrates that, when an alternative and more quantile-specific scheme to handle censoring is easily implementable depending on the chosen regression model, the ICP-based estimators generally perform worse than their competitors, especially when the proportion of censored data is important. In our previously-introduced vocabulary, this is then to be brought into contrast with the quantile-specific, but not model-free, technique of redistribution-of-mass described next.

1.4.2 Redistribution-of-mass without covariates

The technique described here historically emanates from the redistribution-of-mass idea (RM) of Efron (1967), although first employed for quantiles by Portnoy (2003). Starting again from a check-based formulation, the latter procedure may roughly be rallied under the same idea of proposing an appropriate weighting scheme, although the latter is here defined specifically for quantiles. As suggested above, the weighting scheme indeed stems from the observation that data lying above the quantile will in fact have the same impact on the estimation of the latter, regardless of their censoring status. Hence, an appropriate weighting scheme should here avoid altering these censored data. In fact, the only censored data that should be adjusted for in order to prevent underestimation are the ones lying below m_τ, as there is here some ambiguity about where their true, and unobservable value, lies with respect to m_τ.

Before detailing a bit further, we may graphically illustrate the technique by means of our toy example in Figure 1.5. As one can observe, while the censored data lying above the median is left unchanged, the mass of the censored data lying

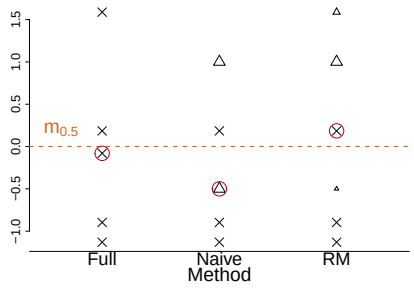

Figure 1.5: Toy example continued with inclusion of the redistribution-of-mass technique.

below the latter has here partly been redistributed to an arbitrary point above the median. Estimating the latter by sorting the (weighted) observations then successfully avoids underestimation in this trivial example.

For ease of presentation, we describe now the procedure more formally but again without any covariates, even though its construction will rapidly be seen to be somewhat clumsy in this context given it includes the c.d.f. of T. Nevertheless, the starting point is that it is obvious that the mass of censored observations such that $F_T(C_i) > \tau$ should not be redistributed. However, for ambiguous observations, that is $\Delta_i = 0$ and $F_T(C_i) < \tau$, the idea is to leave a mass at C_i corresponding to the probability that the unobservable T_i is, indeed, lower than m_τ. In other words, the mass $\mathbb{P}(T_i < m_\tau | T_i > C_i)$ should be left at C_i, and the rest should be redistributed to the right of m_τ. Simplifying a bit, we have:

$$\mathbb{P}(T_i < m_\tau | T_i > C_i) = \frac{\mathbb{P}(C_i < T_i < m_\tau)}{\mathbb{P}(T_i > C_i)} = \frac{\tau - F_T(C_i)}{1 - F_T(C_i)}.$$

This then leads to the following estimator:

$$\widehat{m}_\tau^{RM} = \arg\min_{a \in \mathbb{R}} \sum_{i=1}^{n} \left[w_i(F_T)\rho_\tau(Y_i - a) + (1 - w_i(F_T))\rho_\tau(Y^{+\infty} - a) \right], \quad (1.4.5)$$

where $Y^{+\infty}$ is an arbitrary value sufficiently large to certainly exceed m_τ, and

$$w_i(F_T) = \begin{cases} 1 & \Delta_i = 1 \text{ or } F_T(C_i) > \tau \\ \frac{\tau - F_T(C_i)}{1 - F_T(C_i)} & \Delta_i = 0 \text{ and } F_T(C_i) < \tau. \end{cases}$$

Obviously, in practice F_T is not known and needs to be estimated, leading to estimated weights $w_i(\widehat{F}_T)$, $i = 1, \ldots, n$, that are then inserted in (1.4.5). Again, there is here arguably little motivation for such a procedure without covariates, as requiring to estimate F_T may lead more straightforwardly to an inverse-c.d.f. estimator of m_τ, but the chosen presentation is here sufficient to highlight the main

idea of the technique. Additionally, this also highlights a drawback of the RM weights, as it is observed that including covariates will now require an appropriate estimation of $F_{T|\boldsymbol{X}}$. In particular models, this may somewhat be circumvented at the cost of stringent global model restrictions (e.g. supposing that a linear model holds for every quantile level below the level of interest τ as in Portnoy (2003)), but the general situation will require a flexible nonparametric estimation of the latter conditional distribution. As a consequence, by nature, this strategy only seems appropriate for parametric models as any further extensions to semiparametric or nonparametric regressions would seem questionable given one would have to adopt a preliminary local estimation of the conditional distribution for, in fact, a local estimator of the conditional quantile. This is what prevents us from qualifying the procedure as being completely model-free in opposition to the ICP weighting scheme.

This concludes our summary on the RM technique. We now briefly mention a last alternative modelling strategy proposed in the literature before devoting a few words on different specific regression models reappearing in the remainder of this thesis.

1.4.3 Aiming for higher quantile levels

Although Chapters 2 and 3 are built entirely either on the ICP technique (Chapter 2) or on the shortcomings of both the ICP and the RM techniques (Chapter 3), we briefly mention here an alternative strategy that has, admittedly, engendered a much more modest literature for censored quantile regressions. Without entering into too much detail, the idea may be resumed as an attempt to exploit all observations at hand as if they were complete, and consequently avoid underestimation by appropriately changing the target quantile level in the check-based formulation, hence the title of the section. This idea was developed by Lindgren (1997) and is illustrated by means of our toy example in Figure 1.6.

As can be observed, the underestimation induced by a naive approach is circumvented by aiming for a higher quantile level $\widetilde{\tau}$ than the initial level of interest τ. In opposition to the ICP, this namely signifies that the loss of all censored observations enters the procedure as well. However, while the idea is quite natural, practical implementation is delicate as the method is, as one could expect, highly sensitive to the determination of the appropriate level $\widetilde{\tau}$ that should be targeted and for which an iterative procedure is required. As a result, the technique engendered little to no extensions stemming from the original paper of Lindgren.

We may now focus in more details in the next section on the inclusion of covariates in the previously-described techniques. Analyzing the capabilities of the latter procedures in terms of regression modelling will then eventually help motivate the chosen approaches for the research proposed in this thesis.

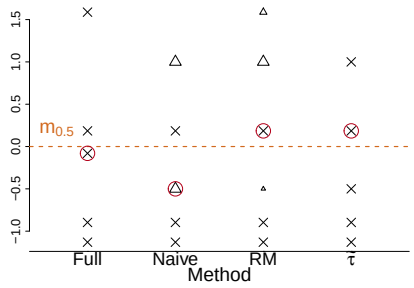

Figure 1.6: Toy example continued with inclusion of the technique of Lindgren (1997), for which all observations may be represented as if they were completely observed, hence the removal of the △.

1.4.4 Including covariate information

We describe in this section an overview of the censored quantile regression models proposed in the literature prior to our work. As one might expect, similarly to the situation for complete observations, the latter literature includes fully parametric, semiparametric and nonparametric approaches, as will be detailed next. To propose an overview that will serve the motivation of our work, our starting point is the diagram exposed in Figure 1.7 and representing a very simplified vision of the literature. Of course, the latter is not intended to encompass every possible suggestion in the literature, but rather focuses on illustrating where our research will fit in this framework. As a consequence, valuable alternatives such as the previously-mentioned procedure of Lindgren, the martingale approach of Peng and Huang (2008), the Laplace formulation of Bottai and Zhang (2010) and Bayesian suggestions are here all disregarded as they will not appear in the following chapters.

Now, as may be observed from Figure 1.7, a first major distinction that we intend to underline once again originates from the duality between check-based and inverse-c.d.f. formulations. Additionally, given that inverse-c.d.f. approaches are in the present literature restricted to nonparametric models, we concentrate here primarily on the much broader research rallied under the check formulation of quantiles. In particular, the two prominent weighting schemes described earlier engendered a fruitful literature on parametric models that will shortly be examined. We will then turn to illustrating the flexibility of the ICP technique by exposing a few supplementary semiparametric and nonparametric models. After having set the scene for the literature, we will then at last be equipped to describe the main contributions of the thesis.

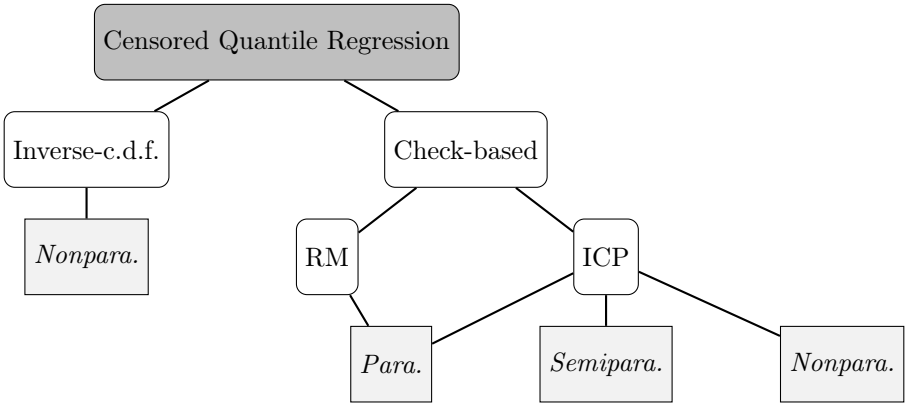

Figure 1.7: Simplistic representation of the censored quantile regression literature prior to our work. Building on the previous section, ICP and RM stand here for regression estimators based on (some form of) inverse-censoring-probability or redistribution-of-mass weighting scheme.

Parametric regression

For parametric regression, and in particular the much exploited linear model, both the ICP and RM techniques are at the cornerstone of numerous papers accounting for various settings encountered in practice. On one hand, ICP-based models, pioneered in this context by Ying et al. (1995) and Bang and Tsiatis (2002), all share in some sense the common observation that (1.4.3) may simply be extended to linear models under the usual conditional independence assumption of T and C given $\boldsymbol{X}$ by

$$\widehat{\beta}_{\tau}^{ICP} = \arg\min_{\beta\in\mathbb{R}^{d+1}} \sum_{i=1}^{n} \rho_{\tau}(Y_i - \beta^{\mathsf{T}}\boldsymbol{X}_i)\frac{\Delta_i}{1-\widehat{G}_C(Y_i - |\boldsymbol{X}_i)}, \tag{1.4.6}$$

where $\widehat{G}_C(\cdot|\boldsymbol{x})$ is an appropriate estimator of $G_C(\cdot|\boldsymbol{x})$. Several papers then further propose to conveniently assume independence between C and $\boldsymbol{X}$ in order to estimate the latter distribution by means of a simple Kaplan-Meier estimator. This is namely adopted in the papers of Ying et al. and Bang and Tsiatis, and in several extensions of the latter as well, see e.g. the previously-described paper of Zhou (2006), the variable selection framework of Shows et al. (2010), the consideration of covariate measurement errors in Ma and Yin (2011) or the implication of time-varying covariates in Gorfine et al. (2017). Alternatively, given that such an assumption of independence is in practice easily prone to violation, several papers suggest for their part either to assume a particular model for the conditional distribution of C given $\boldsymbol{X}$, for instance a Cox model, or simply to estimate the latter nonparametrically using Beran's estimator, see e.g. Leng and Tong (2013).

Overall, even in the simplest linear framework, the ICP technique is thus still at the center of very recent research.

On the other hand, as commented above, an endorsement of the RM weights allows here for a more quantile-specific handling of censored data. For linear regression, simple extension of (1.4.5) then leads to

$$\widehat{\beta}_\tau^{RM} = \arg\min_{\beta\in\mathbb{R}^{d+1}} \sum_{i=1}^{n} \Big[w_i(\widehat{F}_{T|\boldsymbol{X}})\rho_\tau(Y_i - \beta^\mathsf{T}\boldsymbol{X}_i) \\ + (1 - w_i(\widehat{F}_{T|\boldsymbol{X}}))\rho_\tau(Y^{+\infty} - \beta^\mathsf{T}\boldsymbol{X}_i)\Big],$$

where $Y^{+\infty}$ is again an arbitrary value sufficiently large to certainly exceed all $\beta_\tau^\mathsf{T}\boldsymbol{X}_i$, $i = 1, \ldots, n$, and

$$w_i(\widehat{F}_{T|\boldsymbol{X}}) = \begin{cases} 1 & \Delta_i = 1 \text{ or } \widehat{F}_{T|\boldsymbol{X}}(C_i) > \tau \\ \frac{\tau - \widehat{F}_{T|\boldsymbol{X}}(C_i)}{1 - \widehat{F}_{T|\boldsymbol{X}}(C_i)} & \Delta_i = 0 \text{ and } \widehat{F}_{T|\boldsymbol{X}}(C_i) < \tau, \end{cases}$$

for $\widehat{F}_{T|\boldsymbol{X}}$ estimating appropriately $F_{T|\boldsymbol{X}}$. For the latter estimation, the initial suggestion of Portnoy (2003) is to define recursively the involved weights based on the assumption that the regression function is linear for all quantiles levels in $(0, \tau)$, where τ is the quantile level of interest. Relaxing this assumption by only assuming linearity at the level of interest, the papers of Wang and Wang (2009) and Wey et al. (2014) for their part suggest respectively to estimate $F_{T|\boldsymbol{X}}$ using Beran's estimator or survival trees. Overall, the RM strategy has repeatedly been observed to provide better finite sample results than its ICP counterpart for most simulation settings in linear regression. Several extensions building on this success then include variable selection (Wang et al. (2013)), multiple quantile estimation (Tang and Wang (2015)) and cure rate quantile regression (Wu and Yin (2016)). However, the very structure of the RM weights considerably restrains its domain of application outside parametric regression, as exemplified in the next section where the mentioned semiparametric and nonparametric modellings are all purely built on ICP weights when it comes to check-based formulations.

Semiparametric and nonparametric regression

We briefly mention in this section a few suggestions of the literature initiating from the objective of relaxing a pure parametric assumption that may lack in practice the flexibility needed for an adequate modelling of the conditional quantile function. As evoked earlier, for purely nonparametric regression with censored data, both the inverse-c.d.f. and the check-based formulations have been studied in numerous papers, see e.g. Gannoun et al. (2005) and El Ghouch and Van Keilegom (2009) for the latter, and the previously enumerated references for the former. An extension of the inverse-c.d.f. worth mentioning here was further studied somewhat on the

surface in Leconte et al. (2002), where it is suggested to apply the approach on a "double-kernel" version of Beran's estimator, that is, introducing a second kernel in the direction of the response. The latter idea may be attributed to the knowledge, coming from papers of Azzalini (1981), Reiss (1981) and Hansen (2004) to name a few, that smoothing in the direction of the response may result in both finite sample and asymptotic efficiency gains. More details on the subject will be given when necessary in Chapter 4.

Now, while nonparametric methods offer a convenient modelling flexibility, when the dimension of the covariates grows, data analysis must turn to general semiparametric techniques to propose flexible models that are unexposed to the curse of dimensionality. Somewhat surprisingly, the literature on semiparametric regression models in the present context is in fact rather sparse. However, one valuable suggestion mimicking the well-studied single-index model for mean regression was studied in Bücher et al. (2014), where the linear paradigm is relaxed by assuming that

$$m_\tau(\boldsymbol{X}_i) = g_\tau(\beta_\tau^\mathsf{T} \boldsymbol{X}_i),\ i = 1, \ldots, n,$$

where $g_\tau : \mathbb{R} \to \mathbb{R}$ is an unknown smooth link function. The most straightforward route to take censoring into account was here again exploited by Bücher et al., as the suggested estimation procedure relies on opportune ICP weights. The latter estimator will in fact be the main competing procedure of our work in Chapter 2.

This concludes our review on the methodologies for censored quantile regression estimation that were developed prior to our work. We now have all arguments to explicitly detail in the next section the contributions of this thesis to the subject.

1.5 Outline of the thesis

To introduce our contributions in the simplest and most straightforward manner, we propose to replace the diagram of Figure 1.7 with the one reported in Figure 1.8, where our work is clearly highlighted. This will serve as a guideline for this section.

Our starting point, as already multiply evoked, is to consider in Chapter 2 a copula quantile regression estimator stemming from the check-based formulation of conditional quantiles. Resulting in a first paper (De Backer et al. (2017)), the objective of this chapter is in fact twofold: first, for the situation with completely observed responses and extending the recent work of Noh et al. (2013) and Noh et al. (2015), we intend to provide here a simple solution to the shortcomings of a pure parametric copula formulation as highlighted in Dette et al. (2014) and evoked in Section 1.2.3. To that end, building on the ideas of pair-copula constructions, our proposal is to consider a semiparametric estimation scheme for the multivariate copula density itself, for which the chosen pair-copula decomposition is driven by the regression context. The second main objective of this chapter is then to propose a copula-based estimator accommodating for possible censored responses. Given

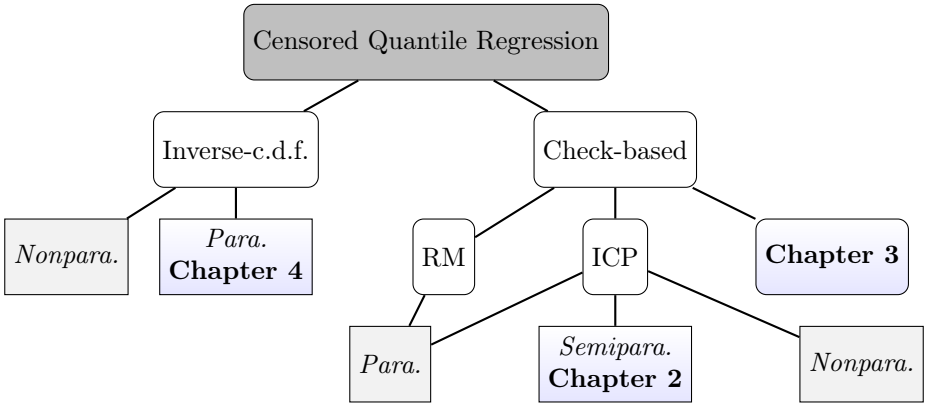

Figure 1.8: Simplistic representation of the censored quantile regression literature with highlighting of our contributions.

the semiparametric context of the chapter, and given the above-described review of the literature, our suggestion will here be based on ICP weights, for which careful introduction will be necessary when handling copulas. This then reveals the choice of the chapter's position in the diagram of Figure 1.8, although it is noted that the procedure would actually fall under the scope of pure nonparametric regression in case of a univariate covariate.

Chapter 3 comes for its part after a reconsideration, following Chapter 2, concerning the handling of censored data in general quantile regression models. In the vocabulary introduced before, our review of the literature indeed suggests that, when proposing a particular model, one has to trade-off between quantile-specific and model-free approaches for the controlling of censored data. The aim of this chapter may then be resumed as an attempt to bridge these two notions into a unified framework for censored quantile regression. To avoid any confusion, we emphasize here that the notion 'quantile-specific' should not automatically be linked to RM weights, as the term was introduced to denote any procedure that actually takes quantile features into account in its handling of censored data. Inspired by the elementary link between quantiles and the check function reported in Section 1.1.1, this reconsideration of censored quantile modelling then leads to the definition of a new loss function that extends the check function by spontaneously accommodating for censoring. Intuitively, using all observations at hand as if they were complete, the loss function is observed to penalize more severely underestimation of the value of the quantile than the usual check function in order to take the latter issue into account in this setting. Hence, by tackling the issues related to censored data at the very level of the quantile loss function, the methodology is qualified as being both quantile-specific and model-free, as it may easily be applied to parametric, semiparametric and nonparametric models. As an illustration of the

latter statement, the procedure is applied to the linear context in a second paper (De Backer et al. (2018a)).

Lastly, while Chapter 3 offers a natural toolbox for numerous extensions outside the linear framework, Chapter 4 digs a little further into the latter context. Two main reasons are driving this choice: first, the obvious explanation comes from the fact that the linear model is arguably the most widely applied model by practitioners, and second, in contrast to the previous chapters, the idea comes here from a more fundamental reconsideration of the almost universal use of check-based modelling in the present context. The latter reconsideration comes indeed from the knowledge that all the proposed check-based estimation procedures require either strong simplifying assumptions, or nonparametric estimation of a conditional distribution (of T given $\boldsymbol{X}$ or of C given $\boldsymbol{X}$). Bearing this observation in mind, along with further arguments that will be developed when necessary, Chapter 4 then proposes what could be qualified as an inverse-c.d.f. approach to linear quantile regression with censored data. To avoid dimensionality issues when considering such an inverse-c.d.f. methodology, and inspired by some check-based literature, the procedure further builds on dimension reduction techniques that were however not reviewed here for sake of brevity. Finally, similarly to Chapter 2, it should be noted that the resulting procedure is in fact new for the situation with complete observations as well, even though the motivation is obviously less straightforward given it induces smoothing for a parametric model with exclusively complete observations. This last chapter leads to a third paper (De Backer et al. (2018b)).

CHAPTER 2

Semiparametric Copula Quantile Regression for Complete or Censored Data

Abstract

In this chapter, a semiparametric copula-based estimator for conditional quantiles is investigated for both complete or right-censored data. In spirit, the methodology is extending the recent work of Noh et al. (2013) and Noh et al. (2015) reported in Section 1.2.3, as the main idea consists in appropriately defining the quantile regression in terms of a multivariate copula and marginal distributions. Prior estimation of the latter and simple plug-in lead to an easily implementable estimator expressed, for both contexts with or without censoring, as a weighted quantile of the observed response variable. In addition, and contrary to the initial suggestion in the literature, a semiparametric estimation scheme for the multivariate copula density is studied, motivated by the possible shortcomings of a purely parametric approach and driven by the regression context. The resulting quantile regression estimator has the valuable property of being automatically monotonic across quantile levels. Furthermore, the copula-based approach allows the analyst to spontaneously take account of common regression concerns, such as interactions between covariates or possible transformations of the latter. From a theoretical prospect, asymptotic normality for both complete and censored data is obtained under classical regularity conditions. Finally, numerical examples as well as a real data application are used to illustrate the validity and finite sample performance of the proposed procedure. This chapter is based on the paper

> De Backer, M., A. El Ghouch, and I. Van Keilegom. *Semiparametric copula quantile regression for complete or censored data.* Electronic Journal of Statistics 11(1): 1660-1698, 2017.

2.1 Introduction

As introduced in Chapter 1, the objective of the present chapter is twofold. First, our concern is here related to the estimation of a quantile regression function where the response variable T is completely observed, for which a wide literature on the subject already includes fully parametric, semiparametric and nonparametric methodologies. When several covariates are to be taken into account, fully parametric methodologies are known to be highly sensitive to model misspecification and may lack the flexibility needed for an adequate modelling. On the other hand, in spite of their great flexibility, fully nonparametric methods such local quantile regression as proposed by Spokoiny et al. (2013) are typically affected by the curse of dimensionality. In light of these restrictions, when the dimension of the covariate is high, data analysis must turn to the semiparametric front of statistical modelling, where estimation procedures such as a single-index regression (Wu et al. (2010), Zhu et al. (2012)) come at hand.

In this context, as recalled in Section 1.2.3, Noh et al. (2013) and Noh et al. (2015) suggested to estimate a regression function based on the copula that defines the dependence structure between the variables of interest. In the previously-introduced vocabulary, stemming from a check-based formulation of conditional quantiles, the central idea of the methodology is to express the conditional quantile function in terms of a copula density and marginal distributions. However, as previously-mentioned, possible shortcomings of the original suggestion of Noh et al. were highlighted by Dette et al. (2014) through several simple settings, where the misspecification of the assumed parametric copula is shown to provide a very poor fit of the true conditional quantile function. Building on these examples, the first contribution of this chapter is then to circumvent such issues by proposing an alternative semiparametric estimation strategy for the copula itself that is motivated by the regression context. The resulting regression estimator is flexible for multi-dimensional data, easy to implement and does not require any iterative procedure in opposition to existing semiparametric alternatives.

The second objective of this chapter is to propose a copula-based methodology in the context of survival analysis, characterized by possible right censoring of T. In this situation, quantile regression becomes attractive as an alternative to popular regression techniques like the Cox proportional hazards model or the accelerated failure time model, as argued in Section 1.3.3. Hence, a notable rise in the literature on censored quantile regression has been observed over the last decades, and, similarly to the uncensored situation, existing literature now includes fully parametric, semiparametric and nonparametric methodologies. With the same motivation for turning to semiparametric methodologies as for situations with complete observations, our interest lies here again in the latter category of statistical models. In this context, an interesting approach prior to our work and briefly recalled in Section 1.4.4 was proposed by Bücher et al. (2014), where a single-index model for the conditional quantile function is studied under the assumption of independence

between the covariates and the censoring variable. The single-index structure assumes that the objective function depends linearly on the covariates through an unknown link function, making the proposed model (i.e. under the aforementioned assumption) insensitive to the curse of dimensionality since the nonparametric part is of dimension one. However, besides the latter single-index model, existing literature on flexible multidimensional methodologies is rather sparse.

Hence, the second main contribution of this chapter is to extend this literature on multivariate quantile regression estimation in the possible presence of censored data by providing a rich, flexible and robust alternative based on the copula function. In essence, the proposed methodology mimics the work of Noh et al. (2013) for complete observations, as the central idea is here again to rewrite the conditional quantile function in terms of marginal distributions and an appropriate copula density using the check-based formulation of conditional quantiles. Employing the proposed multivariate copula density estimation strategy, the resulting regression estimator enjoys the same qualities as for complete observations. Consequently, by taking advantage of copula modelling in regression models, the proposed methodology provides a new class of estimators that allow practitioners to flexibly analyze multidimensional survival data.

The rest of this chapter is organized as follows. Developing a semiparametric copula-based estimation procedure for quantile regression with complete observations is the topic of Section 2.2. Section 2.3 develops the methodology for the copula-based quantile regression estimator when confronted to possible censoring of the responses. The asymptotic properties of the proposed estimators for both complete and censored data are obtained in Section 2.4 and the finite sample performance is illustrated by means of Monte Carlo simulations in Section 2.5, where both the semiparametric copula estimation strategy and the overall performance of our estimator for complete and censored data are investigated. Section 2.6 provides a brief application to real censored data. Lastly, the proofs of our asymptotic properties are deferred to Section 2.8.

2.2 Copula-based estimator for complete data

2.2.1 Background for copula-based quantile regression

We introduce in this section the general context of check-based copula quantile regression prior to our work. Note that much of this section has in fact already been covered in Chapter 1, but is repeated here for ease of reading of the following sections.

Now, let $\boldsymbol{X} = (X_1, \ldots, X_d)^\intercal$ again be a covariate vector of dimension $d \geq 1$, and T be a (time-to-event) response variable with marginal c.d.f. $F_1, \ldots, F_d$ and F_T, respectively. Further recall that we denote by f_j and f_T the densities of $X_j, j = 1, \ldots, d$, and T, respectively. From the pioneering work of Sklar (1959), for a given $\boldsymbol{x} = (x_1, \ldots, x_d)^\intercal$, the c.d.f. of $(T, \boldsymbol{X})$ evaluated at $(t, \boldsymbol{x})$ can be expressed

as $C_{T\boldsymbol{X}}(F_T(t), \boldsymbol{F}(\boldsymbol{x}))$, where $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1), \ldots, F_d(x_d))^\intercal$ and $C_{T\boldsymbol{X}}$ is the unique copula distribution of $(T, \boldsymbol{X})$ defined by $C_{T\boldsymbol{X}}(u_0, u_1, \ldots, u_d) = \mathbb{P}(U_0 \leq u_0, U_1 \leq u_1, \ldots, U_d \leq u_d)$, with $U_0 = F_T(T)$ and $U_j = F_j(X_j)$, $j = 1, \ldots, d$.

The starting point of this chapter is the usual observation that our object of interest, the τ-th conditional quantile function of the dependent variable T given $\boldsymbol{X} = \boldsymbol{x}$, denoted by $m_\tau(\boldsymbol{x})$, allows for a well-known check-based formulation. That is, for any $\tau \in (0, 1)$,

$$m_\tau(\boldsymbol{x}) = \arg\min_a \mathbb{E}\Big[\rho_\tau(T - a)|\boldsymbol{X} = \boldsymbol{x}\Big], \tag{2.2.1}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u \leq 0))$ is the previously-introduced check function, and $\mathbb{1}(\cdot)$ is the indicator function.

To motivate our approach, we suppose in this section that there is no censoring and that we observe an i.i.d. sample $(T_i, \boldsymbol{X}_i), i = 1, \ldots, n$, from $(T, \boldsymbol{X})$. In this context, following the definition of a copula function as detailed in Section 1.2.3, Noh et al. (2015) noted that the conditional quantile function of T given $\boldsymbol{X} = \boldsymbol{x}$ may be expressed as

$$m_\tau(\boldsymbol{x}) = \arg\min_a \mathbb{E}\Big[\rho_\tau(T - a)\, c_{T\boldsymbol{X}}(F_T(T), \boldsymbol{F}(\boldsymbol{x}))\Big], \tag{2.2.2}$$

where $c_{T\boldsymbol{X}}(u_0, \boldsymbol{u}) \equiv c_{T\boldsymbol{X}}(u_0, u_1, \ldots, u_d) = \partial^{d+1} C_{T\boldsymbol{X}}(u_0, u_1, \ldots, u_d)/\partial u_0 \partial u_1 \ldots \partial u_d$ is the copula density corresponding to $C_{T\boldsymbol{X}}$. Consequently, any given estimators $\widehat{c}_{T\boldsymbol{X}}$, $\widehat{F}_T$ and $\widehat{F}_j$ of $c_{T\boldsymbol{X}}$, F_T and F_j, $j = 1, \ldots, d$, respectively, automatically yield an estimator of $m_\tau(\boldsymbol{x})$ given by

$$\widehat{m}_\tau(\boldsymbol{x}) = \arg\min_a \sum_{i=1}^n \rho_\tau(T_i - a)\, \widehat{c}_{T\boldsymbol{X}}(\widehat{F}_T(T_i), \widehat{\boldsymbol{F}}(\boldsymbol{x})), \tag{2.2.3}$$

with $\widehat{\boldsymbol{F}}(\boldsymbol{x}) = (\widehat{F}_1(x_1), \ldots, \widehat{F}_d(x_d))^\intercal$. As indicated earlier, Noh et al. suggest at this stage to estimate the marginals nonparametrically and to consider a parametrization of the copula density, that is, assume that the latter belongs to a certain parametric family of copula densities $\mathcal{C} = \{c(u_0, \boldsymbol{u}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$, for some $p \in \mathbb{N}$.

2.2.2 A motivational one-dimensional example

In this work however, we consider an alternative estimation approach for the copula density, motivated by the issues related to the possible misspecification of the parametric approach. To highlight this shortcoming and illustrate how one may circumvent it, we consider in this section the simplistic example reported by Dette et al. (2014) with a single covariate, where $(T_i, X_{1i}), i = 1, \ldots, n$, are i.i.d. random variables with $T_i = (X_{1i} - 0.5)^2 + \sigma\epsilon_i$, $X_{1i} \sim U[0, 1]$, $\sigma = 0.025$ and $\epsilon_i, i = 1, \ldots, n$, are i.i.d. standard normal random variables. In this situation, where the true

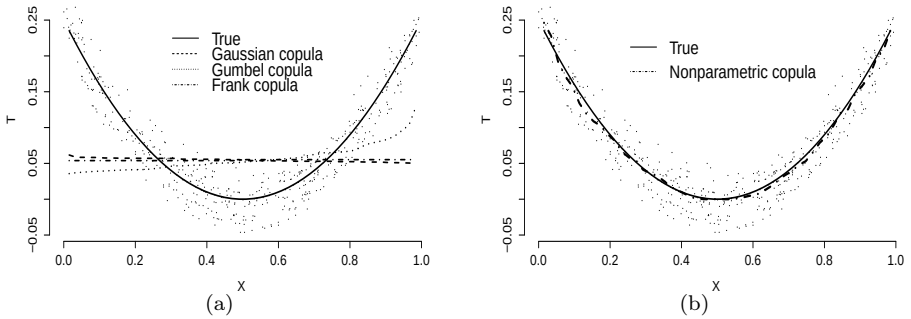

Figure 2.1: Copula-based quantile regression estimates of the one-dimensional minimalistic example. The Gaussian, Gumbel and Frank copulas are used for parametric copula estimation in (a), while (b) depicts the regression fit resulting from a nonparametric estimation of the copula density using the procedure of Geenens et al. (2017) with nearest-neighbor bandwidth selection, implementable via the function `kdecop` (`method = ''TLL2nn''`) of the R package `kdecopula`.

quantile regression function is non-monotonic in the covariate, it is found that most of the common parametric copula families still yield a monotone estimation of the regression function, thereby providing a rather poor fit of the latter. This is illustrated in Figure 2.1a, where the estimation is carried out for $\tau = 0.5$, $n = 500$ and using three common parametric copulas, where estimation of the latter is implemented using the likelihood procedure described in (1.2.2).

Now, as the roots of the above-mentioned limitation are not intrinsic to a copula-based approach, but rather to be attributed to the limited set of parametric copula families existing in the literature, a natural alternative for low dimensional covariates would be to consider a fully nonparametric estimation of the copula density itself. The resulting, and adequate, quantile regression estimation is depicted in Figure 2.1b.

For this approach to be appropriate, a suitable nonparametric estimation of the copula density is required. As reviewed in Chapter 1, in the literature several specific kernel-based methodologies have been proposed in order mainly to correct for well-known bias issues at the boundaries, as the density of interest is only supported on the unit square. These include for instance the mirror reflection method (Gijbels and Mielniczuk (1990)) or the boundary kernel method (Chen and Huang (2007)). In this chapter however, we will adopt the technique proposed by Geenens et al. (2017) that couples the idea of projecting the initial data on an unbounded support with polynomial local-likelihood density estimation. More details about the latter technique have been given in Section 1.2.2 and are therefore omitted here. An exhaustive comparison of existing methodologies for bivariate copula density estimation may be found in Nagler (2014). Furthermore, for completeness it is

worth mentioning that fully nonparametric multidimensional copulas (i.e. $d > 1$) are studied in Hobæk Haff and Segers (2015) and Nagler and Czado (2016). An alternative strategy for the latter multivariate situation will however be considered here and described in the next section.

2.2.3 A semiparametric copula estimator for multivariate covariates

Recalling that we intend to handle multivariate covariates in this chapter, it seems at this point unwise to advocate for a purely nonparametric approach as in the previously-described motivational example. Instead, we will prefer a copula estimation strategy that provides sufficient flexibility to the multidimensional estimator while avoiding dimension related constraints. More specifically, bearing in mind the essence of pair-copula constructions as recalled in Chapter 1, we note that any multivariate copula density can be decomposed into two parts as follows:

$$c_{T\boldsymbol{X}}(F_T(t), \boldsymbol{F}(\boldsymbol{x})) = c_{TX_1}(F_T(t), F_1(x_1)) \times \ldots \times c_{TX_d}(F_T(t), F_d(x_d)) \tag{2.2.4}$$

$$\times\, c_{X_1 \ldots X_d|T}(F_{1|T}(x_1|t), \ldots, F_{d|T}(x_d|t)|t), \tag{2.2.5}$$

where $F_{j|T}$, $j = 1, \ldots, d$, denotes the conditional c.d.f. of X_j given T. The first part of the decomposition contains the product of d bivariate copula densities related to the dependence of T with every covariate, whereas the second part captures the conditional dependence of $\boldsymbol{X}$ given $T = t$. In the general regression context, part (2.2.4) may then be interpreted as the dependence of actual interest since it focuses on the relationship between the response variable with every covariate. On the contrary, part (2.2.5) may be viewed in such framework as a correction parameter for possible (conditional) dependence among covariates that is, arguably, less crucial here. Consequently, a natural reasoning suggests to provide as much flexibility as possible to the modelling of part (2.2.4), while keeping the estimation of part (2.2.5) uncomplicated. We therefore advocate to estimate nonparametrically the d bivariate copulas of interest and, subsequently, exploit standard parametric techniques for the second part of the multivariate copula density. The latter involve, among others, nested Archimedean copulas (see e.g. Hofert and Pham (2013) and Joe (2014)), factor copulas (see e.g. Oh and Patton (2012)) and, arguably the most popular, vine copulas (see e.g. Czado (2010), Joe (2014) and references therein). Note however that, for the estimation of (2.2.5) in practice, one is first advised to adopt the so-called *simplifying assumption* which stipulates that the conditioning on $T = t$ is fully captured by the conditional marginals. In other words, in (2.2.5), the conditional copula itself is not affected by the conditioning on T. This assumption turns out to be the cornerstone of vine copula models as it keeps them tractable for inference and model selection. For more details about this remark and its implications, we refer for instance to the work of Hobæk Haff et al. (2010) and Stöber et al. (2013), as already commented in Section 1.2.2.

Conclusively, in this work we propose to adopt the following detailed procedure for the modelling and estimation of the multivariate copula density:

(1) Based on original observations $(T_i, X_{1i}, \ldots, X_{di}), i = 1, \ldots, n$, construct the so-called pseudo-observations needed for the estimation of (2.2.4), using rescaled versions of empirical distributions:

$$\widehat{U}_{0i} = \frac{1}{n+1} \sum_{k=1}^{n} \mathbb{1}(T_k \leq T_i) \quad \widehat{U}_{ji} = \frac{1}{n+1} \sum_{k=1}^{n} \mathbb{1}(X_{jk} \leq X_{ji}),$$
$$i = 1, \ldots, n, \quad j = 1, \ldots, d,$$

where the factor $1/(n+1)$, commonly adopted in the copula literature, aims at keeping the constructed observations in the interior of $[0, 1]$ as recalled in Chapter 1.

(2) Based on $(\widehat{U}_{0i}, \widehat{U}_{ji}), i = 1, \ldots, n$, estimate each bivariate copula density c_{TX_j}, $j = 1, \ldots, d$, in (2.2.4) using a bivariate kernel density estimator. This can be achieved via the local-polynomial probit methodology of Geenens et al., or any other estimator satisfying assumption (C7) given below.

(3) Compute the pseudo-observations needed for the estimation of (2.2.5) as:

$$\widehat{F}_{j|T}(X_{ji}|T_i) = \int_0^{\widehat{U}_{ji}} \widehat{c}_{TX_j}(\widehat{U}_{0i}, s)\, \mathrm{d}s.$$

This relationship is in fact at the origin of the sequential nature of the vine copula estimation scheme (see e.g. Czado (2010)).

(4) Lastly, for the estimation of $c_{X_1 \ldots X_d|T}$, adopt the simplifying assumption and use standard parametric vine techniques on the dataset $(\widehat{F}_{1|T}(X_{1i}|T_i), \ldots, \widehat{F}_{d|T}(X_{di}|T_i)), i = 1, \ldots, n$.

Finally, to graphically illustrate this procedure in a multivariate example, we consider a simple extension to the previously-described setting with two independent covariates, where (T_i, X_{1i}, X_{2i}), $i = 1, \ldots, n$, are i.i.d. random variables with $T_i = (X_{1i} - 0.5)^2 + (X_{2i} - 0.5)^2 + \sigma\epsilon_i$, $X_{1i} \sim U[0, 1]$, $X_{2i} \sim U[0, 1]$, $\sigma = 0.025$ and $\epsilon_i, i = 1, \ldots, n$, are i.i.d. standard normal random variables. In this situation, a fully parametric vine copula estimation, implemented via the R package `VineCopula`, again yields an inappropriate quantile regression estimation, as depicted in Figure 2.2a for $\tau = 0.5$ and $n = 500$. In opposition, the proposed semiparametric copula density estimation allows the regression estimator to suitably capture the non-monotonic features of both covariates. This can be observed in Figure 2.2b. A more exhaustive analysis of the performance of the proposed estimator with respect to existing competitors will be provided in Section 2.5. For now, with this proposed copula estimation strategy at hand, we may turn in the next section to the consideration of censored responses in our quantile regression context.

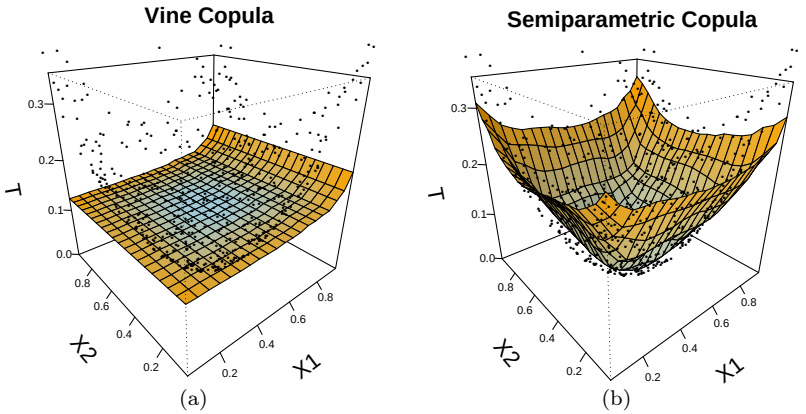

Figure 2.2: Copula-based quantile regression estimates of the two-dimensional minimalistic example. A parametric vine copula estimation is at the basis of the quantile estimation in (a), while (b) illustrates the semiparametric copula estimation approach proposed in this chapter.

2.3 Copula-based estimator for censored data

In the presence of censoring, the estimation equation (2.2.3) becomes inappropriate as we do not fully observe the response variables T_i. Instead, we only observe a sequence of i.i.d. triplets $(Y_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, from $(Y, \Delta, \boldsymbol{X})$, where $Y = \min(T, C)$, $\Delta = \mathbb{1}(T \leq C)$ and C denotes the censoring variable, assumed to be independent of T given $\boldsymbol{X}$. In order to take censoring into account in the estimation procedure, using the vocabulary of Chapter 1, we propose here to adopt an ICP-based approach. That is, similarly to (1.4.1), our first step is to note that for any measurable function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[\varphi(T)|\boldsymbol{X} = \boldsymbol{x}] = \mathbb{E}\left[\varphi(Y)\frac{\Delta}{1 - G_C(Y - |\boldsymbol{x})}\middle|\boldsymbol{X} = \boldsymbol{x}\right], \tag{2.3.1}$$

where $G_C(c|\boldsymbol{x}) = \mathbb{P}(C \leq c|\boldsymbol{X} = \boldsymbol{x})$ denotes the conditional distribution of C given $\boldsymbol{X} = \boldsymbol{x}$. This, along with (2.2.1), suggests a natural way to handle censoring for quantile regression by replacing the function φ with the check function.

At this stage, in the objective of proposing a similar methodology to the one developed in Section 2.2, we propose to work on the obtained conditional expectation in (2.3.1) before considering the introduction of copulas. The underlying rationale is that the latter conditional expectation is in fact the joint conditional expectation of (Y, Δ) given $\boldsymbol{X} = \boldsymbol{x}$. Adopting an analogous reasoning as the one presented by Noh et al. for the uncensored case at this point would therefore result in the insertion of the joint copula of $(Y, \Delta, \boldsymbol{X})$, hence exposing the estimation procedure to

the lack of uniqueness of the copula given that Δ is a discrete (binary) variable. An interesting opinion on several pitfalls appearing for copulas with discrete variables can be found in Embrechts (2009). Additional details may also be found in Genest and Nešlehová (2007).

Instead, the idea is to work on the joint conditional expectation so as to bypass the issues related to the copula of $(Y, \Delta, \boldsymbol{X})$. In short, our intention is to discard the problem by obtaining the copula of $(Y, \boldsymbol{X})$ conditionally on $\Delta = 1$, for which no specific technical difficulties are involved. To that end, using the notations H^u (resp. h^u) as a shorthand for a given distribution (resp. density) conditionally on $\Delta = 1$, first note that

$$\mathbb{E}\bigg[\rho_\tau(Y-a)\frac{\Delta}{1-G_C(Y-|\boldsymbol{x})}\Big|\boldsymbol{X}=\boldsymbol{x}\bigg] = \int_{\mathbb{R}^+}\rho_\tau(y-a)\,\frac{1}{1-G_C(y-|\boldsymbol{x})}\,\mathrm{d}F_{Y,\Delta|\boldsymbol{X}}(y,1|\boldsymbol{x}), \tag{2.3.2}$$

where $F_{Y,\Delta|\boldsymbol{X}}(y,1|\boldsymbol{x}) = \mathbb{P}(Y\le y,\Delta=1|\boldsymbol{X}=\boldsymbol{x}) = p(\boldsymbol{x})\int_{-\infty}^{y} f^u_{Y\boldsymbol{X}}(z,\boldsymbol{x})/f^u(\boldsymbol{x})\,\mathrm{d}z$, with $p(\boldsymbol{x}) = \mathbb{P}(\Delta=1|\boldsymbol{X}=\boldsymbol{x})$ and where $f^u_{Y\boldsymbol{X}}$ and f^u denote the conditional densities of $(Y,\boldsymbol{X})$ and $\boldsymbol{X}$ given $\Delta=1$, respectively. Hence, using the definition of a copula function in a similar spirit as Noh et al., one obtains

$$\mathrm{d}F_{Y,\Delta|\boldsymbol{X}}(y,1|\boldsymbol{x}) = p(\boldsymbol{x})\,\frac{c^u_{Y\boldsymbol{X}}(F^u_Y(y),\boldsymbol{F}^u(\boldsymbol{x}))}{c^u_{\boldsymbol{X}}(\boldsymbol{F}^u(\boldsymbol{x}))}\,f^u_Y(y)\,\mathrm{d}y,$$

where $c^u_{Y\boldsymbol{X}}(u_0,\boldsymbol{u}) = \partial^{d+1}C^u_{Y\boldsymbol{X}}(u_0,u_1,\ldots,u_d)/\partial u_0\partial u_1\ldots\partial u_d$ is the copula density corresponding to the copula $C^u_{Y\boldsymbol{X}}$ of $(Y,\boldsymbol{X}|\Delta=1)$, $c^u_{\boldsymbol{X}}(\boldsymbol{u}) = \partial^d C^u_{\boldsymbol{X}}(u_1,\ldots,u_d)/\partial u_1\ldots\partial u_d$ is the copula density of $(\boldsymbol{X}|\Delta=1)$, and $\boldsymbol{F}^u(\boldsymbol{x}) = (F^u_1(x_1),\ldots,F^u_d(x_d))^\intercal$. Inserting this last expression in (2.3.2), we may write

$$\mathbb{E}[\rho_\tau(T-a)|\boldsymbol{X}=\boldsymbol{x}] = \mathbb{E}\left[\rho_\tau(Y-a)\,\frac{\Delta}{1-G_C(Y-|\boldsymbol{x})}\,\frac{p(\boldsymbol{x})}{\mathbb{P}(\Delta=1)}\,\frac{c^u_{Y\boldsymbol{X}}(F^u_Y(Y),\boldsymbol{F}^u(\boldsymbol{x}))}{c^u_{\boldsymbol{X}}(\boldsymbol{F}^u(\boldsymbol{x}))}\right].$$

Applying this equality in the context of quantile regression, interestingly, one eventually retrieves an expression analogous to (2.2.2):

$$m_\tau(\boldsymbol{x}) = \arg\min_a \mathbb{E}\Big[\rho_\tau(Y-a)\,W(\boldsymbol{x})\,c^u_{Y\boldsymbol{X}}(F^u_Y(Y),\boldsymbol{F}^u(\boldsymbol{x}))\Big], \tag{2.3.3}$$

where $W(\boldsymbol{x}) \equiv \Delta/(1-G_C(Y-|\boldsymbol{x}))$. Note that the copula in question in (2.3.3) is determined by strictly fully observed data. Hence, standard literature on copulas may be manipulated without any censoring related constraints. Given estimators $\widehat{G}_C(\cdot|\boldsymbol{x})$, $\widehat{F}^u_Y$ and $\widehat{F}^u_j$ of $G_C(\cdot|\boldsymbol{x})$, F^u_Y and F^u_j, $j=1,\ldots,d$, satisfying certain high-level conditions which will be given in Section 2.4, this then suggests to estimate

the quantile regression in the presence of censoring by the empirical analogue of (2.3.3), that is

$$\widehat{m}_\tau(\boldsymbol{x}) = \arg\min_a \sum_{i=1}^n \rho_\tau(Y_i - a)\, \widehat{W}_i(\boldsymbol{x})\, \widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})), \tag{2.3.4}$$

where $\widehat{W}_i(\boldsymbol{x}) = \Delta_i/(1-\widehat{G}_C(Y_i-|\boldsymbol{x}))$, and where $\widehat{c}^u_{Y\boldsymbol{X}}$ denotes an estimator of $c^u_{Y\boldsymbol{X}}$ based on the four-step procedure described in Section 2.2.3. Explicitly,

$$\begin{aligned}\widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(y), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) = \widehat{c}^u_{YX_1}(\widehat{F}^u_Y(y), \widehat{F}^u_1(x_1)) \times \ldots \times \widehat{c}^u_{YX_d}(\widehat{F}^u_Y(y), \widehat{F}^u_d(x_d)) \\ \times \widehat{c}^u_{X_1\ldots X_d|Y}(\widehat{F}^u_{1|Y}(x_1|y), \ldots, \widehat{F}^u_{d|Y}(x_d|y)),\end{aligned}$$

where any two-dimensional kernel density estimator may be used for each bivariate copula density $\widehat{c}^u_{YX_j}, j = 1, \ldots, d$, such that condition (C7) of Section 2.4 holds, and where $\widehat{c}^u_{X_1\ldots X_d|Y}$ is estimated by standard parametric vine procedures.

The resulting quantile regression estimator in (2.3.4) may then be viewed as a simple weighted quantile of the observed response variable, and is therefore easy to implement in practice using the efficient quantile regression code developed by Portnoy and Koenker (1997) and Koenker (2005). Nonetheless, in the context of multivariate covariates, the estimation of $G_C(\cdot|\boldsymbol{x})$ requires further assumptions to overcome dimension related issues. Popular choices in the literature include, among others, independence between C and $\boldsymbol{X}$, the Cox model or the single-index model on $C|\boldsymbol{X} = \boldsymbol{x}$. Illustrations of such assumptions are treated in our simulation study.

Lastly, as an interesting property and similarly to the case without censoring, we note that the obtained regression function estimator is automatically monotonic across quantile levels. Applying analogous arguments to the ones adopted in the proof of Theorem 2.5 of Koenker (2005), one can indeed determine that

$$(\tau_2 - \tau_1)(\widehat{m}_{\tau_2}(\boldsymbol{x}) - \widehat{m}_{\tau_1}(\boldsymbol{x})) \sum_{i=1}^n \widehat{W}_i(\boldsymbol{x})\, \widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) \geq 0. \tag{2.3.5}$$

Given that $\widehat{W}_i(\boldsymbol{x})\, \widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) \geq 0$ for all $i = 1, \ldots, n$, this signifies that $\widehat{m}_{\tau_2}(\boldsymbol{x}) \geq \widehat{m}_{\tau_1}(\boldsymbol{x})$ for $\tau_2 \geq \tau_1$.

Conclusively, in parallel to what has been stated for the uncensored case, the resulting estimator $\widehat{m}_\tau(\boldsymbol{x})$ defines a rich class of estimators built on the many different existing methods available in the literature for estimating copula densities and marginal distributions of both complete and censored data.

2.4 Asymptotic Properties

We establish in this section the asymptotic normality of the proposed estimator $\widehat{m}_\tau(\boldsymbol{x})$. To that end, we first report the set of regularity conditions as well as

the required high-level conditions on all estimators involved in the expression of $\widehat{m}_\tau(\boldsymbol{x})$. We then develop an asymptotic representation of our estimator for a general d-variate covariate. As the latter will result in a somewhat unpleasant expression for the asymptotic bias and variance for a general multivariate covariate, and given that the analytical reasoning is similar in spirit, we eventually restrict ourselves to the detailed asymptotic expression for the case $d = 2$.

For a fixed but arbitrary point of interest $\boldsymbol{x}$ in the interior of the support of $\boldsymbol{X}$, denoted by $\text{supp}(\boldsymbol{X})$, let us suppose that there exists a neighborhood $\mathcal{V}_{\boldsymbol{F}^u(\boldsymbol{x})}$ of $\boldsymbol{F}^u(\boldsymbol{x})$ such that the following regularity conditions hold:

(C1) The conditional distribution $F_{T|\boldsymbol{X}}$ of T given $\boldsymbol{X}$ admits a conditional density $f_{T|\boldsymbol{X}}$ that is continuous, strictly positive and bounded uniformly on $\mathbb{R} \times \text{supp}(\boldsymbol{X})$.

(C2) The point of interest $\boldsymbol{x}$ is such that $\boldsymbol{F}^u(\boldsymbol{x}) \in (0,1)^d$ and $\mathbb{P}(\Delta = 1|\boldsymbol{x}) > 0$. Furthermore, $0 < c^u_{\boldsymbol{X}}(\boldsymbol{F}^u(\boldsymbol{x})) < \infty$, $\sup_{t\in\mathbb{R}} c^u_{Y\boldsymbol{X}}(F^u_Y(t), \boldsymbol{F}^u(\boldsymbol{x})) < \infty$ and $\inf_{t\in\mathbb{R}} c^u_{Y\boldsymbol{X}}(F^u_Y(t), \boldsymbol{F}^u(\boldsymbol{x})) > 0$.

(C3) The point $m_\tau(\boldsymbol{x}) \in \mathbb{R}$ satisfies $G_C(m_\tau(\boldsymbol{x}) + \delta|\boldsymbol{x}) < 1$, for some $\delta > 0$.

(C4) Denote $\epsilon \equiv \epsilon(\boldsymbol{x}, \tau) = Y - m_\tau(\boldsymbol{x})$ and define $\psi_\tau(u) = \tau - I(u \leq 0)$. Then,

(i) $\mathbb{E}(|\psi_\tau(\epsilon)|\, W(\boldsymbol{x})) < \infty$.

(ii) $\mathbb{E}\left[\psi_\tau(\epsilon)\, W(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y), \boldsymbol{F}^u(\boldsymbol{x}))\right]^2 < \infty$.

Concerning the high-level conditions, it is assumed that the multivariate copula density $c^u_{Y\boldsymbol{X}}$ is estimated using the proposed four-step strategy of Section 2.2.3, and that, for simplicity, the d bivariate kernel copula estimators of step (2) are based on the same bandwidth $\boldsymbol{H} = h^2\boldsymbol{I}$ for a certain $h > 0$. The following conditions are then assumed to hold:

(C5) The marginal c.d.f. estimators are such that:

(i) $\sup_{t\in\mathbb{R}} \left|\widehat{F}^u_Y(t) - F^u_Y(t)\right| = O_{\mathbb{P}}(n^{-1/2})$.

(ii) $\widehat{\boldsymbol{F}}^u(\boldsymbol{x}) - \boldsymbol{F}^u(\boldsymbol{x}) = O_{\mathbb{P}}(n^{-1/2})$, where $\widehat{\boldsymbol{F}}^u(\boldsymbol{x}) = (\widehat{F}^u_1(x_1), \ldots, \widehat{F}^u_d(x_d))^\intercal$ and $\widehat{F}^u_j$ is an estimator of F^u_j.

(C6) $\sup_{t\leq\tau_{F_Y}} |\widehat{G}_C(t|\boldsymbol{x}) - G_C(t|\boldsymbol{x})| = o_{\mathbb{P}}((nh^2)^{-1/2})$, and $\tau_{F_Y} < \tau_{G_C}$, where $\tau_{F_Y} = \inf\{t : F_Y(t) = 1\}$ and $\tau_{G_C} = \inf\{t : G_C(t) = 1\}$.

(C7) The multivariate copula estimator is such that:

(i) $\sup_{u_0\,\in\,(0,1)} \left|\widehat{c}^u_{YX_j}(u_0, F^u_j(x_j)) - c^u_{YX_j}(u_0, F^u_j(x_j))\right| = o_{\mathbb{P}}(1)$, $j = 1, \ldots, d$, where x_j is the j-th coordinate of $\boldsymbol{x}$.

(ii) $\sup_{u_0 \in (0,1)} \sup_{\boldsymbol{u} \in \mathcal{V}_{\boldsymbol{F}^u(\boldsymbol{x})}} |\partial_j \widehat{c}^u_{Y\boldsymbol{X}}(u_0, \boldsymbol{u})| = O_{\mathbb{P}}(1)$, $j = 1, \ldots, d{+}1$, where ∂_j denotes the partial derivative with respect to the j-th argument.

Assumption (C1) is standard in the context of quantile regression estimation. As for condition (C2), this is similar to assumption (C3)-(i) in Noh et al. (2015) for the simplified case with no censoring, with an additional requirement on the conditional censoring probability that is resulting from the consideration of censored observations. Assumption (C3) is likewise emanating from the proposed handling of censored data, and is rather usual in survival analysis. Note that, in the quantile regression framework, the latter assumption amounts to defining a natural upper bound for the quantile of interest that can be studied. Assumption (C4) reports a set of technical conditions to be met.

As regards conditions (C5)-(C7), (C5) is routinely made in the copula framework. For instance, it is readily satisfied for the empirical distributions when only uncensored observations are taken into account, and their rescaled versions which are prominent in the copula literature and adopted here as well. Assumption (C6) imposes restrictions on the estimator one may consider for the conditional distribution of the censoring variable and is, for instance, fulfilled for a simple Kaplan-Meier estimator for G_C (see e.g. Theorem 2.1 in Chen and Lo (1997) for sufficient and necessary conditions for (C6)). Lastly, assumption (C7) requires the uniform consistency of the bivariate copula density estimator. A crucial observation for the latter assumption is that the property only needs to hold for the fixed point of interest $\boldsymbol{x}$, hereby preventing from exploring the corners of the unit square for which typical copula densities, such as the Gaussian copula, grow unboundedly. We also note that a similar assumption and discussion is to be found in the recent work of Nagler and Vatter (2018), where the aim is to develop a general framework for the use of copulas in estimation equation in a similar spirit as in this chapter. Lastly, while to the best of our knowledge the result of assumption (C7) is not reported as such in the literature for the most common copula density estimators, we note that the latter may also be put in perspective with uniform consistency results for density functions that are defined on compact supports, provided the impact of handling here pseudo-observations instead of true observations does not alter the property. To investigate this perspective, we provide in Section 2.8 an additional result stating the uniform consistency of a simple illustrative kernel copula density estimator based on pseudo-observations with respect to the same estimator but with knowledge of the true marginal distributions. A more detailed discussion on this is deferred to Section 2.8.

We now state the main result of this section holding for a general d-dimensional covariate vector and for all bivariate kernel copula estimators based on the same bandwidth h. In practice, however, it may be recommended to adopt an unconstrained and non-diagonal bandwidth matrix, as is detailed in Section 4 of Geenens et al.. Nevertheless, when considering this general situation, the theoretical results become less tractable while equivalent in nature to this simplified situation.

Theorem 2.1. *Let $h \equiv h_n \to 0$ be the common bandwidth of the d bivariate kernel copula density estimators. For h satisfying $nh^2 \to \infty$ as $n \to \infty$, and under assumptions (C1)-(C7), we have*

$$\left(nh^2\right)^{1/2}\left(\hat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x})\right) =$$
$$\frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}\frac{\left(nh^2\right)^{1/2}}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})$$
$$\times\left[\widehat{c}^{\,u}_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))\right] + o_{\mathbb{P}}(1),$$

where $f_{T|\boldsymbol{X}}$ is the conditional density of T given $\boldsymbol{X}$ and $w(\boldsymbol{x}) = p(\boldsymbol{x})/[\mathbb{P}(\Delta = 1)c^u_{\boldsymbol{X}}(\boldsymbol{F}^u(\boldsymbol{x}))]$.

Theorem 2.1 implies, quite naturally, that the asymptotic behavior of $\hat{m}_\tau(\boldsymbol{x})$ will be characterized by the properties of the copula estimator, specifically through its nonparametric feature, provided that the estimation of $\widehat{G}_C(\cdot|\boldsymbol{x})$ is 'reasonable' when confronted to a multidimensional covariate vector (assumption (C6)). In particular, this suggests that the detailed discussion of Geenens et al. about the asymptotic bias and variance of their distinctive bivariate copula estimators may be transcribed in our context.

Additionally, Theorem 2.1 also covers an asymptotic representation of the copula-based quantile regression estimator when all responses are fully observed. In this situation, one would indeed obtain a similar result for the proposed semiparametric procedure, with the removal of all censoring related terms, that is $w(\boldsymbol{x})$, $W_i(\boldsymbol{x})$, $i = 1, \ldots, n$, and the superfluous conditioning on $\Delta = 1$ for the copula densities and marginal distributions. This results in the following:

$$\left(nh^2\right)^{1/2}\left(\hat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x})\right) =$$
$$\frac{1}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}\frac{\left(nh^2\right)^{1/2}}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)[\widehat{c}_{Y\boldsymbol{X}}(F_Y(Y_i), \boldsymbol{F}(\boldsymbol{x})) - c_{Y\boldsymbol{X}}(F_Y(Y_i), \boldsymbol{F}(\boldsymbol{x}))] + o_{\mathbb{P}}(1).$$

We now consider a detailed asymptotic representation of our estimator for the simplified case where $d = 2$, and for a general nonparametric estimator of the bivariate copula densities. For convenience, we use the notation c^u_k as a shorthand for $c^u_{YX_k}$, and similarly for other functions depending on (Y, Δ, X_k), $k = 1, 2$.

Corollary 2.2. *Suppose that the assumptions of Theorem 2.1 hold for the case $d = 2$. Furthermore, suppose that the bivariate nonparametric copula estimators of c^u_1 and c^u_2 are such that*

$$\left(nh^2\right)^{1/2}\left(\widehat{c}^{\,u}_k(u_0, u_k) - c^u_k(u_0, u_k) - h^2 b_k(u_0, u_k)\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z^{nj}_k(u_0, u_k) + o_{\mathbb{P}}(1),$$
$$\forall u_k \in (0,1),\ \textit{uniformly in } u_0 \in (0,1),\ \textit{for } k = 1, 2,$$

for some some deterministic function $b_k(u_0, u_k)$*, and for some function* $Z_k^{nj}(u_0, u_k)$ *depending on* (Y_j, Δ_j, X_{kj}) *and possibly on* n*, satisfying* $\mathbb{E}(Z_k^{nj}(u_0, u_k)) = 0$*, for all* $u_0, u_k \in (0, 1)$*.*

Define

$$\begin{aligned}
\widetilde{Z}^{nj}(u_0, \boldsymbol{u}) &= \left[Z_1^{nj}(u_0, u_1) c_2^u(u_0, u_2) + Z_2^{nj}(u_0, u_2) c_1^u(u_0, u_1) \right] \\
&\quad \times c^u_{X_1 X_2|Y}(u_1, u_2|u_0), \\
b_{Y\boldsymbol{X}}(u_0, \boldsymbol{u}) &= \left[b_1(u_0, u_1) c_2^u(u_0, u_2) + b_2(u_0, u_2) c_1^u(u_0, u_1) \right] \\
&\quad \times c^u_{X_1 X_2|Y}(u_1, u_2|u_0), \\
\lambda_n\left(Y_i, \Delta_i, \boldsymbol{X}_i, \boldsymbol{x}\right) &= \mathbb{E}\left[\psi_\tau(\epsilon) W(\boldsymbol{x}) \widetilde{Z}^{ni}\left(F^u(Y), \boldsymbol{F}^u(\boldsymbol{x})\right) | Y_i, \Delta_i, \boldsymbol{X}_i \right], i = 1, \ldots, n.
\end{aligned}$$

Suppose furthermore that the following technical conditions hold:

(C8) $\mathbb{E}\left[\psi_\tau(\epsilon) W(\boldsymbol{x}) b_{Y\boldsymbol{X}}\left(F^u(Y), \boldsymbol{F}^u(\boldsymbol{x})\right)\right]^2 < \infty$*.*

(C9) $\mathbb{E}\left[\psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) \widetilde{Z}^{nj}\left(F^u(Y_i), \boldsymbol{F}^u(\boldsymbol{x})\right) \right]^2 = o(n)$ *for all* $i, j = 1, \ldots, n$*, where the expectation is taken with respect to* (Y_i, Δ_i) *and* $(Y_j, \Delta_j, \boldsymbol{X}_j)$*.*

Then, the copula-based quantile regression estimator at any point of interest $\boldsymbol{x}$ *satisfying (C1)-(C9) is such that*

$$\left(nh^2\right)^{1/2} \left(\widehat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x}) - h^2 B(\boldsymbol{x}) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\boldsymbol{x})),$$

where

$$\begin{aligned}
B(\boldsymbol{x}) &= \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})} \mathbb{E}\Big(\psi_\tau(\epsilon) W(\boldsymbol{x}) b_{Y\boldsymbol{X}}(F^u(Y), \boldsymbol{F}^u(\boldsymbol{x})) \Big) \\
\text{and } \sigma^2(\boldsymbol{x}) &= \frac{w^2(\boldsymbol{x})}{f^2_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})} \lim_{n\to\infty} \mathbb{E}\left(\lambda_n(Y, \Delta, \boldsymbol{X}, \boldsymbol{x})^2\right).
\end{aligned}$$

Corollary 2.2 reports the asymptotic normality of our estimator at the expected convergence rate, implied by the nonparametric estimation of the bivariate copula densities. Depending on the choice in step (2) of the kernel density estimator fulfilling the conditions of Corollary 2.2, simple plug-in of the expression of Z_k^{nj}, $k = 1, 2$, in all quantities built upon the latter may then lead to the detailed, although arduous, expressions of the asymptotic bias and variance of the proposed estimator.

Furthermore, as this had yet to be covered, it is worth stressing out that Corollary 2.2 also encompasses the asymptotic normality of the suggested estimator based on semiparametric vine copulas with strictly complete data. Similarly to what has been stated for Theorem 2.1, one is indeed only required to withdraw all censoring related elements from Corollary 2.2 to obtain the expressions of the asymptotic bias and variance of the proposed semiparametric quantile regression estimator for complete observations.

2.5 Simulation Study

In this section, we assess the practical finite-sample performance of the proposed methodology by means of Monte Carlo simulations. For this purpose, we first present a brief numerical study to further motivate the semiparametric copula strategy we intend to adopt for multivariate problems for both complete and censored observations. Secondly, we assess the numerical performance of the copula-based regression estimator for complete observations using the developed copula estimation strategy and compare it with established semiparametric and nonparametric techniques in the literature. Lastly, focusing on survival data, we illustrate the performance of our estimator in (2.3.4) by also showing promising results with respect to competitors in the domain, including when the generated scenario is to the advantage of the latter. All the simulations are carried out using the statistical computing environment R (R Core Team (2017)) and its freely accessible packages.

2.5.1 Assessing the semiparametric copula estimation

This first section aims at numerically illustrating the choice of our semiparametric copula estimation strategy. For this purpose, we will consider here both complete and censored responses as this will result in an interesting analysis of our proposed strategy resulting from the fact that the censoring percentage will have an influence on the number of observations actually entering the copula estimation process. Overall, we consider two distinctive data generating processes (DGP) and compare our methodology with fully parametric and nonparametric procedures one might consider for the estimation of a multivariate copula density. For the general simulation settings, we consider $B = 500$ repetitions of each DGP; three (average) levels of censoring (0%, 30% and 50%), three sample sizes ($n \in \{100, 200, 400\}$) and the quantile level of interest $\tau = 0.3$. As the object of interest here is the copula modelling, when censoring is introduced, we only consider the simple case of independence between the censoring variable and the covariate vector in order to keep the estimation of $\widehat{G}_C$ needed for (2.3.4) uncomplicated, that is, using the Kaplan-Meier estimator reported in (1.3.1). The detailed DGPs are as follows:

- **DGP A**: $(F_T(T), F_1(X_1), F_2(X_2)) \sim$ Gaussian copula with parameters $(\rho_{T1}, \rho_{T2}, \rho_{12}) = (0.3, 0.9, 0.5)$. Given standard uniform marginal distributions for all three variables, the resulting true quantile regression may be calculated as $m_\tau(\boldsymbol{x}) = \Phi(-0.2\Phi^{-1}(x_1) + \Phi^{-1}(x_2) + 0.4\Phi^{-1}(\tau))$, where Φ stands for the standard normal c.d.f. (see Noh et al. (2015)). To include censoring, we introduce the variable $C \sim U[0, M]$, where the parameter M is computed in order to obtain the desired average censoring proportion ($M = 5/3$ for 30% and $M = 1$ for 50%).

- **DGP B**: $(F_T(T), F_1(X_1), F_2(X_2), F_3(X_3)) \sim$ Gaussian copula with parameters $(\rho_{T1}, \rho_{T2}, \rho_{T3}, \rho_{12}, \rho_{13}, \rho_{23}) = (0.3, 0.9, 0.7, 0.5, 0.25, 0.5)$. The resulting

true quantile regression for standard uniform marginal distributions is determined as $m_\tau(\boldsymbol{x}) = \Phi(-0.2\Phi^{-1}(x_1)+0.83\Phi^{-1}(x_2)+0.33\Phi^{-1}(x_2)+0.27\Phi^{-1}(\tau))$. The censoring variable is $C \sim U[0, M]$ ($M = 5/3$ for 30% and $M = 1$ for 50% censoring).

For any general copula-based regression estimator, the marginal distribution estimations are performed, as suggested in Section 2.2.3, using rescaled versions of the empirical distributions:

$$\widehat{F}_Y^u(y) = \frac{1}{n^u+1}\sum_{i=1}^{n}\Delta_i \mathbb{1}(Y_i \leq y),\ \widehat{F}_j^u(x_j) = \frac{1}{n^u+1}\sum_{i=1}^{n}\Delta_i \mathbb{1}(X_{ij} \leq x_j), j = 1, \ldots, d,$$

where $n^u = \sum_{i=1}^{n}\Delta_i$ is the number of uncensored observations.

For the distinctive copula-based estimators, we consider the following:

$\widehat{m}^{SP}_{cop,\,\tau}$: semiparametric estimation strategy detailed in Section 2.2.3. That is, we first estimate the d bivariate copulas of interest employing the probit transformation technique of Geenens et al. (2017) coupled with local likelihood estimation based on quadratic polynomials. To that end, we follow their proposed nearest-neighbor bandwidth selection procedure implemented in the R package `kdecopula`. Concerning the estimation of the d-dimensional copula density (2.2.5), as mentioned above, we apply standard vine techniques using the R package `VineCopula`. Specifically, when $d > 2$ we adopt one automatically selected tree structure for the simplified decomposition of the copula density among many R-vine candidate structures (see Dißmann et al. (2013)), and subsequently determine the appropriate pair-copula families to be selected and parametrically estimated. The selection criterion for bivariate copulas is chosen to be the Akaike information criterion (AIC), which revealed to be adequate in the R-vine context (see Brechmann (2010), chap. 5), and ten potential family candidates, together with their rotations, are considered: eight of them are Archimedian (Clayton, Gumbel, Frank, Joe, Clayton-Gumbel, Joe-Gumbel, Joe-Clayton and Joe-Frank), and the last two are elliptical (Gaussian and Student t).

$\widehat{m}^{NP}_{cop,\,\tau}$: fully nonparametric estimation of the d-dimensional copula using vine techniques. Specifically, while the vine structure is kept identical, here all bivariate building blocks are estimated using the local likelihood technique based on probit-projected data with the bandwidth selection procedure of Geenens et al. (as is studied in Nagler and Czado (2016)). Given its fully nonparametric nature, it should be mentioned that this estimator is not covered by the theoretical results of Section 2.4.

$\widehat{m}^{P}_{cop,\,\tau}$: fully parametric estimation of the d-dimensional copula density, where all bivariate copulas are estimated using the previously enumerated candidate

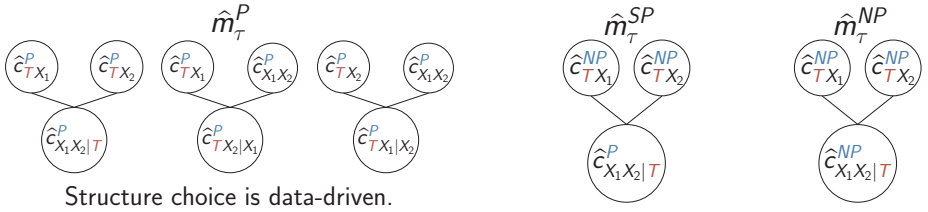

Figure 2.3: Graphical illustration of the three copula-based estimators considered for the situation $d = 2$ in DGP **A** and no censoring (hence no superscript 'u' in the copulas).

families and selection criteria. However, unlike the above-mentioned estimators, we do not force here any structure for the vine decomposition. As a consequence, no explicit distinction is imposed between dependence of interest and noisy dependence. Instead, one data-driven selected structure is adopted, regardless of the arguments of Section 2.2.3. This will allow us to analyse the impact of such dependence distinction in our regression context, as is discussed below. Finally, as it is the case for $\widehat{m}^{NP}_{cop,\,\tau}$, this estimator is not covered by the asymptotic theory of Section 2.4.

Figure 2.3 serves as a graphical support for the description of these copula-based estimators for the particular case of DGP **A** with $d = 2$. Every bivariate copula density is here represented by a node, and the distinction between estimators is highlighted to be originating either from the imposed structure of the decomposition, or from the choice of estimation procedure for each bivariate copula ('P' for parametric, 'NP' for nonparametric). The situation for DGP **B** with $d = 3$ is of course similar but involving more bivariate building blocks and hence more possible decomposition structures.

Now, both DGPs concentrate on the situation where the dependence structure between the response variable and the covariate vector is characterized by a parametric copula. In such circumstances, $\widehat{m}^{P}_{cop,\,\tau}$ will have a critical advantage and may serve in order to evaluate the impact of the nonparametric part of the estimation scheme, especially when the dimension of the covariate vector increases. As a performance criterion, we consider here the empirical integrated mean squared error (IMSE), defined as

$$IMSE(\widehat{m}_\tau(\boldsymbol{x})) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{B}\sum_{b=1}^{B}(\widehat{m}_\tau^{(b)}(\boldsymbol{x}_i) - m_\tau(\boldsymbol{x}_i))^2\right),$$

where $\{\boldsymbol{x}_i, i = 1, \ldots, N\}$ is a generated random sample of size $N = 10$ serving as an evaluation set spread on the domain of $\boldsymbol{X}$, and $\widehat{m}_\tau^{(b)}(\cdot)$ denotes the regression estimation for the b-th simulated sample.

The results of our simulation study are summarized in Table 2.1. Based on these, we detail our analysis in two parts, as the outcomes of our study offer

		DGP **A**					DGP **B**		
n	p_c	$\widehat{m}^{P}_{cop,\,\tau}$	$\widehat{m}^{SP}_{cop,\,\tau}$	$\widehat{m}^{NP}_{cop,\,\tau}$	n	p_c	$\widehat{m}^{P}_{cop,\,\tau}$	$\widehat{m}^{SP}_{cop,\,\tau}$	$\widehat{m}^{NP}_{cop,\,\tau}$
	0	1.442	1.486	2.369		0	1.192	1.416	3.307
100	0.3	2.203	2.173	3.167	100	0.3	2.036	2.459	3.972
	0.5	3.537	3.548	4.135		0.5	5.055	6.552	7.038
	0	0.689	0.737	1.462		0	0.560	0.658	2.609
200	0.3	0.990	1.016	1.905	200	0.3	0.850	0.945	2.857
	0.5	1.887	1.664	2.581		0.5	1.686	2.108	3.647
	0	0.337	0.371	0.987		0	0.259	0.354	2.140
400	0.3	0.503	0.525	1.243	400	0.3	0.403	0.487	2.266
	0.5	0.915	0.863	1.600		0.5	0.738	0.916	2.497

Table 2.1: Simulation results expressed in terms of $IMSE \times 1000$ for the estimation of $m_\tau(\boldsymbol{x})$ in DGP **A** and **B**. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{100, 200, 400\}$, average censoring proportions $p_c \in \{0, 0.3, 0.5\}$ and $\tau = 0.3$.

relevant information on both the copula decomposition choice and the type of bivariate estimators one may adopt in the multivariate setting. Note that, for both DGPs, as the dependence structure is specified by a Gaussian copula, the simplifying assumption intrinsic to the vine decomposition is here applicable (see Theorem 4 in Stöber et al. (2013)). In our context, this means that any observed difference between copula strategies is not to be attributed to a possible violation of the underlying simplifying assumption.

Focusing first on our decomposition strategy, as expected, we note that $\widehat{m}^{P}_{cop,\,\tau}$ globally outperforms $\widehat{m}^{SP}_{cop,\,\tau}$ and $\widehat{m}^{NP}_{cop,\,\tau}$. However, strikingly enough, this is not observed for DGP **A** for different censoring proportions where $\widehat{m}^{SP}_{cop,\,\tau}$ details better results. This is interpreted here as evidence for the validity of our arguments regarding the decomposition choice: as the censoring proportion grows, the number of observations actually entering the copula estimation becomes more moderate, hereby implying two opposite effects in this context. First, the propagation of estimation approximations tends to be more important, signifying that the further we decompose, the more sensitive becomes the estimation of the involved bivariate copulas as these are tributary of the quality of previously estimated bivariate blocks. Using a purely data-driven decomposition may then result in a poor fit of the (conditional) copula of the response variable with one of the covariates, as it is not required that the latter would be primarily treated. This is interpreted as the reason why $\widehat{m}^{SP}_{cop,\,\tau}$ is able to outperform $\widehat{m}^{P}_{cop,\,\tau}$, admittedly by a small amount, when censoring increases for a fixed sample size $n \in \{200, 400\}$, even though the simulated scenario is issued from a purely parametric copula. However, on the other hand, when observations are more scarce, it is well-known that nonparametric estimations become more sensitive than parametric counterparts. This explains why the estimation results for $\widehat{m}^{SP}_{cop,\,\tau}$ are not superior to those of $\widehat{m}^{P}_{cop,\,\tau}$ for $n = 100$ with 50% censoring, as the former require the nonparametric estimation of two

bivariate copulas, whose complexity compared to $\widehat{m}^{P}_{cop,\,\tau}$ seems to override the positive effects of our decomposition choice. Overall, these noteworthy results for DGP **A** illustrate the effectiveness of our proposed copula decomposition in the regression context. When augmenting the covariate vector dimension, the price of estimating now three nonparametric bivariate copulas quite logically exceeds the potential gain of concentrating efforts on the dependence of interest. This is identified in DGP **B**.

Concentrating now on the modelling choice for the noisy dependence, the comparison between $\widehat{m}^{SP}_{cop,\,\tau}$ and $\widehat{m}^{NP}_{cop,\,\tau}$ offers valuable information, particularly in DGP **A**, as the only distinctness here is the estimation of a unique bivariate copula density $c^{u}_{X_1X_2|Y}$. Visibly, the implication of keeping a nonparametric approach for this part seems to be rather severe. This finding clearly also applies when the dimension of the covariate grows.

Conclusively, the recommended copula modelling strategy seems to propose an adequate trade-off between preventing the serious effects of a purely nonparametric estimation and providing the flexibility needed to overcome possible shortcomings associated to a purely parametric alternative. With this judgment at hand, the next section is then devoted to the comparison of our procedure with alternative estimation techniques of the literature.

2.5.2 Comparison with estimation methods for complete responses

In this section, we briefly compare the global performance of the semiparametric copula-based regression estimator for complete observations with several established methodologies in the literature. To that end, we consider the following competing techniques:

$\widehat{m}_{LL,\,\tau}$: Locally polynomial quantile regression estimator studied in Chaudhuri (1991). In this section, we consider a local linear estimator which is implemented with Gaussian kernels in the function `lpqr` of the R package `quantreg`. The bandwidth is selected via leave-one-out cross validation in a set of candidates ranging from 0.05 to 2 by 0.05 increments.

$\widehat{m}_{inv,\,\tau}$: Nonparametric regression estimator studied in Li et al. (2013) based on an inverse-c.d.f. approach by numerically inverting a kernel estimated conditional distribution function. This procedure is implemented in the R package `np` via the function `npreg`. The bandwidth selection is automatically implemented in the function `npcdistbw` of the same library.

$\widehat{m}_{si,\,\tau}$: Single-index regression estimator based on a two stage estimation method as proposed by Zhu et al. (2012). The univariate nonparametric part of the methodology is estimated with a Gaussian kernel and the bandwidth is selected by leave-one-out cross validation in a set of candidates ranging from 0.05 to 0.5 by 0.05 increments.

$\widehat{m}^{P}_{cop,\tau}$: Semiparametric copula quantile regression estimator of Noh et al. (2015) where the copula density is estimated with purely parametric vine techniques as in Section 2.5.1.

$\widehat{m}_{cop,\tau}$: Proposed semiparametric copula quantile regression estimator with the same implementation as in Section 2.5.1. As a remark, we note that considering a Gaussian copula among the potential candidates, although natural in practice, introduces here a slight advantage in terms of information for our procedure and $\widehat{m}^{P}_{cop,\tau}$ given that the settings of this section will be considered with Gaussian distributions. Although our procedure is based on data-driven family selection, and hence does not automatically select the true copula family, a further interesting analysis would then be to consider the removal of the Gaussian copula among the potential candidates, hereby necessarily introducing some misspecification concerns. This is here however left for further investigation, while a similar and more detailed study on the impact of such copula misspecification may already be found in Noh et al. (2015).

These methodologies are compared in the following two DGPs:

- **DGP C**: $(F_T(T), F_1(X_1), F_2(X_2)) \sim$ Gaussian copula with parameters $(\rho_{T1}, \rho_{T2}, \rho_{12}) = (0.9, 0.8, -0.8)$. The marginal distributions are chosen to be standard exponential for T and standard normal for the covariates. For $\beta = (-13/18, 2/9)^{\mathsf{T}}$, the resulting quantile regression may then be determined as $m_\tau(\boldsymbol{x}) = F_T^{-1}\left[\Phi\left(\beta^{\mathsf{T}}\boldsymbol{x} + 0.41\Phi^{-1}(\tau)\right)\right]$.

- **DGP D**: Data simulated from the model

$$T_i = |2X_{1i} - X_{2i}^2 + 0.5|^{0.5} + 0.1X_{4i}^3 X_{5i}(0.5|X_{3i}| + 1) + 0.1\epsilon_i,$$

  where $\epsilon_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, n$, and the five-dimensional covariate $(X_1, \ldots, X_5)$ is simulated from a Gaussian distribution with correlations $(\rho_{12}, \rho_{13}, \ldots, \rho_{45}) = (0.3, 0.4, 0.5, 0.6, 0.7, 0.3, 0.4, 0.5,\ 0.6, 0.7)$ and standard marginals.

The first DGP is a low-dimensional setting where we expect the single-index and both copula-based estimators to perform well given that the true quantile regression is in fact a single-index model and that the data is simulated from a known copula. In opposition, the second DGP explores a higher dimension of the covariates and considers a setting with non-monotonic dependencies between the response and the covariates, for which no parametric copula in the literature is suited as illustrated in Section 2.2. Hence, in this setting $\widehat{m}^{P}_{cop,\tau}$ is expected to perform poorly and both $\widehat{m}_{cop,\tau}$ and $\widehat{m}_{si,\tau}$ are a priori no longer favored over the other methodologies.

To compare the performance of the different procedures, we consider 500 repetitions of each DGP, quantile levels $\tau \in \{0.3, 0.5, 0.7\}$ and sample sizes $n \in \{100, 200, 400\}$. Similarly to the previous section, we then consider the IMSE as

DGP	n	τ	$\widehat{m}_{LL,\tau}$	$\widehat{m}_{inv,\tau}$	$\widehat{m}_{si,\tau}$	$\widehat{m}^{P}_{cop,\tau}$	$\widehat{m}_{cop,\tau}$
C	100	0.3	0.121	0.122	0.104	0.096	0.090
		0.5	0.152	0.157	0.126	0.126	0.121
		0.7	0.212	0.241	0.193	0.180	0.176
	200	0.3	0.070	0.068	0.058	0.041	0.043
		0.5	0.090	0.083	0.073	0.053	0.057
		0.7	0.128	0.128	0.108	0.074	0.082
D	100	0.3	1.507	1.599	6.227	3.207	1.264
		0.5	1.482	1.506	2.587	2.074	0.983
		0.7	1.887	1.676	2.624	2.261	1.198
	400	0.3	0.822	0.674	2.020	3.110	0.852
		0.5	0.726	0.626	1.624	1.890	0.631
		0.7	1.187	0.765	2.170	1.986	0.830

Table 2.2: Simulation results expressed in terms of $IMSE \times 10$ for the estimation of $m_\tau(\boldsymbol{x})$ in DGPs **C** and **D** with respectively two and five covariates. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{100, 200, 400\}$ and with quantiles of interest $\tau \in \{0.3, 0.5, 0.7\}$.

a performance criterion calculated on a testing sample of 20 evaluation points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{20}\}$.

The simulation results of this brief comparative study are presented in Table 2.2. As expected, for the two-dimensional setting we observe that the copula-based and the single-index estimators perform better than the remaining methodologies, as the latter do not take profit of any underlying model assumption here. Furthermore, both copula-based estimators tend to outperform in this setting the single-index estimator for all considered sample sizes and quantile levels. Similarly to Section 2.5.1, we also note that $\widehat{m}_{cop,\tau}$ is here again able to outperform by a slight margin $\widehat{m}^{P}_{cop,\tau}$ for small sample sizes by focusing primarily on the estimation of the dependence of interest. Overall, this suggests that an appropriate semiparametric estimation of the copula density results in a performance of the regression estimator relatively similar to the estimator of Noh et al. when the latter is considered in its most advantageous situation.

Next, the results of DGP **D** numerically illustrate the gain in flexibility of the proposed methodology over $\widehat{m}^{P}_{cop,\tau}$. In this situation, the latter evidently misspecifies the regression model as can be observed from the fact that the IMSE does not tend to diminish much with an increasing sample size. In opposition, $\widehat{m}_{cop,\tau}$ still performs very competitively with respect to the literature, especially for smaller samples. This suggests here again that the copula estimation strategy proposed in this work offers the required flexibility to a copula-based approach for quantile regression. To conclude, this results in a very competitive estimation procedure with respect to the established semiparametric and nonparametric literature.

2.5.3 Comparison with estimation methods for censored responses

The objective of this last section is to provide a comparison study between the proposed copula-based methodology and existing competitors for survival data. On a general note, given that for multidimensional covariates an appropriate estimation of the conditional distribution $G_C(\cdot|\boldsymbol{x})$ for the weights $W(\boldsymbol{x})$ in (2.3.4) may be of crucial influence, we consider distinctive scenarios contrasting in the impact on $\widehat{G}_C(\cdot|\boldsymbol{x})$ in order to provide a sufficiently broad view on the performance of our methodology.

We examine three general simulation models, with $B = 500$ repetitions of each; two (average) levels of censoring (30% and 50%), two sample sizes ($n = 200$ and $n = 400$) and four values for the quantile level of interest ($\tau \in \{0.1, 0.3, 0.5, 0.7\}$). Specifically, we consider general data arising from a Cox regression model. Based on the hazard function, this prominent model for analyzing survival times specifies that $h(t|\boldsymbol{X}_i) = h_0(t)\exp(\beta^\mathsf{T}\boldsymbol{X}_i), i = 1, \ldots, n$, where $h_0(t)$ is the so-called baseline hazard function, as recalled in Chapter 1. In this instance, given a nonnegative time-to-event variable, simple algebraic manipulations show that the general conditional quantile regression can be written as $m_\tau(\boldsymbol{x}) = H_0^{-1}(-\log(1-\tau)\,\exp(-\beta^\mathsf{T}\boldsymbol{x}))$, where $H_0(t) \equiv \int_0^t h_0(s)\mathrm{d}s$. For every proposed setting, the baseline distribution is here set to be standard exponential and consequently, for a given vector $\boldsymbol{x}$ the τ-th conditional quantile function is given by

$$m_\tau(\boldsymbol{x}) = -\log(1-\tau)\,\exp(-\beta^\mathsf{T}\boldsymbol{x}).$$

The distinction between our simulation scenarios is to be found in the three covariate vectors and the dimension of the latter that are taken into account. The first two settings are chosen to be of moderate dimension with $d = 5$ while the third setting exhibits an example with $d = 8$. Focusing first on the case $d = 5$, we consider $\beta = (1, -3/4, 1/2, 1/4, -3/5)^\mathsf{T}$ and 5 covariates $(X_1, \ldots, X_5)$ simulated form a Gaussian copula with parameters $(\rho_{12}, \rho_{13}, \ldots, \rho_{45}) = (0.3, 0.4, 0.5, 0.6, 0.7, 0.3, 0.4, 0.5, 0.6, 0.7)$. To distinguish our scenarios when $d = 5$, we further consider two covariate vectors given by $\boldsymbol{X}^{(1)} = (X_1, X_2, \ldots, X_5)$ and $\boldsymbol{X}^{(2)} = (X_1, \exp(X_2), X_3, X_4, X_5)$, for which the resulting quantile regressions are both single-index models in $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, respectively. This will allow to highlight the performance of our procedure when its competing estimators are both in an ideal situation and a slightly altered version of the latter.

Lastly, for $d = 8$ we consider again a Gaussian copula for the initial covariate vector with parameters $\rho_{ij} = (-0.8)^{|i-j|}, i, j = 1, \ldots, 8$. The parameter vector of the Cox regression is $\beta = (1, -3/4, 1/2, 1/4, -3/5, 4/5, 3/5, -2/5)^\mathsf{T}$ and the covariate vector is $\boldsymbol{X}^{(3)} = (X_1, \exp(X_2), X_3, \ldots, X_7, X_8^2)$. The resulting model is hence a single-index and a Cox regression with respect to $\boldsymbol{X}^{(3)}$.

Model 1: Cox model and single-index model with $d = 5$

We first consider the simple case of data following a Cox regression model issued from covariate vector $\boldsymbol{X}^{(1)}$. The distributions and parameter values of the censoring variables define the following scenarios:

- **DGP E**: $C \sim \text{Exp}(\lambda_C)$, with λ_C independent of $\boldsymbol{X}^{(1)}$. To attain the desired average censoring proportions, we fix $\lambda_C = 0.464$ and $\lambda_C = 1.083$ (corresponding to approximately 30% and 50% censoring on average, respectively). The exact conditional censoring probability given $\boldsymbol{X}^{(1)} = \boldsymbol{x}$ is calculated as $\lambda_C/(\lambda_C + \exp(\beta^\mathsf{T}\boldsymbol{x}))$ and is hence a decreasing function of $\beta^\mathsf{T}\boldsymbol{x}$, making us conjecture better results for higher values of $\beta^\mathsf{T}\boldsymbol{x}$. Note that in this scenario, $G_C(\cdot|\boldsymbol{x})$ boils down to $G_C(\cdot)$, for which the Kaplan-Meier estimator is suited.
- **DGP F**: $C \sim \text{Exp}(\lambda_C(\boldsymbol{x}))$, with $\lambda_C(\boldsymbol{x}) = 3/7 \times \exp(\beta^\mathsf{T}\boldsymbol{x})$ and $\lambda_C(\boldsymbol{x}) = \exp(\beta^\mathsf{T}\boldsymbol{x})$ corresponding to 30% and 50% censoring, respectively. In this scenario, the conditional censoring probability is independent of $\boldsymbol{x}$, and an adequate estimation of $G_C(\cdot|\boldsymbol{x})$ is fulfilled with a Cox model hypothesis for the relationship between the covariates and the censoring variable.

For comparison purposes, we consider the following four estimators:

$\widehat{m}_{cox,\tau}$: parametric estimator exploiting the information related to the parametric Cox model setting. This estimator will serve as a reference for the ideal, yet unknown in practice, situation. Specifically, $\widehat{m}_{cox,\tau}(\boldsymbol{x}) = -\log(1-\tau)\exp(-\widehat{\beta}^\mathsf{T}\boldsymbol{x})/\widehat{\lambda}_{T_0}$, where $\widehat{\beta}$ is estimated by maximum partial likelihood, and $\widehat{\lambda}_{T_0}$ is the maximum likelihood estimator of the exponential baseline distribution.

$\widehat{m}_{si,\tau}$: Single-index regression estimator studied in Bücher et al. (2014), where the censoring distribution is supposed to be independent of $\boldsymbol{x}$. For the univariate nonparametric part of the estimation process, 10 different bandwidths are selected ($h \in \{0.05, 0.1, \ldots, 0.5\}$), and the optimal choice is performed using the described leave-one-out cross-validation procedure. Note that the quantile regression of interest $m_\tau(\boldsymbol{x})$ here is indeed a single-index model in $(X_1, X_2, \ldots, X_5)$. This should provide a critical advantage to the performance of $\widehat{m}_{si,\tau}$.

$\widehat{m}_{cop,\tau}^{(\perp\!\!\!\perp)}$: our copula-based estimator from estimation equation (2.3.4), where the conditional distribution $G_C(\cdot|\boldsymbol{x})$ is supposed to be independent of $\boldsymbol{x}$ and thereby estimated by the classical (unconditional) Kaplan-Meier estimator.

$\widehat{m}_{cop,\tau}^{(cox)}$: our copula-based estimator, where the relationship between the covariates and the censoring variable is supposed to follow a Cox regression model. Namely, $\widehat{G}_C(c|\boldsymbol{x})$ is estimated by $1 - \exp(-c\exp(\widehat{\beta}_C^\mathsf{T}\boldsymbol{x})/\widehat{\lambda}_{C_0})$, where $\widehat{\beta}_C$ is estimated by maximum partial likelihood, and $\widehat{\lambda}_{C_0}$ is the maximum likelihood estimator of the exponential baseline distribution.

Following the arguments of Section 2.2.3 and the results of the previous sections, both copula-based procedures are implemented employing a semiparametric estimation for the d-variate copula built on the aforementioned candidate families and selection criterions.

In order to compare the studied estimators' performance, we consider in this section an integrated version of the median absolute estimation error, that is

$$IMAE(\widehat{m}_\tau(\boldsymbol{x})) = \frac{1}{N}\sum_{i=1}^{N} \text{med}^{(B)}(|\widehat{m}_\tau(\boldsymbol{x}_i) - m_\tau(\boldsymbol{x}_i)|),$$

where $\widehat{m}_\tau$ is a generic estimator of m_τ, $\{\boldsymbol{x}_i, i = 1, \ldots, N\}$ is an evaluation set corresponding to a generated random sample of size $N = 10$, spread on the domain of $\boldsymbol{X}^{(1)}$, and $\text{med}^{(B)}$ denotes the median taken over all $B = 500$ simulations. The choice for this robust L_1-type of measure is motivated by the fact that the optimization routines involved in the single-index procedure may yield very unlikely results with a small probability, hereby strongly disadvantaging the estimator when considering a L_2-type of error measure. The same reasoning is underlying the determination of a robust dispersion measure on estimation errors, taken as the averaged interquartile range. More precisely, we choose $N^{-1}\sum_{i=1}^{N}(Q_3^{(B)}(|\widehat{m}_\tau(\boldsymbol{x}_i) - m_\tau(\boldsymbol{x}_i)|) - Q_1^{(B)}(|\widehat{m}_\tau(\boldsymbol{x}_i) - m_\tau(\boldsymbol{x}_i)|))$, where $Q_3^{(B)}$ and $Q_1^{(B)}$ stand, respectively, for the third and first quartiles taken over all simulations.

The results of our simulation study for this model are reported in Tables 2.3 and 2.4 for 30% and 50% of censoring, respectively. As expected, the Cox regression estimator $\widehat{m}_{cox,\tau}$, serving as a reference case here, outperforms the other estimators for both scenarios and every sample size, quantile level and censoring percentage. More interestingly, while the single-index estimator $\widehat{m}_{si,\tau}$ quite logically displays better overall results than our copula-based estimators, it is worth noticing that the difference seems to fade away when moving to higher quantile levels, especially for higher levels of censoring percentage. This is considered as a first encouraging result for the copula-based estimators as the simulated model is very strongly to the advantage of $\widehat{m}_{si,\tau}$. This indicates that our proposed procedure is flexible enough to compete with $\widehat{m}_{si,\tau}$ in its ideal setup when the number of observations actually entering the estimation scheme becomes moderate, that is, when censoring percentage is important and the quantile level of interest is high. Of course, if this is not the case, there is no a priori reason to believe that the copula-based estimators could outperform $\widehat{m}_{si,\tau}$ when the latter is considered in its optimal setting.

Figure 2.4 serves to illustrate these results for the particular case of DGP **E** with $n = 400$ and for high quantile levels ($\tau \in \{0.5, 0.7\}$). For the sake of brevity, we only report here a graphical comparison between the copula-based estimators and the single-index estimator, as $\widehat{m}_{cox,\tau}$ clearly outperforms the latter. As can be observed from Figure 2.4, the difference between $\widehat{m}_{si,\tau}$ and both copula-based estimators is graphically modest, especially for the highest quantile of interest $\tau = 0.7$, hereby reinforcing the previously discussed results of Tables 2.3 and 2.4. Furthermore,

Censoring = 30%

DGP	n	τ	$\widehat{m}_{cox,\tau}$	$\widehat{m}_{si,\tau}$	$\widehat{m}^{(\perp\!\!\!\perp)}_{cop,\tau}$	$\widehat{m}^{(cox)}_{cop,\tau}$
E	200	0.1	0.016 (0.019)	0.039 (0.053)	0.043 (0.059)	0.043 (0.057)
		0.3	0.054 (0.063)	0.083 (0.103)	0.101 (0.124)	0.102 (0.123)
		0.5	0.105 (0.123)	0.149 (0.185)	0.174 (0.199)	0.175 (0.194)
		0.7	0.183 (0.213)	0.280 (0.373)	0.273 (0.299)	0.279 (0.305)
	400	0.1	0.011 (0.013)	0.030 (0.038)	0.033 (0.044)	0.033 (0.043)
		0.3	0.036 (0.043)	0.070 (0.078)	0.079 (0.097)	0.081 (0.098)
		0.5	0.069 (0.084)	0.115 (0.140)	0.130 (0.160)	0.133 (0.159)
		0.7	0.120 (0.146)	0.221 (0.273)	0.212 (0.251)	0.212 (0.244)
F	200	0.1	0.016 (0.019)	0.037 (0.052)	0.043 (0.059)	0.043 (0.059)
		0.3	0.055 (0.064)	0.082 (0.105)	0.102 (0.129)	0.103 (0.124)
		0.5	0.108 (0.125)	0.152 (0.190)	0.174 (0.197)	0.174 (0.191)
		0.7	0.187 (0.218)	0.301 (0.474)	0.283 (0.310)	0.278 (0.298)
	400	0.1	0.011 (0.014)	0.029 (0.037)	0.035 (0.044)	0.034 (0.045)
		0.3	0.036 (0.046)	0.066 (0.077)	0.082 (0.102)	0.083 (0.098)
		0.5	0.071 (0.089)	0.118 (0.141)	0.138 (0.165)	0.135 (0.159)
		0.7	0.123 (0.155)	0.233 (0.287)	0.222 (0.259)	0.218 (0.241)

Table 2.3: Simulation results expressed in terms of $IMAE$ and the dispersion measure in brackets for the estimation of $m_\tau(\boldsymbol{x})$ with 30% of censoring where the true model is a Cox and single-index model. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{200, 400\}$ and with quantiles of interest $\tau \in \{0.1, 0.3, 0.5, 0.7\}$.

note that all estimators tend to present worse performances for low levels of $\beta^\intercal \boldsymbol{x}$. This is expected in this simulation setting as these levels correspond to the region of $\beta^\intercal \boldsymbol{x}$ where the exact conditional censoring probability is the highest.

Focusing again on Tables 2.3 and 2.4, we further note that, while there is logically a difference between the performances of $\widehat{m}^{(\perp\!\!\!\perp)}_{cop,\tau}$ and $\widehat{m}^{(cox)}_{cop,\tau}$ depending on the simulated scenario, the effect of an appropriate modelling of $\widehat{G}_C(\cdot|\boldsymbol{x})$ seems to be rather limited in this simulation setup for our estimators. Additionally, while both $\widehat{m}_{si,\tau}$ and $\widehat{m}^{(\perp\!\!\!\perp)}_{cop,\tau}$ rely here on the assumption of independence between the censoring variable and the covariate vector, the latter seems to numerically behave better when confronted to the violation of this hypothesis. This can be observed from the comparison between DGP **E** and **F** of the dispersion measures of both estimators, once again especially for high levels of quantile values and censoring percentages. Of course, one could argue that, given the results for the dispersion measure of $\widehat{m}_{si,\tau}$ for high quantiles, this can partly be due to a poor smoothing parameter choice. However, the latter was implemented using the proposed methodology of Bücher et al. (2014), just as a practitioner would have resolved to act.

Censoring = 50%

DGP	n	τ	$\widehat{m}_{cox,\tau}$	$\widehat{m}_{si,\tau}$	$\widehat{m}^{(\perp)}_{cop,\tau}$	$\widehat{m}^{(cox)}_{cop,\tau}$
E	200	0.1	0.019 (0.022)	0.040 (0.056)	0.044 (0.056)	0.044 (0.054)
		0.3	0.063 (0.074)	0.089 (0.111)	0.113 (0.127)	0.115 (0.127)
		0.5	0.123 (0.144)	0.176 (0.242)	0.201 (0.209)	0.205 (0.213)
		0.7	0.213 (0.251)	0.411 (0.402)	0.354 (0.326)	0.362 (0.326)
	400	0.1	0.012 (0.015)	0.031 (0.040)	0.035 (0.042)	0.035 (0.042)
		0.3	0.042 (0.051)	0.071 (0.082)	0.089 (0.106)	0.089 (0.106)
		0.5	0.081 (0.099)	0.136 (0.170)	0.161 (0.179)	0.163 (0.178)
		0.7	0.140 (0.171)	0.299 (0.315)	0.284 (0.290)	0.281 (0.279)
F	200	0.1	0.019 (0.022)	0.040 (0.061)	0.046 (0.059)	0.044 (0.053)
		0.3	0.064 (0.074)	0.090 (0.106)	0.117 (0.129)	0.115 (0.123)
		0.5	0.125 (0.143)	0.202 (0.374)	0.212 (0.218)	0.210 (0.206)
		0.7	0.217 (0.249)	0.526 (0.854)	0.380 (0.341)	0.380 (0.324)
	400	0.1	0.013 (0.015)	0.029 (0.038)	0.038 (0.048)	0.036 (0.045)
		0.3	0.043 (0.052)	0.069 (0.082)	0.099 (0.113)	0.095 (0.105)
		0.5	0.084 (0.102)	0.140 (0.173)	0.177 (0.201)	0.172 (0.174)
		0.7	0.145 (0.176)	0.346 (0.632)	0.306 (0.322)	0.298 (0.283)

Table 2.4: Simulation results expressed in terms of $IMAE$ and the dispersion measure in brackets for the estimation of $m_\tau(\boldsymbol{x})$ with 50% of censoring where the true model is a Cox and single-index model. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{200, 400\}$ and with quantiles of interest $\tau \in \{0.1, 0.3, 0.5, 0.7\}$.

Models 2 & 3: Alteration of a Cox model and single-index model with $d = 5$ and $d = 8$

To pursue our simulation study, we now consider in this section the second and third models, where the time-to-event data is simulated using covariate vectors $\boldsymbol{X}^{(2)}$ and $\boldsymbol{X}^{(3)}$, respectively. Consequently, the resulting quantile regressions are no longer a single-index nor a Cox regression model in $(X_1, X_2, \ldots, X_5)$ for model 2 and $(X_1, X_2, \ldots, X_8)$ for model 3. For comparison purposes, we consider the four estimation procedures described in model 1, given covariate vectors $(X_1, X_2, \ldots, X_5)$ and $(X_1, X_2, \ldots, X_5)$, respectively, hereby implying a slight model bias for both $\widehat{m}_{si,\tau}$ and $\widehat{m}_{cox,\tau}$ as the latter would here require transformations of the covariates. In opposition, both copula-based procedures, which are here constructed using the same previously-described semiparametric modelling strategy, automatically take account of such concerns.

For the sake of brevity, we only consider here the situation where the censoring variable is independent from the covariate vector, as the impact of a dependent scheme has been treated in model 1. Specifically, for model 2 we simulate the censoring variable from an exponential distribution with parameter values 0.208 and 0.486 for approximately 30% and 50% censoring. Concerning model 3, we only report the results for 30% censoring which are attained by simulating here a

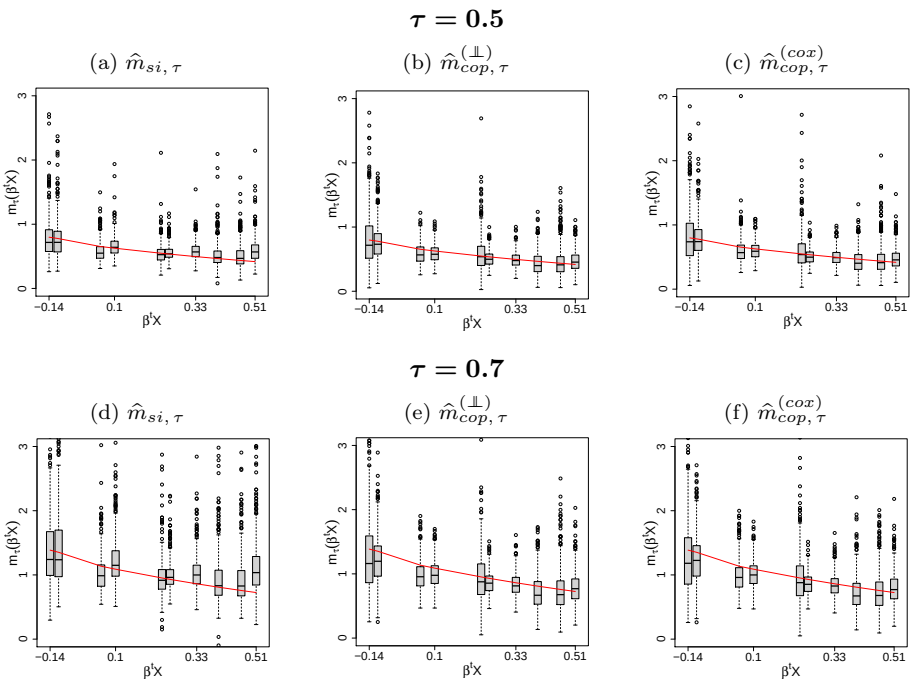

Figure 2.4: Boxplots corresponding to the results in Tables 2.3 and 2.4 of three estimators considered for $m_\tau(\boldsymbol{x})$ for $N = 10$ points spread on the domain of $\beta^\intercal \boldsymbol{X}^{(1)}$ in DGP **E**, and for $\tau \in \{0.5, 0.7\}$. In this setting, the true model is a single-index model and hence favors $\widehat{m}_{si,\tau}$ over both copula-based estimators. All pictures are based on 500 repetitions for $n = 400$ and 30% censoring. The red lines represent the true value of $m_{0.5}(\boldsymbol{x})$ and $m_{0.7}(\boldsymbol{x})$.

censoring variable following an exponential distribution with parameter 0.903.

The resulting performance of the considered estimators are depicted in Tables 2.5 and 2.6, once again in terms of $IMAE$. In Table 2.5, reporting the case $d = 5$, we observe that both copula-based estimators tend to outperform their competitors, with the exception of very low quantiles of interest where the effect of a ('small') misspecification of the underlying model seems to be moderate for both $\widehat{m}_{cox,\tau}$ and $\widehat{m}_{si,\tau}$. As we move away from these low quantile levels, the consequences of misspecifying the model become more severe for the competing estimators, notably for a purely parametric approach ($\widehat{m}_{cox,\tau}$). In contrast, the copula-based approach presents satisfactory results for varying censoring proportions and sample sizes, especially when keeping in mind that the simulated scenario is 'only' a slightly altered version of the ideal scenario for its competitors.

Model 2

n	p_c	τ	$\widehat{m}_{cox,\tau}$	$\widehat{m}_{si,\tau}$	$\widehat{m}^{(\perp)}_{cop,\tau}$	$\widehat{m}^{(cox)}_{cop,\tau}$
200	0.3	0.1	0.115 (0.046)	0.105 (0.142)	0.115 (0.160)	0.115 (0.159)
		0.3	0.391 (0.156)	0.279 (0.273)	0.249 (0.285)	0.250 (0.286)
		0.5	0.760 (0.302)	0.489 (0.538)	0.452 (0.532)	0.454 (0.538)
		0.7	1.319 (0.525)	0.988 (1.023)	0.781 (0.828)	0.783 (0.842)
	0.5	0.1	0.116 (0.054)	0.107 (0.156)	0.115 (0.139)	0.116 (0.143)
		0.3	0.391 (0.182)	0.334 (0.324)	0.303 (0.335)	0.311 (0.338)
		0.5	0.760 (0.359)	0.611 (0.742)	0.574 (0.552)	0.583 (0.558)
		0.7	1.321 (0.615)	1.441 (1.142)	1.053 (0.845)	1.074 (0.865)
400	0.3	0.1	0.115 (0.029)	0.084 (0.101)	0.092 (0.118)	0.092 (0.120)
		0.3	0.388 (0.099)	0.219 (0.222)	0.211 (0.264)	0.212 (0.265)
		0.5	0.753 (0.192)	0.372 (0.401)	0.358 (0.427)	0.359 (0.430)
		0.7	1.308 (0.333)	0.685 (0.844)	0.598 (0.682)	0.604 (0.683)
	0.5	0.1	0.114 (0.035)	0.084 (0.105)	0.096 (0.109)	0.096 (0.112)
		0.3	0.387 (0.118)	0.254 (0.231)	0.221 (0.283)	0.226 (0.285)
		0.5	0.751 (0.228)	0.479 (0.490)	0.431 (0.486)	0.438 (0.488)
		0.7	1.305 (0.397)	0.933 (0.878)	0.850 (0.781)	0.857 (0.793)

Table 2.5: Simulation results expressed in terms of $IMAE$ and the dispersion measure in brackets for the estimation of $m_\tau(\boldsymbol{x})$ for model 2. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{200, 400\}$, censoring proportion $p_c \in \{0.3, 0.5\}$ and with four levels of quantile of interest $\tau \in \{0.1, 0.3, 0.5, 0.7\}$.

Similar findings apply when analyzing the results of Table 2.6 for $d = 8$. In this situation, both copula-based estimator still globally outperform their competitors, although the margin of improvement is relatively smaller than for $d = 5$. This results from the fact that we have here to model a nine-dimensional copula density with no information on the underlying model, while both $\widehat{m}_{si,\tau}$ and $\widehat{m}_{si,\tau}$ rely on model assumptions that are still relatively close to the simulated model. Conclusively, this advocates, here again, for the flexibility of our procedure and its withstanding to misspecification of the underlying model for multidimensional problems when comparing with other semiparametric or fully parametric modelling techniques.

2.6 Real Data Application

We present in this section a brief application of our procedure for censored data by analysing the Colorado Plateau uranium miners cohort data (see e.g. Lubin et al. (1995), Langholz and Goldstein (1996)). The object of the study, for which 3347 Caucasian male miners having worked at least a month in the uranium mines of the Colorado Plateau were followed, is to investigate the risk of lung cancer related to smoking and radon exposure. Hence, the event of interest is defined as

Model 3

p_c	n	τ	$\widehat{m}_{cox,\,\tau}$	$\widehat{m}_{si,\,\tau}$	$\widehat{m}^{(\perp)}_{cop,\,\tau}$	$\widehat{m}^{(cox)}_{cop,\,\tau}$
0.3	200	0.1	0.031 (0.054)	0.037 (0.049)	0.029 (0.037)	0.031 (0.044)
		0.3	0.084 (0.182)	0.124 (0.174)	0.075 (0.080)	0.078 (0.091)
		0.5	0.149 (0.354)	0.241 (0.451)	0.139 (0.139)	0.142 (0.151)
		0.7	0.273 (0.616)	0.419 (0.822)	0.240 (0.222)	0.247 (0.236)
	400	0.1	0.116 (0.035)	0.028 (0.038)	0.023 (0.028)	0.024 (0.030)
		0.3	0.067 (0.118)	0.094 (0.099)	0.062 (0.066)	0.062 (0.069)
		0.5	0.104 (0.230)	0.183 (0.344)	0.114 (0.116)	0.115 (0.122)
		0.7	0.220 (0.400)	0.317 (0.652)	0.194 (0.190)	0.194 (0.197)

Table 2.6: Simulation results expressed in terms of $IMAE$ and the dispersion measure in brackets for the estimation of $m_\tau(\boldsymbol{x})$ for model 3. The number of repetitions operated is $B = 500$ for sample sizes $n \in \{200, 400\}$, censoring proportion $p_c = 0.3$ and with four levels of quantile of interest $\tau \in \{0.1, 0.3, 0.5, 0.7\}$.

the time till lung cancer death (expressed as the logarithm of number of years), which affected a total of 258 miners. Besides failure time, the study also includes information about age at entry to the study, cumulative smoking (in number of packs) and radon exposure (in working level month (WLM)).

As the original data set is prone to heavy censoring (92.3%), and given the illustrative nature of this section, we first define a subsample on which the analysis will be performed. To that end, in order to preserve as best as possible the nature of the population at risk, we decide to define a threshold on the radon exposure above which observations will enter the subsample. The value of the threshold is practically chosen as a trade-off between censoring proportion and actual number of observations that are to be selected. Specifically, we find that by defining a threshold of 2831 WLM on radon exposure, a subsample of 176 observations is constituted, 55 of which were subject to the event of interest. Scatterplots of the selected data are represented in Figure 2.5, where X_1 is the age at entry into the study, X_2 is the cumulative radon exposure and X_3 is the cumulative smoking.

Regarding the data analysis, endorsing the role of a practitioner, we are faced with the choice of an appropriate estimator for the application of quantile regression in this context. We therefore consider the following distinctive candidates:

$\widehat{m}^{(cox)}_{cop,\,\tau}$: Copula-based quantile regression estimator, with Cox regression modelling for $G_C(\cdot|\cdot)$.

$\widehat{m}_{si,\,\tau}$: Single-Index methodology of Bücher et al. (2014).

$\widehat{m}_{cox,\,\tau}$: Semiparametric estimator based on the Cox proportional hazards model. Specifically, $\widehat{m}_{cox,\tau}(\boldsymbol{x}) = \widehat{H}_0^{-1}(-\log(1-\tau)\exp(-\widehat{\beta}^{\mathsf{T}}\boldsymbol{x}))$, where $\widehat{\beta}$ is estimated by maximum partial likelihood, and $\widehat{H}_0$ is the Nelson-Aalen-type estimator of the cumulative baseline hazard.

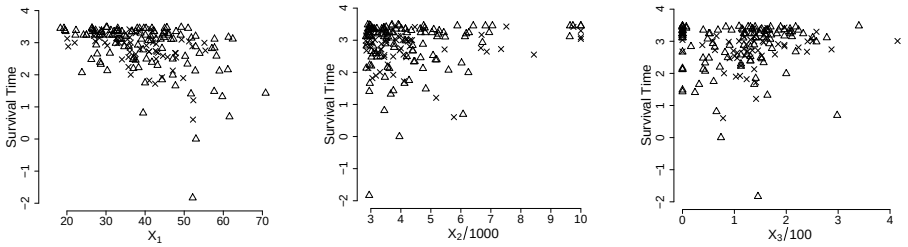

Figure 2.5: Colorado Plateau uranium miners cohort data. Scatter plots of the survival time versus each covariate. Uncensored data points are given by $\times$, while censored observations are represented by $\triangle$.

As a general evaluation measure to compare models for the present data set, we consider the median quantile loss from predicting the τ-th conditional quantile of T for the uncensored observations. In other words, we use the following cross-validated prediction error criterion:

$$\mathrm{PE}\left(\widehat{m}_\tau\right) = \mathrm{med}_{\substack{1\leq i\leq n \\ \Delta_i=1}} \rho_\tau\left(Y_i - \widehat{m}_\tau^{-i}(\boldsymbol{X}_i)\right),$$

where $\widehat{m}_\tau^{-i}$ denotes any estimator of m_τ based on all observations except the i-th.

For the implementation of the copula-based regression estimator, we adopt the methodology of Section 2.2.3 and our simulation study by opting for a semiparametric modelling of the four-variate copula density, where the bivariate copulas of the response variable with each covariate are estimated using the procedure of Geenens et al. (2017) with quadratic polynomials along with the proposed data-driven bandwidth selection scheme. The remaining trivariate noisy copula is, afterwards, estimated using vine techniques with the same candidate families and selection criterion as in Section 2.5. Additionally, a general appropriate hypothesis has to be made for the modelling of the conditional distribution $G_C(\cdot|\cdot)$. As it is shown in the literature that independence is unsuitable for the data of the study (see e.g. Leng and Tong (2013)), we propose to model the conditional distribution through a semiparametric Cox regression for the censoring time with respect to the covariates. Therefore, as is common in practice, the parameter of the regression is estimated by maximum partial likelihood while the Breslow estimator is used for computing the baseline survivor function.

Concerning $\widehat{m}_{si,\tau}$, the bandwidth choice is performed using the proposed leave-one-out cross-validation procedure on normalized covariates with 15 candidates ($h \in \{0.1, 0.2, \ldots, 1.5\}$). Note that, as the latter procedure is computationally costly, it is here assumed that the selected bandwidth is constant for all datasets entering the estimation of $\mathrm{PE}(\widehat{m}_{si,\tau})$, which is of little impact in this context given that only one observation at the time is to be removed from each dataset.

τ	$\mathrm{PE}\left(\widehat{m}^{(cox)}_{cop,\,\tau}\right)$	$\mathrm{PE}\left(\widehat{m}_{si,\,\tau}\right)$	$\mathrm{PE}\left(\widehat{m}_{cox,\,\tau}\right)$
0.1	0.348	0.848	0.946
0.3	0.774	1.471	1.180
0.5	0.785	1.901	1.169
0.7	0.470	2.838	0.967

Table 2.7: Colorado Plateau uranium miners cohort data. Prediction error multiplied by 10 for each censored quantile regression estimator for quantile levels $\tau \in \{0.1, 0.3, 0.5, 0.7\}$.

The results of the evaluation measure are reported in Table 2.7 for four quantile levels of interest $\tau \in \{0.1, 0.3, 0.5, 0.7\}$. It is observed that, in terms of cross-validated prediction error, our copula-based approach depicts quite confidently the best performance for every considered quantile level of interest. By comparison, the single-index structure seems to be inappropriate as such for the studied data set, and should possibly require further work and attention on, for instance, transformations of covariates. Furthermore, its sensitivity to higher quantile levels is, here again, highlighted. Lastly, despite being part of every practitioner's toolbox for survival analysis, the proportional hazards regression estimator is also relatively largely outperformed by the increased flexibility of the copula-based approach. Hence, this example clearly illustrates the ability of our estimator to adapt to the underlying regression structure of the data.

Conclusively, recalling that, from equation (2.3.5), our procedure enjoys the additional valuable property of being automatically monotonic across quantile levels, this real data application plainly highlights the relevance of our estimator for flexible analyses of multivariate censored data.

2.7 Conclusion

In this work we have proposed a semiparametric copula-based quantile regression estimator in the context of potentially right-censored responses. On a general note, for data with or without censoring and motivated by the regression context, a novel semiparametric estimation approach for the implied copula density was studied. Furthermore, in parallel to the procedure of Noh et al. (2015), the proposed regression estimator in this work is obtained as a weighted quantile of the observed response variable, hereby opening the door to the practical use of the quantile regression code developed by Portnoy and Koenker (1997) and Koenker (2005). Asymptotic normality of the resulting estimator for both complete and censored data was obtained with convergence rate determined by the nonparametric estimation of bivariate copula densities. For inferential purposes however, bootstrap procedures are to be preferred as the applicability of the obtained asymptotic normality is restrained by numerous unknown quantities to be estimated in practice.

This bootstrap procedure could be inspired by the wild bootstrap procedure proposed by Feng et al. (2011). Finally, supporting the objective of this work, an extensive simulation study and a real data application have been carried out to illustrate both the validity of the semiparametric copula estimation and the increased flexibility of our procedure in comparison with existing alternatives, especially when no a priori knowledge about the functional form of the quantile regression is available in practice.

2.8 Proofs

We develop in this section the proofs of Section 2.4. To that end, as previously commented in the text, we start by developing a useful proposition related to condition (C7)-(i) where the uniform consistency of the bivariate copula density estimators is required for a fixed $\boldsymbol{x}$. Since the latter assumption is, to the best of our knowledge, not reported as such in the copula literature, we consider here a proposition stating the uniform convergence of a kernel copula density estimator based on the pseudo-observations $(\widehat{F}_T(T_i), \widehat{F}_j(X_{ji}))$, $i = 1, \ldots, n$, for some $j \in \{1, \ldots, d\}$, to the same estimator but constructed with the 'true' copula observations $(F_T(T_i), F_j(X_{ji}))$. This proposition may then be combined with more classical results concerning kernel estimators for density functions defined on compact supports in order for one to retrieve some higher-level conditions, depending on the choice of copula density estimators, that will imply the verification of condition (C7)-(i). Since the latter results are more classical, we only concentrate here on establishing the former in order to ensure that the particularity of handling pseudo-observations for copula density estimation will not alter the desired uniform consistency property.

Now, for ease of presentation, we consider here the example of the bivariate copula density c_{TX_1}, estimated by the simplest kernel density estimator reported in Section 1.2.2 with a product kernel and bandwidth $h \equiv h_n$. Specifically, we define the following estimator based on pseudo-observations:

$$\widehat{c}_{n,TX_1}(u_0, u_1) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{u_0 - \widehat{F}_T(T_i)}{h}\right) K\left(\frac{u_1 - \widehat{F}_1(X_{1i})}{h}\right),$$

where $\widehat{F}_T$ and $\widehat{F}_1$ satisfy (C5). We also define the counterpart of $\widehat{c}_{n,\,TX_1}$ that is based on the 'true' copula observations:

$$\widehat{c}_{TX_1}(u_0, u_1) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{u_0 - F_T(T_i)}{h}\right) K\left(\frac{u_1 - F_1(X_{1i})}{h}\right).$$

Our intention is then to illustrate in the next proposition the negligibility of estimating appropriately the marginals for uniform consistency results when considering a fixed u_1 and the supremum for u_0 over the whole unit interval. As previously mentioned, this result may then be put in perspective for condition (C7)-(i) with

more classical results on kernel estimation for density functions that are defined on bounded domains.

Proposition 2.1. *Consider the fixed point $u_1 = F_1(x_1) \in (0,1)$ with x_1 satisfying (C2) such that $\sup_{u_0 \in [0,1]} c_{TX_1}(u_0, u_1) < \infty$. Let K be a kernel density function of bounded support, bounded variation, and satisfying $\sup_u K(u) < \infty$. For $h \equiv h_n \to 0$ satisfying $nh^2 \to \infty$ when $n \to \infty$, and $c_{TX_1}(u_0, u_1)$ a continuously differentiable function with respect to u_0 on $[0,1]$ for fixed u_1, we have*

$$\sup_{u_0 \in [0,1]} |\widehat{c}_{n,TX_1}(u_0, u_1) - \widehat{c}_{TX_1}(u_0, u_1)| = o_{\mathbb{P}}(1).$$

Proof of Proposition 2.1

Observe first that for every fixed $(u_0, u_1) \in [0,1]^2$, under the assumptions for K, there exist random variables $\widetilde{F}_T(T_i)$ and $\widetilde{F}_1(X_{1i})$, $i = 1, \ldots, n$, such that

$$\begin{aligned}
&\widehat{c}_{n,TX_1}(u_0, u_1) - \widehat{c}_{TX_1}(u_0, u_1) \\
&\quad = -\frac{1}{nh^3} \sum_{i=1}^{n} K' \left(\frac{u_0 - \widetilde{F}_T(T_i)}{h} \right) K \left(\frac{u_1 - \widehat{F}_1(X_{1i})}{h} \right) (\widehat{F}_T(T_i) - F_T(T_i)) \\
&\qquad - \frac{1}{nh^3} \sum_{i=1}^{n} K \left(\frac{u_0 - F_T(T_i)}{h} \right) K' \left(\frac{u_1 - \widetilde{F}_1(X_{1i})}{h} \right) (\widehat{F}_1(X_{1i}) - F_1(X_{1i})) \\
&\quad = A_1(u_0, u_1) + A_2(u_0, u_1).
\end{aligned}$$

We will now focus on establishing that $\sup_{u_0 \in [0,1]} |A_1(u_0, u_1)| \to 0$ when $n \to \infty$, while the development for $A_2(u_0, u_1)$ is equivalent and therefore omitted here.

Now, without loss of generality, suppose that the finite support of K is $[-1, 1]$. We then have that

$$\begin{aligned}
|A_1(u_0, u_1)| \leq \frac{C}{nh^3} \sup_{t \in \mathbb{R}} \left| \widehat{F}_T(t) - F_T(t) \right| \\
\times \sum_{i=1}^{n} \mathbb{1} \left(|u_0 - \widetilde{F}_T(T_i)| \leq h, |u_1 - \widehat{F}_1(X_{1i})| \leq h \right),
\end{aligned}$$

for some finite constant C under our kernel assumptions. By further noting that $\widetilde{F}_T(T_i)$ is between $\widehat{F}_T(T_i)$ and $F_T(T_i)$ for $i = 1, \ldots, n$, we next have that

$$\begin{aligned}
|u_0 - \widetilde{F}_T(T_i)| &\geq |u_0 - F_T(T_i)| - |\widetilde{F}_T(T_i) - F_T(T_i)| \\
&\geq |u_0 - F_T(T_i)| - \sup_{t \in \mathbb{R}} |\widehat{F}_T(t) - F_T(t)|.
\end{aligned}$$

With a similar development for $|u_0 - \widehat{F}_1(X_{1i})|$, we then observe that

$$|A_1(u_0,u_1)| \leq \frac{C}{nh^3} \sup_{t\in\mathbb{R}} \left|\widehat{F}_T(t) - F_T(t)\right| \times \sum_{i=1}^{n} \mathbb{1}\Big(|u_0 - F_T(T_i)| \leq h + \sup_{t\in\mathbb{R}} |\widehat{F}_T(t) - F_T(t)|, \; |u_1 - F_1(X_{1i})| \leq h + \sup_{x_1\in\mathbb{R}} |\widehat{F}_1(x_1) - F_1(x_1)|\Big).$$

Let us now fix some $\epsilon > 0$, and consider δ_n such that $n^{1/2}\delta_n \to \infty$ and $\delta_n/h \to 0$, where the dependance of h on n should be kept in mind although not explicitly exposed in the notations for ease of reading. Notice that such conditions on δ_n are possible if $nh^2 \to \infty$. We may then write,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{u_0\in[0,1]} |A_1(u_0,u_1)| > \epsilon\right) \\
&= \mathbb{P}\left(\sup_{u_0\in[0,1]} |A_1(u_0,u_1)| > \epsilon, \sup_{t\in\mathbb{R}} |\widehat{F}_T(t) - F_T(t)| \leq \delta_n, \sup_{x_1\in\mathbb{R}} |\widehat{F}_1(x_1) - F_1(x_1)| \leq \delta_n\right) \\
&+ \mathbb{P}\left(\sup_{u_0\in[0,1]} |A_1(u_0,u_1)| > \epsilon, \left[\sup_{t\in\mathbb{R}} |\widehat{F}_T(t) - F_T(t)| > \delta_n \text{ or } \sup_{x_1\in\mathbb{R}} |\widehat{F}_1(x_1) - F_1(x_1)| > \delta_n\right]\right).
\end{aligned}$$

From this, we may then report that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{u_0\in[0,1]} |A_1(u_0,u_1)| > \epsilon\Big) \\
&\quad \leq \mathbb{P}\left(\left[\frac{C}{nh^3} \sup_{u_0\in[0,1]} \sum_{i=1}^{n} \mathbb{1}\Big(|u_0 - F_T(T_i)| \leq 2h, |u_1 - F_1(X_{1i})| \leq 2h\Big)\right] > \frac{\epsilon}{\delta_n}\right) \\
&\qquad + \mathbb{P}\left(\sup_{t\in\mathbb{R}} |\widehat{F}_T(t) - F_T(t)| > \delta_n\right) + \mathbb{P}\left(\sup_{x_1\in\mathbb{R}} |\widehat{F}_1(x_1) - F_1(x_1)| > \delta_n\right) \\
&\quad = A_{11} + A_{12} + A_{13}.
\end{aligned}$$

With assumption (C5), we first observe that both A_{12} and A_{13} converge to 0 when $n \to \infty$, since, for A_{12}, $\sup_{t\in\mathbb{R}} |\widehat{F}_T(t) - F_T(t)| = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(\delta_n)$, and similarly with $\widehat{F}_1$ for A_{13}. It then only remains to establish that $A_{11} \to 0$ as well in order to obtain the desired result.

To that end, let us define the following estimator of the bivariate density of $(F_T(T), F_1(X_1))$ with uniform kernel and bandwidth $2h$:

$$\widetilde{c}_{TX_1}(u_0,u_1) = \frac{1}{nh^2} \sum_{i=1}^{n} \mathbb{1}\left(|u_0 - F_T(T_i)| \leq 2h, |u_1 - F_1(X_{1i})| \leq 2h\right).$$

Using a classical result in the literature for the latter type of estimator (see e.g. Silverman (1978)) stating that $\sup_{u_0\in[0,1]}|\widetilde{c}_{TX_1}(u_0,u_1)-c_{TX_1}(u_0,u_1)|=o_{\mathbb{P}}(1)$ under appropriate and common smoothness conditions on c_{TX_1} as a function of u_0 for the fixed u_1, we then eventually have that

$$A_{11}\leq\mathbb{P}\left(\left[\frac{C}{h}\sup_{u_0\in[0,1]}|\widetilde{c}_{TX_1}(u_0,u_1)-c_{TX_1}(u_0,u_1)|+\frac{C}{h}\sup_{u_0\in[0,1]}|c_{TX_1}(u_0,u_1)|\right]>\frac{\epsilon}{\delta_n}\right)$$
$$\leq\mathbb{P}\left(\sup_{u_0\in[0,1]}|\widetilde{c}_{TX_1}(u_0,u_1)-c_{TX_1}(u_0,u_1)|>\frac{\epsilon h}{C\delta_n}-\sup_{u_0\in[0,1]}|c_{TX_1}(u_0,u_1)|\right).$$

Under condition (C2), we have that $\sup_{u_0\in[0,1]}|c_{TX_1}(u_0,u_1)|<\infty$, while the conditions on δ_n ensure that $(\epsilon h)/(C\delta_n)\to\infty$. We thus have that $A_{11}\to 0$ for $n\to\infty$, which concludes the proof. □

We may now report hereafter the proofs of the main results of Section 2.4, Theorem 2.1 and Corollary 2.2.

Proof of Theorem 2.1

Define first

$$\widehat{A}_n(s)=\sum_{i=1}^{n}\left[\rho_\tau(\epsilon_i-s/a_n)-\rho_\tau(\epsilon_i)\right]\widehat{W}_i(\boldsymbol{x})\,\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i),\widehat{\boldsymbol{F}}^u(\boldsymbol{x})),$$

with $\epsilon_i\equiv\epsilon_i(\boldsymbol{x},\tau)=Y_i-m_\tau(\boldsymbol{x})$, and $a_n=\sqrt{nh^2}$. Observe then that, by definition of $\widehat{m}_\tau(\boldsymbol{x})$,

$$a_n(\hat{m}_\tau(\boldsymbol{x})-m_\tau(\boldsymbol{x}))=\arg\min_s\widehat{A}_n(s).$$

Furthermore, given that ρ_τ is a convex function and that $\widehat{W}_i(\boldsymbol{x})\,\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i),\widehat{\boldsymbol{F}}^u(\boldsymbol{x}))$ is positive for all $i=1,\ldots,n$, we have that $s\mapsto\widehat{A}_n(s)$ is convex. The idea is then to develop an expression of $\widehat{A}_n(s)$ leading to the application of the quadratic approximation Lemma of convex functions (Basic Corollary in Hjort and Pollard (1993)). To that end, we have that (Knight's (1998) identity)

$$\rho_\tau(u-v)-\rho_\tau(u)=-v\psi_\tau(u)+R(u,v),$$

with $\psi_\tau(u)=\tau-\mathbb{1}(u\leq 0)$,

$$R(u,v)=\int_0^v(\mathbb{1}(u\leq s)-\mathbb{1}(u\leq 0))\,\mathrm{d}s=(u-v)(\mathbb{1}(u\leq 0)-\mathbb{1}(u\leq v)),$$

and $0\leq R(u,v)\leq|v|$. Hence, we may write

$$\widehat{A}_n(s)=-s\widehat{A}_{1n}+\hat{A}_{2n}(s),$$

with

$$\widehat{A}_{1n} = a_n^{-1} \sum_{i=1}^{n} \psi_\tau(\epsilon_i)\, \widehat{W}_i(\boldsymbol{x})\, \widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})),$$

and

$$\widehat{A}_{2n}(s) = \sum_{i=1}^{n} R(\epsilon_i, s/a_n)\, \widehat{W}_i(\boldsymbol{x})\, \widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})).$$

Focusing first on $\widehat{A}_{2n}(s)$, we will show that

$$\widehat{A}_{2n}(s) = \frac{s^2}{2} w^{-1}(\boldsymbol{x}) f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})\, \frac{n}{a_n^2} + o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right), \tag{A.1}$$

where $f_{T|\boldsymbol{X}}$ is the conditional density of T given $\boldsymbol{X}$ and $w(\boldsymbol{x}) = p(\boldsymbol{x})/[\mathbb{P}(\Delta = 1) c^u_{\boldsymbol{X}}(\boldsymbol{F}^u(\boldsymbol{x}))]$. To that end, we write

$$\widehat{A}_{2n}(s) = A_{2n}(s) + \Delta_{1n}(s) + \Delta_{2n}(s),$$

where

$$\begin{aligned}
A_{2n}(s) &= \sum_{i=1}^{n} R(\epsilon_i, s/a_n)\, W_i(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) \\
\Delta_{1n}(s) &= \sum_{i=1}^{n} R(\epsilon_i, s/a_n)\, (\widehat{W}_i(\boldsymbol{x}) - W_i(\boldsymbol{x}))\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) \\
\Delta_{2n}(s) &= \sum_{i=1}^{n} R(\epsilon_i, s/a_n)\, \widehat{W}_i(\boldsymbol{x})\, [\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))].
\end{aligned}$$

Concentrating on each term, we will show that

$$A_{2n}(s) = \frac{s^2}{2} w^{-1}(\boldsymbol{x}) f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})\, \frac{n}{a_n^2} + o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right) \tag{A.2}$$

$$\Delta_{1n}(s) = o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right) \tag{A.3}$$

$$\Delta_{2n}(s) = o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right). \tag{A.4}$$

For the proof of (A.2), we first establish that

$$\begin{aligned}
\mathbb{E}(A_{2n}(s)) &= n\, \mathbb{E}\left(R(\epsilon_1, s/a_n)\, W_1(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_1), \boldsymbol{F}^u(\boldsymbol{x}))\right) \\
&= n\, w^{-1}(\boldsymbol{x}) \int_0^{s/a_n} (F_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x}) + t|\boldsymbol{x}) - F_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x}))\, \mathrm{d}t,
\end{aligned}$$

where $F_{T|\boldsymbol{X}}$ denotes, as in Section 2.2, the conditional c.d.f. of T given $\boldsymbol{X}$, and where, for the second equality, we used (see Section 2.3) the fact that, for

any measurable function $\varphi : \mathbb{R} \to \mathbb{R}$, we have $\mathbb{E}(\Delta\varphi(Y)\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y), \boldsymbol{F}^u(\boldsymbol{x}))) = w^{-1}(\boldsymbol{x})\mathbb{E}(\Delta\varphi(Y)|\boldsymbol{X} = \boldsymbol{x})$. Next, using the mean-value theorem, we write

$$\begin{aligned} F_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x}) + t|\boldsymbol{x}) - F_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x}) &= t\, f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x}) + \theta t|\boldsymbol{x}) \\ &= t\, f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x}) + t\, R(t), \end{aligned}$$

for some $\theta \in (0, 1)$, with $R(t) = f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x}) + \theta t|\boldsymbol{x}) - f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})$. This yields

$$\mathbb{E}(A_{2n}(s)) = \frac{s^2}{2} w^{-1}(\boldsymbol{x}) f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})\, \frac{n}{a_n^2} + n\, w^{-1}(\boldsymbol{x}) \int_0^{s/a_n} t\, R(t)\mathrm{d}t.$$

From the fact that, under the required bandwidth condition and assumption (C1),

$$\left| \int_0^{s/a_n} t\, R(t)\mathrm{d}t \right| \leq \frac{s^2}{2\, a_n^2} \sup_{|t| \leq |s|/a_n} |R(t)| = o(1/a_n^2),$$

we get that

$$\mathbb{E}(A_{2n}(s)) = \frac{s^2}{2} w^{-1}(\boldsymbol{x}) f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})\, \frac{n}{a_n^2} + o\left(\frac{n}{a_n^2}\right),$$

provided assumption (C2) is satisfied. To conclude the proof of (A.2), it is then sufficient to show that

$$\mathbb{V}\mathrm{ar}(A_{2n}(s)) = o\left(\frac{n}{a_n^2}\right)^2.$$

To that end, observe that, for n sufficiently large, and under assumptions (C1),(C2) and (C3),

$$\begin{aligned} \mathbb{V}\mathrm{ar}(A_{2n}(s)) &\leq n\, \mathbb{E}\left(R(\epsilon_1, s/a_n)\, W_1(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_1), \boldsymbol{F}^u(\boldsymbol{x}))\right)^2 \\ &\leq n\, \mathbb{E}\left(R(\epsilon_1, s/a_n)\, W_1(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_1), \boldsymbol{F}^u(\boldsymbol{x}))\right) \frac{|s|}{a_n} \\ &\quad \times (1 - G_C(m_\tau(\boldsymbol{x}) + \delta)|\boldsymbol{x})^{-1} \sup_{t \in \mathbb{R}} c^u_{Y\boldsymbol{X}}(F^u_Y(t), \boldsymbol{F}^u(\boldsymbol{x})), \text{ for some } \delta > 0 \\ &= O\left(\frac{n}{a_n^3}\right) = o\left(\frac{n}{a_n^2}\right)^2, \end{aligned}$$

as a_n/n converges to 0.

Concentrating now on the proof of (A.3), observe that, for n sufficiently large

$$\begin{aligned} |\Delta_{1n}(s)| &= \left| \sum_{i=1}^n R(\epsilon_i, s/a_n)\, W_i(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))\, \frac{\widehat{G}_C(Y_i - |\boldsymbol{x}) - G_C(Y_i - |\boldsymbol{x})}{1 - \widehat{G}_C(Y_i - |\boldsymbol{x})} \right| \\ &\leq A_{2n}(s) \sup_{t \leq \max_{\substack{1 \leq i \leq n \\ \Delta_i = 1}} Y_i} \frac{|\widehat{G}_C(t|\boldsymbol{x}) - G_C(t|\boldsymbol{x})|}{1 - \widehat{G}_C(t|\boldsymbol{x})} \\ &= o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right), \end{aligned}$$

provided that $\widehat{G}_C(\cdot|\boldsymbol{x})$ satisfies assumption (C6).

Lastly, for the proof of (A.4), note that, for n sufficiently large,

$$\begin{aligned}
|\Delta_{2n}(s)| &= \left| \sum_{i=1}^{n} R(\epsilon_i, s/a_n)\, \widehat{W}_i(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) \right. \\
&\qquad \left. \times \frac{\widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))}{c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))} \right| \\
&\leq (A_{2n}(s) + \Delta_{1n}(s)) \sup_{t\in\mathbb{R}} \frac{|\widehat{c}^u_{Y\boldsymbol{X}}(\widehat{F}^u_Y(t), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(t), \boldsymbol{F}^u(\boldsymbol{x}))|}{c^u_{Y\boldsymbol{X}}(F^u_Y(t), \boldsymbol{F}^u(\boldsymbol{x}))} \\
&= o_{\mathbb{P}}\left(\frac{n}{a_n^2}\right),
\end{aligned}$$

provided that assumption (C7)-(i) is satisfied.

Hence, reassembling (A.2), (A.3) and (A.4) yields

$$\frac{a_n^2}{n}\widehat{A}_n(s) = -s\frac{a_n^2}{n}\widehat{A}_{1n} + \frac{s^2}{2} w^{-1}(\boldsymbol{x}) f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x}) + o_{\mathbb{P}}(1).$$

Therefore, by the quadratic approximation Lemma of a convex function, see e.g. Hjort and Pollard (1993), if $\frac{a_n^2}{n}\hat{A}_{1n} = O_{\mathbb{P}}(1)$, then

$$\begin{aligned}
a_n(\widehat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x})) &= \arg\min_s \frac{a_n^2}{n}\widehat{A}_n(s) \\
&= \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})} \frac{a_n^2}{n}\widehat{A}_{1n} + o_{\mathbb{P}}(1). \qquad \text{(A.5)}
\end{aligned}$$

As a consequence, the asymptotic behavior of our estimator will be driven by the asymptotic expression of $\frac{a_n^2}{n}\widehat{A}_{1n}$. Developing the expression of the latter, we will show that

$$\begin{aligned}
\frac{a_n^2}{n}\widehat{A}_{1n} = \frac{a_n}{n}\sum_{i=1}^{n} &\psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) \\
&\times [\widehat{c}^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))] + o_{\mathbb{P}}(1). \qquad \text{(A.6)}
\end{aligned}$$

To that end, note that $\frac{a_n^2}{n}\widehat{A}_{1n}$ may be decomposed as

$$\begin{aligned}
\frac{a_n^2}{n}\hat{A}_{1n} = \frac{a_n}{n}\sum_{i=1}^{n} &\psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) [\widehat{c}^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))] \\
&+ \Delta_{3n} + \Delta_{4n} + \Delta_{5n} + \Delta_{6n},
\end{aligned}$$

where

$$\Delta_{3n} = \frac{a_n}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)(\widehat{W}_i(\boldsymbol{x}) - W_i(\boldsymbol{x}))\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) \tag{A.7}$$

$$\Delta_{4n} = \frac{a_n}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})[\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - \widehat{c}^{\,u}_{Y\boldsymbol{X}}(F^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x}))] \tag{A.8}$$

$$\Delta_{5n} = \frac{a_n}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})[\widehat{c}^{\,u}_{Y\boldsymbol{X}}(F^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - \widehat{c}^{\,u}_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))] \tag{A.9}$$

$$\Delta_{6n} = \frac{a_n}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})). \tag{A.10}$$

We will show that all these quantities converge to 0 in probability. Starting with Δ_{3n}, we have, for a large n,

$$\begin{aligned}
|\Delta_{3n}| &= \left|\frac{a_n}{n}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x}))\frac{\widehat{G}_C(Y_i - |\boldsymbol{x}) - G_C(Y_i - |\boldsymbol{x})}{1 - \widehat{G}_C(Y_i - |\boldsymbol{x})}\right| \\
&\leq a_n \sup_{t\in\mathbb{R}} \widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(t), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) \sup_{t\leq \max_{\substack{1\leq i\leq n \\ \Delta_i=1}} Y_i} \frac{\left|\widehat{G}_C(t|\boldsymbol{x}) - G_C(t|\boldsymbol{x})\right|}{1 - \widehat{G}_C(t|\boldsymbol{x})} \\
&\quad \times \frac{1}{n}\sum_{i=1}^{n}|\psi_\tau(\epsilon_i)|W_i(\boldsymbol{x}) \\
&= O_{\mathbb{P}}(a_n) \sup_{t\leq \tau_{F_Y}} \frac{\left|\widehat{G}_C(t|\boldsymbol{x}) - G_C(t|\boldsymbol{x})\right|}{1 - \widehat{G}_C(t|\boldsymbol{x})} = o_{\mathbb{P}}(1),
\end{aligned}$$

under the condition that assumptions (C2), (C4)-(i), (C6) and (C7)-(i) are met.

The proofs of (A.8) and (A.9) are very similar in spirit. Hence, for the sake of brevity, we only consider here (A.8). For any $i = 1, \ldots, n$, there exists a $\theta_i \in (0, 1)$ such that,

$$\begin{aligned}
\widehat{c}^{\,u}_{Y\boldsymbol{X}}(\widehat{F}^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) - \widehat{c}^{\,u}_{Y\boldsymbol{X}}(F^u_Y(Y_i), \widehat{\boldsymbol{F}}^u(\boldsymbol{x})) &= (\widehat{F}^u_Y(Y_i) - F^u_Y(Y_i))\times \\
&\quad \partial_1 \widehat{c}^{\,u}_{Y\boldsymbol{X}}\Big(F^u_Y(Y_i) + \theta_i(\widehat{F}^u_Y(Y_i) - F^u_Y(Y_i)),\, \widehat{\boldsymbol{F}}^u(\boldsymbol{x})\Big),
\end{aligned}$$

where ∂_j denotes the partial derivative with respect to the j-th argument. Therefore, for a large n and under the bandwidth requirement,

$$\begin{aligned}
|\Delta_{4n}| &\leq a_n \sup_{t\in\mathbb{R}}|\widehat{F}^u_Y(t) - F^u_Y(t)| \sup_{u_0\,\in\,(0,1)} \sup_{\boldsymbol{u}\in\mathcal{V}_{\boldsymbol{F}^u(\boldsymbol{x})}} |\partial_1\, \widehat{c}^{\,u}_{Y\boldsymbol{X}}(u_0, \boldsymbol{u})| \frac{1}{n}\sum_{i=1}^{n}|\psi_\tau(\epsilon_i)|W_i(\boldsymbol{x}) \\
&= o_{\mathbb{P}}(1),
\end{aligned}$$

provided assumptions (C4)-(i), (C5), and (C7)-(ii) are satisfied. Similarly, $\Delta_{5n} = o_{\mathbb{P}}(1)$.

Lastly, $\Delta_{6n} = o_{\mathbb{P}}(1)$, since

$$\mathbb{E}(\psi_\tau(\epsilon)\, W(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y), \boldsymbol{F}^u(\boldsymbol{x}))) = 0,$$

and, by assumption (C4)-(ii),

$$\frac{1}{n}\sum_{i=1}^{n} \psi_\tau(\epsilon_i)\, W_i(\boldsymbol{x})\, c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) = O_{\mathbb{P}}(n^{-1/2}).$$

Finally, inserting (A.6) in (A.5), we conclude that

$$\begin{aligned} &a_n(\hat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x})) = \\ &\frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}\, \frac{a_n}{n}\sum_{i=1}^{n} \psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) \\ &\qquad\qquad \times [\widehat{c}^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x})) - c^u_{Y\boldsymbol{X}}(F^u_Y(Y_i), \boldsymbol{F}^u(\boldsymbol{x}))] + o_{\mathbb{P}}(1), \end{aligned}$$

which completes the proof. □

Proof of Corollary 2.2

Given that, by the result of Theorem 2.1, the asymptotic behavior of $\widehat{m}_\tau(\boldsymbol{x})$ will be dictated by the expression of the multivariate copula estimator, we will concentrate on the latter. Using the asymptotic expressions of both bivariate copulas along with the multivariate copula decomposition of Section 2.2.3, under the required bandwidth conditions we have that,

$$\begin{aligned} a_n\left(\widehat{c}^u_{Y\boldsymbol{X}}(u_0, \boldsymbol{u}) - c^u_{Y\boldsymbol{X}}(u_0, \boldsymbol{u}) - h^2 b_{Y\boldsymbol{X}}(u_0, \boldsymbol{u})\right) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \widetilde{Z}^{nj}(u_0, \boldsymbol{u}) + o_{\mathbb{P}}(1), \\ &\forall \boldsymbol{u} \in (0,1)^2, \text{ uniformly in } u_0 \in (0,1), \end{aligned}$$

where

$$\begin{aligned} \widetilde{Z}^{nj}(u_0, \boldsymbol{u}) &= \left(Z^{nj}_1(u_0, u_1) c^u_2(u_0, u_2) + Z^{nj}_2(u_0, u_2) c^u_1(u_0, u_1)\right) c^u_{X_1X_2|Y}(u_1, u_2|u_0) \\ b_{Y\boldsymbol{X}}(u_0, \boldsymbol{u}) &= [b_1(u_0, u_1) c^u_2(u_0, u_2) + b_2(u_0, u_2) c^u_1(u_0, u_1)]\, c^u_{X_1X_2|Y}(u_1, u_2|u_0). \end{aligned}$$

Therefore, using the result of Theorem 2.1, we may determine that

$$\begin{aligned} &a_n\Big(\widehat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x}) - h^2 \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})} n^{-1} \sum_{i=1}^{n} \psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) b_{Y\boldsymbol{X}}\left(F^u(Y_i), \boldsymbol{F}^u(\boldsymbol{x})\right)\Big) \\ &= \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})} \sqrt{n} \frac{1}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \psi_\tau(\epsilon_i) W_i(\boldsymbol{x}) \widetilde{Z}^{nj}\left(F^u(Y_i), \boldsymbol{F}^u(\boldsymbol{x})\right) + o_{\mathbb{P}}(1). \end{aligned}$$

Let us define

$$B(\boldsymbol{x}) = \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}\mathbb{E}\left(\psi_\tau(\epsilon)W(\boldsymbol{x})b_{Y\boldsymbol{X}}\left(F^u(Y),\boldsymbol{F}^u(\boldsymbol{x})\right)\right).$$

Provided that assumption (C8) holds, we have

$$\frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}n^{-1}\sum_{i=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})b_{Y\boldsymbol{X}}\left(F^u(Y_i),\boldsymbol{F}^u(\boldsymbol{x})\right) = B(\boldsymbol{x}) + O_{\mathbb{P}}(n^{-1/2}),$$

from which it results that

$$\begin{aligned} &a_n\left(\widehat{m}_\tau(\boldsymbol{x}) - m_\tau(\boldsymbol{x}) - h^2B(\boldsymbol{x})\right) \\ &= \frac{w(\boldsymbol{x})}{f_{T|\boldsymbol{X}}(m_\tau(\boldsymbol{x})|\boldsymbol{x})}\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})\widetilde{Z}^{nj}\left(F^u(Y_i),\boldsymbol{F}^u(\boldsymbol{x})\right) + o_{\mathbb{P}}(1). \end{aligned} \tag{A.11}$$

The last step of the proof is then to simplify the obtained expression on the right hand side of the last equality. To that end, define

$$V_n = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})\widetilde{Z}^{nj}\left(F^u(Y_i),\boldsymbol{F}^u(\boldsymbol{x})\right),$$

and observe that this is a V-statistic with the symmetric kernel

$$\begin{aligned} \Psi_n(V_i,V_j) = \frac{1}{2}\Big[&\psi_\tau(\epsilon_i)W_i(\boldsymbol{x})\widetilde{Z}^{nj}\left(F^u(Y_i),\boldsymbol{F}^u(\boldsymbol{x})\right) \\ &+ \psi_\tau(\epsilon_j)W_j(\boldsymbol{x})\widetilde{Z}^{ni}\left(F^u(Y_j),\boldsymbol{F}^u(\boldsymbol{x})\right)\Big], \end{aligned}$$

where $V_t = (Y_t, \boldsymbol{X}_t, \Delta_t), t = i, j$. As the statistic's kernel depends on n, this suggests to apply Corollary 1 in Martins-Filho and Yao (2006) which establishes the $\sqrt{n}$-equivalence between the V-statistic and the Hájek-projection of its corresponding U-statistic (see e.g. Serfling (1980), page 189). Therefore, as $\mathbb{E}\left(Z_k^{nj}(u_0,u_k)\right) = 0$ for all u_0, u_k, $k = 1, 2$, and $\mathbb{E}\left(\Psi_n(V_i,V_j)\right) = 0$, we have

$$V_n = n^{-1}\sum_{i=1}^{n}\lambda_n\left(Y_i,\Delta_i,\boldsymbol{X}_i,\boldsymbol{x}\right) + o_{\mathbb{P}}(n^{-1/2}), \tag{A.12}$$

where $\lambda_n\left(Y_i,\Delta_i,\boldsymbol{X}_i,\boldsymbol{x}\right) = \mathbb{E}\Big[\psi_\tau(\epsilon)W(\boldsymbol{x})\widetilde{Z}^{ni}\left(F^u(Y),\boldsymbol{F}^u(\boldsymbol{x})\right)|Y_i,\Delta_i,\boldsymbol{X}_i\Big]$, provided that assumption (C9) holds. Lastly, the result of Corollary 2.2 follows readily from the insertion of (A.12) in (A.11) and the application of Lyapunov's central limit theorem. □

CHAPTER 3

An Adapted Loss Function for Censored Quantile Regression

Abstract

In this chapter, we study a novel approach for the estimation of quantiles when facing potential right censoring of the responses. Contrary to the existing literature on the subject reported in Chapter 1, the adopted strategy is here to tackle censoring at the very level of the check loss function usually employed for the computation of quantiles. For interpretation purposes, a simple comparison with the latter reveals how censoring is accounted for in the newly proposed loss function. Subsequently, when considering the inclusion of covariates for conditional quantile estimation, by defining a new general loss function the proposed methodology opens the gate to numerous parametric, semiparametric and nonparametric modelling techniques. In order to illustrate this statement, we consider here the well-studied linear regression under the usual assumption of conditional independence between the true response and the censoring variable. For practical minimization of the studied loss function, we also provide a simple algorithmic procedure shown to yield satisfactory results for the proposed estimator with respect to the existing literature in an extensive simulation study. From a more theoretical prospect, consistency and asymptotic normality of the estimator for linear regression are obtained using several recent results on non-smooth semiparametric estimation equations with an infinite-dimensional nuisance parameter, while numerical examples illustrate the adequateness of a simple bootstrap procedure for inferential purposes. Lastly, an application to a real dataset is used to further illustrate the validity and finite sample performance of the proposed estimator. This chapter is based on the paper

De Backer, M., A. El Ghouch, and I. Van Keilegom. *An adapted loss function for censored quantile regression.* Journal of the American Statistical Association. To appear.

3.1 Introduction

In this chapter, we introduce a reconsideration, following Chapter 2, of the general handling of censored data in quantile regression models. Hence, as a preliminary remark, contrary to chapters 2 and 4 that include results for complete observations as well, the work reported here focuses exclusively on situations where censoring of the response may arise. In this context, our review of the literature reported in Chapter 1 suggests the existence of a certain duality between two prominent weighting schemes when it comes to check-based formulations of conditional quantiles with censored data. We thus start this section by briefly recalling the latter in order to position our work in this framework, and then discuss in more details the existing literature for the particular case of a linear regression, given that our general methodology will be applied to the latter model in the following of this chapter.

Now, as reviewed in Chapter 1, the main rationale of the current quantile regression literature for the handling of censored data may be rallied under the general idea of formulating appropriate weighting schemes, of which the ICP (Section 1.4.1) and RM weights (Section 1.4.2) are undoubtedly most popular. As illustrated in the previous chapter with a copula-based regression model, ICP weights in the context of quantile regression have the great convenience of being easily employed when considering general regression models. However, in situations where alternative strategies are available for the controlling of censored data, ICP-based modelling is often pointed out for its efficiency loss in the estimation of the regression function, given that the calculated quantile loss of only uncensored observations is kept in the procedure (see (1.4.3)). In contrast, the RM-based weighting scheme fully incorporates the quantile context in its handling of censored data, but is on the other hand restricted by construction to the parametric front of regression modelling, as detailed in Section 1.4.4. In light of these shortcomings for both the ICP and RM techniques, this chapter is devoted to the development of a general procedure that attempts at bridging the qualities of both techniques, that is, proposing a unified framework for censored quantile regression that is specifically designed for quantiles in its incorporation of censored data, all the while being easily adapted to general regression models. To illustrate the latter objective, this chapter then considers the well-studied linear regression, for which a quick review of the literature is detailed next for convenience.

The earliest work on linear censored quantile regression may be traced back to the econometric papers of Powell (1984; 1986) with the particular case of fixed censoring, where it is assumed that the censoring variable is always observable. For random right censoring, as it is mostly the case in survival analysis and as is considered throughout this thesis, both the ICP and RM techniques engendered a fruitful literature stemming from the check-based formulation of conditional quantiles. The introduction of RM weights in this context is due to the work of Portnoy (2003), where, as recalled in Chapter 1, the underlying idea is to redis-

tribute the mass of censored data lying under the quantile of interest to artificial outlying observations to the right, as the contribution of each point to the estimation of the quantile regression only depends on the sign of the residual. However, given that the weights to be redistributed are to be determined by the conditional distribution of the survival time given the covariates, Portnoy's estimation scheme was developed under a restrictive global linearity assumption to simplify the procedure, that is, assuming that the regression function is linear for all quantiles levels in $(0, \tau)$, where τ is the quantile level of interest. Relaxing this assumption by only assuming linearity at the level of interest, and exploiting the same idea of redistribution-of-mass, Wang and Wang (2009) proposed a locally weighted estimation scheme based on Beran's local Kaplan-Meier estimator (see (1.3.3)) for the conditional distribution of the variable of interest instead of Portnoy's global assumption. More recently, Wey et al. (2014) proposed to exploit, in the same methodology, survival trees instead of Beran's estimator for the required conditional distribution estimation in an attempt to apprehend better multivariate and categorical covariates. Overall, the redistribution-of-mass methodology proved to be a valuable approach for parametric censored quantile regression and lead to further research such as variable selection (Wang et al. (2013)), multiple quantile estimation (Tang and Wang (2015)) and cure rate quantile regression (Wu and Yin (2016)). As for the ICP weighting scheme, the latter technique was adapted to the linear quantile regression through different versions by Ying et al. (1995), Bang and Tsiatis (2002), Shows et al. (2010) and Leng and Tong (2013) to cite a few. These procedures may be linked in spirit to the notorious work of Koul et al. (1981) and Zhou (1992) on mean regression models, and are often illustrated to be numerically outperformed in the parametric context by their RM-based counterparts.

As previously evoked, in this chapter we aim at providing a new insight into censored quantile regression by tackling the problem of censoring through an alternative strategy to the above-discussed literature. In fact, instead of handling weighting schemes or substitute quantile levels as in the paper of Lindgren (1997) recalled in Chapter 1, our suggestion is to account for censoring at the very level of the check loss function used in quantile regression. Intuitively, when confronted to right-censored datasets, the proposed loss function will be observed to penalize more severely underestimation of the value of the regression than the usual check function. The methodology then allows to plainly exploit the information of all the observations at hand, hence avoiding any estimation efficiency loss in contrast with the ICP weights. However, as a possible inconvenience, the studied loss function is observed to be no longer convex, which then requires the proposal of an efficient algorithmic procedure to appropriately minimize the resulting objective function. In this work, a simple adjustment of the Majorize-Minimize (MM) algorithm as proposed by Hunter and Lange (2000) is therefore examined. Apart from this, it is worth stressing out that, by concentrating on a censored version of the check function, the proposed strategy has the potential to engender multiple extensions to the linear context considered here. The latter could for instance concern pa-

rametric models and variable selection in penalized regression, but one could also easily accommodate the strategy to nonparametric and semiparametric models. In short, handling censoring at the level of the loss function allows both to take the quantile context into account in the consideration of censored observations and to define a general framework for censored regression modelling. Therefore, the overall objective of this work is to illustrate this novel general estimation strategy in a simple and well-studied linear context in order to analyze its behavior in comparison with the existing literature, and hence guide any potential further research on the advantages and pitfalls of this strategy.

The rest of this chapter is organized as follows. Motivating the origin of the newly proposed loss function first for sample quantiles, and then illustrating its accommodation for linear regression is the topic of Section 3.2. Investigating further the linear context, consistency and asymptotic normality of the proposed estimator are obtained in Section 3.3. Section 3.4 provides a detailed adaptation of an MM algorithm to practically implement the procedure, and the finite sample performance of the latter is illustrated by means of Monte Carlo simulations in Section 3.5. Section 3.6 highlights a brief application to real data. Lastly, Section 3.7 briefly concludes this chapter while the proofs of Section 3.3 are deferred to Section 3.8.

3.2 The Proposed Methodology

3.2.1 Quantiles without covariates

To motivate and illustrate the proposed methodology, we introduce in this section a simple one-dimensional example and consider the problem of finding the τ-th quantile m_τ of a survival time variable T, or some transformation of the latter. For notational convenience, recall that for any $\tau \in (0,1)$, the τ-th quantile is defined as $m_\tau = \inf\{t : F_T(t) \geq \tau\}$, where F_T is the c.d.f. of T. As further recalled in Chapter 1, and using a slightly different notation for convenience with what will follow, it is well-known that this problem is equivalent to solving

$$m_\tau = \arg\min_a \mathbb{E}\Big[\rho_\tau(a;T)\Big], \tag{3.2.1}$$

where $\rho_\tau(a;T) = (T-a)(\tau - \mathbb{1}(T \leq a))$ is the check function, and $\mathbb{1}(\cdot)$ is the indicator function.

Let us suppose now that the variable of interest T is subject to censoring, that is, instead of observing T, one only observes (Y, Δ), where $Y = \min(T, C)$, $\Delta = \mathbb{1}(T \leq C)$ and C denotes the censoring variable assumed for now to be independent of T, with c.d.f. G_C. Our objective is now to provide an analogue reasoning to the one exposed in Section 1.1.1 concerning the link between the definition of quantiles and the check loss function. To that end, our starting point is to note that, in this context, finding m_τ is equivalent to finding the root in a of

the equation

$$\mathbb{E}\Big[\mathbb{1}(Y > a) - \bar{G}_C(a)(1-\tau)\Big] = 0, \tag{3.2.2}$$

with $\bar{G}_C(a) := 1 - G_C(a) = \mathbb{P}(C > a)$. For this equivalence to hold, it is required that T is independent of C, and that $\bar{G}_C(m_\tau) > 0$. Note that this latter condition, routinely made in survival analysis, also amounts to establishing a natural upper bound for the quantile of interest that can be studied in the presence of censored data.

Now, starting from (3.2.2), mimicking the reasoning behind the check function ρ_τ recalled in Section 1.1.1 leads us to define the following function in a, for a given $y \in \mathbb{R}$ and G_C:

$$\varphi_\tau(a; y, G_C) = \rho_\tau(a; y) - (1-\tau)\int_0^a G_C(s)\,\mathrm{d}s,$$

which is equal to $\int_a^y \big(\mathbb{1}(y > s) - \bar{G}_C(s)(1-\tau)\big)\,\mathrm{d}s$, up to an additive constant with respect to a. Our claim is that the function φ_τ actually is an extended version of the check function ρ_τ, such that the effect of censoring is handled at the very level of the loss function through a correcting term $(1-\tau)\int_0^a G_C(s)\mathrm{d}s$. To support this statement, first note that when all observations are complete, the above-defined function trivially boils down to the check function. Next, observe that, for all values of $a \neq y$, $\partial\varphi_\tau(a; y, G_C)/\partial a = \bar{G}_C(a)(1-\tau) - \mathbb{1}(y > a)$. Since $0 \leq \bar{G}_C(a) \leq 1$, $\forall a$, this suggests that for a value $y < \tau_{G_C} = \inf\{t : G_C(t) = 1\}$, $a \mapsto \varphi_\tau(a; y, G_C)$ is strictly decreasing on $(-\infty, y)$, strictly increasing on $[y, \tau_{G_C})$, and hence with global minimum value at $a = y$, exactly as the check function. Now, searching for the minimum value in a of $\mathbb{E}[\varphi_\tau(a; Y, G_C)]$ reveals to be corresponding to the formulation of (3.2.2), as

$$\frac{\partial}{\partial a}\mathbb{E}\Big[\varphi_\tau(a; Y, G_C)\Big] = \bar{G}_C(a)(1-\tau) - \mathbb{P}(Y > a).$$

From these observations, one may conclude that φ_τ is an adapted version of the check function ρ_τ accounting for potential incompleteness of observations. Therefore, in the presence of censoring, we define the logical counterpart of (3.2.1):

$$m_\tau = \arg\min_a \mathbb{E}\Big[\varphi_\tau(a; Y, G_C)\Big]. \tag{3.2.3}$$

Accordingly, based on an i.i.d. sample $(Y_i, \Delta_i), i = 1, \ldots, n$, of (Y, Δ), and given an estimator $\widehat{G}_C$ of G_C, we define a natural estimator of m_τ as the empirical version of (3.2.3):

$$\widehat{m}_\tau = \arg\min_a \sum_{i=1}^n \varphi_\tau(a; Y_i, \widehat{G}_C). \tag{3.2.4}$$

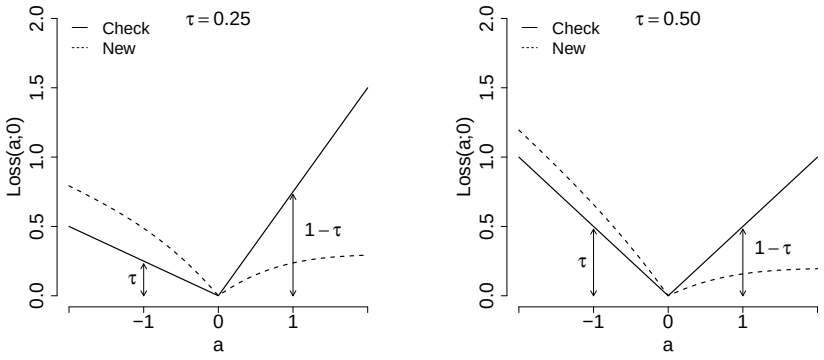

Figure 3.1: Comparison of the check function $a \mapsto \rho_\tau(a;0)$ with the censored loss function $a \mapsto \varphi_\tau(a;0,G_C)$ for $\tau = \{0.25, 0.5\}$ and $C \sim \mathcal{N}(0,1)$. Note that, for comparison purposes and contrary to the usual case exposed in Figure 1.1, the check function is here not represented in terms of residuals $(y-a)$.

Note that $\widehat{m}_\tau$ exploits all observations in the estimation process, regardless of their censoring status. From the above-described reasoning, it also naturally appears that this general loss function may then be extended outside basic sample quantiles to numerous regression models, as will be illustrated next for the linear model.

To gain further insight into the newly proposed loss function and understand how censoring is accounted for through the regulating term $(1-\tau)\int_0^a G_C(s)\mathrm{d}s$, we consider in Figure 3.1 a graphical illustration for the situation $y = 0$ with two quantile levels $\tau = \{0.25, 0.5\}$ and a standard normal distribution for the censoring variable. Note that in this situation, the correcting term of φ_τ will always be positive for $a > 0$, while negative for $a < 0$. Intuitively, this simply suggests that φ_τ penalizes more heavily underestimation of the value of the quantile than the usual check function to correct for the fact that the response may not be fully observed. Similarly, as can be graphically observed, a smaller loss is assigned for overestimation in order to stimulate the latter when using incomplete data. Note also that the penalization will logically be impacted by the distribution G_C, as the amount of alteration of the loss function will be all the more important given that $G_C(m_\tau) = \mathbb{P}(C \leq m_\tau)$ is large. This is in line with one's intuition that censoring should especially be corrected for when it is susceptible of altering the estimation of a quantile, that is, when censoring is likely to happen to data points lying below m_τ.

Finally, an important feature that should be noted, appearing clearly in Figure 3.1, is that $a \mapsto \varphi_\tau(a;y,G_C)$ is in fact no longer convex. Therefore, special care is

called for the numerical optimization required in (3.2.4) as the objective function may allow for local minima, hereby possibly harming the search for the global one. In light of this, we provide in Section 3.4 an adapted version of the Majorize-Minimize (MM) algorithm as proposed by Hunter and Lange (2000) that proves to be satisfactory given the simulation results of Section 3.5, even in more complex multivariate regression situations.

3.2.2 Quantiles with covariates: linear regression

To apply the newly proposed loss function in a regression context, we propose to examine the well-studied linear regression model. Specifically, we consider the inclusion of a covariate vector $\boldsymbol{X}$ of dimension $d+1, d \geq 1$, whose first component corresponding to the intercept is taken to be 1. Then, based on an i.i.d. sample $(T_i, \boldsymbol{X}_i), i = 1, \ldots, n$, from $(T, \boldsymbol{X})$, it is assumed that

$$m_\tau(\boldsymbol{X}_i) = \beta_\tau^\mathsf{T} \boldsymbol{X}_i, \tag{3.2.5}$$

where $m_\tau(\boldsymbol{x}) = \inf\{t : F_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \geq \tau\}$ is the τ-th conditional quantile of T given $\boldsymbol{X} = \boldsymbol{x}$, $F_{T|\boldsymbol{X}}$ denotes the conditional c.d.f. of T given $\boldsymbol{X}$, and β_τ is the $(d+1)$-dimensional unknown quantile coefficient vector. Similarly to sample quantiles and as recalled in (1.1.8), it is well known that the estimation of β_τ in the uncensored case is obtained by minimizing the objective function

$$Q_n(\beta) = \sum_{i=1}^{n} \rho_\tau(\beta^\mathsf{T} \boldsymbol{X}_i; T_i). \tag{3.2.6}$$

Under random censoring, the statistical problem now consists of estimating the true value of β_τ based on i.i.d. triplets $(Y_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, from $(Y, \Delta, \boldsymbol{X})$, where, as usual, $Y = \min(T, C)$, $\Delta = \mathbb{1}(T \leq C)$ and C is assumed here to be independent of T given $\boldsymbol{X}$. In this context, analogous arguments to those previously described lead us to define β_τ as the minimizer of the function

$$Q^c(\beta) = \mathbb{E}\Big[\varphi_\tau\left(\beta^\mathsf{T} \boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\Big],$$

where the superscript c explicitly expresses the presence of censoring, and where $G_C(\cdot|\boldsymbol{x})$ denotes the conditional distribution of C given $\boldsymbol{X} = \boldsymbol{x}$. Subsequently, we propose to estimate β_τ by minimizing the following objective function:

$$\begin{aligned} Q_n^c(\beta) &= \sum_{i=1}^{n} \varphi_\tau\left(\beta^\mathsf{T} \boldsymbol{X}_i; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i)\right) \\ &= \sum_{i=1}^{n} \left\{ \rho_\tau(\beta^\mathsf{T} \boldsymbol{X}_i; Y_i) - (1-\tau) \int_0^{\beta^\mathsf{T} \boldsymbol{X}_i} \widehat{G}_C(s|\boldsymbol{X}_i)\, \mathrm{d}s \right\}, \end{aligned} \tag{3.2.7}$$

where $\widehat{G}_C(\cdot|\boldsymbol{x})$ is a consistent estimator of $G_C(\cdot|\boldsymbol{x})$. Popular choices for the latter include, among others, the Kaplan-Meier estimator if one assumes independence between C and $\boldsymbol{X}$, a Cox regression model or Beran's prominent conditional Kaplan-Meier estimator. Recall from Chapter 1 that the latter takes the following form for Nadaraya-Watson type of weights:

$$\widehat{G}_C(c|\boldsymbol{x}) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{B_{ni}(\boldsymbol{x})}{\sum_{j=1}^{n} \mathbb{1}(Y_i \leq Y_j) B_{nj}(\boldsymbol{x})} \right\}^{\mathbb{1}(Y_i \leq c,\, \Delta_i = 0)}, \tag{3.2.8}$$

where $B_{nj}(\boldsymbol{x}) = K\left((\boldsymbol{x} - \boldsymbol{X}_j)/h_n\right) / \sum_{k=1}^{n} K\left((\boldsymbol{x} - \boldsymbol{X}_k)/h_n\right)$ is a sequence of weights depending on some multivariate kernel density function K and a positive bandwidth parameter h_n.

Similarly to sample quantiles, formulation (3.2.7) allows one to plainly extract the information of every observation at hand, even if confronted to incompleteness of the latter. In comparison, the initial inverse-censoring-probability strategy of Koul et al. (1981) has to trade-off between appropriate estimation of G_C and efficiency loss through the use of Δ in the adopted weights. Therefore, one could expect that the proposed methodology will be more efficient than competitors constructed on the basic inverse-censoring-probability strategy, especially when the censoring proportion is large. In the same spirit, one might expect that the above-described estimator provides in this linear context an interesting alternative to the redistribution-of-mass approach when confronted to high censoring percentages, as the latter will inevitably suffer from low sample size for the required estimation of $F_{T|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ in this situation. However, a possible drawback of the formulation (3.2.7) compared to its competitors is that the estimation of $G_C(\cdot|\boldsymbol{x})$ will affect in our context all observations, hereby possibly harming the global estimation process in case of low sample size for the estimation of the latter distribution. These preliminary remarks will be empirically investigated in Section 3.5.

3.3 Large Sample Properties

We establish in this section both the consistency and asymptotic normality of the proposed estimator $\widehat{\beta}_\tau$ defined as the minimizer of $Q_n^c(\beta)$ in (3.2.7). To that end, we require that the following assumptions hold:

(C1) The support $\mathrm{supp}(\boldsymbol{X})$ of $\boldsymbol{X}$ is contained in a compact subset of $\mathbb{R}^{d+1}$, and the variance-covariance matrix of $\boldsymbol{X}$ is positive definite.

(C2) For β in a neighborhood $\mathcal{B}$ of β_τ, $\inf_{\beta \in \mathcal{B}} \inf_{\boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{x}|\boldsymbol{x}) > 0$, where $f_{T|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ denotes the conditional density function of T given $\boldsymbol{X} = \boldsymbol{x}$. Furthermore, $\sup_{t,\, \boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})} f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) < \infty$.

(C3) Define the (possibly infinite) time $\tau_{F_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})} = \inf\{t : F_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = 1\}$, where $F_{Y|\boldsymbol{X}}$ designates the conditional c.d.f. of Y given $\boldsymbol{X}$. Suppose first that there

exists a real number $\upsilon < \tau_{F_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})}$ for all $\boldsymbol{x}$ in $\text{supp}(\boldsymbol{X})$. Denote next by $\mathcal{G}$ the class of functions $G(t,\boldsymbol{x}) :]-\infty,\upsilon] \times \text{supp}(\boldsymbol{X}) \to [0,1]$ of bounded variation with respect to t (uniformly in $\boldsymbol{x}$) that have first-order partial derivatives with respect to $\boldsymbol{x}$ of bounded variation in t (uniformly in $\boldsymbol{x}$), and bounded (uniformly in t) second-order partial derivatives with respect to $\boldsymbol{x}$ which are uniformly in t Lipschitz of order η for some $0 < \eta < 1$. Suppose that $G_C \in \mathcal{G}$.

(C4) For every $\boldsymbol{x} \in \text{supp}(\boldsymbol{X})$ and for $\beta \in \mathcal{B}$, the point $\beta^\mathsf{T}\boldsymbol{x} \in \mathbb{R}$ lies below υ.

(C5) The estimator $\widehat{G}_C$ of G_C satisfies:

(i) $\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})} \sup_{y\leq\upsilon} \left|\widehat{G}_C(y|\boldsymbol{x}) - G_C(y|\boldsymbol{x})\right| = o_\mathbb{P}(1)$ and $\mathbb{P}(\widehat{G}_C \in \mathcal{G}) \to 1$ as $n \to \infty$.

(ii) Uniformly in $\beta \in \mathcal{B}$,

$$\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\left(\widehat{G}_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) - G_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right)\right] = n^{-1}\sum_{i=1}^{n} \boldsymbol{X}_i \xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i) + o_\mathbb{P}(n^{-1/2}),$$

where the notation $\mathbb{E}_{\boldsymbol{X}}$ explicitly indicates that the expectation is taken with respect to the distribution of $\boldsymbol{X}$, and where $\xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i)$, $i = 1, \ldots, n$, are i.i.d. random vectors with mean 0 and

$$\sup_{\beta\in\mathcal{B}} \mathbb{E}\left[||\xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i)||^2\right] < \infty.$$

Before commenting our main assumptions, it should be noted that our proofs both rely on a crucial result of Lopez (2011) for the class $\mathcal{G}$ defined in assumption (C3). The latter result is constructed upon a dimension reduction technique of the form $G_C(\cdot|\boldsymbol{X}) = G_C(\cdot|g(\boldsymbol{X}))$ for some function $g : \mathbb{R}^{d+1} \to \mathbb{R}$. As a consequence, our proofs are here valid as such under the same hypothesis or, more simply, for univariate covariates.

Now, assumptions (C1) and (C2) are standard in the context of quantile regression estimation for both complete and censored observations in order namely to guarantee the uniqueness of β_τ. Assumption (C3) defines a general class of functions embedding G_C coming from the work of Lopez (2011). The latter paper develops the theory of bracketing numbers (defined in Van der Vaart and Wellner (1996, p. 83)) associated to the class $\mathcal{G}$ on which part of the proofs rely. Assumption (C4) is required for both the uniqueness of β_τ and the asymptotic properties of $\widehat{\beta}_\tau$. Note that (C4) also encompasses the remark following equation (3.2.2) in Section 3.2.1, as it defines here in the regression context a natural upper bound for the quantile of interest that can be studied when considering censored responses. Assumption (C5)-(i) requires in its first part the uniform consistency of the

censoring distribution estimator and is for instance fulfilled with a conditional, respectively unconditional, Kaplan-Meier estimator as shown in Van Keilegom and Akritas (1999), respectively Gill (1983), under namely suitable bandwidth conditions for the former. The second part of (C5)-(i) may also be verified using for instance a conditional Kaplan-Meier estimator under appropriate bandwidth and kernel conditions as stated in Lemma 6.2 of Lopez (2011). Finally, assumption (C5)-(ii) is requested for the asymptotic distribution of our estimator and requires implicitly a general linear representation of $\widehat{G}_C$. Stating such an assumption allows us not to be restrictive for the estimator one wishes to consider for G_C. It should of course be noted that, when selecting a particular form of estimator, such an assumption can in fact be deduced from appropriate higher-level conditions. To illustrate this statement, Lemma 3.1 in Section 3.8 exposes how condition (C5)-(ii) may be deduced from bandwidth and kernel conditions if one adopts Beran's estimator as described in (3.2.8). A similar development for the appropriate transcription of (C5)-(ii) in the case of a Kaplan-Meier estimator for G_C may follow the exact same arguments along with the linear representation proposed in Lo and Singh (1986).

We now state in Theorem 3.1 the consistency of our proposed estimator. Its proof is deferred to Section 3.8.

Theorem 3.1. *For a given quantile level $0 < \tau < 1$, define $\widehat{\beta}_\tau$ as the minimizer of $Q_n^c(\beta)$ in (3.2.7). Assume that the censoring time C is conditionally independent of the survival time T given the covariates $\boldsymbol{X}$, and that the triples $(Y_i, \Delta_i, \boldsymbol{X}_i), i = 1, \ldots, n$, form an i.i.d. multivariate random sample. Then, under assumptions (C1)-(C5)-(i),*

$$\widehat{\beta}_\tau \to \beta_\tau$$

in probability, as $n \to \infty$.

The proof of Theorem 3.1 relies heavily on the work of Delsol and Van Keilegom (2015) on non-smooth semiparametric estimating equations with an infinite-dimensional nuisance parameter. The first key requirement to apply this work to our framework is the uniform consistency of the conditional censoring distribution estimator. The second key requirement concerns the class $\mathcal{G}$ for which it has to be shown that the latter is 'well-behaved' in terms of size, which is represented by the notion of bracketing number, in order for $\widehat{\beta}_\tau$ to be consistent when constructed on a preliminary estimator $\widehat{G}_C \in \mathcal{G}$. For this part, we rely on a crucial result of Lopez (2011) which establishes a bound on bracketing numbers for the class $\mathcal{G}$ under the above-mentioned dimension reduction technique.

Next, we report in the following theorem the limiting distribution of our proposed estimator $\hat{\beta}_\tau$. For this result, we rely on Proposition 2 in Birke et al. (2017) which slightly modifies Theorem 2 of Chen et al. (2003), and where the key requirement is now the linear representation of the nuisance parameter implied by assumption (C5)-(ii). We refer to Section 3.8 for a small discussion on the use of distinctive tools for the proofs of our theorems.

Theorem 3.2. *For a given* $0 < \tau < 1$*, under the assumptions of Theorem 3.1 and (C5)-(ii),*

$$n^{1/2}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma_1^{-1}\Sigma\Gamma_1^{-1}),$$

where $\Gamma_1 = \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^\intercal f_{T|\boldsymbol{X}}(\beta_\tau^\intercal \boldsymbol{X}|\boldsymbol{X})(1 - G_C(\beta_\tau^\intercal \boldsymbol{X}|\boldsymbol{X}))\right]$*, and* $\Sigma = Cov(\Lambda_i)$ *with* Λ_i *defined in equation* (B.5) *of Section 3.8.*

It is noted from Theorem 3.2 that, as anticipated and similarly for instance to the procedure of Wang and Wang (2009), the obtained asymptotic result reduces to the usual limiting distribution for linear quantile regression in case of strictly complete observations (see e.g. Theorem 4.1 in Koenker (2005)). A similar remark holds true when considering small quantile levels of interest along with censoring mechanisms that strictly prohibit censored data from appearing below the quantile regression region of interest, i.e. $G_C(\beta^\intercal \boldsymbol{x}|\boldsymbol{x}) = 0,\ \forall \boldsymbol{x} \in \text{supp}(\boldsymbol{X})$ and β in the neighborhood $\mathcal{B}$ of β_τ, given that $\widehat{G}_C$ will in this case have no impact on the asymptotic behavior of our estimator.

Now, practical inferential use of Theorem 3.2 on β_τ is made difficult by the complex form of the asymptotic covariance matrix, as the latter involves several unknown quantities that are challenging to accurately estimate in practice. Therefore, for inferential purposes we propose to adopt the simple percentile bootstrap through resampling the triples $(Y_i, \Delta_i, \boldsymbol{X}_i), i = 1, \ldots, n$, with replacement. Specifically, 95% bootstrap confidence intervals for β_τ may easily be constructed using the 2.5th and 97.5th percentiles of the bootstrap coefficients, after having drawn a sufficient amount of bootstrap samples. This technique was repeatedly shown to be satisfactory in the literature (e.g. Portnoy (2003), Wang and Wang (2009), Leng and Tong (2013)) and its validity for the proposed procedure will be illustrated in Section 3.5.

3.4 Minimization Algorithm

As discussed in Section 3.2, application of the newly proposed methodology for linear regression has to account for the computational aspect of the mathematical minimization of (3.2.7). In particular, in addition to the classical differentiability issue of quantile regression, the nonconvexity of $a \mapsto \varphi_\tau(a; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i))$ requires an efficient optimization routine for the global minimization of the objective function. Given that the newly proposed loss function may be seen as an extension of the check function, we suggest in this section to adapt an existing methodology for uncensored data, for which much work has been provided for the linear context as current procedures include, among others, the Interior point algorithm, Simplex algorithms and the MM algorithm. We therefore first briefly review in this section one particular optimization procedure for complete observations, and then discuss how the latter may be adapted to the practical minimization of (3.2.7).

Primarily due to its numerical robustness and ease of adaptation, we propose to investigate in our context the use of an MM, or Majorize-Minimize, algorithm as first introduced by Hunter and Lange (2000) for quantile regression. As suggested by its denomination, the rationale of the MM algorithm is to operate in two steps: the objective function to be minimized is first majorized by an appropriate surrogate function, which is then in turn minimized in the second step in order to define the next iterate of the algorithm. By doing so, a difficult optimization problem is to be replaced by a simpler one, with iteration counting as the price to pay for this substitution. More formally, in the quantile regression context with complete observations, letting $\beta^{(m)}$ denote the m-th iterate in finding the minimum of $Q_n(\beta)$ defined in (3.2.6), Hunter and Lange propose in the first step to majorize $Q_n(\beta)$ by a surrogate function $\xi_n(\beta|\beta_{(m)}) : \mathbb{R}^{(d+1)} \times \mathbb{R}^{(d+1)} \to \mathbb{R}$ such that $\xi_n(\beta|\beta_{(m)}) \geq Q_n(\beta)$, for all β and $\xi_n(\beta_{(m)}|\beta_{(m)}) = Q_n(\beta_{(m)})$. Specifically, exploiting the fact that majorization relations are closed under the formation of sums, Hunter and Lange suggest to majorize $Q_n(\beta)$ by the sum for each $i = 1, \ldots, n$, of the unique quadratic curve tangent to the graph of $\rho_\tau(\beta_{(m)}^\mathsf{T} \boldsymbol{X}_i; T_i)$ at $\beta_{(m)}^\mathsf{T} \boldsymbol{X}_i$, yielding

$$\xi_n(\beta|\beta_{(m)}) = \sum_{i=1}^{n} \left\{ \frac{(T_i - \beta^\mathsf{T} \boldsymbol{X}_i)^2}{4(\epsilon + |T_i - \beta_{(m)}^\mathsf{T} \boldsymbol{X}_i|)} + \left(\tau - \frac{1}{2}\right)(T_i - \beta^\mathsf{T} \boldsymbol{X}_i) \right\} + c_\tau,$$

where c_τ is a constant such that $\xi_n(\beta_{(m)}|\beta_{(m)}) = Q_n(\beta_{(m)})$, and $\epsilon > 0$ is a small perturbation to be selected in order to avoid issues with possible zero residuals for iteration m. The explicit minimizer of $\xi_n(\beta|\beta_{(m)})$ with respect to β then simply becomes the next iterate $\beta_{(m+1)}$ in the second step of the MM algorithm. As an interesting property, called 'descent property', it can be shown that the MM algorithm automatically drives the objective function downhill, that is $\xi_n(\beta_{(m+1)}|\beta_{(m)}) \leq \xi_n(\beta_{(m)}|\beta_{(m)})$. A further convenience for the use of the MM algorithm for linear quantile regression is that convergence to the unique minimum of (3.2.6) is guaranteed provided $\boldsymbol{X}$ is of full rank, as reported in Proposition 4 of Hunter and Lange.

Applying now the same arguments with the objective of finding the minimizer of $Q_n^c(\beta)$ defined in (3.2.7), we note that $a \mapsto \varphi_\tau(a; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i))$ is the difference of two convex functions. In the search of a surrogate function, this suggests then to explicitly use the quadratic form of $\xi_n(\beta|\beta_{(m)})$ for the first part of φ_τ, and exploit the convexity of the second part as recommended in the 'Tricks of the Trade' section of Hunter and Lange (2004). Explicitly, for this second part we have that for all $i = 1, \ldots, n$,

$$\int_0^{\beta^\mathsf{T} \boldsymbol{X}_i} \widehat{G}_C(s|\boldsymbol{X}_i)\, \mathrm{d}s \geq \int_0^{\beta_{(m)}^\mathsf{T} \boldsymbol{X}_i} \widehat{G}_C(s|\boldsymbol{X}_i)\, \mathrm{d}s + (\beta^\mathsf{T} \boldsymbol{X}_i - \beta_{(m)}^\mathsf{T} \boldsymbol{X}_i) \widehat{G}_C(\beta_{(m)}^\mathsf{T} \boldsymbol{X}_i | \boldsymbol{X}_i).$$

Recombining this with the expression of $\xi_n(\beta|\beta_{(m)})$ yields the following surrogate function for $Q_n^c(\beta)$:

$$\xi_n^c(\beta|\beta_{(m)}) = \sum_{i=1}^{n} \left\{ \frac{(Y_i - \beta^\mathsf{T}\boldsymbol{X}_i)^2}{4(\epsilon + |Y_i - \beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|)} + \left(\tau - \frac{1}{2}\right)(Y_i - \beta^\mathsf{T}\boldsymbol{X}_i) \right.$$
$$- (1-\tau)\int_0^{\beta_{(m)}^\mathsf{T}\boldsymbol{X}_i} \widehat{G}_C(s|\boldsymbol{X}_i)\,\mathrm{d}s$$
$$\left. - (1-\tau)(\beta^\mathsf{T}\boldsymbol{X}_i - \beta_{(m)}^\mathsf{T}\boldsymbol{X}_i)\widehat{G}_C(\beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i) \right\} + \tilde{c}_\tau,$$

where $\tilde{c}_\tau$ is a constant such that $\xi_n^c(\beta_{(m)}|\beta_{(m)}) = Q_n^c(\beta_{(m)})$. For the second step of the algorithm, let $\mathcal{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n]$ be the $(d+1) \times n$ matrix of covariates, and $\mathcal{Y}^\mathsf{T} = (Y_1, \ldots, Y_n)$ be the vector of observed responses. Minimizing $\xi_n^c(\beta|\beta_{(m)})$ with respect to β to obtain $\beta_{(m+1)}$ then yields the following succinct result:

$$-\mathcal{X}\mathcal{A}_{(m)}\mathcal{Y} + \mathcal{X}\mathcal{A}_{(m)}\mathcal{X}^\mathsf{T}\beta_{(m+1)} - \mathcal{X}\mathcal{D} = \mathcal{X}\mathcal{E}_{(m)}, \tag{3.4.1}$$

where $\mathcal{A}_{(m)}$ is a $n \times n$ diagonal matrix with i-th diagonal entry $1/[2(\epsilon - |Y_i - \beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|)]$, $\mathcal{D}$ is a $n \times 1$ vector of $(\tau - 1/2)$, and $\mathcal{E}_{(m)}$ is a $n \times 1$ vector with i-th entry $(1-\tau)\widehat{G}_C(\beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i)$. Solving (3.4.1) with respect to $\beta_{(m+1)}$ yields the explicit iteration of the MM algorithm at step $(m+1)$, which is a simple adaptation of the MM algorithm for linear quantile regression with complete observations as there is only a supplementary term $\mathcal{E}_{(m)}$ appearing at every iteration. The proposed algorithm may then be resumed as follows:

Algorithm

Step 0. Given an initial estimate of β_τ denoted by $\beta_{(0)}$, set $m = 0$. Select a small tolerance value δ, for instance $\delta = 10^{-9}$, and choose ϵ such that $\epsilon \ln \epsilon \approx -\delta/n$.

Step 1. Estimate $G_C(\cdot|\boldsymbol{X})$ and calculate $\mathcal{E}_{(m)}$ as $(1-\tau)\widehat{G}_C(\beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i), i = 1, \ldots, n$. Calculate $\mathcal{A}_{(m)}$ and set

$$\beta_{(m+1)} = \left(\mathcal{X}\mathcal{A}_{(m)}\mathcal{X}^\mathsf{T}\right)^{-1}\mathcal{X}\left(\mathcal{A}_{(m)}\mathcal{Y} + \mathcal{D} + \mathcal{E}_{(m)}\right).$$

Step 2. If $||\beta_{(m+1)} - \beta_{(m)}|| > \delta$ or $\left|\xi_n^c(\beta_{(m+1)}|\beta_{(m)}) - \xi_n^c(\beta_{(m)}|\beta_{(m)})\right| > \delta$, replace m by $m+1$ and return to step 1.

Concerning the choice of $\beta_{(0)}$, special care is to be brought to the latter given the nonconvexity of φ_τ and hence the possibility of capturing only a local minimum, if there is any. In an attempt to mend this obstacle, similarly to Leng and Tong (2013), we propose to adopt the estimator of Bang and Tsiatis (2002) as $\beta_{(0)}$, given

the consistency of the latter. Hence, one could hope that, by already providing a decent estimation of β_τ, $\beta_{(0)}$ should be close to the global minimum of $Q_n^c(\beta)$, hereby possibly avoiding the ambush of potential local minima. Furthermore, given the need at each iteration m of the algorithm to evaluate $\widehat{G}_C(\beta_{(m)}^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i), i = 1, \ldots, n$, a decent estimator $\beta_{(0)}$ may also, to some extent, prevent too many evaluations of $\widehat{G}_C(\cdot|\boldsymbol{X})$ in the right tail, for which classical estimators are known to be typically unstable due to sparsity of the data. On the other hand, as noted by Leng and Tong, given that roughly only uncensored observations are handled for $\beta_{(0)}$, the latter estimator may severely suffer from an efficiency loss in the estimation, hereby motivating the use of an estimator exploiting all observations at hand such as the one proposed in this work. Nonetheless, as there is no guarantee that the algorithm starting from this $\beta_{(0)}$ will avoid converging to a possible local minimum, it is preferable to restart the latter as suggested by Hunter and Lange, in this case for instance by adding small perturbations to $\beta_{(0)}$ to verify the stability of the solution.

Finally, note that, analogously to the remarks of Section 3.2, the described algorithmic procedure of this section based on the MM algorithm is again easily adaptable to broader parametric, nonparametric and semiparametric models. For instance, considering a general parametric model, a convenient minimization algorithmic procedure would only require an adaptation of equation (3.4.1) to define the appropriate iterative process one could readily implement for the desired model.

3.5 Simulation Study

In this section, we assess the finite sample performance of the proposed methodology by means of Monte Carlo simulations. As previously mentioned, given that the adapted loss function may engender further modelling techniques beyond the simple linear framework, we are mainly interested here in comparing the performance of our estimator (NEW) with the two prominent weighting schemes for quantile regression with censored data: the redistribution-of-mass (RM), which will be embodied here by Wang and Wang's estimator, and the inverse-censoring-probability (ICP), represented here by Bang and Tsiatis' estimator. Based on the choice of these competitors, as a preliminary remark that has repeatedly been illustrated in the literature, while ICP may be subject to possible improvement for linear models, RM provides in this context a very competitive estimator, often taken as primary reference. Lastly, to grasp the impact of censoring, we also include an omniscient procedure (*Omni*) using all the observations $T_i, i = 1, \ldots, n$, as if they were available in practice for the estimation of β_τ.

The four procedures are compared in four main data generating processes (DGP) where the survival time is systematically independent of the censoring time given the covariates. The first three DGPs are taken to be univariate settings, while the fourth setting considers an example with four covariates. Given that both RM

and NEW will be implemented with local estimators of conditional distributions as will be detailed below, the latter setting will be of interest in order to analyze the performance of both procedures when confronted to multivariate covariates. Note that, in their paper on variable selection using the redistribution-of-mass, Wang et al. (2013) also tolerate as many as four covariates for the determination of the involved local weights.

The first DGP is taken from Wang and Wang (2009) and Leng and Tong (2013), and presents a simple setting which is linear in all quantile levels and where all estimation procedures in the literature perform relatively well for the estimation of the regression coefficients. Hence, a basic requirement for a newly proposed methodology as the one considered here would be to perform appropriately as well in this particular setting. Next, the second and third DGP, partly inspired by Wang and Wang, are taken to be linear only in the quantile level of interest, hereby ruling out the use of estimators assuming global linearity such as Portnoy's or the martingale-based estimator of Peng and Huang (2008), despite their relative robustness illustrated for instance in Wang and Wang and Leng and Tong. For the second DGP, the censoring variable C is taken to be independent of the covariate, while the third DGP covers a dependent scenario. This will allow to explore the potential impact of a conditional censoring distribution. Note however that, in all four settings, the proposed procedure NEW will repeatedly be implemented considering a conditional censoring distribution even when the simulated settings do not present any influence of the covariate on the latter distribution.

Concerning the implementation of these procedures, we first use the `rq` function of the R library `quantreg` for *Omni*. ICP is likewise implemented with `rq` by incorporating in the function the weights $\Delta_i/(1-\widehat{G}_C(Y_i)), i = 1, \ldots, n$, where $\widehat{G}_C$ is the Kaplan-Meier estimator of G_C. As for RM, the estimator is implemented using Wang and Wang's code, available on their websites. For the estimation of local weights, we use the biquadratic kernel $K(x) = (15/16)(1-x^2)^2\mathbb{1}(|x| \leq 1)$ for the univariate settings as recommended in Wang and Wang, while an eighth-order kernel $K(x) = (1/13)(1-x^2)(35-385x^2+1001x^4-715x^6)\mathbb{1}(|x| \leq 1)$ is used for the four-variate example, just as in Wang et al.. The selection of the required bandwidth is further implemented using the 5-fold cross validation (see e.g. section 7.10 of Hastie et al. (2001)) suggested in Wang and Wang, which we briefly review here for an arbitrary estimator $\widehat{\beta}_\tau$ of β_τ given the technique will be applied to our newly proposed procedure as well. Specifically, for each candidate bandwidth h, we first randomly separate the data into 5 roughly equal-sized and non-overlapping parts. For each part $j = 1, \ldots, 5$, β_τ is then estimated using all observations but the ones in part j, and the quality of fit of the model is determined by calculating the following prediction error on the complete observations in the set $\mathcal{J}$ corresponding to all observations of part j:

$$\mathrm{PE}_j(h) = \sum_{\substack{i \in \mathcal{J} \\ \Delta_i = 1}} \rho_\tau(Y_i - \widehat{\beta}^{\mathsf{T}}_{\tau\,(-j)} \boldsymbol{X}_i),$$

where the notation $(-j)$ indicates estimation using all observations except the ones in the j-th part. The procedure is repeated and averaged over $j = 1, \ldots, 5$, and the bandwidth yielding the smallest average prediction error is then selected among all candidates.

Now, returning to the specific implementation of RM, for the first three settings 15 candidate bandwidths are considered, equally ranging from 0.05 to 0.5. As for the multivariate DGP, given the computational cost of searching for tuning parameters across multiple dimensions, only one initial simulated dataset serves for the determination of the bandwidth among 15 candidates equally ranging from 0.5 to 2. Lastly, NEW is implemented using the algorithm of Section 3.4 where, for every DGP we adopt Beran's local Kaplan-Meier estimator as $\widehat{G}_C(\cdot|\boldsymbol{X})$. Equivalently to RM, the latter is based either on the biquadratic or the eighth-order kernel depending on the dimension of the covariate. Finally, as already mentioned, the required bandwidth is likewise computed via the above-described 5-fold cross validation at each iteration or on one initial dataset. This suggests that for the last DGP, the performance of both RM and NEW could probably be improved if the bandwidths were adapted to each simulation, although none of the estimators is here favored over the other in the analysis we intend to provide.

DGP 1.

The following model is generated:

$$T_i = \beta_0 + \beta_1 X_i + \eta_i,$$

where $\beta_0 = 3$, $\beta_1 = 5$, $X_1, \ldots, X_n$ are i.i.d. $U[0,1]$ variables and $\eta_1, \ldots, \eta_n$ are i.i.d. $\mathcal{N}(0,1)$ variables. The censoring variables $C_1, \ldots, C_n$ are simulated from $U[0, M]$, with M calibrated to attain the desired censoring proportion (15% or 40%) at the median. We note that the choice of M also implies that condition (C4) is not violated for the true β_τ, where $\tau = 0.5$ in this case. The latter consideration regarding condition (C4) repeatedly applies to all the following DGPs and for all the considered levels of τ.

DGP 2.

The following model is generated:

$$T_i = \beta_0 + \beta_1 X_i + \left(3 + (X_i - 0.5)^2\right)\left(\eta_i - \Phi^{-1}(\tau)\right),$$

where $\beta_0 = 1$, $\beta_1 = 0.1$, $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0,1)$ variables, $\eta_1, \ldots, \eta_n$ are also i.i.d. $\mathcal{N}(0,1)$ variables and Φ^{-1} denotes the quantile function of the standard normal distribution. The censoring variables $C_1, \ldots, C_n$ are simulated from $U[m, M]$, where $m = -5/3$ for $\tau = 0.3$, $m = -3$ for $\tau = 0.5$, $m = -3.5$ for $\tau = 0.7$, and M is chosen to attain the desired censoring proportions (30% and 60%) depending on the considered quantile levels.

DGP 3.

The model to generate $T_i, i = 1, \ldots, n$, is the same as for DGP 2 but with parameters $\beta_0 = 1$ and $\beta_1 = 1$. Additionally, in this setting the censoring variable is generated from the model $C_i = \gamma_0 + \gamma_1 X_i + \upsilon_i, i = 1, \ldots, n$, where $\gamma_0 = 1$, $\gamma_1 = 1$, and $\upsilon_1, \ldots, \upsilon_n$ are i.i.d. variables from $U[m, M]$, where $m = -2$ for 30% censoring, $m = -4$ for 60% censoring, and M is calibrated to obtain the desired censoring proportion at the median.

DGP 4.

The following model, inspired by Shows et al. (2010), is generated:

$$T_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \eta_i,$$

where $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1.5$, $\beta_4 = 2$, X_{ji} are i.i.d. standard normal variables for $i = 1, \ldots, n$, and $j = 1, \ldots, 4$, and $\eta_1, \ldots, \eta_n$, are i.i.d. variables simulated from $t(5)$. The variables $C_1, \ldots, C_n$ are, here again, simulated from $U[m, M]$, where $m = 0$ for 30% censoring, $m = -2$ for 60% censoring, and M is chosen to obtain the desired censoring proportion for $\tau = 0.5$.

For all simulation settings, we consider $B = 500$ repetitions of each DGP, four average censoring proportions ($p_c \in \{15\%, 40\%\}$ for DGP 1 and $p_c \in \{30\%, 60\%\}$ for DGP 2, 3 and 4), and sample sizes $n \in \{100, 200, 500\}$. The quantile levels of interest are chosen among $\{0.3, 0.5, 0.7\}$ depending on the DGP and the censoring proportion of interest. Estimators are compared in terms of bias, root mean squared errors (RMSE), and median absolute error (MAE). We further include an overall criterion taken from a prediction point of view: the mean absolute deviation (MAD), given for an estimator $\widehat{\beta}$ by $n^{-1} \sum_{i=1}^{n} |\widehat{\beta}^\top \boldsymbol{X}_i - \beta_\tau^\top \boldsymbol{X}_i|$.

Table 3.1 reports the results of our simulation study for DGP 1. As expected, all estimation procedures perform very similarly in this basic setting, although ICP already exhibits some relative difficulties to compete when confronted to higher censoring proportions. Concerning NEW, we note that in this example the procedure is relatively robust to the low sample size for computing the required local Kaplan-Meier estimator, as observed for $n = 100$ and $p_c = 15\%$. Furthermore, even though the simulated scenario does not account for any influence of the covariate on the censoring distribution, our simulation experience with the present DGP suggests that the estimator is relatively robust to the smoothing parameter to be selected when using Beran's estimator $\widehat{G}_C(\cdot|X)$. This robustness of the results towards both the sample size and the smoothing parameter may be considered as an encouraging observation for the newly proposed estimator given that all observations are affected by the conditional censoring distribution estimator through the studied adapted loss function, as mentioned in Section 3.2.

For visual examination of the results for this DGP, a graphical illustration of the latter is proposed in Figure 3.2 where only the scenarios with 40% censoring are reported for sake of brevity. As expected, apart from slightly worse results for

			Bias		RMSE		MAE		
n	p_c	Method	β_0	β_1	β_0	β_1	β_0	β_1	MAD
100	15%	*Omni*	0.007	0.000	0.255	0.425	0.168	0.281	0.135
		NEW	0.013	-0.006	0.267	0.454	0.175	0.282	0.147
		RM	0.009	-0.004	0.267	0.459	0.173	0.286	0.148
		ICP	0.010	-0.006	0.270	0.456	0.174	0.283	0.148
	40%	*Omni*	0.007	0.000	0.255	0.425	0.168	0.281	0.135
		NEW	0.009	0.013	0.298	0.558	0.183	0.343	0.177
		RM	0.001	0.003	0.297	0.552	0.184	0.361	0.176
		ICP	0.009	-0.001	0.311	0.583	0.185	0.384	0.181
200	15%	*Omni*	0.009	-0.006	0.169	0.298	0.108	0.204	0.095
		NEW	0.013	-0.009	0.182	0.323	0.126	0.227	0.104
		RM	0.011	-0.011	0.184	0.325	0.127	0.230	0.105
		ICP	0.013	-0.014	0.182	0.322	0.129	0.226	0.104
	40%	*Omni*	0.009	-0.006	0.169	0.298	0.108	0.204	0.095
		NEW	0.014	-0.011	0.206	0.390	0.133	0.267	0.124
		RM	0.008	-0.025	0.204	0.387	0.134	0.268	0.123
		ICP	0.016	-0.027	0.217	0.403	0.148	0.276	0.127

Table 3.1: Simulation results for DGP 1 expressed in terms of bias, RMSE, MAE and MAD averaged over $B = 500$ repetitions. Average censoring proportions p_c are taken in $\{15\%, 40\%\}$ at the median, and sample sizes are taken in $\{100, 200\}$.

ICP already mentioned previously, there is little distinction between the procedures based on censored data, as these coherently mimic the omniscient procedure but

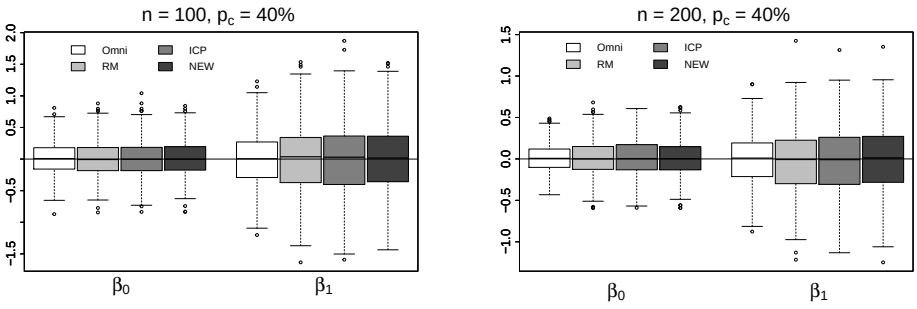

Figure 3.2: Boxplots relative to DGP 1 and Table 3.1 for the estimation of β_0 and β_1, both centered (i.e. depicting $\widehat{\beta}_\tau - \beta_\tau$), for the different estimation procedures with $\tau = 0.5$, $p_c = 40\%$ and $n \in \{100, 200\}$.

			Bias		RMSE		MAE		
p_c	τ	Method	β_0	β_1	β_0	β_1	β_0	β_1	MAD
30%	0.3	*Omni*	0.021	0.014	0.373	0.507	0.244	0.337	0.466
		NEW	0.014	0.031	0.388	0.509	0.258	0.350	0.479
		RM	-0.116	0.057	0.400	0.486	0.265	0.321	0.473
		ICP	-0.071	0.140	0.400	0.582	0.264	0.363	0.513
	0.7	*Omni*	0.014	-0.013	0.371	0.502	0.243	0.332	0.456
		NEW	-0.176	0.049	0.445	0.502	0.293	0.324	0.538
		RM	-0.403	0.104	0.560	0.456	0.404	0.311	0.581
		ICP	-0.317	0.218	0.570	0.732	0.417	0.504	0.691
60%	0.3	*Omni*	0.021	0.014	0.373	0.507	0.244	0.337	0.466
		NEW	-0.058	0.075	0.390	0.503	0.268	0.338	0.473
		RM	-0.517	0.102	0.627	0.414	0.524	0.290	0.613
		ICP	-1.097	0.442	1.208	0.792	1.083	0.502	1.204
	0.5	*Omni*	0.002	0.009	0.350	0.473	0.226	0.308	0.432
		NEW	-0.269	0.113	0.504	0.564	0.339	0.353	0.565
		RM	-0.744	0.190	0.828	0.475	0.760	0.346	0.815
		ICP	-1.387	0.477	1.487	0.856	1.395	0.566	1.486

Table 3.2: Simulation results for DGP 2 with sample size $n = 200$, expressed in terms of bias, RMSE, MAE and MAD averaged over $B = 500$ repetitions. Average censoring proportions p_c are taken in $\{30\%, 60\%\}$, and quantile levels τ are chosen among $\{0.3, 0.5, 0.7\}$.

with anticipated larger variance results due to their handling of incomplete data.

Moving to a more challenging setting with higher censoring proportions, Table 3.2 reports the results of our simulation study for DGP 2 with $n = 200$ observations at each iteration. Due to identifiability issues as encountered in condition (C4), the quantile level $\tau = 0.7$ is not considered here when simulating as much as 60% censoring, whereas the quantile level $\tau = 0.5$ is left aside for 30% censoring for the sake of brevity. From these results, we observe that both RM and ICP present substantial biases in the parameter estimation for this DGP, particularly for the intercept. In contrast, NEW exhibits difficulties in terms of bias only for the most complicated situations considered here, that is, for high quantile levels with respect to the censoring proportion. These bias results concerning the estimation of β_0 naturally impact the RMSE and MAE of both RM and ICP. However, while ICP presents considerable biases for both parameters to be estimated, RM quite surprisingly seems to compensate its difficulties for estimating β_0 by an excellent estimation of β_1, especially with respect to *Omni*. There is little intuition to explain this particular behavior in this more complicated DGP, although part of the explanation could probably be related to the general arguments concerning the redistribution-of-mass pointed out by Wey et al. (2014) in their Supplementary

			Bias		RMSE		MAE		
n	p_c	Method	β_0	β_1	β_0	β_1	β_0	β_1	MAD
200	30%	*Omni*	0.007	-0.016	0.346	0.421	0.238	0.292	0.403
		NEW	-0.030	-0.085	0.367	0.444	0.242	0.328	0.435
		RM	-0.202	-0.201	0.392	0.463	0.284	0.334	0.467
		ICP	-0.146	-0.660	0.440	0.909	0.297	0.655	0.725
	60%	*Omni*	0.007	-0.016	0.346	0.421	0.238	0.292	0.403
		NEW	-0.302	-0.340	0.490	0.659	0.369	0.509	0.640
		RM	-0.737	-0.444	0.806	0.618	0.744	0.498	0.854
		ICP	-1.373	-1.330	1.469	1.600	1.408	1.220	1.771
500	30%	*Omni*	-0.001	0.017	0.225	0.286	0.160	0.207	0.273
		NEW	-0.021	-0.043	0.235	0.316	0.165	0.228	0.297
		RM	-0.180	-0.172	0.284	0.330	0.200	0.250	0.338
		ICP	-0.101	-0.723	0.287	0.850	0.202	0.709	0.656
	60%	*Omni*	-0.001	0.017	0.225	0.286	0.160	0.207	0.273
		NEW	-0.159	-0.206	0.340	0.493	0.243	0.390	0.458
		RM	-0.638	-0.360	0.676	0.472	0.644	0.398	0.705
		ICP	-1.298	-1.529	1.343	1.711	1.303	1.376	1.755

Table 3.3: Simulation results for DGP 3 for the estimation of the median with sample size $n \in \{200, 500\}$ and average censoring proportions $p_c \in \{30\%, 60\%\}$. Results are expressed in terms of bias, RMSE, MAE and MAD averaged over $B = 500$ repetitions.

Materials to which the interested reader is referred. Additionally, similar ambiguous results for the latter technique were observed when considering an alternative estimator for $F_{T|\boldsymbol{X}}$, for instance using survival trees as proposed in Wey et al. and using the R code available on their website. In opposition to these results, NEW here satisfactorily mimics the behavior of the omniscient procedure. Simulations for larger sample sizes reveal the same patterns here, and are therefore omitted.

Next, to grasp the impact of a conditional censoring distribution, Table 3.3 highlights the results of DGP 3 for both sample sizes as well as both censoring percentages. For the sake of brevity, we only report here the case $\tau = 0.5$. Similar findings to those previously described still apply here, as the results of NEW still convincingly mimic the benchmark omniscient estimator in comparison to RM and ICP, especially when considering difficult estimation scenarios. These considerations are naturally reflected when it comes to prediction, as NEW confidently outperforms in this setting both considered weighting schemes.

Figure 3.3 serves to illustrate these results for both DGP 2 and 3. Contrary to Figure 3.2, there is here a clear distinction in the behaviors of all three estimation procedures based on censored data, as the satisfying robustness of NEW with respect to bias results in more challenging estimation scenarios is clearly exhibited. Comparing now specifically NEW and RM, the cost for this notable improvement for NEW seems to translate in a slight increase in variance results, although comparison is here delicate due to the particular behavior of the latter estimator with

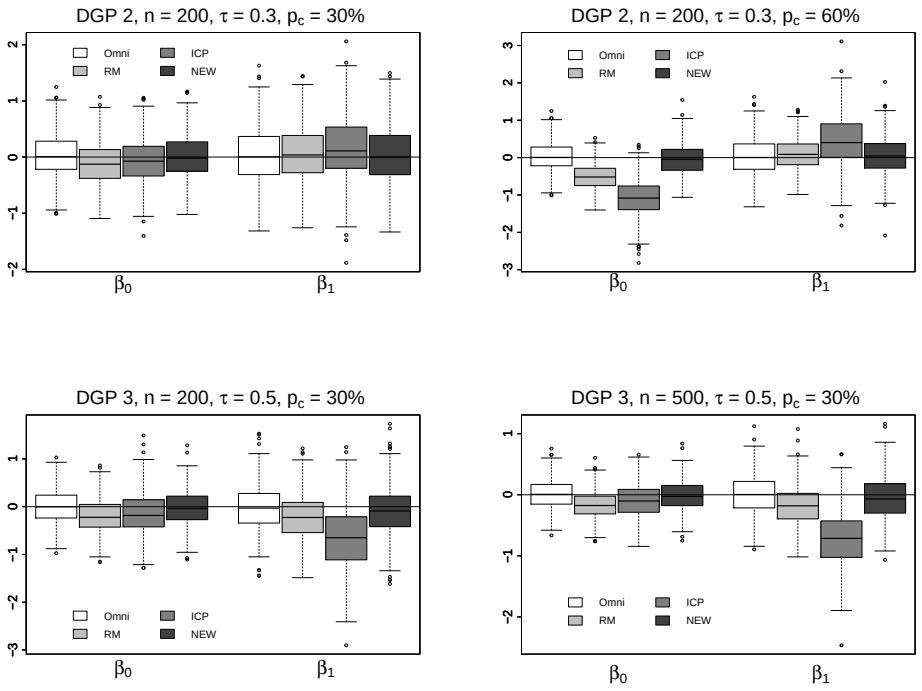

Figure 3.3: Boxplots relative to DGP 2 and Table 3.2 (top), and DGP 3 and Table 3.3 (bottom). Estimation of β_0 and β_1 is reported, both centered (i.e. depicting $\widehat{\beta}_\tau - \beta_\tau$), for the different procedures with $\tau \in \{0.3, 0.5\}$, $p_c \in \{30\%, 60\%\}$ and $n \in \{200, 500\}$.

respect to the benchmark omniscient procedure. On the other hand, Figure 3.3 highlights again the confident dominance of NEW over ICP in these DGPs.

Moving to multivariate covariates, Table 3.4 reports the results of DGP 4 where four covariates are considered for $\tau = 0.5$. For ease of presentation, the results of this DGP are aggregated with respect to the parameters. More specifically, the reported bias stands here for the sum over all parameters of bias taken in absolute value, RMSE stands for the root of mean squared errors summed over all parameters, and MAD stands for the sum over all parameters of median absolute error. Note that in this DGP, and similarly to DGP 1, Portnoy's or Peng and Huang's estimators would a priori be the most appropriate candidates for estimating β_τ given their inherent global linear assumption. However, as already mentioned earlier, we are mainly interested here in comparing the proposed procedure with the two weighting techniques that have engendered further research in alternative models. One such model is the multivariate example of variable selection in linear

		$n = 200$				$n = 500$			
p_c	Method	Bias	RMSE	MAE	MAD	Bias	RMSE	MAE	MAD
30%	*Omni*	0.018	0.213	0.320	0.177	0.014	0.135	0.202	0.112
	NEW	0.062	0.270	0.405	0.220	0.140	0.184	0.281	0.150
	RM	0.192	0.268	0.409	0.219	0.257	0.206	0.337	0.167
	ICP	0.271	0.352	0.547	0.283	0.252	0.248	0.379	0.199
60%	*Omni*	0.018	0.213	0.320	0.177	0.014	0.135	0.202	0.112
	NEW	0.224	0.394	0.595	0.304	0.313	0.293	0.457	0.237
	RM	0.581	0.448	0.731	0.364	0.559	0.364	0.608	0.296
	ICP	1.044	0.716	1.153	0.560	1.017	0.621	1.096	0.493

Table 3.4: Simulation results for DGP 4 for the estimation of the median. Results are in terms of aggregated over the parameters bias, RMSE, MAE and MAD averaged over $B = 500$ repetitions.

regression for which ICP and RM were adapted in Shows et al. (2010) and Wang et al. (2013), respectively.

As can be observed, the obtained results exhibit analogous patterns to those of the univariate settings; NEW takes advantage of superior bias results to outperform its competitors especially when the proportion of censored responses is important. As a consequence, the predictive potential of NEW also outperforms its competitors in this multivariate setting. Finally, it is worth stressing out that for this DGP, the estimator of Wey et al. may suppositionally appear more appropriate to embody the technique of redistribution-of-mass than Wang and Wang's estimator, as the former was proposed to avoid kernel smoothing of $F_{T|\boldsymbol{X}}$ when confronted to multivariate covariates. However, despite using the available code and tuning the parameters with the recommendations of the paper, the results of Wey et al.'s estimator were here subject to numerical instabilities when studying high censoring proportions, hereby preventing an objective comparison of the methodologies for the considered DGP. Additionally, for lower censoring proportions, the obtained results were analogous to those of RM, and are therefore omitted here in Table 3.4.

Lastly, to evaluate the effectiveness of the percentile bootstrap described in Section 3.3 for the proposed procedure, we compare the performance of the latter with the percentile bootstrap of RM for DGPs 1 and 2. Table 3.5 reports the empirical coverage probability and empirical mean length of the confidence intervals obtained with 300 bootstrapped samples at each of 500 iterations. The nominal level is taken to be 0.95, and for both procedures the bandwidths are fixed at 0.05 for DGP 1 and 0.10 for DGP 2. As can first be observed, for the simplest DGP 1 both procedures yield coverage probabilities adequately close to the chosen nominal level and with analogous mean empirical length of the confidence intervals. Secondly, for a more complicated scenario with namely more censoring, we observe here that NEW still appropriately approaches the nominal level while RM exhibits

			ECP				EML			
			RM		NEW		RM		NEW	
DGP	p_c	n	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
1	15%	100	0.954	0.952	0.948	0.952	1.049	1.894	1.052	1.902
		200	0.936	0.950	0.938	0.960	0.752	1.342	0.752	1.346
	40%	100	0.956	0.958	0.966	0.970	1.162	2.258	1.187	2.292
		200	0.940	0.952	0.940	0.954	0.839	1.587	0.837	1.589
2	30%	100	0.956	0.956	0.960	0.948	2.085	2.692	2.218	2.835
		200	0.922	0.946	0.972	0.954	1.490	1.872	1.543	1.990
	60%	100	0.758	0.944	0.968	0.942	1.864	2.315	2.301	2.892
		200	0.756	0.950	0.960	0.966	1.463	1.716	1.609	1.970

Table 3.5: Bootstrap results for DGP 1 ($\tau = 0.5$) and 2 ($\tau = 0.3$) based on 500 simulations with 300 bootstrap samples. ECP and EML stand respectively for empirical coverage probability and empirical mean length, for a nominal level of 0.95. Average censoring proportions p_c are taken in $\{15\%, 30\%, 40\%, 60\%\}$.

difficulties for the latter. We note that this could partly be due to an inappropriate choice of bandwidth. Nevertheless, these results suggest here that the percentile bootstrap represents a satisfactory tool for inference purposes when considering the estimator NEW.

To conclude, from the described simulation results we observe that NEW provides a valuable complement to the literature, especially in cases where the latter seems to fail when confronted to high censoring percentages and quantile levels. Moreover, for the simpler setting considered here, NEW offers similar results to those of RM which are often taken as primary reference. These considerations are believed to be encouraging for the practical use of the newly proposed loss function for quantile regression models.

3.6 Real Data Analysis

As an illustration, we propose to apply the developed methodology to the Channing House dataset which is readily available in the R package `boot`. The data consists of 462 records of residents living at the retirement center 'Channing House' during the period January 1964 to July 1975. One of the purposes of the study was to assess the difference between the survival of men and women under the retirement center program, taking age of entry into account. To that end, the dataset contains information about the survival time (in months) of individuals, their sex and their age (in months) at the time they entered the retirement center. At the end of the study, only 176 residents taken into account experienced the event of interest, resulting in approximately 62% of censoring. Further information on the dataset may be found for instance in Hyde (1980), while a graphical representation of the latter

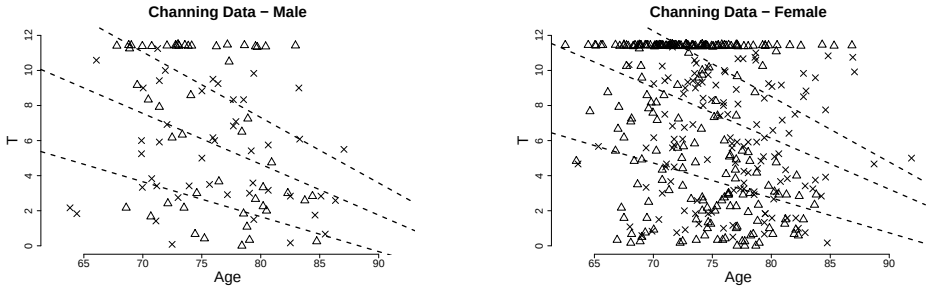

Figure 3.4: Channing House data. Uncensored data points are given by $\times$, while censored observations are represented by $\triangle$. The dotted lines represents the quantile regression estimation using the proposed methodology taking both covariates into account for $\tau \in \{0.1, 0.3, 0.5\}$ (more details given below).

is given by each sex in Figure 3.4 along with a first depiction of estimated quantile regressions for quantile levels $\tau \in \{0.1, 0.3, 0.5\}$ using our proposed methodology. Details about the implementation of the latter are given below.

Now, embracing the role of a practitioner, as a preliminary remark in the objective of fitting a quantile regression model to the data, we note that the responses are subject to a considerable amount of censoring, hereby ruling out a confident use of inverse-censoring-probability-based methodologies, and in particular Bang and Tsiatis' estimator as illustrated in our simulations. Therefore, we will only consider for this analysis Wang and Wang's estimator as a point of comparison for the proposed methodology.

In order to examine the effects of the covariates on several quantile levels of the survival time, we now consider first a sequence of quantiles equally ranging from 0.1 to 0.5 with 0.05 increments. For every quantile level, we estimate β_τ with both Wang and Wang's estimator and the proposed methodology. Implementation of the latter is carried out using the algorithm of Section 3.4, where, as in Section 3.5, we adopt Beran's estimator on each sex for $\widehat{G}_C$. Wang and Wang's estimator is likewise implemented with Beran's estimator on each sex for $\widehat{F}_{T|\boldsymbol{X}}$. Concerning the bandwidths of both estimators, equivalently to our simulation study, the latter are chosen using 5-fold cross validation on the normalized covariate age among 15 candidates equally ranging from 0.05 to 1.5. Finally, for both procedures, 95% confidence intervals are constructed using the percentile bootstrap described in Section 3.3 based on 300 bootstrap samples. For computational convenience, all bootstrap estimations for both methodologies are based on the same bandwidths as for the initial dataset.

Estimation results are presented in Figure 3.5. A first consideration one may acknowledge from the presented figure is the noticeable numerical stability offered

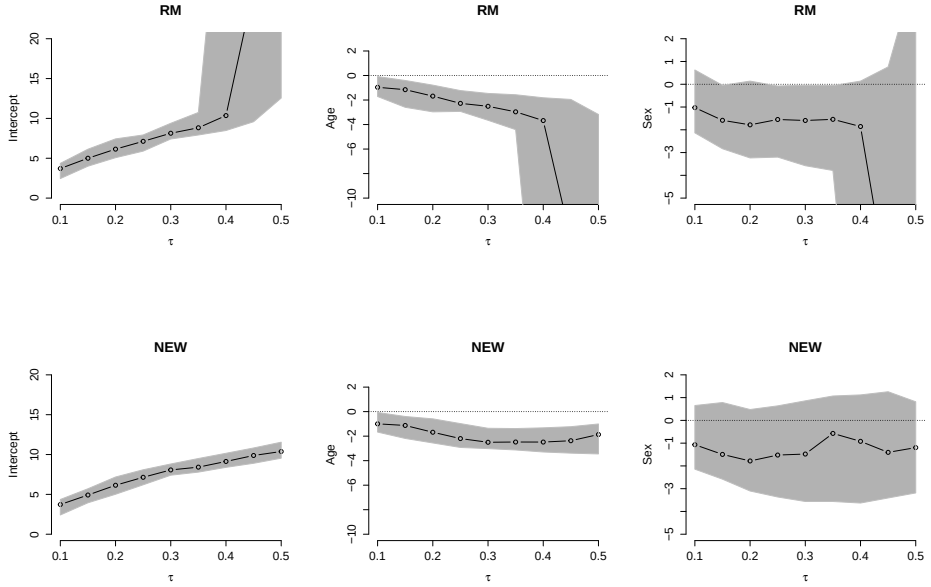

Figure 3.5: Channing House data. Estimated quantile coefficients for regressing the survival time (in years) on sex and age at entry (standardized). The top row reports the results for Wang and Wang's estimator (RM), while the bottom row depicts results obtained with the present methodology (NEW). Shaded areas correspond to 95% confidence intervals based on 300 bootstrap samples.

by the proposed methodology in comparison to Wang and Wang's estimator when estimating relatively high quantile levels with respect to the censoring proportion. This appears as a natural consequence of requiring for our procedure the estimation of G_C rather than $F_{T|\boldsymbol{X}}$, hereby highlighting again the potential complement our estimator may provide to the literature for relatively highly censored datasets. Apart from this first examination, considering only quantile levels below the region of numerical instability of RM, both methods suggest unsurprisingly a significant negative effect of age at entry on the survival time, while there seems to be a discrepancy on the significance of a gender effect for the present dataset even though the upper confidence band of RM is very close to fully incorporating the value 0. Lastly, examining the bootstrap confidence intervals, we observe that confidence intervals for NEW tend to be a little smaller than for RM except for the covariate sex. Furthermore, it is worth stressing out that the width of the NEW's confidence intervals does not seem here to fluctuate much with the increasing quantile levels of interest, hereby exposing a confident resistance to the risk of estimating higher

Method	$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.2$	$\tau = 0.25$	$\tau = 0.3$	$\tau = 0.35$	$\tau = 0.4$
RM	0.474	0.647	0.876	1.112	1.225	1.522	2.103
	(0.02)	(0.04)	(0.07)	(0.10)	(0.10)	(0.11)	(0.10)
NEW	0.474	0.650	0.853	1.065	1.225	1.437	1.774
	(0.02)	(0.03)	(0.05)	(0.05)	(0.11)	(0.11)	(0.10)

Table 3.6: Channing House data. Median prediction error over 200 cross validations, with standard deviations of the latter in parentheses, for quantile levels $\tau \in \{0.1, \ldots, 0.4\}$ for both RM and NEW.

quantiles with a consequent proportion of censored responses.

To further examine the performance of both procedures for the present dataset, we consider a prediction study based on cross validation. Specifically, we first randomly split the data into a training set of 350 observations and a testing set of 112 observations. For quantile levels ranging only from 0.1 to 0.4 by 0.05 increments given the instability of RM beyond $\tau = 0.4$, we then estimate the quantile coefficients for each method based on the training set, and consequently evaluate the predictive performance of the τ-th conditional quantile of fully observed responses in the testing set. As an evaluation measure, we therefore consider the following prediction error:

$$\mathrm{PE} = \mathrm{med}_{\substack{i \in \text{testing} \\ \Delta_i = 1}} \rho_\tau \left(Y_i - \widehat{\beta}_\tau^\intercal \boldsymbol{X}_i \right),$$

where $\widehat{\beta}_\tau$ is an estimator of β_τ based on the training set, and $Y_i, i = 1, \ldots, n$, denote the observed survival times. For robust results, this cross validation is repeated 200 times and the median of all PE's is reported in Table 3.6. Similarly to the above-described bootstrap confidence intervals, for computational consideration, the bandwidths of both NEW and RM are selected by cross validation on one initial training set. Results suggest that for this dataset, the prediction accuracies of both procedures are very similar, with a slight overall advantage for NEW. Given the further improvement of NEW over RM for quantile levels lying above $\tau = 0.4$, this observation may then advocate here for the practical choice of the proposed methodology for a robust quantile regression analysis.

3.7 Conclusion

In this work we have proposed an adapted version of the check function for evaluating sample quantiles, or more generally quantile regression, when confronted to possible right censored responses. By tackling the problem of incomplete data at the level of the loss function through an adjustment of the underlying penalization for under- and overestimation of the true quantile value, the proposed loss function allows to plainly take profit of all observations at hand. This has to be contrasted for instance to the prominent inverse-censoring-probability weighting scheme

of Koul et al. (1981).

As an illustration of the newly proposed loss function for regression models, we proposed to consider the particular and well-studied case of a simple linear regression, for which consistency and asymptotic normality of the quantile coefficient estimator were obtained using recent results on non-smooth semiparametric estimation equations with an infinite-dimensional nuisance parameter. For practical implementation, a detailed adaptation of the MM algorithm was then proposed, and an extensive simulation study was carried out to illustrate the finite sample performance of the methodology. From the latter study, the proposed estimator was observed to perform competitively with respect to the established technique of redistribution-of-mass, especially when considering highly censored datasets. This consideration was highlighted in a real data application as well. Furthermore, for simpler settings, the proposed estimator revealed to be interestingly robust to both the sample size and the smoothing parameter to be selected for a preliminary, and possibly local, estimator of the censoring distribution on which the procedure is built. Lastly, for multivariate settings, while the present work illustrates the use of Beran's estimator for the conditional censoring distribution, we note that alternative modelling techniques for the latter may be more appropriate to avoid the curse of dimensionality, such as a Cox model, a single-index model or survival trees for instance. Such developments deserve further analysis.

To conclude, considering now a broader picture than strictly parametric regression, the proposed results also illustrate the potential of the studied loss function for improvement of various modelling techniques that are currently built upon the inverse-censoring-probability weights, such as single-index regression or the copula-based regression framework of Chapter 2 to name a few. This provides encouraging perspectives for enhancement of flexible censored quantile regression models.

3.8 Proofs

We provide in this section the proofs of Section 3.3. Consistency of the proposed estimator relies heavily on the work of Delsol and Van Keilegom (2015) (DVK hereafter) on non-smooth semiparametric M-estimation problems, while the proof of asymptotic normality relies on a modified version of Theorem 2 of Chen et al. (2003) proposed in Birke et al. (2017). This distinction in the tools used for consistency and asymptotic normality is here to be attributed to the particular nature of our loss function. In order to show in the first place that our estimator is weakly consistent, we indeed need to rely here on the arduous work of Delsol and Van Keilegom given that the proposed loss function is no longer strictly convex, and is hence not immune to potential local minima. However, under the conditions that will allow us to establish weak consistency, the need to rely on the work of Delsol and Van Keilegom for establishing asymptotic normality in a second step vanishes in our particular context, as we may conveniently suppose that we avoid possible local minima for a sufficiently large n and for a sufficiently small neighborhood around the true parameter vector. As a consequence, the proof of asymptotic normality will only need to consider shrinking neighborhoods around the true β_τ and G_C instead of the whole spaces $\mathcal{B}$ and $\mathcal{G}$ as for consistency, hereby only requiring the application of (some version of) the work of Chen et al. (2003).

Now, as a preliminary remark and as already mentioned in Section 3.3, note that the following proofs are built on a crucial result of Lopez (2011) for the class $\mathcal{G}$ in (C3). This implies that our proofs have to be read as such under similar conditions as for Lopez, that is, assuming the existence of a function $g : \mathbb{R}^{d+1} \to \mathbb{R}$ such that $G_C(\cdot|\boldsymbol{X}) = G_C(\cdot|g(\boldsymbol{X}))$, or simply considering the case of a univariate covariate. For ease of reading, the proofs are written considering the latter case.

Proof of Theorem 3.1. In Theorem 1 of DVK, five high level conditions (A1)-(A5) are developed under which an M-estimator is consistent in a general semiparametric maximization problem. We therefore only need to verify the latter conditions to prove the consistency of $\widehat{\beta}_\tau$. For notational convenience with respect to the work of DVK, let us first rewrite $\widehat{\beta}_\tau$ as

$$\widehat{\beta}_\tau = \arg\max_{\beta \in \mathcal{B}} \sum_{i=1}^{n} \psi_\tau \left(\beta^\mathsf{T} \boldsymbol{X}_i; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i) \right), \tag{B.1}$$

where $\psi_\tau(a; y, G) = -\rho_\tau(a; y) + (1-\tau) \int_0^a G(s)\, \mathrm{d}s$ and where $\mathcal{B}$ is a fixed compact parameter space, taken to be the neighborhood of β_τ mentioned in our assumptions. For further convenience and without any loss of generality, the proof is written considering the response variable Y to be positive. This consideration is solely done in the purpose of being coherent with the arbitrary set value of 0 in the correcting term of ψ_τ with respect to v defined in assumption (C3). For instance, one could also easily consider in the following a strictly negative variable Y, but this would possibly require the arbitrary chosen constant 0 to be replaced by any

constant below v as we only wish to control the behavior of the nuisance parameter $\widehat{G}_C(\cdot|\boldsymbol{X})$ below this v.

Next, starting with condition (A1) in DVK, note that the latter is readily satisfied in our framework by construction of $\widehat{\beta}_\tau$. Furthermore, using the definition of $\mathcal{G}$ in (C3) as the space embedding the nuisance parameter G_C, and equiping the latter with the distance $d_{\mathcal{G}}(G_1, G_2) = \sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})} \sup_{y\leq v} |G_1(y|\boldsymbol{x}) - G_2(y|\boldsymbol{x})|$ for any $G_1, G_2 \in \mathcal{G}$, note that (A3) in DVK is straightforwardly satisfied as well provided assumption (C5)-(i) holds. We therefore only need to verify here conditions (A2), (A4) and (A5).

Starting with the identifiability condition (A2) ensuring the uniqueness of β_τ, we need to verify that for any $\epsilon > 0$,

$$\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\psi_\tau\left(\beta_\tau^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right) - \psi_\tau\left(\beta^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\right] > 0,$$

where $||\cdot||$ denotes the Euclidean distance. To that end, using the definition of φ_τ, we can show that

$$\begin{aligned}
&\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\psi_\tau\left(\beta_\tau^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right) - \psi_\tau\left(\beta^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\right] \\
&= \inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\int_{\beta^\mathsf{T}\boldsymbol{X}}^{\beta_\tau^\mathsf{T}\boldsymbol{X}} \left(\mathbb{1}(Y \geq s) - (1-\tau)\bar{G}_C(s|\boldsymbol{X})\right) \mathrm{d}s\right] \\
&= \inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}_{\boldsymbol{X}}\left[\int_{\beta^\mathsf{T}\boldsymbol{X}}^{\beta_\tau^\mathsf{T}\boldsymbol{X}} (1 - G_C(s|\boldsymbol{X}))(\tau - F_{T|\boldsymbol{X}}(s|\boldsymbol{X})) \,\mathrm{d}s\right].
\end{aligned}$$

Under assumptions (C1), (C2) and (C4), the latter expectation is observed to be strictly positive, hereby ensuring condition (A2) is satisfied.

Next, for (A4) to hold, it suffices by Remark 1(ii) in DVK and assumption (C3) to show that the class

$$\mathcal{F} = \left\{(y, \boldsymbol{x}) \mapsto \psi_\tau\left(\beta^\mathsf{T}\boldsymbol{x}; y, G(\cdot|\boldsymbol{x})\right) : \beta \in \mathcal{B}, G \in \mathcal{G}\right\}$$

is Glivenko-Cantelli. For this, by Theorem 2.4.1 in Van der Vaart and Wellner (1996), we need to prove that for all $\epsilon > 0$, the ϵ-bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_1(P))$ of the class $\mathcal{F}$ with respect to the L_1 probability measure on $(Y, \boldsymbol{X})$ is finite. To that end, let $\psi_\tau = \psi_{\tau 1} + \psi_{\tau 2} + \psi_{\tau 3}$, where

$$\begin{aligned}
\psi_{\tau 1}\left(\beta^\mathsf{T}\boldsymbol{x}; y, G(\cdot|\boldsymbol{x})\right) &= -\tau(y - \beta^\mathsf{T}\boldsymbol{x}) \\
\psi_{\tau 2}\left(\beta^\mathsf{T}\boldsymbol{x}; y, G(\cdot|\boldsymbol{x})\right) &= (y - \beta^\mathsf{T}\boldsymbol{x})\mathbb{1}(y < \beta^\mathsf{T}\boldsymbol{x}) \\
\psi_{\tau 3}\left(\beta^\mathsf{T}\boldsymbol{x}; y, G(\cdot|\boldsymbol{x})\right) &= (1-\tau)\int_0^{\beta^\mathsf{T}\boldsymbol{x}} G(s|\boldsymbol{x})\,\mathrm{d}s,
\end{aligned}$$

and let $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$ denote the classes induced by $\psi_{\tau 1}, \psi_{\tau 2}$ and $\psi_{\tau 3}$, respectively.

From this decomposition, it is easy to see that

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) \leq \prod_{j=1}^{3} N_{[]}(\epsilon, \mathcal{F}_j, L_1(P)). \tag{B.2}$$

Now, for the classes $\mathcal{F}_1$ and $\mathcal{F}_2$, suppose for simplicity and without loss of generality that all coordinates of $\boldsymbol{x}$ are positive, and define $M_\epsilon = O(\epsilon^{-2})$ pairs $(\beta_k^L, \beta_k^U), k = 1, \ldots, M_\epsilon$, that cover $\mathcal{B}$, assumed to be compact by (C1), such that $(\beta_k^{L\intercal}\boldsymbol{x}, \beta_k^{U\intercal}\boldsymbol{x})$ define brackets of length $\epsilon\tau/(1-\tau)$ for the class $\{\boldsymbol{x} \mapsto \beta^\intercal \boldsymbol{x} : \beta \in \mathcal{B}\}$ with respect to the L_1-norm. Then, it is straightforward that $N_{[]}(\epsilon, \mathcal{F}_j, L_1(P)) \leq K_j \epsilon^{-2}$ for some finite constants $K_j > 0, j = 1, 2$, which, combined with (B.2), suggests we only have to verify that $N_{[]}(\epsilon, \mathcal{F}_3, L_1(P))$ is bounded in order to prove that condition (A4) holds in our framework.

To that end, by Lemma 6.1 in Lopez (2011) which extends Theorem 2.7.5 in Van der Vaart and Wellner, first note that there exist $N_\epsilon \leq \exp(K_3 \epsilon^{-2/(1+\eta)})$ brackets $(\underline{G}_j, \overline{G}_j), j = 1, \ldots, N_\epsilon$, for a finite constant $K_3 > 0$ such that, under (C3) and for all $G \in \mathcal{G}$, there exists $j = 1, \ldots, N_\epsilon$, for which $\underline{G}_j \leq G \leq \overline{G}_j$, and

$$\int_{\text{supp}(\boldsymbol{X})} \int_0^{\upsilon} \left| \overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x}) \right| \mathrm{d}s \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) < \epsilon, \tag{B.3}$$

where $F_{\boldsymbol{X}}(\boldsymbol{x})$ denotes the c.d.f. of $\boldsymbol{X}$. From this result, our claim for (A4) to hold is that brackets for $\mathcal{F}_3$ are given by $(\underline{\zeta}_{jk}, \overline{\zeta}_{jk}), j = 1, \ldots, N_\epsilon$, $k = 1, \ldots, M_\epsilon$, where

$$\underline{\zeta}_{jk}(\boldsymbol{x}) = (1-\tau) \int_0^{\beta_k^{L\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \, \mathrm{d}s,$$
$$\overline{\zeta}_{jk}(\boldsymbol{x}) = (1-\tau) \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \overline{G}_j(s|\boldsymbol{x}) \, \mathrm{d}s.$$

For this claim to hold, as it is straightforward to verify that for all $\zeta \in \mathcal{F}_3$ there exist $j = 1, \ldots, N_\epsilon$, and $k = 1, \ldots, M_\epsilon$, such that $\underline{\zeta}_{jk} \leq \zeta \leq \overline{\zeta}_{jk}$, we only need to show that

$$\int_{\text{supp}(\boldsymbol{X})} \left| \overline{\zeta}_{jk}(\boldsymbol{x}) - \underline{\zeta}_{jk}(\boldsymbol{x}) \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) < \epsilon, \quad j = 1, \ldots, N_\epsilon, \; k = 1, \ldots, M_\epsilon.$$

To that end, developing the expressions of $\underline{\zeta}_{jk}$ and $\overline{\zeta}_{jk}$, we have that

$$\begin{aligned}
&\int_{\text{supp}(\boldsymbol{X})} \left| \overline{\zeta}_{jk}(\boldsymbol{x}) - \underline{\zeta}_{jk}(\boldsymbol{x}) \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\qquad = (1-\tau) \int_{\text{supp}(\boldsymbol{X})} \left| \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \overline{G}_j(s|\boldsymbol{x}) \mathrm{d}s - \int_0^{\beta_k^{L\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \mathrm{d}s \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}),
\end{aligned}$$

where the latter expression can be bounded above by $T_1 + T_2$ where

$$T_1 = (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\left|\int_0^{\beta_k^{U\intercal}\boldsymbol{x}}\overline{G}_j(s|\boldsymbol{x})\mathrm{d}s - \int_0^{\beta_k^{U\intercal}\boldsymbol{x}}\underline{G}_j(s|\boldsymbol{x})\mathrm{d}s\right|\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}),$$

$$T_2 = (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\left|\int_0^{\beta_k^{U\intercal}\boldsymbol{x}}\underline{G}_j(s|\boldsymbol{x})\mathrm{d}s - \int_0^{\beta_k^{L\intercal}\boldsymbol{x}}\underline{G}_j(s|\boldsymbol{x})\mathrm{d}s\right|\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}).$$

We will now show that both T_1 and T_2 can be bounded above such that their sum is bounded by ϵ. Starting with T_1, we have that

$$\begin{aligned}T_1 &\leq (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\int_0^{\beta_k^{U\intercal}\boldsymbol{x}}\left|\overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x})\right|\mathrm{d}s\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\\ &\leq (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\int_0^{\upsilon}\left|\overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x})\right|\mathrm{d}s\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \leq (1-\tau)\epsilon,\end{aligned}$$

using assumption (C4) and (B.3) for the second and last inequalities, respectively. Concentrating now on T_2, we have that

$$\begin{aligned}T_2 &\leq (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\int_{\beta_k^{L\intercal}\boldsymbol{x}}^{\beta_k^{U\intercal}\boldsymbol{x}}\left|\underline{G}_j(s|\boldsymbol{x})\right|\mathrm{d}s\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\\ &\leq (1-\tau)\int_{\text{supp}(\boldsymbol{X})}\left|\beta_k^{U\intercal}\boldsymbol{x} - \beta_k^{L\intercal}\boldsymbol{x}\right|\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \leq \tau\epsilon,\end{aligned}$$

given the brackets induced by $(\beta_k^{L\intercal}, \beta_k^{U\intercal})$ for the class $\{\boldsymbol{x} \mapsto \beta^\intercal\boldsymbol{x} : \beta \in \mathcal{B}\}$ with respect to the L_1-norm. This completes the proof that $N_{[]}(\epsilon, \mathcal{F}_3, L_1(P))$ is bounded. Hence, we conclude that $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) = O(\exp(K_3\epsilon^{-2/(1+\eta)}))$, from which it follows that condition (A4) holds.

Lastly, for condition (A5), we need to establish that

$$\lim_{d_{\mathcal{G}}(G,G_C)\to 0}\sup_{\beta\in\mathcal{B}}\left|\mathbb{E}\left[\psi_\tau\left(\beta^\intercal\boldsymbol{X}; Y, G(\cdot|\boldsymbol{X})\right) - \psi_\tau\left(\beta^\intercal\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\right]\right| = 0.$$

To that end, note that

$$\begin{aligned}&\sup_{\beta\in\mathcal{B}}\left|\mathbb{E}\left[\psi_\tau\left(\beta^\intercal\boldsymbol{X}; Y, G(\cdot|\boldsymbol{X})\right) - \psi_\tau\left(\beta^\intercal\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\right]\right|\\ &\qquad\leq (1-\tau)\sup_{\beta\in\mathcal{B}}\mathbb{E}_{\boldsymbol{X}}\left[\int_0^{\beta^\intercal\boldsymbol{X}}\left|G(s|\boldsymbol{X}) - G_C(s|\boldsymbol{X})\right|\mathrm{d}s\right].\end{aligned}$$

Under assumption (C4), this expression can then in turn be bounded above by

$$\begin{aligned}&(1-\tau)\int_0^{\upsilon}\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}\left|G(s|\boldsymbol{x}) - G_C(s|\boldsymbol{x})\right|\mathrm{d}s\\ &\qquad\leq (1-\tau)\,\upsilon\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}\sup_{y\leq\upsilon}\left|G(y|\boldsymbol{x}) - G_C(y|\boldsymbol{x})\right|,\end{aligned}$$

which converges to 0 when $d_{\mathcal{G}}(G, G_C) \to 0$, provided assumption (C3) holds. This completes the proof that (A5) holds in our framework. Hence the assumptions of Theorem 1 in DVK are met, from which the weak consistency of $\widehat{\beta}_\tau$ follows. □

Before developing the proof of Theorem 3.2, we now illustrate in the following Lemma how one could replace the general condition (C5)-(ii) in our assumptions by appropriate bandwidth and kernel conditions when considering the particular case of Beran's estimator described in (3.2.8).

Lemma 3.1. *Suppose conditions (C2) and (C3) hold, and that the following assumptions for Beran's estimator in* (3.2.8) *hold as well:*

(C6) The bandwidth h_n satisfies $h_n = O(n^{-\nu})$, for $1/4 < \nu < 1/3$.

(C7) The kernel function $K(\cdot) \geq 0$ is compactly supported and Lipschitz continuous of order 1. Furthermore, $\int K(u)\,\mathrm{d}u = 1$, $\int uK(u)\,\mathrm{d}u = 0$ and $\int K^2(u)\,\mathrm{d}u < \infty$.

Then, uniformly in $\beta \in \mathcal{B}$,

$$\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\left(\widehat{G}_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) - G_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right)\right] = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}_i|\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),$$

where ξ takes the following form for a response variable still considered to be positive:

$$\xi(Y_i, \Delta_i, t|\boldsymbol{x}) = (1 - G_C(t|\boldsymbol{x}))\left[\int_0^{Y_i\wedge t}\frac{-\mathrm{d}F_{Y,\Delta}(s,0|\boldsymbol{x})}{\{1 - F_{Y|\boldsymbol{X}}(s|\boldsymbol{x})\}^2} + \frac{(1-\Delta_i)\mathbb{1}(Y_i \leq t)}{1 - F_{Y|\boldsymbol{X}}(Y_i|\boldsymbol{x})}\right], \tag{B.4}$$

where $F_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = \mathbb{P}(Y \leq t|\boldsymbol{X} = \boldsymbol{x})$ and $F_{Y,\Delta}(t, 0|\boldsymbol{x}) = \mathbb{P}(Y \leq t, \Delta = 0|\boldsymbol{X} = \boldsymbol{x})$.

Proof of Lemma 3.1. Using the i.i.d. expansion of $\widehat{G}_C(c|\boldsymbol{x})$ uniformly in c and $\boldsymbol{x}$, under conditions (C2), (C3), (C6) and (C7), we have (see e.g. Gonzalez-Manteiga and Cadarso-Suarez (1994) or Van Keilegom and Veraverbeke (1997a)):

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}\Big[\boldsymbol{X}\Big(\widehat{G}_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) - G_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\Big)\Big] \\
&= (nh_n)^{-1}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\sum_{i=1}^{n}\frac{K\left(\frac{\boldsymbol{X}-\boldsymbol{X}_i}{h_n}\right)}{(nh_n)^{-1}\sum_{j=1}^{n}K\left(\frac{\boldsymbol{X}-\boldsymbol{X}_j}{h_n}\right)}\xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right] + o_{\mathbb{P}}(n^{-1/2}) \\
&= (nh_n)^{-1}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\boldsymbol{x}K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\xi(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{x}|\boldsymbol{x})\mathrm{d}\boldsymbol{x} + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}$$

where ξ is described in (B.4). Observing that the function $\xi(Y_i, \Delta_i, t|\boldsymbol{x})$ involves an indicator function with respect to t, standard change of variables and a modulus of continuity argument of the empirical distribution function (see Theorem 2.14 in Stute (1982)) combined with assumption (C7) then lead to the desired result. □

Proof of Theorem 3.2. Given that $\widehat{\beta}_\tau$ is shown to be weakly consistent in Theorem 3.1 and that $\widehat{G}_C$ is assumed by (C5)-(i) to be a uniformly consistent estimator of G_C, to prove that our proposed estimator is asymptotically normally distributed we now restrict the spaces $\mathcal{B}$ and $\mathcal{G}$ to shrinking neighborhoods around the true β_τ and G_C in order to avoid possible local minima. That is, we define the spaces $\mathcal{B}_\delta = \{\beta \in \mathcal{B} : ||\beta - \beta_\tau|| \leq \delta_n\}$ and $\mathcal{G}_\delta = \{G \in \mathcal{G} : d_{\mathcal{G}}(G, G_C) \leq \delta_n\}$ for some $\delta_n = o(1)$. In this context, we rely for our proof on the work of Birke et al. (2017) which slightly adapts Theorem 2 of Chen et al. (2003).

We therefore need to verify conditions (C.1)-(C.6) of Proposition 2 in Birke et al. (2017) in order to establish the asymptotic normality of our proposed estimator. To that end, let us first define $M_n(\beta, G) = n^{-1} \sum_{i=1}^n m(Y_i, \boldsymbol{X}_i, \beta, G)$, where

$$m(Y_i, \boldsymbol{X}_i, \beta, G) = \boldsymbol{X}_i \left((1-\tau)(1 - G(\beta^\mathsf{T} \boldsymbol{X}_i | \boldsymbol{X}_i)) - \mathbb{1}(Y_i > \beta^\mathsf{T} \boldsymbol{X}_i)\right).$$

Furthermore, let

$$\begin{aligned} M(\beta, G) &= \mathbb{E}\left[m(Y, \boldsymbol{X}, \beta, G)\right] \\ &= \mathbb{E}_{\boldsymbol{X}}\Big[\boldsymbol{X}\Big\{(1-\tau)(1 - G(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})) \\ &\qquad - (1 - F_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}))(1 - G_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}))\Big\}\Big], \end{aligned}$$

and observe that $M(\beta_\tau, G_C) = 0$.

We now verify the conditions of Proposition 2 in Birke et al.. First, note that (C.1) trivially holds by construction of our estimator. Next, for $\beta \in \mathcal{B}_\delta$ let $\Gamma_1(\beta, G_C)$ denote the ordinary derivative of $M(\beta, G_C)$ with respect to β, that is,

$$\begin{aligned} \Gamma_1(\beta, G_C) &:= \frac{\partial M(\beta, G_C)}{\partial \beta} \\ &= \mathbb{E}_{\boldsymbol{X}}\Big[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\Big\{f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})(1 - G_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})) \\ &\qquad + g_C(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})(\tau - F_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}))\Big\}\Big], \end{aligned}$$

where $g_C(\cdot|\boldsymbol{x})$ denotes the density of C conditionally on $\boldsymbol{X} = \boldsymbol{x}$. Under assumptions (C1)-(C4), $\Gamma_1(\beta, G_C)$ is then observed to be continuous and of full rank at β_τ. Hence, condition (C.2) is satisfied in our framework.

For condition (C.3), define first for all $\beta \in \mathcal{B}_\delta$ the functional derivative of

$M(\beta, G)$ at G_C in the direction $[G - G_C]$ as

$$\begin{aligned}\Gamma_2(\beta, G_C)[G - G_C] &:= \lim_{\eta \to 0} \frac{1}{\eta} \left[M(\beta, G_C + \eta(G - G_C)) - M(\beta, G_C) \right] \\ &= (1 - \tau)\mathbb{E}_{\boldsymbol{X}} \left[\boldsymbol{X} \left(G_C(\beta^\intercal \boldsymbol{X} | \boldsymbol{X}) - G(\beta^\intercal \boldsymbol{X} | \boldsymbol{X}) \right) \right].\end{aligned}$$

We then observe that for all $(\beta, G) \in \mathcal{B}_\delta \times \mathcal{G}_\delta$, $M(\beta, G)$ is linear in G since $M(\beta, G) - M(\beta, G_C) - \Gamma_2(\beta, G_C)[G - G_C] = 0$. This verifies the first part of (C.3). For the second part, we have to show that $||\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] - \Gamma_2(\beta_\tau, G_C)[\widehat{G}_C - G_C]|| = O_{\mathbb{P}}(n^{-1/2})$. To prove this, using assumption (C5)-(ii), we have uniformly in $\beta \in \mathcal{B}_\delta$ that

$$\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] = n^{-1} \sum_{i=1}^{n} \boldsymbol{\phi}(Y_i, \Delta_i, \beta^\intercal \boldsymbol{X}_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),$$

where $\boldsymbol{\phi}(Y_i, \Delta_i, t, \boldsymbol{x}) = -(1 - \tau)\, \boldsymbol{x}\, \xi(Y_i, \Delta_i, t | \boldsymbol{x})$. An application of the central limit theorem then implies that $n^{1/2}\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{V})$ where $\boldsymbol{V}$ is finite under assumptions (C1) and (C5)-(ii). This, in turn, implies that condition (C.3) in Birke et al. is satisfied in our context.

Next, condition (C.4) in Birke et al. is readily satisfied in our context by assumption (C5). To establish that condition (C.5) holds as well, we need to verify the conditions of Theorem 3 in Chen et al. (2003). To that end, let $m = m_c + m_{lc}$, where

$$\begin{aligned} m_c(Y, \boldsymbol{X}, \beta, G) &= \boldsymbol{X}(1 - \tau)(1 - G(\beta^\intercal \boldsymbol{X} | \boldsymbol{X})) \\ m_{lc}(Y, \boldsymbol{X}, \beta, G) &= -\boldsymbol{X} \mathbb{1}(Y > \beta^\intercal \boldsymbol{X}). \end{aligned}$$

Then, condition (3.1) is easily observed to hold for some $s_{1j}, s_j \in (0, 1]$ and $r = 2$ under assumption (C1). For condition (3.2), as m_{lc} does not depend here on G, we first note via the proof of Theorem 3 in Chen et al. that the constant s_j controlling for the regularity of the nuisance parameter and appearing in condition (3.2) may in fact be replaced by the constant s_{1j} already appearing in condition (3.1). Therefore, we will verify condition (3.2) with respect to s_{1j} instead of s_j as initially stated in Chen et al.. To that end, first it can be observed that for all positive values $\epsilon_n = o(1)$,

$$\begin{aligned}\sup_{\beta_\star : ||\beta - \beta_\star|| \leq \epsilon_n} & |\mathbb{1}(Y > \beta_\star^\intercal \boldsymbol{X}) - \mathbb{1}(Y > \beta^\intercal \boldsymbol{X})| \\ &\leq \mathbb{1}(Y > \beta^\intercal \boldsymbol{X} - \epsilon_n ||\boldsymbol{X}||) - \mathbb{1}(Y > \beta^\intercal \boldsymbol{X} + \epsilon_n ||\boldsymbol{X}||).\end{aligned}$$

Hence, we have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\beta_\star:||\beta-\beta_\star||\leq\epsilon_n}\left|\left|\boldsymbol{X}\left(\mathbb{1}(Y>\beta_\star^\intercal\boldsymbol{X})-\mathbb{1}(Y>\beta^\intercal\boldsymbol{X})\right)\right|\right|^2\right]\\
&\leq\mathbb{E}\left[||\boldsymbol{X}||^2\Big\{\mathbb{1}(Y>\beta^\intercal\boldsymbol{X}-\epsilon_n||\boldsymbol{X}||)-\mathbb{1}(Y>\beta^\intercal\boldsymbol{X}+\epsilon_n||\boldsymbol{X}||)\Big\}\right]\\
&=\mathbb{E}_{\boldsymbol{X}}\Big[||\boldsymbol{X}||^2\Big\{H\left(\beta^\intercal\boldsymbol{X}+\epsilon_n||\boldsymbol{X}||\,\Big|\boldsymbol{X}\right)-H\left(\beta^\intercal\boldsymbol{X}-\epsilon_n||\boldsymbol{X}||\,\Big|\boldsymbol{X}\right)\Big\}\Big]\\
&\leq\mathbb{E}_{\boldsymbol{X}}\left[||\boldsymbol{X}||^3K_4\epsilon_n\right],
\end{aligned}$$

for some finite constant K_4 under assumptions (C2) and (C3). Hence, provided assumption (C1) is satisfied, we may observe that condition (3.2) holds for $s_{1j}=1/2$. For the last condition of Theorem 3 in Chen et al., for $\epsilon>0$ denote first by $N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})$ the covering number (Van der Vaart and Wellner (1996, p. 83)) of the class $\mathcal{G}_\delta$ under the sup-norm metric we consider on the latter with a slight abuse of notation. Now, keeping in mind that $N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})\leq N_{[]}(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})$, and since all the functions in the class $\mathcal{G}_\delta$ have values between 0 and 1 by (C3), we first observe that only one ϵ-bracket suffices to cover $\mathcal{G}_\delta$ if $\epsilon>1$. Then, using Lemma 6.1 in Lopez (2011) for a bound on the bracketing number for the case $\epsilon\leq 1$, we have that

$$\begin{aligned}
\int_0^\infty\sqrt{\log N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})}\,\mathrm{d}\epsilon&\leq\int_0^1\sqrt{\log N_{[]}(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})}\,\mathrm{d}\epsilon\\
&\leq K_5\int_0^1\epsilon^{-\frac{1}{1+\eta}}\,\mathrm{d}\epsilon\\
&<\infty,
\end{aligned}$$

for some finite constant K_5, hereby satisfying condition (3.3) in Chen et al. for $s_j=1$. It then follows from their Theorem 3 that condition (C.5) in Birke et al. holds in our context.

For the final condition (C.6), we need here to establish that $n^{1/2}(M_n(\beta_\tau,G_C)+\Gamma_2(\beta_\tau,G_C)[\widehat{G}_C-G_C])$ converges to a normal distribution $\mathcal{N}(0,\Sigma)$ for some positive definite matrix Σ. Recalling that $M_n(\beta_\tau,G_C)=n^{-1}\sum_{i=1}^n m(Y_i,\boldsymbol{X}_i,\beta_\tau,G_C)$ is the average of independent random vectors with mean 0, this follows easily using the same arguments as for the verification of condition (C.3) for the particular case of $\beta=\beta_\tau$. Hence, we obtain

$$n^{1/2}\left[M_n(\beta_\tau,G_C)+\Gamma_2(\beta_\tau,G_C)[\widehat{G}_C-G_C]\right]\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma),$$

where $\Sigma=Cov(\Lambda_i)$ with

$$\Lambda_i=m(Y_i,\boldsymbol{X}_i,\beta_\tau,G_C)-(1-\tau)\,\boldsymbol{X}_i\,\xi(Y_i,\Delta_i,\beta_\tau^\intercal\boldsymbol{X}_i|\boldsymbol{X}_i). \tag{B.5}$$

Theorem 3.2 then follows directly from an application of Proposition 2 in Birke et al., hereby concluding the proof.

□

CHAPTER 4

Linear Censored Quantile Regression: A Novel Minimum-Distance Approach

Abstract

In this work, we investigate a new procedure for the estimation of a linear quantile regression with possibly right censored responses. Contrary to the main literature on the subject, we propose in this context to circumvent the formulation of conditional quantiles through the so-called check loss function that stems from the notorious work of Koenker and Bassett (1978). Instead, our suggestion is here to estimate the quantile coefficients by minimizing an alternative measure of distance. In fact, our approach could be qualified as a generalization in a parametric regression framework of the technique consisting in inverting the conditional distribution of the response given the covariates, motivated by the knowledge that the main literature for censored data already relies on some nonparametric conditional distribution estimation as well. The ideas of effective dimension reduction are then exploited in order to accommodate for higher-dimensional settings as well in this context. Extensive numerical results then suggest that such an approach provides a strongly competitive procedure to the classical approaches based on the check function, in fact both for complete and censored observations. From a theoretical prospect, both consistency and asymptotic normality of the proposed estimator for linear regression are obtained under classical regularity conditions. As a by-product, several asymptotic results on some 'double-kernel' version of the conditional Kaplan-Meier distribution estimator based on effective dimension reduction, and its corresponding density estimator, are also obtained and may be of interest on their own. A brief application of our procedure to quasar data then serves to further highlight the relevance of the latter for quantile regression estimation with censored data. This chapter is based on the paper

> De Backer, M., A. El Ghouch, and I. Van Keilegom. *Linear Censored Quantile Regression: A Novel Minimum-Distance Approach*. Submitted.

4.1 Introduction

In this chapter, based on some insights following Chapter 3, we dig further into the particular linear quantile regression framework. As a preliminary remark, while the proposed methodology could, debatably, solely be devoted to the estimation of a linear model for complete observations, our primary motivation comes again from the context of survival analysis, where possible right-censoring of the response variable may occur. In this situation, our intention is indeed to provide an alternative estimation procedure for the unknown regression coefficients, by taking a step back from the classical approach to estimating conditional quantiles based on the check loss function. For convenience, and in order to clearly contrast our approach with the latter, we start by briefly recalling here the main existing estimation strategies prior to this work. As exposed in Chapter 1 through Figure 1.8, the latter may essentially be decomposed into three main modelling categories, all boiling down to the initial formulation of Koenker and Bassett (1978) in case of no censoring.

As multiply evoked in the previous chapters, the first category of modelling techniques starting from the check-based formulation of quantile regression is based on the so-called inverse-censoring-probability (ICP) weighting scheme, notoriously introduced in mean regression by Koul et al. (1981) and Zhou (1992). Briefly described in its general version in (1.4.6), this approach stems from the observation that the expected conditional check loss of the unobservable response can easily be recovered from the expected loss of the actually observed response, by only keeping the loss of uncensored observations in the procedure and correctly adjusting through the conditional distribution of the censoring variable C given the covariates $\boldsymbol{X}$. As a result, for a flexible modelling and without additional assumptions, smoothing of the latter distribution is required which inevitably introduces a practical limit to the dimension of the covariate one may consider. Additionally, the basic ICP approach suffers from a more fundamental shortfall already exposed earlier, as it fails to take profit of the robustness of quantile regression in its handling of censored observations in opposition to the second strategy recalled below.

This second main modelling technique stems from the perception that all censored observations are not fundamentally to be treated similarly in the context of quantiles, as developed in more details in Section 1.4.2. Employing the idea of redistribution-of-mass (RM), Wang and Wang (2009) provided for this strategy a flexible estimation procedure for the linear regression coefficients, based on the nonparametric estimation of the conditional distribution of the unobservable true response T given $\boldsymbol{X}$. Similarly to the ICP technique, this thus suggests that a flexible modelling here also requires smoothing across the covariates, despite the initial model being parametric.

The last main technique to adapt the check function approach to censored quantile regression may be resumed into two attempts of employing all observations at hand as if they were complete, and appropriately change the 'target' in the check-based formulation. The first attempt, as previously-evoked in Section 1.4.3, goes

back to Lindgren (1997) and suggests to use all observed responses while avoiding underestimation of the quantile function through aiming for a higher quantile level than the actual level of interest. While the idea is quite natural, practical implementation also requires smoothing conditional distributions. Additionally, an iterative procedure is required to determine the appropriate quantile level that should be targeted. As a result, the literature based on this approach is noticeably more modest. The second attempt of changing the 'target' was developed in Chapter 3, where it is suggested to work on the loss function itself instead of considering weighting schemes or alternative quantile levels. The methodology results in a simple adaptation of the check function and has a broad application range in terms of modelling potential. However, similarly to the above-described literature, for flexible modelling in a linear framework one is again conducted to consider smoothing a conditional distribution (of C given $\boldsymbol{X}$) even though the initial model is parametric.

In short, these three main approaches all share two basic common features: they are 'check-function-based' and are limited to moderate covariate dimension if flexibility is required without additional assumptions. This then raises the legitimate question of whether a check-based approach, although natural from a modeling point of view, is the most appropriate for censored quantile regression given the natural covariate dimension constraint one faces without involving additional assumptions. In fact, as mentioned above, this question may also be considered for complete observations as one may wonder if a check-based approach is systematically to be considered as a gold standard for linear regression with moderate number of covariates, although the motivation is here arguably less obvious than for censored data.

An additional argument may be considered to challenge the supremacy of check-function-based approaches, as censored quantile regressions are, by essence, more constrained to central quantile levels than for complete observations, at least regarding upper quantile levels. This comes from a basic identifiability condition requiring that not all censoring occurs below the quantile function with probability one, hereby introducing a natural upper bound to the quantile level one may consider in practice. It then seems natural to wonder if check-based approaches are best suited to make the most of this 'centrality of the quantile level' constraint.

With these considerations in mind, the main contribution of this work is to propose an alternative to the above-mentioned literature on linear censored quantile regression by circumventing the check-based modelling. In fact, as exposed in the simplistic representation of Figure 1.8, our procedure may be seen as a very simple adaptation to linear regression of the well-known inverse cumulative distribution function technique (inverse-c.d.f.), which first estimates the conditional c.d.f. of the response variable given the covariates, and then recovers the quantile function by inversion. As mentioned in Chapter 1, this approach is well-studied in the literature on nonparametric quantile regression (see e.g. for complete observations Yu and Jones (1998), Li and Racine (2008) and Li and Racine (2017)), which are,

by construction, limited to small covariate dimensions. Although this technique was, to the best of our knowledge, never studied for linear regression with complete observations given it induces smoothing for a parametric model, we note that flexible censored quantile regression seems an appropriate candidate to carry out the extension of an inverse-c.d.f. approach to a linear model given the inherent smoothing arguments of the literature described above. Additionally, numerical results of the nonparametric literature for complete observations (see e.g. Yu and Jones (1998)) suggest that inverse-c.d.f. approaches tend, quite naturally, to outperform their check-based competitors when it comes to the estimation of quantile functions for central quantile levels. From the argumentation described above, this again suggests that an inverse-c.d.f. approach may be a valuable alternative to the existing literature for censored data. Lastly, to accommodate for higher dimensional covariates, similarly to the existing literature we introduce here an additional assumption to overcome the curse of dimensionality, and suggest to adopt a general dimension reduction approach as in Wang et al. (2013). The resulting estimator is then observed for both small and moderate covariate dimensions to perform very competitively with respect to its check-based competitors, hereby providing a simple and valuable alternative for robust censored quantile regression estimation.

The rest of the chapter is as follows. Section 4.2 is devoted to the newly proposed estimation procedure. Consistency and asymptotic normality of the estimator are obtained in Section 4.3 under classical assumptions. Additionally, some new asymptotic results on conditional distribution and density estimators with censored responses and covariates estimated through dimension reduction techniques are also provided as a by-product, given these are required for establishing the results of Section 4.3. An extensive Monte Carlo simulation study is then conducted in Section 4.4, where the, sometimes spectacular, differences between the main existing estimators and our new procedure are clearly exposed to originate from the duality between check-function-based and inverse-c.d.f. approaches. Section 4.5 highlights a brief application to real data. Lastly, Section 4.6 brings a conclusion to the chapter while the proofs of Section 4.3 are deferred to Section 4.7.

4.2 The Proposed Methodology

We develop in this section the proposed methodology for a censored linear quantile regression model. To that end, we start by recalling for convenience a few generalities on conditional quantile modelling for complete observations that will help us motivate our estimation procedure in the next sections.

4.2.1 Background on quantile regression modelling for complete observations

Similarly to the rest of this thesis, let us first denote by $m_\tau(\boldsymbol{x})$, for any $\tau \in (0,1)$, the τ-th conditional quantile function of a continuous response variable T, or some

monotone transformation of the latter, given $\boldsymbol{X} = \boldsymbol{x}$, where $\boldsymbol{X}$ is a covariate vector of dimension $d \geq 1$. Specifically,

$$m_\tau(\boldsymbol{x}) = \inf\{t : F_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \geq \tau\}, \tag{4.2.1}$$

where $F_{T|\boldsymbol{X}}$ denotes the conditional c.d.f. of T given $\boldsymbol{X}$. As already exploited throughout the previous chapters, Koenker and Bassett proposed an equivalent formulation for conditional quantiles as resulting from a minimization problem given by

$$m_\tau(\boldsymbol{x}) = \arg\min_a \mathbb{E}\Big[\rho_\tau(T-a)|\boldsymbol{X} = \boldsymbol{x}\Big], \tag{4.2.2}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u \leq 0))$ is the "check" loss function, and $\mathbb{1}(\cdot)$ is the indicator function. The attractiveness of this check-based formulation can be seen as being, at least, twofold. First, expressing the problem as minimizing a conditional expected loss allows one to retrieve a similar, and hence familiar, framework as for mean regressions, and second, the computational features of minimizing the check function turn out to be very tractable, even for higher dimensions of the covariate (see e.g. Koenker (2005)).

However, in situations where these two characteristics are not fundamental, formulation (4.2.1) may also emerge as a satisfying starting point to estimate conditional quantiles in practice. For instance, a wide literature on nonparametric quantile regression for complete observations (see e.g. Yu and Jones (1998), Li and Racine (2008) and Li et al. (2013)), where the dimension of the covariate is restricted by construction, is based on (4.2.1) and hence proposes to estimate $m_\tau(\boldsymbol{x})$ by $\widehat{m}_\tau(\boldsymbol{x}) = \inf\{t : \widehat{F}(t|\boldsymbol{x}) \geq \tau\}$, where $\widehat{F}$ is an appropriate estimator of $F_{T|\boldsymbol{X}}$. As recalled in Chapter 1, the latter technique is often referred to as 'inverse-c.d.f.' for rather obvious reasons. Interestingly, the numerical results in this literature globally suggest that the latter method tends to outperform its check-based counterpart, especially for 'central' quantile levels where one is not required to estimate $F_{T|\boldsymbol{X}}$ in the tails of the response. For example, Yu and Jones explicitly advocate for the use of the inverse-c.d.f. version of their local linear estimator rather than the check-based version. One might then conjecture that, for problems where the dimension of the covariate is intrinsically restrained, an inverse-c.d.f. approach may numerically perform better than check-based competitors. This statement motivates the next section.

4.2.2 Censored linear regression with low-dimensional covariates

Now, recall that we intend to work in this chapter, and similarly to Chapter 3, on a linear regression model. That is, we assume that

$$m_\tau(\boldsymbol{X}_i) = \beta_\tau^\mathsf{T} \boldsymbol{X}_i, \tag{4.2.3}$$

for $i = 1, \ldots, n$, where $\boldsymbol{X}$ is now a $(d+1)$-dimensional covariate vector whose first component coinciding with the intercept is taken to be 1, and β_τ is the $(d+1)$-dimensional unknown quantile coefficient vector. Furthermore, we assume to be confronted to censored data, that is, instead of completely observing the response T, we only observe $Y = \min(T, C)$, where C is the censoring variable. Our objective is then to estimate the true value of β_τ based on i.i.d. triplets $(Y_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, from $(Y, \Delta, \boldsymbol{X})$, where $\Delta = \mathbb{1}(T \leq C)$ and T and C are assumed to be conditionally independent given the covariate $\boldsymbol{X}$.

In this context, bearing in mind that the previously-introduced check-based estimators in the literature already involve the estimation of conditional distributions, and observing that for all $i = 1, \ldots, n$, $F_{T|\boldsymbol{X}}(\beta_\tau^\intercal \boldsymbol{X}_i | \boldsymbol{X}_i) = \tau$, a natural extension of the inverse-c.d.f. technique is to estimate β_τ by

$$\widehat{\beta}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \left(\widehat{F}(\beta_\tau^\intercal \boldsymbol{X}_i | \boldsymbol{X}_i) - \tau \right)^2, \tag{4.2.4}$$

where $\widehat{F}$ is an appropriate estimator of $F_{T|\boldsymbol{X}}$. Note that (4.2.4) considers a sum of distances that are, in absolute value, necessarily smaller than one. Hence, while a L_1-distance is usually employed in the context of quantile regression to control for arbitrary large distances, we propose in this chapter to use the computationally less complicated squared distance as reported in (4.2.4). Additionally, we suggest here to estimate $F_{T|\boldsymbol{X}}$ nonparametrically using a 'double-kernel' approach, as first introduced for censored data by Leconte et al. (2002). Before motivating this choice, let us first denote the i-th order statistic of the uncensored responses by $Y_{(i)}^u$, $n^u = \sum_{i=1}^n \Delta_i$, and let $H(t) = \int_{-\infty}^t \widetilde{K}(u)\,\mathrm{d}u$ for some kernel density $\widetilde{K}$. We then propose to estimate $F_{T|\boldsymbol{X}}$ by $\widehat{F}^s_{T|\boldsymbol{X}}$, where

$$\begin{aligned} \widehat{F}^s_{T|\boldsymbol{X}}(t|\boldsymbol{x}) &= \int_{\mathbb{R}} H\left(\frac{t-u}{h_T}\right) \mathrm{d}\widehat{F}_{T|\boldsymbol{X}}(u|\boldsymbol{x}) \\ &= \sum_{i=1}^{n^u} \left(\widehat{F}_{T|\boldsymbol{X}}(Y_{(i)}^u|\boldsymbol{x}) - \widehat{F}_{T|\boldsymbol{X}}(Y_{(i-1)}^u|\boldsymbol{x}) \right) H\left(\frac{t - Y_{(i)}^u}{h_T}\right), \end{aligned} \tag{4.2.5}$$

where h_T is a positive bandwidth parameter, $Y_{(0)}^u = 0$ and $\widehat{F}_{T|\boldsymbol{X}}$ is Beran's local Kaplan-Meier estimator (Beran (1981)), recalled to be defined as

$$\widehat{F}_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{B_{ni}(\boldsymbol{x})}{\sum_{j=1}^n \mathbb{1}(Y_i \leq Y_j) B_{nj}(\boldsymbol{x})} \right\}^{\mathbb{1}(Y_i \leq t,\, \Delta_i = 1)}, \tag{4.2.6}$$

for a sequence of weights $B_{nj}(\boldsymbol{x})$ adding up to 1. Similarly to Wang and Wang (2009), Leng and Tong (2013) and the procedure of Chapter 3, we adopt here

Nadaraya-Watson type of weights given by

$$B_{nj}(\boldsymbol{x}) = \frac{K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_{\boldsymbol{X}}}\right)}{\sum_{k=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_k}{h_{\boldsymbol{X}}}\right)}, j = 1, \ldots, n,$$

where K_d is some multivariate kernel density function, $K_d(\boldsymbol{u}/h_{\boldsymbol{X}}) = K_d(u_1/h_{\boldsymbol{X}}, \ldots, u_d/h_{\boldsymbol{X}})$, and $h_{\boldsymbol{X}}$ is a positive bandwidth parameter converging to zero as $n \to \infty$. Furthermore, as commonly endorsed in the literature, in the following we will adopt product kernels $K_d(u_1, \ldots, u_d) = \prod_{i=1}^{d} K(u_i)$, where K is a univariate kernel function. As a remark, note that $\widehat{F}^s_{T|\boldsymbol{X}}$ may be read in (4.2.5) similarly to the kernel version of the empirical c.d.f., except the jumps n^{-1} are here naturally replaced by the jumps of Beran's estimator. Additionally, when there is no censoring, we point out that the estimation procedure through (4.2.4) and (4.2.5) also offers a new alternative to quantile regression estimation for complete observations, with $\widehat{F}_{T|\boldsymbol{X}}$ boiling down to the kernel estimator of Stone (1977).

Now, while the bandwidth in the '$\boldsymbol{X}$-direction' in (4.2.5) plays essentially the same role as the bandwidths in the literature on censored quantile regression, adding a second one in the 'T-direction' might seem unappealing at first. However, note that this approach is similar for instance to Yu and Jones and Li and Racine, and the advantage is here twofold: first, from a pure optimization point of view, it allows for defining a smooth function in the 'T-direction' as well, the latter being simpler to optimize when inserted in (4.2.4) than a crude step function, and second, this choice is motivated by the knowledge coming from a large literature on both conditional and unconditional distribution estimation that smoothing in this direction as well can produce both asymptotic and finite sample efficiency gains by reducing variance at the cost of an increase in bias; see Azzalini (1981), Reiss (1981) and Hansen (2004) to cite a few. One might then conjecture that such implications may spread to our procedure as well. Finally, our practical experience also suggests that the overall procedure is, as one might expect, not very sensitive to the value of h_T, as will be illustrated in Section 4.5. We nevertheless discuss in Section 4.4 a practical procedure to select both bandwidths.

We complete this section by providing a last comment on the analysis of our estimator, as we observe that the proposed methodology forces one to evaluate the double-kernel estimator of $F_{T|\boldsymbol{X}}$ in the conditional part in all $\boldsymbol{X}_i$, $i = 1, \ldots, n$. This implies that we have to evaluate $\widehat{F}^s_{T|\boldsymbol{X}}$ in tail regions of the covariates as well, where the latter estimator may possibly be instable due to sparcity of the data or boundedness of the domain. While the effect of this might already be attenuated through the summation over all observations in our estimation procedure, we note that a possible remedy and extension could be to suitably weigh or even trim some observations in (4.2.4), although the impact of this strategy is left for further investigation.

4.2.3 Censored linear regression for higher-dimensional covariates

When the dimension of the covariates becomes too large to reasonably estimate β_τ using (4.2.4) and (4.2.5), similarly to the existing literature on censored quantile regression, we inevitably need to introduce an additional assumption in our model. In this work, in the same spirit as Wang et al., we propose to adopt a very general global dimension reduction assumption. The latter states that all the information regarding the dependence of T on $\boldsymbol{X}$ is in fact contained in q linear combinations of $\boldsymbol{X}$, that is

$$T \perp\!\!\!\perp \boldsymbol{X} | (\gamma_1^\mathsf{T} \boldsymbol{X}, \ldots, \gamma_q^\mathsf{T} \boldsymbol{X}), \tag{4.2.7}$$

where $\perp\!\!\!\perp$ stands for independence, $q < d+1$ for an effective dimension reduction (EDR), and where the γ's are unknown $(d+1)$-dimensional linearly independent vectors. Such an assumption was first introduced in the pioneering work of Li (1991) for complete observations, and is qualified in the latter paper as the weakest form of assumption one may invoke in the hope that a low-dimensional projection of a covariate vector may contain all the regression information of $T|\boldsymbol{X}$, given no assumptions are made on the actual functional form of the dependence structure. In fact, it is easily seen that identifiability of the γ's is not implied by (4.2.7) as such given that conditioning on any affine transformation of $(\gamma_1^\mathsf{T} \boldsymbol{X}, \ldots, \gamma_q^\mathsf{T} \boldsymbol{X})$ will be equivalent to conditioning simply on the latter. Hence, only the q-dimensional linear subspace of $\mathbb{R}^q$ spanned by the γ's is identifiable. The latter is often referred to as EDR space, and any direction belonging to the latter is called an EDR-direction. Estimation of EDR-directions is a fertile domain of research on its own; see e.g. Cook (1998) or Cook and Ni (2005) for the case of complete observations. For censored data, estimation of the EDR-directions can be carried out using for instance the sliced inverse regression methodology described in Li et al. (1999), or the hazard-based minimum average variance estimation method of Xia et al. (2010). Both methodologies are shown in the latter works to yield $n^{1/2}$-consistent estimators $\widehat{\gamma}_{0,j}$ of a set of EDR-directions $\gamma_{0,j}$, for $j = 1, \ldots, q$.

Next, denoting $\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,q})^\mathsf{T}$ with $Z_{i,j} = \gamma_{0,j}^\mathsf{T} \boldsymbol{X}_i$, $i = 1 \ldots, n$, $j = 1, \ldots, q$, the direct implication of (4.2.7) in our context is that we may restrict ourselves in our procedure to the estimation of $F_{T|\boldsymbol{Z}}$, the conditional distribution of T given the projections of $\boldsymbol{X}$, instead of $F_{T|\boldsymbol{X}}$. That is, for higher-dimensional covariates, given $(\widehat{\gamma}_{0,1}, \ldots, \widehat{\gamma}_{0,q})$ a $n^{1/2}$-consistent estimator of $(\gamma_{0,1}, \ldots, \gamma_{0,q})$ as will be required in Section 4.3, we propose to estimate β_τ by

$$\widehat{\beta}_\tau = \arg\min_\beta \sum_{i=1}^n \left(\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta^\mathsf{T} \boldsymbol{X}_i | \widehat{\boldsymbol{Z}}_i) - \tau \right)^2, \tag{4.2.8}$$

where $\widehat{\boldsymbol{Z}}_i = (\widehat{Z}_{i,1}, \ldots, \widehat{Z}_{i,q})^\mathsf{T}$ with $\widehat{Z}_{i,j} = \widehat{\gamma}_{0,j}^\mathsf{T} \boldsymbol{X}_i$, $j = 1, \ldots, q$, $i = 1 \ldots, n$, and where $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ is the double-kernel estimator of $F_{T|\boldsymbol{Z}}$, as described in (4.2.5) but replacing $\boldsymbol{X}$ with its estimated projections. Lastly, the selection of the value of q,

representing the last unknown in (4.2.8), has naturally been the subject of numerous proposals in the literature on both complete and censored observations. In practice, it is worth mentioning that the choice of q is encouraged to be done along with visual inspection and examination of the eigenvalues involved in the methodologies for estimating the γ_0's (see e.g. Remark 5.2 in Li (1991)). Nevertheless, for an automated choice as will be the case in our simulation studies, we mention here as in Wang et al. the Chi-squared test of Li, or the more computationally intensive cross validation method of Xia et al.. The numerical performance of our resulting estimator for β_τ will then be studied in Section 4.4.

4.3 Large Sample Properties

We establish in this section both the consistency and asymptotic normality of our proposed estimator $\widehat{\beta}_\tau$ defined in (4.2.8). To that end, we start by reporting the set of regularity conditions that are assumed to hold in order to prove the desired results. As a preliminary remark and to avoid any confusion, we emphasize here on the fact that some assumptions (in particular assumptions (C5) and (C9) given below) are written considering explicitly the case $q = 1$ in (4.2.8). This is for reading convenience only, as the assumptions may easily be written for a general $q > 1$ but the notations are unnecessarily more involved, that is, instead of conditioning on a one-dimensional variable, one would have to write down the same assumptions but conditionally on q variables. The remaining assumptions where the dimension q has a more important role, such as the bandwidth assumptions, are of course written for a general value of q. Finally, some auxiliary results which may be of interest on their own concerning the estimator $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ and its corresponding density estimator are also established, but deferred to Section 4.7 for sake of brevity. The set of assumptions is then the following:

(C1) The support $\mathrm{supp}(\boldsymbol{X})$ of $\boldsymbol{X}$ is contained in a compact subset of $\mathbb{R}^{d+1}$, and the variance-covariance matrix of $\boldsymbol{X}$ is positive definite.

(C2) There exists a neighborhood $\mathcal{B}$ of β_τ such that for $\beta \in \mathcal{B}$, $\inf_{\beta\in\mathcal{B}} \inf_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{x}|\boldsymbol{x}) > 0$, where $f_{T|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ denotes the conditional density function of T given $\boldsymbol{X} = \boldsymbol{x}$.

(C3) There exist EDR-directions $\gamma_{0,j}$, $j = 1, \ldots, q$, belonging to a subset Ξ of $\mathbb{R}^{d+1}$ such that $\widehat{\gamma}_{0,j} - \gamma_{0,j} = O_{\mathbb{P}}(n^{-1/2})$, $j = 1, \ldots, q$.

(C4) The univariate kernel functions $K(\cdot)$ and $\widetilde{K}(\cdot)$ are compactly supported. Furthermore, $K(\cdot)$ is a continuously differentiable function of order ν satisfying $\int K(u)\mathrm{d}u = 1$, $\int K^2(u)\mathrm{d}u < \infty$ and $\int u^j K(u)\mathrm{d}u = 0$ for $j < \nu$, where $\nu \geq 2$ is an integer. $\widetilde{K}(\cdot)$ is a continuously differentiable function satisfying $\int \widetilde{K}(u)\mathrm{d}u = 1$, $\int u\widetilde{K}(u)\mathrm{d}u = 0$ and $\int \widetilde{K}^2(u)\mathrm{d}u < \infty$.

(C5) Define the (possibly infinite) time $\tau_{F_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})} = \inf\{t : F_{Y|\boldsymbol{X}}(t|\boldsymbol{x}) = 1\}$, where $F_{Y|\boldsymbol{X}}$ designates the conditional c.d.f. of Y given $\boldsymbol{X}$. Suppose first that there exists a real number $\upsilon < \tau_{F_{Y|\boldsymbol{X}}(\cdot|\boldsymbol{x})}$ for all $\boldsymbol{x}$ in $\mathrm{supp}(\boldsymbol{X})$. Define next $\boldsymbol{Z} = \gamma_0^\mathsf{T}\boldsymbol{X}$ and recall that we simplify the notations here by writing down only the case $q = 1$. For a finite value M, denote then by $\mathcal{F}$ the class of functions $F(t, \boldsymbol{z}) :]-\infty, \upsilon] \times \mathrm{supp}(\boldsymbol{Z}) \to [0, M]$ that have bounded second order derivatives with respect to t (uniformly in $\boldsymbol{z}$), have partial derivatives of order $(\nu - 1)$, for ν in (C4), with respect to $\boldsymbol{z}$ of bounded variation in t (uniformly in $\boldsymbol{z}$), and have bounded (uniformly in t) ν-th order partial derivatives with respect to $\boldsymbol{z}$ which are, uniformly in t, Lipschitz of order η for some $0 < \eta < 1$. Then:

- (i) Define by $\mathcal{F}_F$ the class of distributions $(t, \boldsymbol{x}) \mapsto F_{T|\gamma^\intercal\boldsymbol{X}}(t|\gamma^\mathsf{T}\boldsymbol{x})$ for some $\gamma \in \Xi$, such that $(t, \boldsymbol{z}) \mapsto F_{T|\gamma^\intercal\boldsymbol{X}}(t|\boldsymbol{z})$ belongs to $\mathcal{F}$. Suppose that $F_{T|\boldsymbol{Z}} \in \mathcal{F}_F$.
- (ii) Define by $\mathcal{F}_f$ the class of densities $(t, \boldsymbol{x}) \mapsto f_{T|\gamma^\intercal\boldsymbol{X}}(t|\gamma^\mathsf{T}\boldsymbol{x})$ for some $\gamma \in \Xi$, such that $(t, \boldsymbol{z}) \mapsto f_{T|\gamma^\intercal\boldsymbol{X}}(t|\boldsymbol{z})$ belongs to $\mathcal{F}$. Suppose that $f_{T|\boldsymbol{Z}} \in \mathcal{F}_f$ where $f_{T|\boldsymbol{Z}}$ is the conditional density corresponding to $F_{T|\boldsymbol{Z}}$.

(C6) The first-order partial derivative of $G_C(t|\boldsymbol{z})$ with respect to t is uniformly bounded with respect to both t and $\boldsymbol{z}$. In addition, $G_C(t|\boldsymbol{z})$ has bounded (uniformly in t) partial derivatives with respect to $\boldsymbol{z}$ up to order ν, where ν is in (C4).

(C7) For every $\boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})$ and for $\beta \in \mathcal{B}$ defined in (C2), the point $\beta^\mathsf{T}\boldsymbol{x} \in \mathbb{R}$ lies below υ defined in (C5).

(C8) For some finite constant $C > 0$, the bandwidths $h_{\boldsymbol{X}}$ and h_T satisfy:

- (i) $h_{\boldsymbol{X}} \equiv Cn^{-u}$ with $0 < u < 1/(q+2)$, and $h_T \equiv Cn^{-w}$ with $w < 1 - qu$ and $w \leq u(\nu + 1)$ where ν is in (C4).
- (ii) $h_{\boldsymbol{X}} \equiv Cn^{-u}$ with $1/(2\nu) < u < 1/(3q)$ for $\nu > 2q$ and $h_T \equiv Cn^{-w}$ with $1/4 < w < 1/2 - qu$.

(C9) The following technical conditions, written for $q = 1$ for ease of reading, are assumed to hold:

- (i) For $\gamma \in \Xi$, let $L_{Y|\gamma^\intercal\boldsymbol{X}}(t|\gamma^\mathsf{T}\boldsymbol{x}) = F_{Y|\gamma^\intercal\boldsymbol{X}}(t|\gamma^\mathsf{T}\boldsymbol{x}) f_{\gamma^\intercal\boldsymbol{X}}(\gamma^\mathsf{T}\boldsymbol{x})$ where $F_{Y|\gamma^\intercal\boldsymbol{X}}$ is the conditional c.d.f. of Y given $\gamma^\mathsf{T}\boldsymbol{X}$, and $f_{\gamma^\intercal\boldsymbol{X}}$ denotes the density of $\gamma^\mathsf{T}\boldsymbol{X}$. Then, $\sup_{t \leq \upsilon} \sup_{\boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})} \sup_{\gamma \in \Xi} \nabla_\gamma L_{Y|\gamma^\intercal\boldsymbol{X}}(t|\gamma^\mathsf{T}\boldsymbol{x}) < \infty$, where υ is defined in (C5) and where $\nabla_\gamma g(\gamma)$ denotes the vector of partial derivatives of a function $g(\gamma)$ with respect to all occurrences of γ.
- (ii) Let $f_{\boldsymbol{X}|\gamma_0^\intercal\boldsymbol{X}}$ denote the conditional density of $\boldsymbol{X}$ given $\gamma_0^\mathsf{T}\boldsymbol{X}$. Then, the partial derivatives of $f_{\boldsymbol{X}|\gamma_0^\intercal\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{z})$ with respect to $\boldsymbol{x}$ up to order ν are uniformly bounded with respect to both $\boldsymbol{x}$ and $\boldsymbol{z}$.

We briefly comment here the reported set of assumptions before stating the main theoretical results of this section. First, note that assumptions (C1) and (C2) are typical in the literature on both complete and censored quantile regression, as these are namely required for the uniqueness of β_τ. Condition (C3) for its part requires the $n^{1/2}$-consistency of the estimators for EDR-directions. The latter condition serves to control for the impact of handling estimated observations when establishing the asymptotic properties of $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ and its corresponding density estimator, on which our proofs will rely. In particular, condition (C3) will be observed to imply, namely through Lemma 4.1, that the estimation of the EDR-directions will be of no impact on the asymptotic distribution of $\widehat{\beta}_\tau$. Note also that this assumption is similar to assumption A4-(i) in Wang et al. (2013) and is for instance verified for the estimation procedures of Li et al. (1999) or Xia et al. (2010) under appropriate higher-level conditions, as already mentioned earlier. The latter assumptions typically involve, first, smoothness conditions on the involved conditional distributions that are in coherence with our assumptions, and second, either the so-called and well-studied 'linear condition' associated to sliced inverse regression (Li et al.), or additional conditions associated to kernel smoothing techniques (Xia et al.).

Next, assumption (C4) reports conventional conditions on both kernel functions used in our framework that are needed in order to establish appropriate asymptotic properties of $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ and its corresponding density estimator, with a higher-order kernel possibly required for the '$\boldsymbol{X}$-direction', as is usual in multivariate kernel frameworks. Assumption (C5) for its part defines general classes of functions embedding the true $F_{T|\boldsymbol{Z}}$ and its corresponding density $f_{T|\boldsymbol{Z}}$. The definition of these classes comes as a more restricted version of a general class definition reported in the work of Lopez (2011), where it is established that the latter is in some sense well-behaved in terms of size, which is represented by the notion of bracketing numbers (see e.g. Van der Vaart and Wellner (1996, p. 83)). Note that this condition is somewhat similar to condition (C3) in Chapter 3, but extended here for the inclusion of the dimension reduction framework. Conditions (C6) and (C7) are likewise classical in the literature on censored quantile regression, and may for instance be also found in the work of Chapter 3 and Wang and Wang (2009) to name a few. Additionally, assumption (C7) is to be brought in parallel with previously-developed comments on constraints on the quantile level τ when confronted to censored data, as the latter condition is observed to define a natural upper bound on the quantile level one may consider in this context. Assumption (C8) then specifies the conditions on both bandwidths used in our estimation procedure, where (i) is required for establishing the consistency of $\widehat{\beta}_\tau$, and the stronger condition (ii) is needed for asymptotic normality. Lastly, (C9) considers two purely technical conditions that are originating, for (i), from the handling of estimated observations in order to establish consistency, and required for the asymptotic normality for (ii).

We now state in the following theorem the consistency of our proposed estimator $\widehat{\beta}_\tau$, for which the proof is deferred to Section 4.7.

Theorem 4.1. *Assume that the censoring time C is conditionally independent of the survival time T given the covariates $\boldsymbol{X}$, and that the triples $(Y_i, \Delta_i, \boldsymbol{X}_i), i = 1, \ldots, n$, form an i.i.d. multivariate random sample. Then, under assumptions (C1)-(C8)-(i) and (C9)-(i), for a given quantile level $0 < \tau < 1$ and for $\widehat{\beta}_\tau$ defined in* (4.2.8)*,*

$$\widehat{\beta}_\tau \to \beta_\tau,$$

in probability, as $n \to \infty$.

The proof of Theorem 4.1 is built on the general empirical-process-based theory of Chen et al. (2003), for which the key requirement is a uniform consistency result for the infinite-dimensional nuisance parameter appearing in the estimation procedure. More precisely, our proof reveals that the latter uniform consistency result is required in our context for both $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ and $\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}$, where $\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}$ is the double-kernel density estimator corresponding to $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$. As a consequence, these results are thus established in Section 4.7 under the above-reported assumptions, and may be of interest on their own as they extend previous work of Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Akritas (1999) to name a few, by considering the inclusion of two smoothing directions and multivariate estimated observations through a dimension reduction technique.

The next theorem reports the limiting distribution of our estimator $\widehat{\beta}_\tau$, for which the crucial requirement is now a linear representation of the infinite-dimensional nuisance parameter on which the estimation procedure is built. For this proof, relying again on the work of Chen et al. and contrary to the proof of the consistency, it is observed that only the estimation of $F_{T|\boldsymbol{Z}}$ will influence the asymptotic covariance matrix of our estimator, and hence only a linear representation for $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ is here required. The latter is again reported in Section 4.7 and extends previous work of, for instance, Gonzalez-Manteiga and Cadarso-Suarez (1994), Van Keilegom and Veraverbeke (1997a) and Du and Akritas (2002).

Theorem 4.2. *For a given $0 < \tau < 1$, under the assumptions of Theorem 4.1, (C8)-(ii) and (C9)-(ii),*

$$n^{1/2}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma_1^{-1}\Sigma\Gamma_1^{-1}),$$

where $\Gamma_1 = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\intercal f^2_{T|\boldsymbol{X}}(\beta_\tau^\intercal \boldsymbol{X}|\boldsymbol{X})]$, and $\Sigma = Cov(g_i(\gamma_0^\intercal \boldsymbol{X}_i))$ with $g_i(\cdot)$ defined in (D.5) *in Section 4.7.*

As a preliminary remark, we point out that Theorem 4.2 is here written explicitly for the case $q = 1$, although this is for notational convenience only, similarly to some of our assumptions listed above. Of course, the latter theorem, and in particular the expression of the function $g_i(\cdot)$ may easily be considered for the general case of $q > 1$, but this would require the inconvenient introduction of several notations for the functions involved in $g_i(\cdot)$ that would be conditional on q

arguments instead of simply one as reported here. In any case, while Theorem 4.2 reports the asymptotic behavior of our estimator, when inference is of interest a general consensus in the censored quantile regression literature is to advocate for bootstrap procedures instead, given that the asymptotic covariance matrix depends on several unknown quantities that are cumbersome to estimate in practice. Similarly to Portnoy (2003), Wang and Wang (2009) and the procedure of Chapter 3 for instance, we therefore consider here for inference a simple percentile bootstrap procedure, where 95% bootstrap confidence intervals for β_τ are constructed by taking the 2.5th and 97.5th percentiles of the bootstrap coefficients, obtained by drawing a sufficient amount of bootstrap samples through resampling $(Y_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, with replacement. The validity of the latter technique in our framework will be empirically investigated in Section 4.4.3.

4.4 Simulation Study

This section is devoted to the finite sample performance of our estimator through Monte Carlo simulations for both small and larger dimensions of the covariates. To that end, focusing first on univariate settings, we start by describing the different estimators that will enter our simulation study along with their practical implementation for replicability of the results, and develop next the different considered data generating processes (DGP). Multivariate settings will then be considered in the second part of this section, while a simple illustration of the effectiveness of the bootstrap procedure proposed in Section 4.3 will complete the section. All of our simulations are carried out using the statistical computing environment R (R Core Team (2017)).

4.4.1 Small-dimensional covariates

As our main motivation comes from the world of censored data, we first outline the main competing procedures we consider for the estimation of a censored linear quantile regression with a covariate of dimension $d = 1$. The choice of these particular estimators is inspired by the introduction of this chapter, resulting in the following:

ICP: Check-based estimator of Bang and Tsiatis (2002) embodying, as in Chapter 3, the ICP-technique described earlier. In opposition to the other procedures, this estimator further assumes that the censoring variable is independent of the covariate. As a result, implementation is carried out using the `rq` function in the R library `quantreg` with incorporation of the weights $\Delta_i/(1-\widehat{G}_C(Y_i))$, $i = 1, \ldots, n$, where G_C is the c.d.f. of C and $\widehat{G}_C$ is the Kaplan-Meier estimator of G_C. The latter is simply recovered for instance from (4.2.6) by replacing the weights $B_{ni}(\boldsymbol{x})$ with n^{-1}, and Δ_i with $1-\Delta_i$ for all $i = 1, \ldots, n$.

RM: Check-based estimator of Wang and Wang representing the redistribution-of-mass technique. The estimator is implemented using Wang and Wang's code, available on their websites, with use of the recommended biquadratic kernel $K(x) = (15/16)(1-x^2)^2\mathbb{1}(|x| \leq 1)$ for the estimation of the required local weights. Furthermore, the bandwidth selection is implemented using the 5-fold cross validation as suggested in Wang and Wang and also implemented in Chapter 3, with candidate bandwidth ranging from 0.05 to 0.5 by 0.05 increments.

AC: 'Adapted-check' estimator of Chapter 3, for which we apply the proposed MM algorithm of Section 3.4 along with Beran's estimator for $G_C(\cdot|\boldsymbol{X})$ with a biquadratic kernel as well. The bandwidth selection is, here again, implemented using a 5-fold cross validation procedure on candidates ranging from 0.05 to 0.5 by 0.05 increments.

NEW: Newly proposed estimator (NEW) in (4.2.4)-(4.2.5), where both implied kernel density functions are chosen to be biquadratic as well. A practical procedure for the choice of both bandwidths involved in the estimator is described below.

For the required bandwidths in NEW, although Leconte et al. (2002) suggest a rule-of-thumb when considering only the estimation of $F_{T|\boldsymbol{X}}$, we advocate here for a simple extension of the usual cross validation procedures used in the literature on censored quantile regression, by considering a matrix of candidate bandwidths instead of simply a vector of candidates when facing only one smoothing direction. More precisely, for each candidate pair of bandwidths $(h_T, h_{\boldsymbol{X}})$, we start by randomly breaking the observations into 5 non-overlapping and roughly equal-sized parts. For each part $j = 1, \ldots, 5$, we then estimate β_τ via (4.2.4)-(4.2.5) using the observations in all the parts but the j-th, and determine the quality of fit of the model by computing the following prediction error on the complete observations in part j:

$$\mathrm{PE}_j(h_T, h_{\boldsymbol{X}}) = \sum_{\substack{i \in \mathcal{J} \\ \Delta_i = 1}} \rho_\tau(Y_i - \widehat{\beta}^{\mathsf{T}}_{\tau\,(-j)} \boldsymbol{X}_i),$$

where $\mathcal{J}$ is the set of all observations in part j and the notation $(-j)$ explicitly indicates that the estimation is carried out using all observations except the ones in part j. The procedure is repeated and averaged over $j = 1, \ldots, 5$, and the pair of bandwidths yielding the smallest average prediction error is then selected among all candidates. Now, in the spirit of the procedure of this chapter, we also note that an alternative prediction error of the type $\mathrm{PE}_j(h_T, h_{\boldsymbol{X}}) = \sum_{i \in \mathcal{J}} (\widehat{F}^{s,\,(-j)}_{T|\boldsymbol{X}}(\widehat{\beta}^{\mathsf{T}}_{\tau\,(-j)} \boldsymbol{X}_i | \boldsymbol{X}_i) - \tau)^2$ may be considered for the cross validation procedure instead, where the notation $(-j)$ is again reflecting estimation using all observations except the ones in part j. This alternative criterion is however observed to be more computationally demanding since it would require for each part

j the evaluation of $\widehat{F}^{s,(-j)}_{T|\boldsymbol{X}}$ in $\#\mathcal{J}$ new points, where $\#$ denotes the cardinal of a set, hereby also requiring through (4.2.5) the consideration of the local Kaplan-Meier estimator in much more new evaluation points. Additionally, one may also conjecture that extrapolation issues, possibly appearing when evaluating $\widehat{F}^{s,(-j)}_{T|\boldsymbol{X}}$ in new points in the conditional part, will be of more concern than possible extrapolation for a check-based prediction error. As a consequence, this strategy is here disfavored over a check-based proposal when considering Monte Carlo simulations.

Now, this more computationally intensive choice of a cross validation procedure in comparison with the rule-of-thumb of Leconte et al. is motivated by the observation that our bandwidths should primarily serve to the best possible estimation of β_τ in our model, which need not coincide with the bandwidths leading to best possible estimation of $F_{T|\boldsymbol{X}}$ taken on its own. Finally, for the candidate bandwidths in practice, as already commented, our experience first suggests that the performance of NEW is not very sensitive to the choice of h_T. Hence, we consider for this direction in our DGPs always a set of candidates ranging from 0.25 to 1.5 by 0.25 increments. As for the candidates for $h_{\boldsymbol{X}}$, these will vary from one simulation setting to another, given the computational cost of adding an extra bandwidth, and based on a few extra iterations in order to evaluate the order of magnitude in which one is more likely to observe the smallest cross validated prediction error. Hence, the exact range of bandwidth candidates for $h_{\boldsymbol{X}}$ will be given below, depending on the simulated scenario.

Now, as previously-mentioned, one of our main goals is to highlight that the numerical differences that will be observed between the above-described estimators emanates from the distinction between check-based and inverse-c.d.f. approaches. Consequently, and in order to analyze the impact of censored data as well, we will also include two omniscient estimators assuming knowledge of all the T_i, $i = 1, \ldots, n$. Our results will therefore incorporate the following procedures as well:

$\mathcal{O}_{\rho_\tau}$: Omniscient and check-based estimator of Koenker and Bassett, for which practical implementation is carried out using the function `rq` in the R library `quantreg`.

$\mathcal{O}_{NEW}$: Omniscient newly proposed estimator described in (4.2.4)-(4.2.5) where $\widehat{F}_{T|\boldsymbol{X}}$ is the well-known estimator of Stone and $Y_i = T_i, i = 1, \ldots, n$. The practical implementation is analogous to that of NEW.

These various estimation procedures are then compared in this section in two different DGPs, where the underlying model is either linear in all quantile levels or only at the τ-th quantile level of interest. As a preliminary remark, we mention here that all simulation settings are constructed with censoring in accordance with assumption (C7) for the true values of β_τ. Specifically, the following DGPs are considered:

DGP 1

The simulated model, repeatedly illustrated in the literature in Wang and Wang, Leng and Tong and Chapter 3, assumes

$$T_i = \beta_0 + \beta_1 X_i + (\eta_i - \Phi^{-1}(\tau)),$$

where $\beta_0 = 3$, $\beta_1 = 5$, $X_1, \ldots, X_n$ are i.i.d. $U[0,1]$ variables, Φ^{-1} is the quantile function of the standard normal distribution and $\eta_1, \ldots, \eta_n$ are i.i.d. $\mathcal{N}(0,1)$ variables with $\eta_i \perp\!\!\!\perp X_i, i = 1, \ldots, n$. The censoring variables $C_1, \ldots, C_n$ are independent of the covariate and are simulated from $U[0, M]$, with M calibrated to attain the desired censoring proportion at each quantile level of interest.

DGP 2

The model is again taken from the papers of Wang and Wang, and Leng and Tong, assuming now

$$T_i = \beta_0 + \beta_1 X_i + (0.2 + 2(X_i - 0.5)^2)(\eta_i - \Phi^{-1}(\tau)),$$

where $\beta_0 = 2$, $\beta_1 = 1$, $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0,1)$ variables and $\eta_1, \ldots, \eta_n$ are i.i.d. $\mathcal{N}(0,1)$ variables with $\eta_i \perp\!\!\!\perp X_i, i = 1, \ldots, n$. The censoring variables $C_1, \ldots, C_n$ are again independent of the covariate and simulated from $U[0, M]$, with M calibrated to attain the desired censoring proportion at the quantile levels of interest.

For these simulation settings, we examine two average censoring proportions $p_c \in \{15\%, 40\%\}$, two quantile levels of interest $\tau \in \{0.3, 0.5\}$, and sample sizes $n \in \{100, 200\}$. Based on $B = 500$ repetitions of each DGP, the different estimators are then compared in terms of root mean squared errors (RMSE) and median absolute error (MAE), which are standard criteria in the literature for fairly comparing different estimators that are based on very distinctive estimation strategies. Similarly to Chapter 3, additional graphical illustrations will also be provided in order to cover alternative criteria one could have reported in the tables, such as bias and variance results. Lastly, for robust results with regard to the optimization routine involved in the procedures, simulations are here presented by removing a few iterations for all estimators, based on the settings for each estimator that lead to the one percent worst mean absolute deviation (MAD) results, where the latter is defined for an estimator $\widehat{\beta}$ by $n^{-1} \sum_{i=1}^{n} |\widehat{\beta}^{\intercal} \boldsymbol{X}_i - \beta_\tau^{\intercal} \boldsymbol{X}_i|$.

We start our analysis by reporting in Table 4.1 the simulation results relative to DGP 1 for both censored and complete data, and where only the estimation of the median is here exposed for sake of brevity. For this DGP, candidate bandwidths for NEW and $\mathcal{O}_{NEW}$ are taken in $h_{\boldsymbol{X}} \in \{0.05, 0.1, \ldots, 0.25\}$. As can be observed, NEW exhibits for this trivial example slightly better results in terms of RMSE than its considered competitors, although the difference may seem modest at first sight. It is however further noted that, for small censoring proportions, NEW is here able to compete with $\mathcal{O}_{\rho_\tau}$, the omniscient check-based procedure. This is due

		$n = 100$				$n = 200$			
		RMSE		MAE		RMSE		MAE	
p_c	Method	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
15%	ICP	0.255	0.471	0.179	0.326	0.193	0.340	0.136	0.220
	RM	0.246	0.462	0.182	0.325	0.194	0.337	0.136	0.232
	AC	0.246	0.460	0.173	0.323	0.194	0.339	0.134	0.223
	NEW	0.239	0.443	0.158	0.292	0.175	0.309	0.122	0.225
	$\mathcal{O}_{\rho_\tau}$	0.238	0.427	0.170	0.290	0.177	0.311	0.131	0.209
	$\mathcal{O}_{NEW}$	0.221	0.397	0.152	0.267	0.165	0.287	0.120	0.202
40%	ICP	0.291	0.557	0.196	0.366	0.214	0.385	0.150	0.264
	RM	0.280	0.532	0.195	0.342	0.202	0.365	0.139	0.239
	AC	0.285	0.538	0.195	0.350	0.201	0.366	0.140	0.244
	NEW	0.269	0.517	0.179	0.344	0.193	0.360	0.135	0.258
	$\mathcal{O}_{\rho_\tau}$	0.238	0.427	0.170	0.290	0.177	0.311	0.131	0.209
	$\mathcal{O}_{NEW}$	0.221	0.397	0.152	0.267	0.165	0.287	0.120	0.202

Table 4.1: Simulation results for DGP 1 for both censored and complete observations, with $\tau = 0.5$, sample size $n = \{100, 200\}$, and average censoring proportions p_c taken in $\{15\%, 40\%\}$.

to the slight numerical advantage of the inverse-c.d.f. approach on which NEW is constructed in opposition to $\mathcal{O}_{\rho_\tau}$, as can be observed from the comparison between $\mathcal{O}_{\rho_\tau}$ and $\mathcal{O}_{NEW}$. A visual inspection of the obtained results, as depicted in Figure 4.1, brings further knowledge as it is observed that NEW's competitive RMSE results seem to hold despite slightly larger bias results for finite samples than its competitors. This then suggests that the proposed estimation strategy coupled

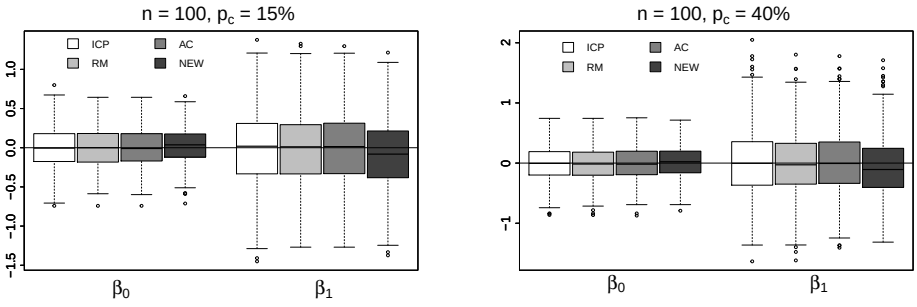

Figure 4.1: Boxplots relative to DGP 1 and Table 4.1 (left) for the estimation of β_0 and β_1, both centered (i.e. depicting $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $\tau = 0.5$, $n = 100$ and $p_c \in \{15\%, 40\%\}$.

		$p_c = 15\%$				$p_c = 40\%$			
		RMSE		MAE		RMSE		MAE	
τ	Method	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
0.3	ICP	0.229	0.449	0.152	0.312	0.314	0.552	0.206	0.352
	RM	0.214	0.407	0.141	0.286	0.228	0.420	0.154	0.286
	AC	0.223	0.424	0.145	0.300	0.229	0.461	0.154	0.313
	NEW	0.181	0.309	0.119	0.212	0.205	0.307	0.136	0.212
	$\mathcal{O}_{\rho_\tau}$	0.214	0.407	0.146	0.288	0.214	0.407	0.146	0.288
	$\mathcal{O}_{NEW}$	0.169	0.309	0.112	0.209	0.169	0.309	0.112	0.209
0.5	ICP	0.213	0.439	0.146	0.295	0.282	0.500	0.193	0.350
	RM	0.201	0.400	0.131	0.261	0.205	0.396	0.133	0.262
	AC	0.208	0.420	0.138	0.262	0.218	0.440	0.148	0.296
	NEW	0.142	0.288	0.092	0.200	0.150	0.304	0.104	0.203
	$\mathcal{O}_{\rho_\tau}$	0.204	0.402	0.134	0.274	0.204	0.402	0.134	0.274
	$\mathcal{O}_{NEW}$	0.145	0.292	0.090	0.193	0.145	0.292	0.090	0.193

Table 4.2: Simulation results for DGP 2 for both censored and complete responses, with $n = 100$ observations, $\tau \in \{0.3, 0.5\}$, and average censoring proportions p_c taken in $\{15\%, 40\%\}$.

with a double-kernel approach noticeably reduces the variability for the estimation of the parameters across simulated datasets in comparison with its competitors. These observations are in line for instance with the literature on smoothed (in the 'T-direction') distribution function estimation, as already mentioned in Section 4.2.

We now consider the slightly more elaborate DGP 2, where the true model is only linear at the τ-th quantile level of interest. For this DGP, candidate bandwidths for NEW and $\mathcal{O}_{NEW}$ are now taken in $h_{\boldsymbol{X}} \in \{0.2, 0.25, ..., 0.7\}$. Table 4.2 reports the obtained results for $n = 100$ observations at each iteration. Ignoring momentarily the results for NEW, a first observation is that RM outperforms AC and ICP when using this exact original setting from Wang and Wang. Hence, since RM is constructed upon estimation of $F_{T|\boldsymbol{X}}$ in opposition to AC and ICP, we may presuppose that this DGP favors modelling strategies that require estimation of $F_{T|\boldsymbol{X}}$ instead of G_C. In the context of this chapter, our main interest then becomes comparing RM with NEW since both procedures are based on a suitable estimation of the same conditional distribution. As can be observed, the results for NEW provide here a spectacular improvement over RM, and by extension AC and ICP. In fact, the improvement is such that NEW even exhibits strikingly better results than $\mathcal{O}_{\rho_\tau}$ for all considered quantile levels and censoring proportions. This is of course again due to the advantage of using an inverse-c.d.f. approach, as one can observe from comparing the results of $\mathcal{O}_{\rho_\tau}$ and $\mathcal{O}_{NEW}$. Still, this suggests that a dramatic improvement in finite sample performance can be observed in settings where RM is often considered as a primary reference in the literature. Figure 4.2 illustrates graphically this statement by depicting the evident improvement of NEW in terms of variance over its three considered competitors handling censored responses in

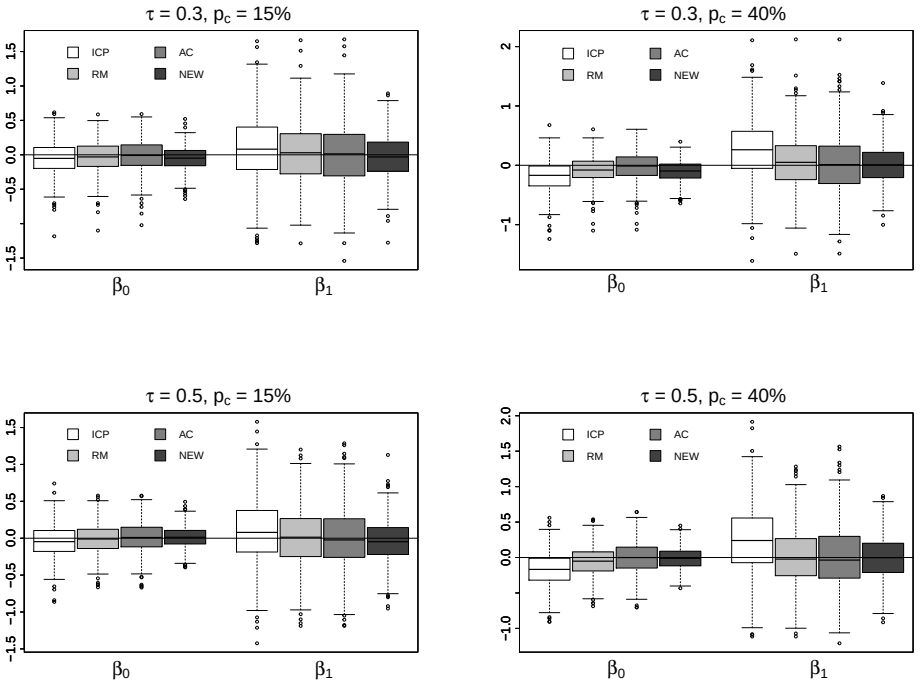

Figure 4.2: Boxplots relative to DGP 2 and Table 4.2 for the estimation of β_0 and β_1, both centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $n = 100$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

this DGP. Results for larger sample sizes exhibit the same singular patterns and are therefore omitted for this DGP.

Recalling that both DGPs are taken from the literature and are hence not designed to particularly favor here our procedure, we conclude this section by conjecturing from the obtained results that for low-dimensional regression models with or without censored data, an inverse-c.d.f. approach provides very competitive finite sample results in comparison with check-based formulations, with the possibility of a dramatic improvement in terms of variance for more elaborate DGPs. The next section will then be devoted to the question of whether this statement still holds for higher dimensional-covariates, in particular for possibly censored responses.

4.4.2 Multidimensional covariates

For multivariate problems, we consider again two different DGPs where the true model is either linear in all quantile levels or only at the τ-th level of interest. Specifically, we consider:

DGP 3

The simulated model is inspired from the paper of Wang et al. (2013):

$$T_i = \beta_\tau^\mathsf{T} \boldsymbol{X}_i + (\eta_i - \Phi^{-1}(\tau)),$$

where $\beta_\tau = (1, 1.5, 0.7, 1, -0.5)^\mathsf{T}$, X_{ji} are i.i.d. $U[-1, 1]$ variables for $i = 1, \ldots, n$, and $j = 1, \ldots, 4$, and $\eta_1, \ldots, \eta_n$ are i.i.d. $\mathcal{N}(0, 1)$ variables independent from the covariates. The censoring variables $C_1, \ldots, C_n$ are independent of the covariates as well and are simulated from $U[-2, M]$, with M calibrated to attain the desired censoring proportion at each quantile level of interest.

DGP 4

The last simulated model assumes:

$$T_i = \beta_\tau^\mathsf{T} \boldsymbol{X}_i + (0.2 + (\gamma^\mathsf{T} \boldsymbol{X}_i)^2)(\eta_i - \Phi^{-1}(\tau)),$$

where $\beta = (0.5, 0.5, 0.5, 0.5, 0.5)^\mathsf{T}$, $\gamma = (-1, 1, 0.5, 0, -1)^\mathsf{T}$, X_{ji} are i.i.d. $U[-1, 1]$ variables for $i = 1, \ldots, n$, and $j = 1, \ldots, 4$, and $\eta_1, \ldots, \eta_n$ are i.i.d. $\mathcal{N}(0, 1)$ variables independent from the covariates. The censoring variables $C_1, \ldots, C_n$ are again independent of the covariates and simulated from $U[-2, M]$, with M calibrated to attain the desired censoring proportion at each quantile level of interest.

To account for the multivariate feature of these DGPs, we first need to slightly adjust some of the competing procedures in order to provide a fair comparison. In particular, Wang and Wang's estimator for RM is here suitably replaced by the estimator of Wang et al. without the variable selection feature included in the latter. Hence, for the estimation of the involved local weights, the estimator here includes the same dimension reduction assumption as in Section 4.2.3. Estimation of $\gamma_{0,j}$, $j = 1, \ldots, q$, is carried out using the sliced inverse regression (SIR) approach of Li et al. (1999), with R code for instance available on the webpage of Huixia Wang. Alternatively, one could consider the minimum average variance estimation of Xia et al. (2010), but results are very comparable and SIR is computationally more convenient, as observed and also commented in Wang et al.. For an automatic choice of q across simulations, we adopt as in Wang et al. the Chi-squared test of Li et al. with 5% significance level. The corresponding R code is again available on the webpage of Huixia Wang, and the resulting estimator will henceforth be denoted as $\mathrm{RM}_{\hat{q}}$ to highlight both the dimension reduction specificity and the data-driven choice of q. Lastly, since the maximal number of estimated indices in the

simulated DGPs is 4, we consider here a fourth-order kernel $K(x) = (105/64)(1 - 5x^2 + 7x^4 - 3x^6)\mathbb{1}(|x| \leq 1)$, and the bandwidth is chosen via 5-fold cross validation with candidates ranging from 0.1 to 2 by 0.1 increments.

The ICP and AC estimators are, for their part, the same as for the small dimension simulations as no specific dimension reduction technique is either required or evoked in these procedures. A small adjustment is however considered for AC, as we implement here the procedure with a fourth-order kernel, similarly to $\text{RM}_{\hat{q}}$. The bandwidth is chosen equivalently to the univariate settings, with the same candidates as for $\text{RM}_{\hat{q}}$.

Lastly, for our newly proposed procedure, we consider here two versions of the estimator described in (4.2.8): $\text{NEW}_{\hat{q}}$ will denote the proposed estimator with data-driven choice of q, while NEW_q will stand for the same procedure but forcing one to use the true number of required indices in each DGP, that is, $q = 1$ in DGP 3 and $q = 2$ in DGP 4. Implementation of the dimension reduction related elements is carried out using the same tools as for $\text{RM}_{\hat{q}}$. Finally, a biquadratic kernel density is chosen for the univariate T-direction, a fourth-order kernel is adopted for the $\boldsymbol{X}$-direction just as for $\text{RM}_{\hat{q}}$, and bandwidth selection is performed using cross validation with $h_T \in \{0.25, 0.5, \ldots, 1.5\}$ and $h_{\boldsymbol{X}} \in \{0.2, 0.4, \ldots, 1\}$.

For all simulation settings, we consider again $B = 500$ repetitions of each DGP, and the procedures are now compared in terms of aggregated results across parameters for ease of reading, that is, RMSE will stand for the root of mean squared errors summed over all parameters, and MAE will stand for the sum over all parameters of median absolute error. Furthermore, similarly as for the simulations of Chapter 3, we consider for these multivariate settings an additional criterion taken from a prediction point of view, as the procedures will also be compared in terms of mean absolute deviation (MAD) defined earlier. Equivalently to the univariate settings, the reported results are here again trimmed based on the one percent worst simulation results for each procedure in terms of MAD.

We start by depicting in Table 4.3 and Figure 4.3 the obtained results for DGP 3 with two average censoring proportions $p_c \in \{15\%, 30\%\}$, sample sizes $n \in \{100, 200\}$, and only for the particular case of $\tau = 0.5$ for sake of brevity. The following observations may then be provided: first, as can be observed from the top row of Figure 4.3, the general pattern reported in the univariate settings for an inverse-c.d.f. technique with additional smoothing in the response direction is again observed, as $\text{NEW}_{\hat{q}}$ exhibits moderately larger bias results than its check-based competitors but with noticeably smaller variance. As a result, both NEW_q and $\text{NEW}_{\hat{q}}$ confidently outperform in this setting the check-based procedures in terms of RMSE and MAE as reported in Table 4.3. The same considerations are reflected through the reported prediction criterion, as can be observed from both Table 4.3 and the bottom line of Figure 4.3 where the absolute deviations of all $B = 500$ simulations are depicted instead of only the MAD.

Regarding the considered competitors for censored data, as repeatedly reported in the literature as well, it is further observed in Table 4.3 that ICP exhibits

$p_C = 15\%$					$p_C = 30\%$				
n	Method	RMSE	MAE	MAD	n	Method	RMSE	MAE	MAD
	ICP	0.502	0.746	0.238		ICP	0.565	0.850	0.266
	$\text{RM}_{\hat{q}}$	0.487	0.712	0.232		$\text{RM}_{\hat{q}}$	0.526	0.794	0.252
	AC	0.490	0.711	0.233		AC	0.529	0.791	0.251
100	$\text{NEW}_{\hat{q}}$	0.479	0.698	0.228	100	$\text{NEW}_{\hat{q}}$	0.517	0.774	0.246
	NEW_q	0.478	0.698	0.228		NEW_q	0.515	0.771	0.245
	$\mathcal{O}_{\rho_\tau}$	0.462	0.717	0.220		$\mathcal{O}_{\rho_\tau}$	0.462	0.717	0.220
	$\mathcal{O}_{NEW}$	0.447	0.682	0.212		$\mathcal{O}_{NEW}$	0.447	0.682	0.212
	ICP	0.346	0.500	0.166		ICP	0.391	0.589	0.188
	$\text{RM}_{\hat{q}}$	0.342	0.499	0.165		$\text{RM}_{\hat{q}}$	0.368	0.566	0.177
	AC	0.341	0.489	0.164		AC	0.370	0.563	0.180
200	$\text{NEW}_{\hat{q}}$	0.323	0.475	0.155	200	$\text{NEW}_{\hat{q}}$	0.350	0.534	0.168
	NEW_q	0.323	0.475	0.155		NEW_q	0.347	0.531	0.166
	$\mathcal{O}_{\rho_\tau}$	0.323	0.485	0.156		$\mathcal{O}_{\rho_\tau}$	0.323	0.485	0.156
	$\mathcal{O}_{NEW}$	0.300	0.439	0.145		$\mathcal{O}_{NEW}$	0.300	0.439	0.145

Table 4.3: Simulation results for DGP 3 for both complete and censored data with $\tau = 0.5$ and $n \in \{100, 200\}$.

difficulties in comparison with $\text{RM}_{\hat{q}}$ and AC. The latter two perform here very similarly despite $\text{RM}_{\hat{q}}$ attempting to take profit of a possible dimension reduction in the smoothing of the conditional distribution involved in the procedure. Next, we note that for this DGP, there seems to be little influence in our newly proposed procedure of choosing q with the Chi-squared test of Li et al. rather than imposing $q = 1$. A similar observation is provided in Wang et al. from which this DGP is inspired. Lastly, although aggregating the results makes the difference between procedures at first seem less spectacular than for DGP 2, it is noted that both NEW_q and $\text{NEW}_{\hat{q}}$ are here again in some situations able to compete with the omniscient procedure $\mathcal{O}_{\rho_\tau}$, particularly when the censoring proportion is low and with higher sample sizes, as can for instance be observed from the absolute deviation results depicted in Figure 4.3. Similarly to the univariate settings, these results are again to be attributed to the advantages of an inverse-c.d.f. approach in this context, as can be deduced from the comparison with $\mathcal{O}_{NEW}$. The latter notation stands here for the omniscient newly proposed estimator with dimension reduction implemented similarly to $\text{RM}_{\hat{q}}$ and $\text{NEW}_{\hat{q}}$, but with all complete responses at hand.

We now turn to the last simulated setting, for which the results are reported in Table 4.4 and Figure 4.4. For sake of brevity, only the sample size $n = 200$ is exposed here. Several comments are to be brought: first, we observe again that in terms of RMSE, MAE and MAD, the newly proposed procedure performs very competitively with respect to its check-based estimators, the margin being especially important for the more central quantile level $\tau = 0.5$. For small censoring proportions, the improvement is again such that $\text{NEW}_{\hat{q}}$ is here able to outperform $\mathcal{O}_{\rho_\tau}$ for both considered quantile levels. Among the check-based procedures handling censored data, similarly to the other DGPs it is further observed that ICP performs rather poorly for higher censoring proportions, while $\text{RM}_{\hat{q}}$ seems for this

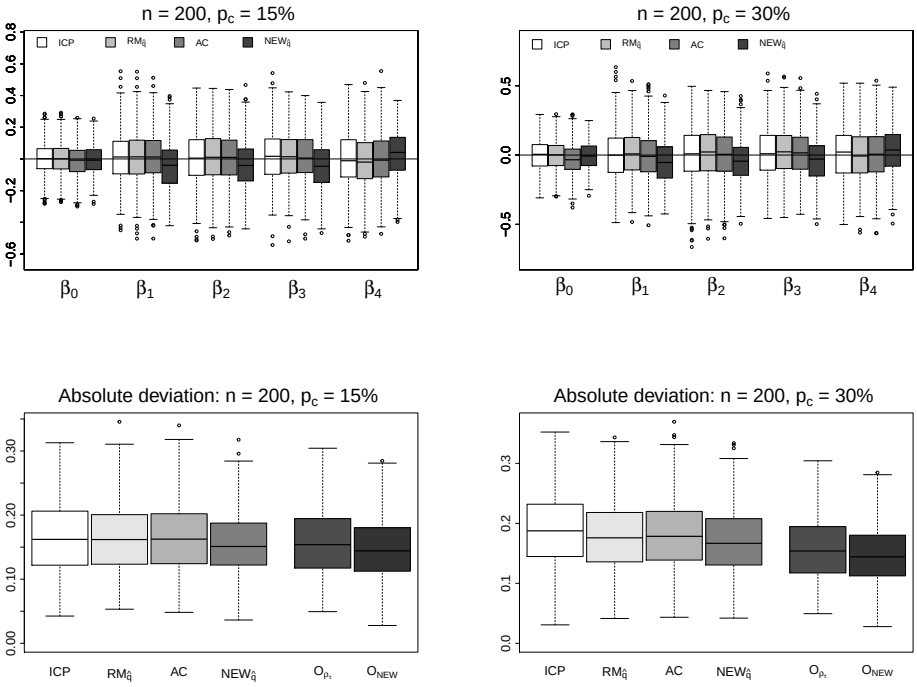

Figure 4.3: Boxplots relative to DGP 3. The top line reports the estimation of β_τ centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the competitors based on censored data with $\tau = 0.5$, $n = 200$ and $p_c \in \{15\%, 30\%\}$. The bottom line reports for all estimation procedures the absolute deviation results obtained under the same settings.

DGP to take advantage over AC of its dimension reduction feature. A visual examination of the results reported in Figure 4.4 also highlights that all procedures tend to exhibit more difficulties for the estimation of β_1 and β_4, as these correspond to the covariates for which the associated γ's are the largest in absolute value. Hence, for these covariates, it is quite natural for the procedures to estimate more difficultly the true signal β_τ in our model taking into account the noise induced by γ.

Next, we note that $\text{NEW}_{\hat{q}}$ and NEW_q seem to perform again relatively comparably, although the difference is more pronounced than for DGP 3 where the true number of required projections was $q = 1$. From the MAE results and our practical experience, this is actually observed to result from a few iterations for which the automated choice in our simulations of the value of q through the Chi-squared test lead to $\hat{q} = 1$. In this situation, we leave the possibility of automatically choosing

τ	Method	RMSE	MAE	MAD	τ	Method	RMSE	MAE	MAD
		$pc = 15\%$					$pc = 30\%$		
	ICP	0.399	0.609	0.198		ICP	0.465	0.704	0.231
	$\text{RM}_{\hat{q}}$	0.401	0.617	0.200		$\text{RM}_{\hat{q}}$	0.431	0.625	0.212
	AC	0.395	0.613	0.197		AC	0.440	0.654	0.218
0.3	$\text{NEW}_{\hat{q}}$	0.381	0.565	0.192	0.3	$\text{NEW}_{\hat{q}}$	0.429	0.621	0.212
	NEW_q	0.365	0.560	0.191		NEW_q	0.414	0.618	0.208
	$\mathcal{O}_{\rho_\tau}$	0.388	0.586	0.192		$\mathcal{O}_{\rho_\tau}$	0.388	0.586	0.192
	$\mathcal{O}_{NEW}$	0.359	0.534	0.183		$\mathcal{O}_{NEW}$	0.359	0.534	0.183
	ICP	0.410	0.601	0.200		ICP	0.506	0.741	0.247
	$\text{RM}_{\hat{q}}$	0.407	0.597	0.200		$\text{RM}_{\hat{q}}$	0.436	0.645	0.212
	AC	0.406	0.600	0.199		AC	0.453	0.667	0.224
0.5	$\text{NEW}_{\hat{q}}$	0.350	0.511	0.172	0.5	$\text{NEW}_{\hat{q}}$	0.391	0.587	0.200
	NEW_q	0.343	0.507	0.171		NEW_q	0.381	0.580	0.192
	$\mathcal{O}_{\rho_\tau}$	0.387	0.572	0.191		$\mathcal{O}_{\rho_\tau}$	0.387	0.572	0.191
	$\mathcal{O}_{NEW}$	0.341	0.492	0.167		$\mathcal{O}_{NEW}$	0.341	0.492	0.167

Table 4.4: Simulation results for DGP 4 for both complete and censored data with $n = 200$ and $\tau \in \{0.3, 0.5\}$.

only the projection $\gamma^\mathsf{T}\boldsymbol{X}$ in our DGP for which one coefficient of γ is equal to 0, and leave out the projection relative to $\beta_\tau^\mathsf{T}\boldsymbol{X}$. As a result, considering only the former projection nearly neglects the conditional influence in our procedure of the covariate X_3 for which the estimated coefficient for γ is close to 0, which explains why these simulation iterations present erratic results with respect to the majority of iterations, especially for the estimation of the β associated to X_3. We note however that, in practice and as already commented in Section 4.2, the choice of q is encouraged to be done along with visual inspection and examination of the eigenvalues involved in the SIR method (see e.g. Li (1991) and the different tools reported in the R package `edrGraphicalTools`). Nevertheless, despite this observation, the inverse-c.d.f. strategy is again illustrated to globally outperform its check-based counterparts, as is also sustained by the results for complete observations.

Bringing back the results of DGPs 1-4, we conclude this section by postulating that the newly proposed procedure offers a significant improvement over the current check-based literature in terms of variance for finite sample estimation of a linear quantile regression, as repeatedly illustrated through the presented tables. These encouraging results may then advocate for the practical use of an inverse-c.d.f. estimation approach for censored linear quantile regression.

4.4.3 Illustration of the percentile bootstrap

For inference purposes, we end this section by providing a brief illustration of the validity of the proposed bootstrap procedure for our estimator. To that end, restricting ourselves to the two univariate DGPs studied in this chapter, we compare the performance of the bootstrap procedure for NEW with the bootstrap procedures of RM and AC. We choose to leave out the results for ICP for sake of

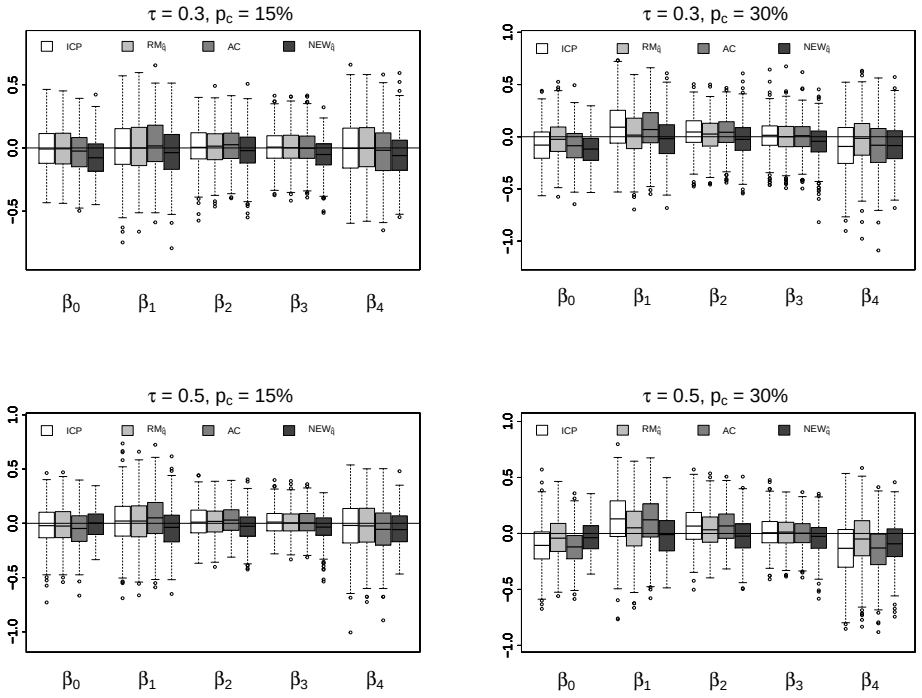

Figure 4.4: Boxplots relative to DGP 4 and Table 4.4 for the estimation of β_τ centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $n = 200$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

brevity and given the clear dominance of all other estimators over the latter. For all procedures, bandwidths are chosen for each of 500 iterations via the same cross validation procedures as in our previous simulations on the original sample, and are then kept fixed for the 300 bootstrapped samples considered at each iteration. With a nominal level of 0.95 and a unique sample size $n = 100$, Table 4.5 then reports the empirical mean length and empirical coverage probability (in brackets) of the confidence intervals obtained from this simulation study.

Concentrating first on the empirical coverage probabilities, we observe that all three procedures perform relatively similarly given the adequately close values to the chosen nominal level. This serves as a first indication that the percentile bootstrap represents a satisfactory tool for inference purposes when considering the estimator NEW. Secondly, we observe that similar findings as for our previous simulation study are here confirmed in the bootstrap results, as the empirical mean lengths of NEW are noticeably smaller than for its competitors, especially for

			RM		AC		NEW	
DGP	p_c	τ	β_0	β_1	β_0	β_1	β_0	β_1
1	15%	0.3	1.112 (0.954)	2.018 (0.964)	1.102 (0.958)	2.010 (0.962)	1.120 (0.944)	1.968 (0.952)
		0.5	1.042 (0.960)	1.909 (0.962)	1.036 (0.952)	1.892 (0.964)	0.963 (0.950)	1.712 (0.932)
	40%	0.3	1.250 (0.954)	2.420 (0.980)	1.218 (0.950)	2.343 (0.968)	1.230 (0.946)	2.309 (0.958)
		0.5	1.153 (0.960)	2.277 (0.972)	1.146 (0.964)	2.242 (0.974)	1.078 (0.944)	2.030 (0.954)
2	15%	0.3	0.883 (0.932)	1.696 (0.924)	0.887 (0.938)	1.718 (0.918)	0.641 (0.932)	1.101 (0.946)
		0.5	0.836 (0.936)	1.638 (0.940)	0.860 (0.940)	1.694 (0.936)	0.541 (0.948)	1.013 (0.934)
	40%	0.3	0.898 (0.942)	1.701 (0.944)	0.909 (0.936)	1.814 (0.940)	0.702 (0.922)	1.201 (0.970)
		0.5	0.819 (0.934)	1.639 (0.946)	0.883 (0.952)	1.838 (0.948)	0.597 (0.968)	1.126 (0.970)

Table 4.5: Bootstrap results for DGP 1 and 2 expressed in terms of empirical mean length and empirical coverage probability (in brackets) , based on 500 simulations with 300 bootstrap samples. The nominal level is 0.95, with sample size $n = 100$, $p_c \in \{15\%, 40\%\}$, and $\tau \in \{0.3, 0.5\}$.

DGP 2. This again suggests that a reduction in variance results may be expected with an inverse-c.d.f. approach for the present context. Graphically, Figure 4.5 illustrates this statement by reporting boxplots of all 500 iterations of the length of the bootstrap confidence intervals instead of only the mean as in Table 4.5. For comparison purposes, the results for $\mathcal{O}_{\rho_\tau}$ are here also reported, and highlight again the visible improvement of NEW over its check-based competitors. Note that for the latter estimator, further refinements of the percentile bootstrap have been considered in the literature (see e.g. Section 3.9 in Koenker (2005)), but these do not serve the comparison we intend to provide here. Overall, these results confirm here the validity of the simple percentile bootstrap procedure for our newly proposed estimator.

4.5 Real Data Analysis

We propose in this section a brief illustrative application of our methodology to real data coming from an astronomical study of quasars, and start here by describing the general context of the latter data. Quasars are the brightest objects in the universe and consist of a super-massive black hole surrounded by an orbiting accretion disk of gas. As gas falls into the black hole, it is heated up and emits thermal radiation

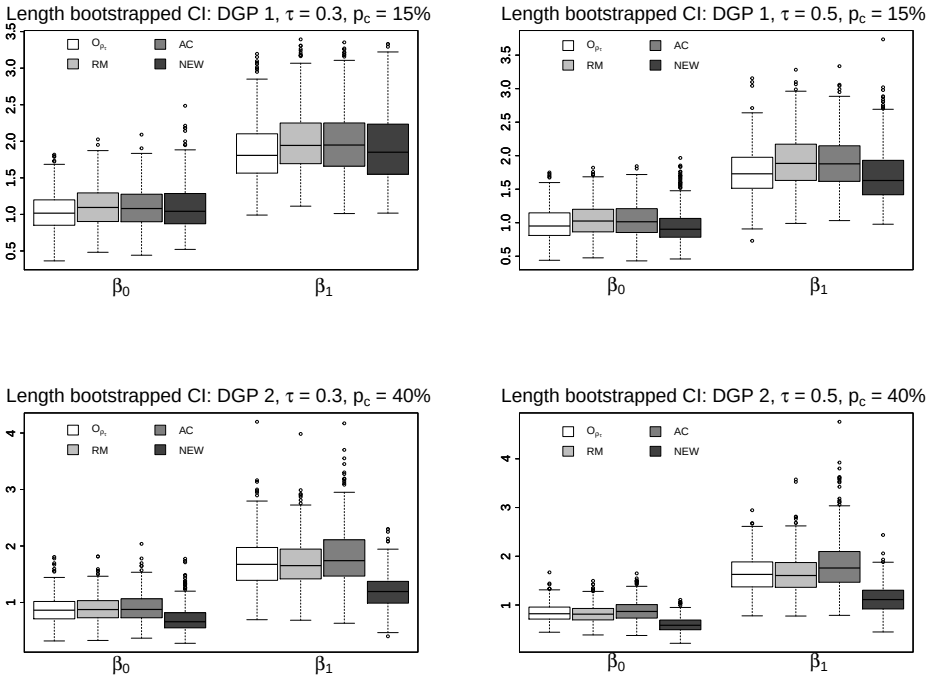

Figure 4.5: Boxplots of bootstrapped confidence interval length relative to DGP 1 and 2. Sample size is $n = 100$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

that spans the spectrum, making quasars bright in the visible spectrum as well as in X-rays. In the last decades, astronomers have cataloged a large amount of quasars based on optical considerations, many of which have also been observed with modern telescopes in the X-rays. This then offers the possibility to study in detail the relations between UV and X-ray properties of optically selected quasars, which are believed to be insightful for understanding the structure and physics of quasars nuclear regions.

In this work, we consider the data described in Table 2 of Vignali et al. (2003), reporting information of 206 radio-quiet and optically selected quasars. In particular, we illustrate here our methodology on the relationship between $l_{UV} = \log L_{2500\text{Å}}$ and $l_X = \log L_{2\text{keV}}$, where $L_{2500\text{Å}}$ and $L_{2\text{keV}}$ denote the rest-frame 2500 Å and 2 keV luminosity densities, respectively. Due to technical limitations, 70 of the 206 values of l_X are only observed through upper bounds (see Table 1 in Vignali et al.), and are hence left-censored. To transform these observations back in the right-censored context of this chapter, we replace $l_{X,i}$ by

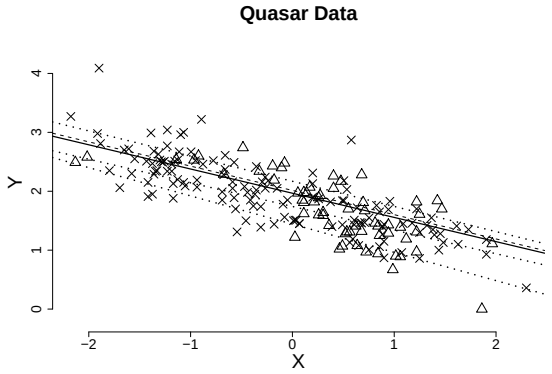

Figure 4.6: Quasar data of Vignali et al. (2003). Uncensored data points are given by $\times$, while censored observations are represented by $\triangle$. The dashed line represents the Buckley-James mean regression estimation, while the plain line represents the estimated median regression using NEW. For completeness, the estimated quantile regressions for NEW with $\tau \in \{0.1, 0.3, 0.7\}$ are also depicted on the graph in dotted lines.

$Y_i = \max_{1 \leq j \leq 206}(l_{X,j}) - l_{X,i}$, $i = 1, \ldots, 206$, while the variable X will henceforth denote l_{UV} (centered and standardized) for notational convenience. The resulting dataset is plotted in Figure 4.6.

Now, we note that several authors in the literature on parametric censored regression have already analyzed (a sub-sample of) the present dataset, although focusing solely on the conditional mean, see e.g. Heuchenne and Van Keilegom (2007) and Pardo-Fernández et al. (2007). We propose here to extend their analysis by considering the effect of various quantile levels on the relationship between X and the complete and unobservable response. Additionally, we will also consider the parallel between results for median regression with the usually employed mean regression. We start here by first considering a sequence of quantile levels ranging from 0.20 to 0.65 by 0.05 increments, and propose to fit a linear quantile regression for each increment using our proposed methodology. Implementation is carried out similarly to Section 4.4, with candidates bandwidths $h_{\boldsymbol{X}} \in \{0.20, 0.25, \ldots, 0.60\}$ and $h_T \in \{0.10, 0.15, \ldots, 0.30\}$.

Before investigating the resulting estimations, we propose here to analyze the sensitivity of our methodology to the matrix of considered candidate bandwidths. In order to do so, taking the examples of $\tau = 0.3$ and $\tau = 0.5$, we propose to fit the linear regression using every possible combination of $(h_T, h_{\boldsymbol{X}})$, and plot in Figure 4.7 the resulting values of the estimated intercept and slope. We focus first on the left part of Figure 4.7, illustrating in red the estimated values of the intercept using our proposed methodology for $\tau = 0.3$ (bottom) and $\tau = 0.5$ (top). For comparison, we also plot the estimated values using RM in blue and ICP in green, the latter procedures being of course independent either to the second bandwidth

h_T (RM) or to both bandwidths (ICP). As can be observed, both NEW and RM methodologies are quite robust to the values of their bandwidth(s), with a little more fluctuations for the smallest values of $h_{\boldsymbol{X}}$, as one may have expected. Both methodologies are also observed here to produce similar intercept estimates for the different values of τ, while ICP proposes a somewhat lower estimation for the latter, especially for the case $\tau = 0.5$. We now turn to the right part of Figure 4.7, representing only for the case $\tau = 0.3$ the estimated values of the slope for every combination of $(h_T, h_{\boldsymbol{X}})$. The case $\tau = 0.5$ is here left out for visual convenience, as the slope estimates are very close from one quantile level to the other. For comparison purposes, we here again include the estimated slopes using RM in blue and ICP in green. We again observe from the resulting plot that NEW and RM are both quite regular in their slope estimation for every possible bandwidth(s). Additionally, NEW and RM are here again relatively close in comparison with ICP. Altogether, these figures suggest that the choice of bandwidth(s) for both NEW and RM has pleasantly very limited impact for the present dataset.

We now turn to the estimation of the regression parameters for different incremental values of τ, and present the obtained results in Figure 4.6 and Figure 4.8. Bandwidths are here selected using the same cross validation procedure as in Section 4.4, while 95% confidence intervals depicted in Figure 4.8 are constructed upon 300 bootstrap samples using the percentile bootstrap illustrated earlier where,

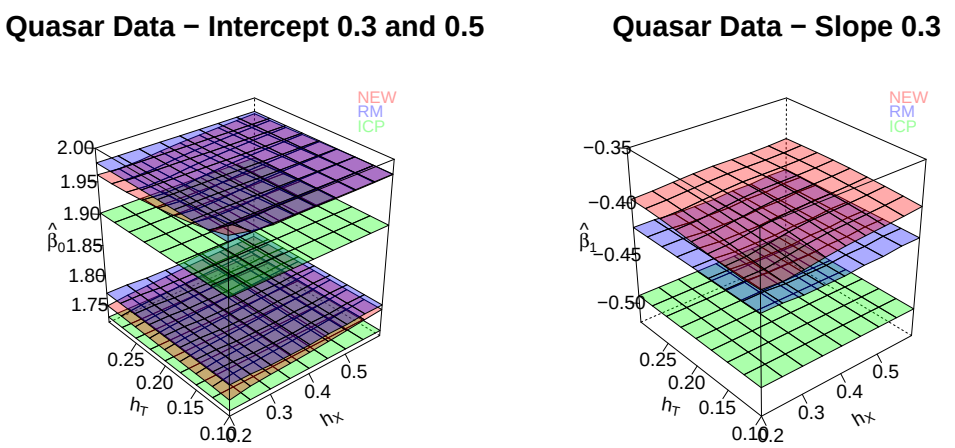

Figure 4.7: Quasar data, bandwidth sensitivity. The left plot represents estimation of the intercept for $\tau = 0.3$ (bottom) and $\tau = 0.5$ (top) for NEW (red), RM (blue) and ICP (green), for every pair of $(h_T, h_{\boldsymbol{X}})$ considered (RM being independent to h_T and ICP being independent to both h_T and $h_{\boldsymbol{X}}$). The right plot illustrates the estimation of the slope for $\tau = 0.3$ for every combination of h_T and $h_{\boldsymbol{X}}$.

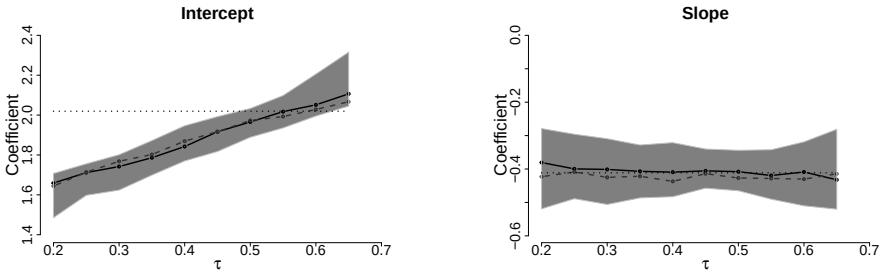

Figure 4.8: Quasar data. Estimated quantiles coefficients for regressing Y on X. Shaded areas correspond to 95% confidence intervals based on 300 bootstrap samples. Black lines correspond to the estimator NEW, grey lines correspond to RM, and dotted lines correspond to the Buckley-James mean regression estimation.

for computational convenience, bandwidths are kept fixed for all bootstrap samples based on the original dataset. For comparison purposes, the coefficient estimations for RM are also depicted in Figure 4.8. Additionally, the mean regression estimation for the present dataset is also depicted in the latter figure using the procedure of Buckley and James (1979). As can be acknowledged, both quantile estimation procedures offer relatively similar results by suggesting first that the regression coefficients are significantly different from 0 for all considered values of τ, as one could have visually anticipated. More interestingly, both procedures also suggest that the slope of the linear relation is close to being constant across the point cloud, with an estimated quantile coefficient value that is very similar to the mean regression estimation, as can be observed both from Figure 4.6 and Figure 4.8. Hence, the quantile regressions are revealed here to be close to being perfectly parallel across values of τ, although formal testing of this assumption is here left out (see e.g. Section 3 in Koenker (2005)). This represents nevertheless an information and property for the present quasar dataset that mean regressions are, by nature, not able to expose. As a remark, a further step for the analysis of the present dataset could then be to consider a composite quantile regression estimation (see e.g. Tang and Wang (2015) for censored data), where it is explicitly assumed in the model that the effects of covariates in a range of quantile levels only differ through the intercept term. Taking profit of such additional information on the regression function may quite naturally result in improvements of the estimation efficiency with respect to fitting each quantile level separately, although the accommodation of our procedure to this context is left for further research. Finally, returning to Figure 4.8 and regarding the confidence intervals of NEW, it is observed that the widths of the latter are quiet constant for the intercept, while somewhat smaller for more central quantile levels for the estimation of the slope. The latter observation

is in line with one's expectations regarding an inverse-c.d.f. approach. Altogether, in addition to exhibiting the relative robustness of our methodology towards bandwidths selection, this brief real data example illustrates how a simple application of our methodology may extend the analysis of linear censored regression models while being in agreement with previous work regarding the central tendency of the data.

4.6 Conclusion

In this work, we have introduced a reconsideration of the use of check-based modelling in the context of linear quantile regression with censored data. Our underlying interest was indeed motivated by the knowledge that existing methodologies in the literature already need to rely for their handling of censored observations on appropriate nonparametric estimators of conditional distributions, even though the regression model is purely parametric. Based on this consideration, we investigated the application of an inverse-c.d.f. approach in the present context, rather than a usual check-based formulation. Our resulting estimation procedure was then built on a double-kernel estimator of the conditional distribution of the response variable given the covariates. In order to appropriately accommodate for multivariate covariates, the application of a dimension reduction technique in this framework was further considered. Overall, the resulting quantile regression estimator was observed, in an extensive simulation study, to offer a very competitive procedure, characterized by a notable decrease in variance results with respect to established check-based formulations. Furthermore, the latter finding was also observed to apply to the situation with only complete observations at hand, although the argumentation for embracing an inverse-c.d.f. approach is here less obvious, apart from the latter numerical results. For inference, a simple percentile bootstrap procedure was further illustrated to provide satisfactory results with respect to the literature. From a theoretical point of view, consistency and asymptotic normality of our proposed estimator for linear regression were obtained under classical regularity requirements. Additionally, as a by-product, several asymptotic results such as uniform consistency and linear representation were also developed for the double-kernel estimators of the conditional distribution and density of the censored response given the covariates, with consideration of the dimension reduction framework. Lastly, a brief application to astronomical data was proposed and advocates, together with our simulation results, for the practical choice of an inverse-c.d.f. approach when considering a robust quantile regression analysis with censored data, especially for central quantile levels of interest and when the proportion of censored observations in the dataset is reasonable with respect to the latter.

4.7 Proofs

We provide in this section the proofs of Section 4.3, along with additional theoretical results that are needed for the latter proofs. In particular, we start by considering the development of two lemmas; the first one states the uniform consistency of both the double-kernel version of Beran's estimator with dimension reduction and its corresponding density estimator, while the second one reports a linear representation of the smoothed version of Beran's estimator only.

Lemma 4.1. *Assume conditions (C1), (C3)-(C6), (C8)-(i) and (C9)-(i) hold. Then, for* $1 \leq q \leq d+1$*:*

(a) Writing $\boldsymbol{z} = (z_1, \ldots, z_q)^\mathsf{T} = (\gamma_{0,1}^\mathsf{T}\boldsymbol{x}, \ldots, \gamma_{0,q}^\mathsf{T}\boldsymbol{x})^\mathsf{T}$*, and* $\widehat{\boldsymbol{z}} = (\widehat{\gamma}_{0,1}^\mathsf{T}\boldsymbol{x}, \ldots, \widehat{\gamma}_{0,q}^\mathsf{T}\boldsymbol{x})^\mathsf{T}$*:*

$$\begin{aligned}\sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{X}}(t|\boldsymbol{x})| &= \sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})| \\ &= O_{\mathbb{P}}\left((\log n/(nh^q_{\boldsymbol{X}}))^{1/2} + h^\nu_{\boldsymbol{X}} + h^2_T\right).\end{aligned}$$

(b) Writing $f_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \frac{\partial}{\partial t}F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})$ *and* $\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) = \frac{\partial}{\partial t}\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ *:*

$$\begin{aligned}\sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - f_{T|\boldsymbol{X}}(t|\boldsymbol{x})| &= \sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})| \\ &= O_{\mathbb{P}}\left((\log n/(nh^q_{\boldsymbol{X}}h_T))^{1/2} + h^\nu_{\boldsymbol{X}} + h^2_T\right).\end{aligned}$$

Lemma 4.2. *Assume the conditions of Lemma 4.1 and (C8)-(ii) hold. Then, for* $t \leq \upsilon$*,* $\boldsymbol{x} \in supp(\boldsymbol{X})$ *and* $1 \leq q \leq d+1$*, writing* $F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) = \mathbb{P}(Y \leq t|\boldsymbol{Z} = \boldsymbol{z})$ *and* $F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z}) = \mathbb{P}(Y \leq t, \Delta = 1|\boldsymbol{Z} = \boldsymbol{z})$*:*

$$\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \sum_{i=1}^n B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + O_{\mathbb{P}}\left((\log n/(nh^q_{\boldsymbol{X}}))^{3/4} + h^\nu_{\boldsymbol{X}} + h^2_T\right),$$

where

$$\xi(Y_i, \Delta_i, t|\boldsymbol{z}) = (1 - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}))\left[\int_0^{Y_i\wedge t}\frac{-\mathrm{d}F_{Y,1|\boldsymbol{Z}}(s|\boldsymbol{z})}{\{1 - F_{Y|\boldsymbol{Z}}(s|\boldsymbol{z})\}^2} + \frac{\Delta_i\mathbb{1}(Y_i \leq t)}{1 - F_{Y|\boldsymbol{Z}}(Y_i|\boldsymbol{z})}\right]. \tag{D.1}$$

To simplify the presentation, note that the proofs are here written considering the situation $q = 1$. The case $q > 1$ may be considered using the exact same developments, but the notations are more involved.

Proof of Lemma 4.1. For part (a), we first write

$$\begin{aligned}\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \{\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} + \{\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} \\ &= (T_1) + (T_2),\end{aligned}$$

and treat these terms separately. Starting with (T_1), by Proposition 4.3 in Van Keilegom and Akritas (1999), along with the same arguments as for the proof of the Lemma 4.2 on which we concentrate below, we have under our kernel assumptions:

$$\sup_{t\leq v}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})-F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|= O_{\mathbb{P}}\left(\left(\log n/(nh_{\boldsymbol{X}})\right)^{1/2}+h^{\nu}_{\boldsymbol{X}}+h^2_T\right).$$

We now turn to (T_2), and sketch here the main parts for proving the negligibility of this term under our assumptions. To that end, we define first

$$\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})=\sum_{i=1}^n\frac{K\left(\frac{\widehat{\boldsymbol{z}}-\widehat{\boldsymbol{Z}}_i}{h_{\boldsymbol{X}}}\right)}{\sum_{k=1}^n K\left(\frac{\widehat{\boldsymbol{z}}-\widehat{\boldsymbol{Z}}_k}{h_{\boldsymbol{X}}}\right)}\mathbb{1}(Y_i\leq t),$$

$$\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})=\sum_{i=1}^n\frac{K\left(\frac{\widehat{\boldsymbol{z}}-\widehat{\boldsymbol{Z}}_i}{h_{\boldsymbol{X}}}\right)}{\sum_{k=1}^n K\left(\frac{\widehat{\boldsymbol{z}}-\widehat{\boldsymbol{Z}}_k}{h_{\boldsymbol{X}}}\right)}\mathbb{1}(Y_i\leq t,\Delta_i=1).$$

Now, as $\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ and $\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ are to be seen as the building blocks for $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$, showing negligibility of $\sup_{t\leq v}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-\widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})|$, and similarly for $\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}$, implies that the same will hold for $\sup_{t\leq v}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|$. Hence, we focus here on showing the result for $\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$. Specifically, we write

$$\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})=\frac{\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})}{\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}})},$$

where $\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ is an estimator of $L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})=F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})$, with $f_{\boldsymbol{Z}}$ denoting the density of $\boldsymbol{Z}$. We then have

$$\begin{aligned}
&\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-\widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})\\
&\qquad=\frac{1}{nh_{\boldsymbol{X}}}\sum_{i=1}^n\left\{K\left(\frac{\widehat{\gamma}_0^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)-K\left(\frac{\gamma_0^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)\right\}\mathbb{1}(Y_i\leq t)\\
&\qquad=\frac{1}{nh^2_{\boldsymbol{X}}}\sum_{i=1}^n K'\left(\frac{\xi^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)(\widehat{\gamma}_0-\gamma_0)^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)\mathbb{1}(Y_i\leq t)\\
&\qquad=(\widehat{\gamma}_0-\gamma_0)^{\mathsf{T}}\nabla_\gamma\,\widehat{L}_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x})\Big|_{\gamma=\xi},
\end{aligned}$$

for some ξ between $\widehat{\gamma}_0$ and γ_0. Now, by similar arguments as in Proposition 4.3 in Van Keilegom and Akritas (1999) and under assumption (C8)-(i), we have that

$$\sup_{t\leq v}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}\sup_{\gamma\in\Xi}\left\|\nabla_\gamma\,\widehat{L}_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x})-\nabla_\gamma\,L_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x})\right\|=o_{\mathbb{P}}(1).$$

Along with assumption (C9)-(i), we then observe that

$$\sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in \mathrm{supp}(\boldsymbol{X})}|\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-\widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})|= O_{\mathbb{P}}(||\widehat{\gamma}_0-\gamma_0||)=O_{\mathbb{P}}(n^{-1/2}),$$

by assumption (C3). This concludes the proof that (T_2) is negligible under our assumptions, and hence completes the proof of the first part of Lemma 4.1.

For the proof of part (b), we decompose again

$$\begin{aligned}\widehat{f}^{s}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})&=\{\widehat{f}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})-f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\}+\{\widehat{f}^{s}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})-\widehat{f}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\}\\&=(T_3)+(T_4),\end{aligned}$$

and treat these terms separately. For the leading term (T_3), first note that

$$\widehat{f}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})=h_T^{-1}\int\widetilde{K}\left(\frac{t-s}{h_T}\right)\mathrm{d}\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}),$$

from which we decompose (T_3) as follows:

$$\begin{aligned}\widehat{f}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})-f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})&=h_T^{-1}\int\widetilde{K}\left(\frac{t-s}{h_T}\right)\mathrm{d}\left(\widehat{F}_{T|\boldsymbol{Z}}(s|\boldsymbol{z})-F_{T|\boldsymbol{Z}}(s|\boldsymbol{z})\right)\\&\quad+h_T^{-1}\int\widetilde{K}\left(\frac{t-s}{h_T}\right)\mathrm{d}F_{T|\boldsymbol{Z}}(s|\boldsymbol{z})-f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\\&=(T_{31})+(T_{32}).\end{aligned}$$

We start by treating the term (T_{31}). Integrating by parts and using standard change of variables, this term is developed as

$$\begin{aligned}(T_{31})&=h_T^{-1}\int\left\{\widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})-F_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})\right\}\widetilde{K}'(u)\,\mathrm{d}u\\&=h_T^{-1}\int\Big\{\widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})-\mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z}))\\&\qquad\qquad-\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})+\mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}))\Big\}\widetilde{K}'(u)\,\mathrm{d}u\\&\quad+h_T^{-1}\int\Big\{\mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z}))-F_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})\\&\qquad\qquad-\mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}))+F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\Big\}\widetilde{K}'(u)\,\mathrm{d}u\\&\quad+h_T^{-1}\left\{\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})-F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\right\}\int\widetilde{K}'(u)\,\mathrm{d}u\\&=(T_{311})+(T_{312})+(T_{313}),\end{aligned}$$

where the last term (T_{313}) is straightforwardly equal to 0 by assumption (C4). We now treat the remaining terms (T_{311}) and (T_{312}) separately. Starting with (T_{311}),

by Theorem 3(b) in Van Keilegom and Veraverbeke (1996), for some finite constant K_1 we then have that

$$(T_{311}) \leq h_T^{-1} K_1 \sup_{\substack{t,s \leq v \\ |t-s| \leq h_T}} \left| \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - \mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})) - \widehat{F}_{T|\boldsymbol{Z}}(s|\boldsymbol{z}) + \mathbb{E}(\widehat{F}_{T|\boldsymbol{Z}}(s|\boldsymbol{z})) \right|$$
$$= O_{\mathbb{P}}\left((\log n/(nh_{\boldsymbol{X}} h_T))^{1/2} \right).$$

Concerning now the term (T_{312}), following the work of Van Keilegom and Veraverbeke (1997a) with consideration here of higher-order kernels, we have that

$$(T_{312}) = h_T^{-1} \int \Big\{ b(t - uh_T|\boldsymbol{z}) h_{\boldsymbol{X}}^{\nu} + O(h_{\boldsymbol{X}}^{\nu+1}) + O(n^{-1}) \\ - b(t|\boldsymbol{z}) h_{\boldsymbol{X}}^{\nu} + O(h_{\boldsymbol{X}}^{\nu+1}) + O(n^{-1}) \Big\} \widetilde{K}'(u)\, \mathrm{d}u,$$

where the function b is in the expression (3.3) in Van Keilegom and Veraverbeke (1997a), and where their term $o(h_{\boldsymbol{X}}^{\nu})$ is actually observed to be equal to $O(h_{\boldsymbol{X}}^{\nu+1})$. Hence, under Lipschitz continuous requirements for the function $b(t|\boldsymbol{z})$ with respect to t that are induced by assumptions (C4)-(C6), we have that

$$(T_{312}) = h_T^{-1} O\left(h_T h_{\boldsymbol{X}}^{\nu} + h_{\boldsymbol{X}}^{\nu+1} + n^{-1} \right) \\ = O(h_{\boldsymbol{X}}^{\nu}),$$

where the last equality holds by our bandwidths assumptions. We thus observe that (T_{31}) is $O_{\mathbb{P}}((\log n/(nh_{\boldsymbol{X}} h_T))^{1/2} + h_{\boldsymbol{X}}^{\nu})$.

Now, moving on to the term (T_{32}), we write

$$(T_{32}) = h_T^{-1} \int \widetilde{K}\left(\frac{t-s}{h_T} \right) \left\{ f_{T|\boldsymbol{Z}}(s|\boldsymbol{z}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) \right\} \mathrm{d}s.$$

Standard Taylor expansion and change of variables under assumptions (C4) and (C5) then show that this term is $O(h_T^2)$. Hence, bringing back (T_{31}) and (T_{32}), we note that $\sup_{t \leq v} \sup_{\boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})} |(T_3)|$ is $O_{\mathbb{P}}((\log n/(nh_{\boldsymbol{X}} h_T))^{1/2} + h_{\boldsymbol{X}}^{\nu} + h_T^2)$.

It then remains to show the negligibility of (T_4) under our assumptions. For this, using integration by parts and standard change of variables, we have

$$\widehat{f}^{\,s}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{f}^{\,s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = h_T^{-1} \int \left\{ \widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t - uh_T|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z}) \right\} \widetilde{K}'(u)\, \mathrm{d}u \\ \leq \sup_{t \leq v} \sup_{\boldsymbol{x} \in \mathrm{supp}(\boldsymbol{X})} |\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|\, h_T^{-1} \int \widetilde{K}'(u)\, \mathrm{d}u.$$

Under assumption (C4) and using similar arguments as for part (a), the latter term is observed to be negligible with respect to (T_3) under assumption (C8)-(i), hereby concluding the proof. □

Proof of Lemma 4.2. Similarly to the proof of Lemma 4.1, we consider the decomposition

$$\begin{aligned}\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \{\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} + \{\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} \\ &= (T_5) + (T_6),\end{aligned}$$

and treat these terms separately. Starting with (T_5), integrating by parts we first have

$$\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \int \widehat{F}_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z})\widetilde{K}(u)\,\mathrm{d}u. \tag{D.2}$$

Now, using the i.i.d. expansion of $\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})$ uniformly in t and $\boldsymbol{z}$ (see e.g. Theorem 3.2 in Du and Akritas (2002)), we have:

$$\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + R_n(t, \boldsymbol{z}), \tag{D.3}$$

where $\sup_{t\leq\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|R_n(t,\boldsymbol{z})| = O_{\mathbb{P}}\left((\log n)^{3/4}(nh_{\boldsymbol{X}})^{-3/4} + h^{\nu}_{\boldsymbol{X}}\right)$, and ξ is defined in (D.1). Inserting (D.3) in (D.2), we then have

$$\begin{aligned}&\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) \\ &\quad= \int \Big\{F_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z}) + \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t - uh_T|\boldsymbol{z}) \\ &\qquad\qquad + R_n(t - uh_T, \boldsymbol{z})\Big\}\widetilde{K}(u)\mathrm{d}u \\ &\quad= F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) + \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + O_{\mathbb{P}}\left((\log n/(nh_{\boldsymbol{X}}))^{3/4} + h^{\nu}_{\boldsymbol{X}} + h^2_T\right),\end{aligned}$$

using the smoothness of $F_{T|\boldsymbol{Z}}$ and a modulus-of-continuity argument of the empirical distribution function for the function ξ (see Theorem 2.14 in Stute (1982)), along with the assumptions on $\widetilde{K}$ in (C4).

Lastly, concerning part (T_6), note that this term will be $O_{\mathbb{P}}(n^{-1/2})$ by assumption (C3), and is hence asymptotically negligible with respect to the main term in part (T_5). This concludes the proof. □

We now state a third and purely technical lemma employed both for consistency and asymptotic normality of $\widehat{\beta}_\tau$, and start by first introducing the notations used in the latter. For ease of reading, the lemma is here written considering the case $q = 1$. As already commented above, the case $q > 1$ may be considered using the same developments, but the notations are more involved. Now, for notational

convenience with respect to the general framework of Chen et al. (2003) on which the proofs of our following theorems are built, we first define:

$$\begin{aligned} M_n(\beta; F_\gamma, f_\gamma) &= n^{-1}\sum_{i=1}^{n} m(\boldsymbol{X}_i, \beta; F_\gamma, f_\gamma) \\ &= n^{-1}\sum_{i=1}^{n} 2\boldsymbol{X}_i f(\beta^\mathsf{T}\boldsymbol{X}_i|\gamma^\mathsf{T}\boldsymbol{X}_i)\left(F(\beta^\mathsf{T}\boldsymbol{X}_i|\gamma^\mathsf{T}\boldsymbol{X}_i) - \tau\right). \end{aligned}$$

Furthermore, we let

$$\begin{aligned} M(\beta; F_\gamma, f_\gamma) &= \mathbb{E}\left[m(\boldsymbol{X}, \beta; F_\gamma, f_\gamma)\right] \\ &= \mathbb{E}\left[2\boldsymbol{X} f(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X})\left(F(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - \tau\right)\right], \end{aligned}$$

and, conveniently denoting $F_{T|\gamma_0^\mathsf{T}\boldsymbol{X}}$ and $f_{T|\gamma_0^\mathsf{T}\boldsymbol{X}}$ by F_0 and f_0, respectively, we note that $M(\beta_\tau; F_0, f_0) = M_n(\beta_\tau; F_0, f_0) = 0$. We further denote by $||\cdot||$ the Euclidean distance and

$$\begin{aligned} d_{\mathcal{F}_F}(F_{\gamma_1}, F^\star_{\gamma_2}) &= \sup_{t\le\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})} |F(t|\gamma_1^\mathsf{T}\boldsymbol{x}) - F^\star(t|\gamma_2^\mathsf{T}\boldsymbol{x})| \\ d_{\mathcal{F}_f}(f_{\gamma_1}, f^\star_{\gamma_2}) &= \sup_{t\le\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})} |f(t|\gamma_1^\mathsf{T}\boldsymbol{x}) - f^\star(t|\gamma_2^\mathsf{T}\boldsymbol{x})|, \end{aligned}$$

for any $F_{\gamma_1}, F^\star_{\gamma_2} \in \mathcal{F}_F$ and $f_{\gamma_1}, f^\star_{\gamma_2} \in \mathcal{F}_f$, where $\mathcal{F}_F$ and $\mathcal{F}_f$ are defined in our assumptions.

Lemma 4.3. *Under assumptions (C1), (C5) and (C7), we have for all positive* $\delta_n = o(1)$*,*

$$\sup_{||\beta-\beta_\tau||\le\delta_n}\ \sup_{d_{\mathcal{F}_F}(F_\gamma,F_0)\le\delta_n}\ \sup_{d_{\mathcal{F}_f}(f_\gamma,f_0)\le\delta_n} ||M_n(\beta; F_\gamma, f_\gamma) - M(\beta; F_\gamma, f_\gamma)|| = o_{\mathbb{P}}(n^{-1/2}).$$

Proof of Lemma 4.3. Note that this result is similar to condition (2.5') in Chen et al. with $M_n(\beta_\tau; F_0, f_0) = 0$ in our context. Hence, it suffices here to verify the conditions (3.1) and (3.3) in their Theorem 3. To that end, starting with (3.1), we note that for $(\beta_1, F_{\gamma_1}, f_{\gamma_1}) \in \mathcal{B}\times\mathcal{F}_F\times\mathcal{F}_f$ and $(\beta_2, F^\star_{\gamma_2}, f^\star_{\gamma_2}) \in \mathcal{B}\times\mathcal{F}_F\times\mathcal{F}_f$, we have

$$\begin{aligned} &||m(\boldsymbol{x}, \beta_1; F_{\gamma_1}, f_{\gamma_1}) - m(\boldsymbol{x}, \beta_2; F^\star_{\gamma_2}, f^\star_{\gamma_2})|| \\ &\quad= \Big|\Big|2\boldsymbol{x}\Big(f(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_1^\mathsf{T}\boldsymbol{x})\{F(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_1^\mathsf{T}\boldsymbol{x}) - \tau\} - f^\star(\beta_2^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x})\{F^\star(\beta_2^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x}) - \tau\}\Big)\Big|\Big| \\ &\quad\le K_1\Big(||F(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_1^\mathsf{T}\boldsymbol{x}) - F^\star(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x})||+||F^\star(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x}) - F^\star(\beta_2^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x})|| \\ &\qquad\quad + ||f(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_1^\mathsf{T}\boldsymbol{x}) - f^\star(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x})||+||f^\star(\beta_1^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x}) - f^\star(\beta_2^\mathsf{T}\boldsymbol{x}|\gamma_2^\mathsf{T}\boldsymbol{x})||\Big) \\ &\quad\le K_2\Big(d_{\mathcal{F}_F}(F_{\gamma_1}, F^\star_{\gamma_2}) + d_{\mathcal{F}_f}(f_{\gamma_1}, f^\star_{\gamma_2}) + ||\beta_1 - \beta_2||\Big), \end{aligned}$$

for some finite constants K_1 and K_2 under assumptions (C1), (C5) and (C7), hereby verifying condition (3.1) in Chen et al. with $r = 2$.

Next, for condition (3.3), for $\epsilon > 0$ we denote first by $N(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}})$ the covering number (Van der Vaart and Wellner (1996, p. 83)) of the class $\mathcal{F}_F$ under the sup-norm metric we consider on the latter, and similarly for the class $\mathcal{F}_f$. The proof is then written for the class $\mathcal{F}_F$, but the same arguments hold for the class $\mathcal{F}_f$ and are here thus omitted.

Now, we recall that $N(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F}) \leq N_{[]}(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F})$, where the notation $N_{[]}(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F})$ denotes the ϵ-bracketing number of the class $\mathcal{F}_F$ with respect to the same sup-norm metric (Van der Vaart and Wellner (1996, p. 83)). This suggests we may verify here condition (3.3) by concentrating on bracketing numbers rather than covering numbers. Now, since all the functions in the class $\mathcal{F}_F$ have values between 0 and 1 (and between 0 and M for $\mathcal{F}_f$ as stated in assumption (C5)), we then observe that only one ϵ-bracket suffices to cover $\mathcal{F}_F$ if $\epsilon > 1$. Using Lemma 6.1 in Lopez (2011) for a bound on the bracketing number for the case $\epsilon \leq 1$, we then have that

$$\begin{aligned}\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F})}\,\mathrm{d}\epsilon &\leq \int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F})}\,\mathrm{d}\epsilon \\ &\leq K \int_0^1 \epsilon^{-\frac{1}{1+\eta}}\,\mathrm{d}\epsilon \\ &< \infty\,,\end{aligned}$$

for some finite constant K, hereby satisfying condition (3.3) in Chen et al. for $s_j = 1$. An application of Theorem 3 in Chen et al. then concludes the proof. □

We may now at last turn to the proofs of the two main results reported in Section 4.3.

Proof of Theorem 4.1. To prove $\widehat{\beta}_\tau$ is a weakly consistent estimator of β_τ, we choose to verify the five high level conditions (1.1)-(1.5) stated in Theorem 1 of Chen et al. (2003). Starting with condition (1.1), note that the latter is readily satisfied in our framework by construction of $\widehat{\beta}_\tau$, while condition (1.3) holds simply under assumption (C5). Furthermore, using the definitions of $\mathcal{F}_F$ and $\mathcal{F}_f$ in (C5) as the spaces embedding the nuisance parameters $F_{T|\boldsymbol{Z}}$ and $f_{T|\boldsymbol{Z}}$, and equiping the latter with the distances $d_{\mathcal{F}_F}(F_{\gamma_1}, F^\star_{\gamma_2})$ and $d_{\mathcal{F}_f}(f_{\gamma_1}, f^\star_{\gamma_2})$ as in Lemma 4.3, note that (1.4) in Chen et al. is straightforwardly satisfied by Lemma 4.1 under assumption (C8)-(i). As for condition (1.5), this is a weaker version of condition (2.5) in Chen et al. which is verified similarly as for Lemma 4.3. It therefore only remains to verify here condition (1.2) required for the uniqueness of β_τ in our model.

Hence, recalling that $F_{T|\boldsymbol{Z}}(t|\gamma_0^\mathsf{T}\boldsymbol{x}) = F_{T|\boldsymbol{X}}(t|\boldsymbol{x})$ for any t and $\boldsymbol{x}$, we need to

verify that for any $\epsilon > 0$, $\inf_{||\beta-\beta_\tau||>\epsilon} ||M(\beta; F_0, f_0)|| > 0$. To that end, note that

$$\begin{aligned}
&\inf_{||\beta-\beta_\tau||>\epsilon} \Big|\Big|\mathbb{E}\Big[\boldsymbol{X} f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left(F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau\right)\Big]\Big|\Big| \\
&= \inf_{||\beta-\beta_\tau||>\epsilon} \Big|\Big|\mathbb{E}\Big[\boldsymbol{X} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\left(F_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) - F_{T|\boldsymbol{X}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right)\Big]\Big|\Big| \\
&= \inf_{||\beta-\beta_\tau||>\epsilon} \left|\left|\int_{\mathrm{supp}(\boldsymbol{X})} \boldsymbol{x} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{x}|\boldsymbol{x}) \int_{\beta_\tau^\mathsf{T}\boldsymbol{x}}^{\beta^\mathsf{T}\boldsymbol{x}} f_{T|\boldsymbol{X}}(t|\boldsymbol{x})\mathrm{d}t\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\right|\right|,
\end{aligned}$$

where $F_{\boldsymbol{X}}(\boldsymbol{x})$ denotes the c.d.f. of $\boldsymbol{X}$. Under assumptions (C1) and (C2), the latter quantity is observed to be strictly positive, hereby ensuring condition (1.2) is satisfied. Hence the assumptions of Theorem 1 in Chen et al. are met, from which the weak consistency of $\widehat{\beta}_\tau$ follows. □

Proof of Theorem 4.2. The proof consists in verifying the six high-level conditions depicted in Theorem 2 of Chen et al. (2003). To that end, similarly to the context of the latter theorem, we replace in this section the spaces $\mathcal{B}$, $\mathcal{F}_F$ and $\mathcal{F}_f$ by shrinking neighborhoods around the true β_τ, $F_{T|\boldsymbol{Z}}$ and $f_{T|\boldsymbol{Z}}$. Specifically, we define the spaces $\mathcal{B}_\delta = \{\beta \in \mathcal{B} : ||\beta - \beta_\tau|| \leq \delta_n\}$, $\mathcal{F}_{F_\delta} = \left\{F \in \mathcal{F}_F : d_{\mathcal{F}_F}(F, F_{T|\boldsymbol{Z}}) \leq \delta_n\right\}$ and $\mathcal{F}_{f_\delta} = \left\{f \in \mathcal{F}_f : d_{\mathcal{F}_f}(f, f_{T|\boldsymbol{Z}}) \leq \delta_n\right\}$ for some $\delta_n = o(1)$. Furthermore, for notational convenience with the work of Chen et al., we use the abbreviations $s_\gamma = (F_\gamma, f_\gamma)$, $s_0 = (F_0, f_0)$, $\widehat{s} = (\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}, \widehat{f}^s_{T|\widehat{\boldsymbol{Z}}})$, and rewrite $M_n(\beta; s_\gamma) = M_n(\beta; F_\gamma, f_\gamma)$ and $M(\beta; s_\gamma) = M(\beta; F_\gamma, f_\gamma)$. Lastly, we define the following norm on the vector of nuisance parameters: $d_{\mathcal{S}_\delta}(s_{\gamma_1}, s^\star_{\gamma_2}) = \max(d_{\mathcal{F}_{F_\delta}}(F_{\gamma_1}, F^\star_{\gamma_2}), d_{\mathcal{F}_{f_\delta}}(f_{\gamma_1}, f^\star_{\gamma_2}))$.

Now, starting with condition (2.1), note that the latter is readily satisfied in our framework by construction of our estimator. Next, for condition (2.2), we first need to determine the expression of the ordinary derivative of $M(\beta; s_0)$ with respect to β calculated at β_τ, denoted by $\Gamma_1(\beta; s_0)$:

$$\begin{aligned}
\Gamma_1(\beta; s_0) &:= \frac{\partial M(\beta; s_0)}{\partial \beta} \\
&= \mathbb{E}\Big[2\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\Big(f^2_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \\
&\qquad + f'_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left\{F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau\right\}\Big)\Big],
\end{aligned}$$

where $f'_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \partial/\partial t\, f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})$. In particular, recalling that $F_{T|\boldsymbol{Z}}(t|\gamma_0^\mathsf{T}\boldsymbol{x}) = F_{T|\boldsymbol{X}}(t|\boldsymbol{x})$ for any t and $\boldsymbol{x}$, we have under assumption (C5)

$$\Gamma_1(\beta_\tau; s_0) = 2\,\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^\mathsf{T} f^2_{T|\boldsymbol{X}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right].$$

Under assumptions (C1), (C2), (C5) and (C7), $\Gamma_1(\beta; s_0)$ is thus observed to be continuous and of full rank at β_τ, hereby verifying condition (2.2).

Next, for condition (2.3), we first need to determine for all $\beta \in \mathcal{B}_\delta, F \in \mathcal{F}_{F_\delta}$ and $f \in \mathcal{F}_{f_\delta}$ the functional derivative of $M(\beta; s_0)$ at s_0 in the direction $[s_\gamma - s_0]$. Denoting the latter by $\Gamma_2(\beta; s_0)[s_\gamma - s_0]$, we have

$$\begin{aligned}
&\Gamma_2(\beta; s_0)[s_\gamma - s_0] \\
&\quad := \lim_{\eta \to 0} \frac{1}{\eta} \left[M(\beta; s_0 + \eta(s_\gamma - s_0)) - M(\beta; s_0) \right] \\
&\quad = 2\,\mathbb{E}\Big[\boldsymbol{X} \Big(f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \left\{ F(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \right\} \\
&\qquad\qquad + \left\{ f(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \right\} \left\{ F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau \right\} \Big) \Big].
\end{aligned}$$

In particular, for $\beta = \beta_\tau$ we have

$$\Gamma_2(\beta_\tau; s_0)[s_\gamma - s_0] = 2\,\mathbb{E}\left[\boldsymbol{X} f_{T|\boldsymbol{X}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) \left\{ F(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - \tau \right\} \right].$$

Now, for condition (2.3)-(i), we have for all $\beta \in \mathcal{B}_\delta, F \in \mathcal{F}_{F_\delta}$ and $f \in \mathcal{F}_{f_\delta}$ that

$$\begin{aligned}
&||M(\beta; s_\gamma) - M(\beta; s_0) - \Gamma_2(\beta; s_0)[s_\gamma - s_0]|| \\
&\quad = \Big|\Big| \mathbb{E}\Big[2\boldsymbol{X} \left\{ f(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \right\} \\
&\qquad\qquad \times \left\{ F(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \right\} \Big] \Big|\Big| \\
&\quad \leq K_1\, d_{\mathcal{F}_{F_\delta}}(F_\gamma, F_0)\, d_{\mathcal{F}_{f_\delta}}(f_\gamma, f_0) \\
&\quad \leq K_2\, d^2_{\mathcal{S}_\delta}(s_\gamma, s_0), \qquad \text{(D.4)}
\end{aligned}$$

for some finite constants K_1, K_2, hereby veryfying (2.3)-(i). As for condition (2.3)-(ii), using a Taylor expansion with assumptions (C1), (C5) and (C7), we have

$$||\Gamma_2(\beta; s_0)[s_\gamma - s_0] - \Gamma_2(\beta_\tau; s_0)[s_\gamma - s_0]|| \leq ||\beta - \beta_\tau|| o(1).$$

Moving now to conditions (2.4) and (2.5), note that the first part of (2.4) will follow by similar arguments as in Lemma 6.2 in Lopez (2011). As for the second part of this condition, this follows readily from Lemma 4.1 provided assumption (C8)-(ii) holds. Condition (2.5) is, for its part, verified by Lemma 4.3 and Remark 2 in Chen et al..

Lastly, for condition (2.6), recalling that $M_n(\beta_\tau; s_0) = 0$, we only need to prove that

$$n^{1/2} \left[\Gamma_2(\beta_\tau; s_0)[\widehat{s} - s_0] \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

for some positive definite matrix Σ. To that end, using Lemma 4.2, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X} f_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \left\{ \widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\widehat{\gamma}_0^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) \right\} \right] \\
&= (nh_{\boldsymbol{X}})^{-1} \mathbb{E}_{\gamma_0^\mathsf{T}\boldsymbol{X}} \left[\sum_{i=1}^n \frac{K\left(\frac{\gamma_0^\mathsf{T}(\boldsymbol{X} - \boldsymbol{X}_i)}{h_{\boldsymbol{X}}} \right)}{(nh_{\boldsymbol{X}})^{-1} \sum_{k=1}^n K\left(\frac{\gamma_0^\mathsf{T}(\boldsymbol{X} - \boldsymbol{X}_k)}{h_{\boldsymbol{X}}} \right)} g_i(\gamma_0^\mathsf{T}\boldsymbol{X}) \right] + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}$$

provided assumption (C8)-(ii) holds, where the notations $\mathbb{E}_{\boldsymbol{X}}$ and $\mathbb{E}_{\gamma_0^\intercal \boldsymbol{X}}$ explicitly indicate that the expectations are taken with respect to the distributions of $\boldsymbol{X}$ and $\gamma_0^\intercal \boldsymbol{X}$, respectively, and where

$$g_i(u) = \int_{\mathrm{supp}(\boldsymbol{X})} \boldsymbol{x} f_{T|\boldsymbol{Z}}(\beta_\tau^\intercal \boldsymbol{x}|u)\xi(Y_i, \Delta_i, \beta_\tau^\intercal \boldsymbol{x}|u) f_{\boldsymbol{X}|\gamma_0^\intercal \boldsymbol{X}}(\boldsymbol{x}|u)\,\mathrm{d}\boldsymbol{x}. \tag{D.5}$$

Simplifying a bit and writing $\boldsymbol{z} = \gamma_0^\intercal \boldsymbol{x}$ and $\boldsymbol{Z}_i = \gamma_0^\intercal \boldsymbol{X}_i, i = 1, \ldots, n$, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}}\Big[\boldsymbol{X} f_{T|\boldsymbol{Z}}(\beta_\tau^\intercal \boldsymbol{X}|\gamma_0^\intercal \boldsymbol{X}) \left\{ \widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta_\tau^\intercal \boldsymbol{X}|\widehat{\gamma}_0^\intercal \boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta_\tau^\intercal \boldsymbol{X}|\gamma_0^\intercal \boldsymbol{X}) \right\} \Big] \\ = (nh_{\boldsymbol{X}})^{-1} \sum_{i=1}^n \int_{-\infty}^{+\infty} g_i(\boldsymbol{z}) K\left(\frac{\boldsymbol{z} - \boldsymbol{Z}_i}{h_{\boldsymbol{X}}} \right) \mathrm{d}\boldsymbol{z} + o_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

Using standard change of variables and a Taylor expansion for g under assumption (C9)-(ii), we then have

$$\begin{aligned}\Gamma_2(\beta_\tau; s_0)[\widehat{s} - s_0] &= 2\, n^{-1} \sum_{i=1}^n g_i(\gamma_0^\intercal \boldsymbol{X}_i) + O(h_{\boldsymbol{X}}^\nu) + o_{\mathbb{P}}(n^{-1/2}) \\ &= 2\, n^{-1} \sum_{i=1}^n g_i(\gamma_0^\intercal \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

provided assumptions (C4) and (C8)-(ii) hold. An application of the central limit theorem under our assumptions then leads to the desired result, hereby concluding the proof. □

CHAPTER 5

Discussion and Extensions

In this thesis, we have explored and developed several techniques to estimate a quantile regression model in the presence of complete and, more particularly, censored response variables. Censoring is one of the many phenomena encountered in practice that may prevent the observer from completely detecting the data of interest, and is in particular often associated to time-to-event variables given the latter are, by nature, not measured instantaneously and hence exposed to many potential interferences during the time required for gathering observations. As a result, many interesting applications have to cope in their data analysis with the knowledge that the observed responses are, in fact, only the minimum value between the true response of interest and a random censoring variable. The latter mechanism corresponds to a particular form of censoring known as random right censoring, to which our attention has been drawn throughout this thesis. Providing appropriate quantile regression modelling and subsequent estimation in this context has then been the main motivation of our work, for which we devote this last chapter to the summary of some general conclusions following our journey before dedicating a few lines to potential extensions one might consider for further research in the domain.

5.1 General conclusion

Our investigation of quantile regression models started in Chapter 2 with the consideration of a copula-based multivariate regression modelling. Several conveniences have been driving here the motivation for employing copulas in a general regression framework, as the latter approach namely allows to straightforwardly avoid common regression issues such as the examination of transformed covariates or the potential inclusion of interactions among the latter. Additionally, for the specific quantile context, a copula-based approach allows the analyst to further ensure

that the regression estimations will satisfyingly be automatically monotonic across quantile levels, as observed both in the work of Noh et al. (2015) and the developments of Chapter 2.

With these incentives at hand, Chapter 2 provides two main fronts for contributions to the literature on semiparametric quantile regression with both complete and censored data. Starting with the situation for complete response observations, the first line of work of this chapter was to provide a bypass to the shortcomings related to the misspecification of the copula family when considering a parametrization of the latter, as was initially suggested in the works of Noh et al. (2013) and Noh et al. (2015). A somewhat surprisingly strong limitation for the latter approach was indeed unveiled by Dette et al. (2014) and further reported in the motivational example of Section 2.2.2, where the approach of Noh et al. using several common parametric copulas was illustrated to provide a very poor fit for the estimation of a regression function that is non-monotonic in a one-dimensional covariate. In this context, and supported by several other examples, the main finding of Dette et al. consisted in establishing that most commonly used parametric copula families are in fact unable to reproduce the non-monotonic feature of the regression function. Hence, the need for an alternative copula estimation strategy to remedy this shortcoming motivated the first part of Chapter 2. Our subsequent suggestion consisted in combining the ideas of pair-copula decomposition with the observation that a nonparametric estimation of the copula density may, as expected, result in an appropriate regression estimation in case of low-dimensional covariates. Therefore, turning to higher-dimensional problems, we advocated for a decomposition into two main parts of the multivariate copula density that needs to be estimated in a regression model with d covariates: the first part contains the product of the d bivariate copula densities modelling the dependence between the response variable and each covariate at hand, while the second part contains a d-variate conditional copula density relative to the dependence between the covariates, conditionally on the response variable. In a regression context, our simple argument was then to provide as much flexibility as needed for the estimation of the first part, given that the latter incorporates the dependence of actual interest and for which nonparametric estimation is best suited. On the other hand, a parametric modelling was advised for the second part of the decomposition in order to account for dimensionality concerns. Our bivariate simplistic example of Section 2.2.3 and the small simulation study of Section 2.5.1 both tend to support the adequateness of such an approach for copula-based regression estimation. For quantile regressions, the resulting estimator was then observed to provide a strong competitor to the existing literature with complete observations, as highlighted through the brief simulation results of Section 2.5.2.

The second contribution of Chapter 2 concerned the inclusion of potential right censored responses in the problem, for which the literature on semiparametric quantile regression estimation is rather sparse. Similarly to the single-index methodology of Bücher et al. (2014), our handling of censoring was then dealt with through the

use of inverse-censoring-probability (ICP) weights under the assumption of conditional independence between the response and the censoring variable given the covariates. Such ICP weights offer in a first step the convenient possibility of recovering the conditional expected loss of the true response by considering instead the conditional expected loss of the actually observed responses, with an appropriate weighting adjustment that incorporates the censoring indicator. In the context of copula-based regression, in order to avoid having to cope with copulas for censored variables and copulas with discrete variables, our suggestion was then in a second step to rewrite the problem such that the latter censoring indicator is in fact moved to the conditional part of the copula density entering the procedure. By doing so, one eventually retrieves a very analogous expression for conditional quantiles to the case of strictly complete observations. This then straightforwardly suggests that a similar approach to the one described above may be exploited for the required estimation of the multivariate copula density, by focusing exclusively on the responses for which it is known that the latter are indeed uncensored.
From a theoretical point of view, asymptotic normality for the proposed copula-based quantile regression estimator was obtained for both complete and censored responses, where the convergence rate is driven, regardless of the dimension d of the covariate, by the nonparametric estimation of bivariate copula densities in the first part of our copula estimation decomposition. Finally, for the analysis of finite-sample behavior of the proposed estimator for censored data, an extensive simulation study was provided where the increased flexibility of the procedure over its competitors was illustrated by depicting, first, competitive results in settings that favor the literature, and second, outperforming results in slightly altered versions of the previous settings where the literature is hence no longer favored.
However, in an attempt to provide at this point a constructive self-criticism for the methodology, the obtained promising finite-sample results may be put into perspective with the knowledge that ICP weights, although convenient from a modelling point of view, do not exactly take the quantile context into account for the handling of censored observations, as multiply evoked throughout this thesis. Additionally, our particular association of ICP weights with copulas resulted in the formulation of conditional quantiles through a copula density that is conditional on the censoring indicator, hence exposing the procedure to possible estimation difficulties in case of an important censoring proportion for too low sample size. These considerations then motivated the search in this thesis for an alternative, and more specific, handling of censoring in the general context of quantiles.

The developments of Chapter 3 then resulted from this search for a more quantile-specific consideration of censoring. For the latter objective, a review of the literature revealed that prior work on the subject mainly focuses on two different strategies in order to account for censored responses with unconditional quantiles, both based on a check-based formulation of the form $m_\tau = \arg\min_a \mathbb{E}[\rho_\tau(T-a)]$; first, a wide literature is devoted to correcting the 'data part' of the latter formula through the expression of appropriate weighting schemes, of which ICP weights and

the technique of redistribution-of-mass (RM) are predominant, and second, a much more modest literature proposes to focus on the quantile level τ in the check-based formula by targeting a higher level the initial level τ with complete observations. From this examination, a natural idea in order to account for censoring was then to focus on the third and last element of the check-based formula, that is, the loss function itself. This was precisely the approach of Chapter 3.

In order to provide a loss function that automatically accounts for censoring, our suggestion was then to mimic the link between the check function and the definition of quantiles. For censored data and under the usual assumption of conditional independence between the response and the censoring variable given the covariates, this resulted in a loss function that naturally generalizes the check function by introducing a correcting term for censoring, where the amount of alteration from the check function is all the more important given that the probability of censoring occurring below the value of the quantile is important. This is pleasantly in line with the observation that the phenomenon of censoring in the context of quantiles should be accounted for especially when it is susceptible of generating an underestimation of the value of the quantile.

In the objective of illustrating the potential of the newly proposed loss function, our attention was then drawn in particular to the linear regression model, given that the latter is subject to estimation suggestions in the literature based on both ICP and RM weights. Our resulting estimator in this context was then straightforwardly defined as the direct counterpart of the estimator for complete observations, where the the check function is simply replaced by with its newly proposed censored version. For theoretical interests in this context, based on several recent works on non-smooth semiparametric estimation equations with an infinite-dimensional nuisance parameter, both consistency and asymptotic normality were then established for the quantile coefficient estimator based on the adapted loss function.

Lastly, for practical implementation of our procedure in this context, and given that the proposed loss function is no longer convex, a simple adaptation of an established algorithm for quantile regression with complete observations was proposed. Simulation results subsequently revealed two major insights; first, the proposed procedure confidently outperforms the basic ICP weighting scheme in this linear context, especially for higher censoring proportions. This is argued to come as a natural consequence of discarding the use of the censoring indicator in the numerator of the check-based formulation. The second main insight is that the procedure also provides a valuable competitor to the RM weighting strategy, often taken as primary reference in this linear context. The latter statement is again especially emphasized when the censoring proportion is high in the dataset. Taking one step back from the linear framework, these results then make us conjecture that the proposed estimation strategy naturally opens the door to enhancements of numerous alternative modelling techniques in the censored quantile regression literature, of which the most obvious are those that currently exploit the basic ICP weighting approach, as will be illustrated in the next section.

Nevertheless, as far as linear quantile regression is concerned, a general observation emphasized by this work as well is that censored data drives the most prolific approaches to rely in their estimation procedure on a preliminary nonparametric estimation of a conditional distribution function, even though the model is parametrically specified. In particular, while the RM weighting scheme requires the estimation of the conditional distribution of the response given the covariates, the adapted check function is observed to require for its part the estimation of the conditional distribution of the censoring variable given the covariates. This then raised the legitimate question of whether such approaches may be improved by plainly acknowledging the need to rely on a conditional distribution estimator in a linear quantile framework with censored data, hereby introducing the work of Chapter 4.

In Chapter 4, we thus proposed to take a step back from the classical approach for estimating a censored linear quantile regression by circumventing a check-based formulation of conditional quantiles. Instead, our suggestion was to extend to a parametric framework a well-studied technique usually employed for nonparametric regression that consists in first estimating the conditional distribution of the response given the covariates, and then inverting it in a second step in order to recover the conditional quantiles. Since censoring transcribes, as mentioned above, in a linear context into the need of estimating a conditional distribution, we observed that such a technique could then offer a new insight into censored linear quantile regression estimation. An additional and more moderate argument was also invoked to motivate such an approach, as the latter is observed in the nonparametric regression literature to usually numerically outperform its check-based counterparts, especially for quantile levels that are not 'too distant' from the median. This is then to be put in parallel with a classical identifiablity constraint for censored data that prevents one from confidently studying upper quantile levels, and forcing instead to consider lower and more central quantile levels.
Hence, rather than minimizing the conventional check loss function, it was suggested in this work to estimate the linear regression coefficients by minimizing the distance between the quantile level of interest and an estimator of the conditional distribution of the response given the covariates, evaluated at the value of the linear function for each observation. For optimization concerns, and driven by further knowledge coming from the literature, a 'double-kernel' approach was then suggested for the nonparametric estimation of the conditional distribution, that is, the latter estimator accounts for kernel smoothing both in the directions of the covariates and the response. In small dimensional settings, the resulting linear regression estimator was then observed through Monte Carlo simulations to provide strikingly competitive results with respect to its check-based counterparts for censored data. In fact, the consideration of complete observations in the simulation study further revealed that the improvement of the proposed estimator over its competitors in some settings is such that the procedure is even able to compete, when the censoring proportion is no too high, with a check-based estimator

for strictly complete observations, especially for the estimation of the median. A general pattern sustaining the obtained results seems to be a reduction in variance of the proposed estimator across simulated datasets, admittedly at the cost of a small increase in bias. These observations are in fact in line with the results in the literature on nonparametric quantile regression based on inverting a double-kernel estimator of the conditional distribution of the response given the covariates.
For higher-dimensional settings and similarly to some existing literature on censored quantile regression, the investigation of a global dimension reduction technique was further considered for the required nonparametric conditional distribution estimator. For theoretical concerns for the resulting linear regression estimator, both consistency and asymptotic normality were then here again established under classical regularity conditions. As a by-product of these results, the work of this chapter also required the establishment of several new asymptotic results for the double-kernel estimators of the conditional distribution and density of the censored response given the covariates, where the covariates are considered along with the dimension reduction framework.
Lastly, numerical results for multivariate settings then highlighted similar patterns as for low-dimensional settings, as a decrease in variance results were observed to lead to global outperforming results of the proposed linear regression estimator with respect to its competitors. Overall, the work of this chapter then suggests that for linear quantile regression, the proposed technique offers a very valuable estimation procedure with respect to the usually employed check-based formulation of conditional quantiles, especially here when the censoring proportion is moderate in the dataset.

These few lines complete this section on some general insights and conclusions following the work of this thesis. We now devote the following and last section to some perspectives offered by the latter for potential further research in the domain.

5.2 Future research

As for any doctoral studies, the end of the presented work is rather regulated by limited time than by the lack of research ideas and interests. To illustrate this statement, this section is devoted to the development of a few directions for further research that are believed to be potentially worth exploring. Similarly to the main body of this thesis, we first mention that we focus here primarily on the investigation of possible estimation procedures in censored quantile regression, hereby deliberately leaving out the development of related statistical methodologies such as goodness-of-fit testing. Now, our objective in this section is twofold: first, we consider the development of potential estimation procedures in the context of censored quantile regression where the observed data have the same structure as throughout this thesis, before devoting in a second step a few lines to the consideration of additional features in the data that will influence the estimation procedures. As a preliminary remark, we also point out that the approach of Chapter 3 consisting in

appropriately taking incompleteness of the data into account at the level of the loss function could, in spirit, be further extended to alternative phenomena that interfere with a complete observation of the data. The latter could for instance result in the presence of *left-truncation*, where one only observes data typically satisfying a certain boundary condition for entering the sample, or in *double-censoring* as intended by Turnbull (1974), where left and right censoring are both present in the dataset. The main challenge to mimic the approach of Chapter 3 would then be, based on the particularities of the considered context, to find a similar link between the development of the check function for complete observations in (1.1.3) and the development of the new loss function for censored data starting from (3.2.2). This consideration is here however left for further investigation.

5.2.1 Related procedures in censored quantile regression

Our investigation of further estimation procedures in censored quantile regression is considered in this section in particular through two examples, although the new loss function of Chapter 3 offers by nature undoubtedly a great potential for many other extensions in the domain. To that end, our starting point is the evident bridging of copula-based regression modelling with the adapted loss function for censored data in a first proposal. In a second suggestion, and as a further illustration of the flexibility of the adapted check function, we then mention a few words on the consideration of a variable selection feature in the parametric regression framework.

An alternative censored copula quantile regression estimator

The most straightforward extension of our work is here briefly developed, and consists in combining in the context of censored quantile regression the advantages of a copula-based formulation with the handling of censoring through the newly proposed loss function. To that end, using the same notations and similar arguments as in Chapters 2 and 3 for a regression model with d covariates, we have that

$$\begin{aligned} m_\tau(\boldsymbol{x}) &= \arg\min_a \mathbb{E}\Big[\varphi_\tau(a; Y, G_C(\cdot|\boldsymbol{x}))\Big|\boldsymbol{X} = \boldsymbol{x}\Big] \\ &= \arg\min_a \mathbb{E}\Big[\varphi_\tau(a; Y, G_C(\cdot|\boldsymbol{x}))\, c_{Y\boldsymbol{X}}(F_Y(Y), \boldsymbol{F}(\boldsymbol{x}))\Big], \end{aligned}$$

where F_Y is the c.d.f. of Y, $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1), \ldots, F_d(x_d))^\mathsf{T}$ and $c_{Y\boldsymbol{X}}$ is the unique copula density of $(Y, \boldsymbol{X})$. Consequently, a simple estimator of $m_\tau(\boldsymbol{x})$ is naturally given by

$$\widehat{m}_\tau(\boldsymbol{x}) = \arg\min_a \sum_{i=1}^n \varphi_\tau(a; Y_i, \widehat{G}_C(\cdot|\boldsymbol{x}))\, \widehat{c}_{Y\boldsymbol{X}}(\widehat{F}_Y(Y_i), \widehat{\boldsymbol{F}}(\boldsymbol{x})), \tag{5.2.1}$$

where all unknown quantities are hence replaced by appropriate estimators as in the previous chapters. On paper, one could expect such an approach to offer

three main advantages over the estimator of Chapter 2 where ICP weights were exploited for the handling of censoring; first, the adapted check function discards the use of the censoring indicator in the numerator of the estimation equation, which is of particular interest when the censoring proportion is high in the dataset, as illustrated namely in the numerical results of Chapters 3 and 4. The second advantage concerns the estimation of the copula density, for which no conditioning on the censoring indicator is here implied, thereby allowing in contrast to the procedure of Chapter 2 to exploit all observations at hand for this part as well. Finally, a transcription to this setting of the MM algorithm exploited in Chapter 3 reveals that the numerical minimization of (5.2.1) will only require the evaluation of $\widehat{G}_C(\cdot|\boldsymbol{x})$ in every $\widehat{m}_\tau^{(k)}(\boldsymbol{x})$, where $\widehat{m}_\tau^{(k)}(\boldsymbol{x})$ is the estimation of $m_\tau(\boldsymbol{x})$ at iteration k of the algorithm. This is again to be contrasted with the use of ICP weights inducing the assessment of $\widehat{G}_C(\cdot|\boldsymbol{x})$ in every $Y_i, i = 1, \ldots, n$, hereby forcing the evaluation of the latter estimator in the right tail as well, for which classical estimators are unstable due to sparse observations.

In practice however, early simulation results somewhat surprisingly did not reveal striking improvements over the procedure of Chapter 2. Surely, a major challenge that needs to be addressed in order to refute this finding concerns the handling of $\widehat{G}_C(\cdot|\boldsymbol{x})$ in (5.2.1), for which a flexible modelling and subsequent estimation will be defying when the dimension d is too large to confidently rely on Beran's estimator as in Chapter 3. A suggestion that merits further investigation would be to apply in this setting a similar dimension reduction technique as in Chapter 4, that is, estimate $G_C(\cdot|\boldsymbol{x})$ under the assumption that $C \perp\!\!\!\perp \boldsymbol{X}|(\gamma_{0,1}^\intercal \boldsymbol{X}, \ldots, \gamma_{0,q}^\intercal \boldsymbol{X})$, where the γ_0's are linearly independent vectors. A simpler, but arguably less flexible alternative would be to consider a Cox regression modelling between the covariates and the censoring variable, similarly as in Chapter 2.

Variable selection in censored quantile regression

Our second suggestion for further work leaves a copula-based modelling aside and digs instead further into the linear context exploited through the adapted loss function of Chapter 3. In particular, in order to account for higher-dimensional settings in this framework, our attention is drawn to the issue of variable selection appearing in situations where the number of covariates at hand is large but only a subset of the latter are actually relevant in the regression model. A common technique in this context is to introduce a penalty function in the estimation procedure that forces the estimated coefficients of irrelevant variables to be shrunk towards 0, hereby allowing in fact estimation and variable selection at once. Among those penalty functions, perhaps the most widely spread nowadays are the Lasso (Tibshirani (1996)), SCAD ('smoothly clipped absolute deviations', Fan and Li (2001)) and Adaptive Lasso (Zou (2006)), all of which have been adapted to various regression settings in a rich literature. For the particular case of censored quantile regression, both ICP and RM weighting schemes have been exploited in association

with an adaptive lasso penalization in Shows et al. (2010) and Wang et al. (2013), respectively. For a check-based modelling that easily accommodates such a penalty, the simulation results of Chapter 3 then make us conjecture that a valuable complement to the latter literature, especially when the censoring proportion is important, would be to consider an estimator of the form:

$$\widehat{\beta}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \varphi_\tau \left(\beta^\intercal \boldsymbol{X}_i ; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i) \right) + n\lambda \sum_{j=1}^{d+1} \frac{|\beta_j|}{|\widetilde{\beta}_j|}, \tag{5.2.2}$$

where the coefficient relative to the intercept is left unpenalized, λ is a tuning parameter controlling for the amount of shrinkage and $\widetilde{\beta}$ is a first-step estimator of β_τ with no penalty function (i.e. when $\lambda = 0$). Introducing adaptive weights for penalizing the coefficients differently in contrast with the lasso penalty allows for shrinking at a higher speed the estimated coefficients for which the unpenalized estimation is already close to 0. A more formal discussion on the use of adaptive lasso weights is referred to the so-called 'oracle' property of Zou (2006).

For practical minimization of (5.2.2), a simple accommodation of the MM algorithm of Chapter 3 may then be invoked by using a similar fashion as the approach of Hunter and Li (2005) for the handling of the penalty term. This results in a very comparable algorithm as for Chapter 3, where the 'Step 1' of the latter only needs to accommodate for an additional term corresponding to the penalty part. Based on this, the application of the dimension reduction technique of Chapter 4 could then be investigated for the estimation of $G_C(\cdot|\boldsymbol{x})$, similarly to the paper of Wang et al. (2013) where the latter technique is applied for the estimation of $F_{T|\boldsymbol{X}}$ in the RM weights. Lastly, note that a similar variable selection procedure could also be examined in association with the inverse-c.d.f. estimation strategy of Chapter 4, although the numerical aspects would here probably require further work.

5.2.2 Towards additional features in the data

To close this thesis, and as an additional direction for potential further research, we consider in this section an example of additional features in the data that would alter the estimation procedures developed throughout this thesis. In particular, we consider the often encountered example of survival data incorporating some observations that may be completely immune to experiencing the event of interest. This may for instance happen in cancer studies, where some patients are in fact ultimately cured and hence no-longer exposed to the event of interest. In order to account for this particularity, an interesting literature has emerged under the natural denomination of *cure* models, where the population is divided into two groups of subjects: the *cured* ones, for which the time-to-event of interest is infinite, and the *susceptible* ones. The main challenge in cure models with random right censoring then comes from the consideration that one never knows if a censored observation is in fact cured or not, hence hindering the identification of cured

individuals in the dataset. As a result, when investigation of the relationship between covariates and the response of the susceptible population is of interest, standard tools as those developed throughout this thesis have to be adapted in order to control for censored observations that may in fact be cured. We therefore briefly discuss in this section a possible accommodation of our estimation procedures for censored quantile regression in the presence of a cure fraction. A complete and comprehensive review on the literature on cure models may be found in Amico and Van Keilegom (2018).

Our attention is drawn here to a particular class of cure models, called mixture cure models and first suggested by Boag (1949). Introducing the indicator η that takes the value 1 if an observation is susceptible and 0 if cured, the latter model writes the conditional c.d.f. of the whole population $F_{T^\star|\boldsymbol{X},\boldsymbol{Z}}$ as

$$F_{T^\star|\boldsymbol{X},\boldsymbol{Z}}(t|\boldsymbol{x},\boldsymbol{z}) = p(\boldsymbol{z})F_{T|\boldsymbol{X}}(t|\boldsymbol{x}),$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are two vectors of covariates possibly sharing some, or all components, $p(\boldsymbol{z}) = \mathbb{P}(\eta = 1|\boldsymbol{Z} = \boldsymbol{z})$ is the conditional probability of belonging to the uncured group and $F_{T|\boldsymbol{X}}$ is the conditional c.d.f. of the susceptible individuals, i.e. $F_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = \mathbb{P}(T \leq t|\boldsymbol{X} = \boldsymbol{x}, \eta = 1)$. This formulation of the mixture cure model thus explicitly expresses the conditional c.d.f. of the whole population as a mixture of the conditional c.d.f. of both cured and uncured groups, for which it is assumed that $T = \infty$. An advantage of such a formulation is then that it allows for a different influence of different covariates on the conditional probability of being susceptible and the conditional c.d.f. of the response of interest. For simplicity of notation however, and without loss of generality, we assume in the following that both covariate vectors are in fact identical, and consequently only write this section considering the vector $\boldsymbol{X}$.

The statistical problem we aim at addressing here then becomes the estimation of a quantile regression model for the susceptible population based on the i.i.d. data $(Y_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, where the notations are similar to the previous chapters. Note that this objective has already been addressed in existing work by Wu and Yin (2016), on which we build some of the following developments. For sake of brevity and similarly to the work of Wu and Yin, we only consider here a linear quantile regression, deliberately leaving out the development of a copula-based modelling that could for instance easily alleviate such an assumption if necessary.

Now, the basic starting point of Wu and Yin is to note that, if one were to know all the values of $\eta_i, i = 1, \ldots, n$, estimating the quantile regression would be trivially accomplished by keeping in the estimation procedure only the observations in the set $\{i : \eta_i = 1, i = 1, \ldots, n\}$, as these represent here the population of interest. A first observation in order to take the latter statement into account is to note that complete observations ($\Delta_i = 1$) are automatically uncured ($\eta_i = 1$), and hence their value of η is trivially known. The ambiguity then only comes from censored observations, for which the susceptible indicator is not known and one thus first estimates the conditional probability that these observations indeed belong to the

susceptible group, denoted by $p_0(\boldsymbol{x}_i) \equiv \mathbb{P}(\eta = 1|\boldsymbol{X} = \boldsymbol{x}_i, \Delta_i = 0)$. In particular, as is relatively common in the mixture cure model literaturee, the work of Wu and Yin proposes a logistic regression modelling for this part, where the coefficients of the latter are estimated through an iterative procedure that is here left out for sake of brevity. Then, in a second step, Bernoulli sampling (i.e. sampling from a Bernoulli distribution with success probability given by the first step estimation of $p_0(\boldsymbol{x}_i)$) is employed to label each censored observation as cured or not. The quantile regression estimation is then computed using the RM technique of Wang and Wang (2009) for all (estimated) uncured observations and, for robustness, the sampling of the second step is repeated multiple times and the quantile regression estimates are averaged in order to obtain the final coefficient estimators.

Now, in an attempt to enrich this literature based on the results of this thesis, we mention here two main suggestions that could be investigated in more detail in future research. Our first and somewhat straightforward suggestion mimics the procedure of Wu and Yin, only to employ the adapted check function of Chapter 3 in the second step of the estimation procedure. That is, let first $\widehat{\eta}_i^{(k)}$ be either 1 if $\Delta_i = 1$ or the result of the k-th Bernoulli sampling with success probability $\widehat{p}_0(\boldsymbol{x}_i)$ if $\Delta_i = 0$, where $\widehat{p}_0(\boldsymbol{x}_i)$ is estimated in a first step. Then, the estimation of the regression coefficient for the k-th imputed Bernoulli sample could be carried out by

$$\widehat{\beta}_\tau^{(k)} = \arg\min_\beta \sum_{i=1}^n \widehat{\eta}_i^{(k)} \varphi_\tau \left(\beta^\intercal \boldsymbol{X}_i; Y_i, \widehat{G}_C^{(k)}(\cdot|\boldsymbol{X}_i)\right),$$

where $\widehat{G}_C^{(k)}(\cdot|\boldsymbol{x})$ is the estimator of the conditional censoring distribution based on observations $\{i : \widehat{\eta}_{i(k)} = 1, i = 1, \ldots, n\}$. When the procedure is repeated K times, the final estimator is then given by $\widehat{\beta}_\tau = K^{-1} \sum_{k=1}^K \widehat{\beta}_\tau^{(k)}$. Based on the results of Chapter 3, we may then conjecture that this procedure may provide a valuable complement to the estimator of Wu and Yin, especially when the censoring proportion is important in the dataset.

Our second suggestion takes a little more distance from the procedure of Wu and Yin, and is rather based on combining the technique of Chapter 4 with the ideas of López-Cheda et al. (2017). Similarly to the latter paper, our starting point is here indeed to define a nonparametric estimator of $F_{T|\boldsymbol{X}}$ as follows:

$$\widehat{F}_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = \frac{\widehat{F}_{T^\star|\boldsymbol{X}}(t|\boldsymbol{x})}{\widehat{p}(\boldsymbol{x})},$$

where $p(\boldsymbol{x})$ is of course strictly positive is order to allow for uncured observations, and $\widehat{F}_{T^\star|\boldsymbol{X}}$ and $\widehat{p}(\boldsymbol{x})$ are both nonparametric estimator of $F_{T^\star|\boldsymbol{X}}$ and $p(\boldsymbol{x})$, respectively. Now, while López-Cheda et al. suggest to estimate $\widehat{F}_{T^\star|\boldsymbol{X}}$ employing a classical Beran estimator for all observations at hand, following the developments of Chapter 4 we advocate here naturally for a double-kernel version of the latter, as reported in (4.2.5). As for the estimation of $\widehat{p}(\boldsymbol{x})$, this can be carried out using

the suggestion of Xu and Peng (2014), similarly as in López-Cheda et al.. Following the procedure of Chapter 4, this would then suggest to estimate the quantile coefficients by

$$\widehat{\beta}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \left(\widehat{F}_{T^\star|\boldsymbol{X}}(\beta^\mathsf{T} \boldsymbol{X}_i | \boldsymbol{X}_i) - \widehat{p}(\boldsymbol{X}_i)\, \tau \right)^2 .$$

For interpretation, we observe that the presence of a cure fraction in the data, implying a point mass at infinity with positive probability and hence that $\lim_{t\to\infty} F_{T^\star|\boldsymbol{X}}(t|\boldsymbol{X}_i) = p(\boldsymbol{X}_i) < 1$, $i = 1, \ldots, n$, forces to consider smaller quantile levels than τ when calculating a smoothed version of Beran's estimator based on both cured and uncured group of subjects. This is somewhat in line with the intuition that the quantile level will be overestimated if ignoring the proportion of cured observations in the data, as the set of censored observations is 'artificially' inflated by the presence of the latter and hence overly-compensated in the estimator $\widehat{F}_{T^\star|\boldsymbol{X}}$ with respect to the population of interest. Based on the numerical results of Chapter 4, we may then conjecture that such an approach, if more thoroughly investigated, could provide satisfying results for the practical estimation of a censored quantile regression model with the presence of a cure fraction. This last presumption ultimately closes this section on the establishment of further research perspectives based on the work of this thesis.

References

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. **Insurance: Mathematics and economics**, 44(2):182–198, 2009.

M. Amico and I. Van Keilegom. Cure models in survival analysis. **Ann. Rev. Statist. Applic.**, 5:311–342, 2018.

A. Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. **Biometrika**, 68(1):326–328, 1981.

H. Bang and A. Tsiatis. Median regression with censored cost data. **Biometrics**, 58(3):643–649, 2002.

R. Beran. Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley, 1981.

M. Birke, S. Van Bellegem, and I. Van Keilegom. Semi-parametric estimation in a single-index model with endogenous variables. **Scandinavian Journal of Statistics**, 44(1):168–191, 2017.

J. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. **Journal of the Royal Statistical Society. Series B (Methodological)**, 11(1):15–53, 1949.

M. Bottai and J. Zhang. Laplace regression with censored data. **Biometrical Journal**, 52(4):487–503, 2010.

E. Bouyé and M. Salmon. Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets. **The European Journal of Finance**, 15 (7-8):721–750, 2009.

E. C. Brechmann. Truncated and simplified regular vines and their applications. Masterarbeit, Technische Universität München, 2010.

A. Bücher, A. El Ghouch, and I. Van Keilegom. Single-index quantile regression models for censored data. **Submitted**, 2014.

J. Buckley and I. James. Linear regression with censored data. **Biometrika**, 66 (3):429–436, 1979.

A. Charpentier, J.-D. Fermanian, and O. Scaillet. **The estimation of copulas: Theory and practice.** J. Rank, ed., Risk Books, 2006.

P. Chaudhuri. Global nonparametric estimation of conditional quantile functions and their derivatives. **Journal of Multivariate Analysis**, 39(2):246–269, 1991.

K. Chen and S.-H. Lo. On the rate of uniform convergence of the product-limit estimator: strong and weak laws. **The Annals of Statistics**, 25(3):1050–1087, 1997.

S.X. Chen and T.-M. Huang. Nonparametric estimation of copula functions for dependence modelling. **Canadian Journal of Statistics**, 35:265–282, 2007.

X. Chen and Y. Fan. Estimation of copula-based semiparametric time series models. **Journal of Econometrics**, 130(2):307–335, 2006.

X. Chen, O. Linton, and I. Van Keilegom. Estimation of semiparametric models when the criterion function is not smooth. **Econometrica**, 71(5):1591–1608, 2003.

X. Chen, R. Koenker, and Z. Xiao. Copula-based nonlinear quantile autoregression. **The Econometrics Journal**, 12(s1), 2009.

R. D. Cook. **Regression graphics**. Wiley., 1998.

R. D. Cook and L. Ni. Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. **J. Amer. Statist. Assoc.**, 100(470):410–428, 2005.

D.R. Cox. Regression models and life-tables. **Journal of the Royal Statistical Society. Series B (Methodological)**, 34:187–220, 1972.

C. Czado. **Workshop on Copula Theory and its Applications**. Springer, 2010.

D.M. Dabrowska. Uniform consistency of the kernel conditional Kaplan-Meier estimate. **The Annals of Statistics**, pages 1157–1167, 1989.

M. De Backer, A. El Ghouch, and I. Van Keilegom. Semiparametric copula quantile regression for complete or censored data. **Electronic Journal of Statistics**, 11(1):1660–1698, 2017.

M. De Backer, A. El Ghouch, and I. Van Keilegom. An adapted loss function for censored quantile regression. **J. Amer. Statist. Assoc.**, To appear, 2018a.

M. De Backer, A. El Ghouch, and I. Van Keilegom. Linear censored quantile regression: A novel minimum-distance approach. **Submitted**, 2018b.

P. Deheuvels. La fonction de dépendance empirique et ses propriétés. **Acad. Roy. Belg. - Bull. Cl. Sci.**, pages 274–292, 1979.

L. Delsol and I. Van Keilegom. Semiparametric M-estimation with non-smooth criterion functions. Technical report, Université catholique de Louvain, 2015. URL http://www.uclouvain.be/en-369695.html.

H. Dette, R. Van Hecke, and S. Volgushev. Some comments on copula-based regression. **J. Amer. Statist. Assoc.**, 109(507):1319–1324, 2014.

J. Dißmann, E. C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. **Computational Statistics & Data Analysis**, 59(0):52–69, 2013.

K.A. Doksum and M. Gasko. On a correspondence between models in binary regression analysis and in survival analysis. **International Statistical Review**, pages 243–252, 1990.

Y. Du and M.G. Akritas. I.i.d. representations of the conditional Kaplan-Meier process for arbitrary distributions. **Math. Meth. Statist.**, 11:152–182, 2002.

B. Efron. The two-sample problem with censored data. **Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics**, IV:831–853, 1967.

A. El Ghouch and I. Van Keilegom. Local linear quantile regression with dependent censored data. **Statistica Sinica**, 19:1621–1640, 2009.

J. B. Elsner, J. P. Kossin, and T. H. Jagger. The increasing intensity of the strongest tropical cyclones. **Nature**, 455(7209):92–95, 2008.

P. Embrechts. Copulas: A personal view. **Journal of Risk and Insurance**, 76(3):639–650, 2009.

E. Engel. Die lebenkosten Belgischerarbeiter-familien fruher und jetzt. **International Statistical Institute Bulletin**, 9:1–125, 1857.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. **J. Amer. Statist. Assoc.**, 96(456):1348–1360, 2001.

X. Feng, X. He, and J. Hu. Wild bootstrap for quantile regression. **Biometrika**, 98(4):995–999, 2011.

J.-D. Fermanian, D. Radulovic, and M. Wegkamp. Weak convergence of empirical copula processes. **Bernoulli**, 10(5):847–860, 2004.

A. Gannoun, J. Saracco, A. Yuan, and G.E. Bonney. Non-parametric quantile regression with censored data. **Scandinavian Journal of Statistics**, 32(4): 527–550, 2005.

G. Geenens, A. Charpentier, and D. Paindaveine. Probit transformation for non-parametric kernel estimation of the copula density. **Bernoulli**, 23(3):1848–1873, 2017.

C. Genest and J. Nešlehová. A primer on copulas for count data. **The ASTIN Bulletin**, 37(2):475–515, 2007.

C. Genest, K. Ghoudi, and L-P Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. **Biometrika**, 82(3):543–552, 1995.

C. Genest, M. Gendron, and M. Bourdeau-Brien. The advent of copulas in finance. **The European Journal of Finance**, 15(7-8):609–618, 2009a.

C. Genest, E. Masiello, and K. Tribouley. Estimating copula densities through wavelets. **Insurance: Mathematics and Economics**, 44(2):170–181, 2009b.

I. Gijbels and J. Mielniczuk. Estimating the density of a copula function. **Communications in Statistics - Theory and Methods**, 19(2):445–464, 1990.

I. Gijbels, M. Omelka, and N. Veraverbeke. Nonparametric testing for no covariate effects in conditional copulas. **Statistics**, 51(3):475–509, 2017.

R. Gill. Large sample behaviour of the product-limit estimator on the whole line. **The Annals of Statistics**, pages 49–58, 1983.

W. Gonzalez-Manteiga and C. Cadarso-Suarez. Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. **Journal of Nonparametric Statistics**, 4(1):65–78, 1994.

M. Gorfine, Y. Goldberg, and Y. Ritov. A quantile regression model for failure-time data with time-dependent covariates. **Biostatistics**, 18(1):132–146, 2017.

B. E. Hansen. Nonparametric estimation of smooth conditional distributions. Technical report, University of Winconsin, 2004.

T. Hastie, R. Tibshirani, and J. Friedman. **The elements of statistical learning**, volume 1. Springer series in statistics New York, 2001.

C. Heuchenne and I. Van Keilegom. Polynomial regression with censored data based on preliminary nonparametric estimation. **Annals of the Institute of Statistical Mathematics**, 59(2):273–297, 2007.

N. L. Hjort and D. Pollard. Asymptotics for minimisers of convex processes. Technical report, Yale University, 1993.

I. Hobæk Haff and J. Segers. Nonparametric estimation of pair-copula constructions with the empirical pair-copula. **Computational Statistics & Data Analysis**, 84:1–13, 2015.

I. Hobæk Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction - simply useful or too simplistic? **Journal of Multivariate Analysis**, 101(5): 1296–1310, 2010.

M. Hofert and D. Pham. Densities of nested Archimedian copulas. **Journal of Multivariate Analysis**, 118:37–52, 2013.

M. Hubert, I. Gijbels, and D. Vanpaemel. Reducing the mean squared error of quantile-based estimators by smoothing. **Test**, 22(3):448–465, 2013.

D. R. Hunter and K. Lange. Quantile regression via an MM algorithm. **Journal of Computational and Graphical Statistics**, 9(1):60–77, 2000.

D. R. Hunter and K. Lange. A tutorial on MM algorithms. **The American Statistician**, 58(1):30–37, 2004.

D.R. Hunter and R. Li. Variable selection using MM algorithms. **The Annals of Statistics**, 33(4):1617–1642, 2005.

J. Hyde. Testing survival with incomplete observations. **Biostatistics Casebook**, 1980.

H. Joe. **Multivariate Models and Dependence Concepts**. Cambridge Univ. Press., 1997.

H. Joe. **Dependence Modeling with Copulas.** Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2014.

J.D. Kalbfleisch and R.L. Prentice. **The Statistical Analysis of Failure Time Data, second edition**. Wiley Series in Probability and Statistics, 2002.

E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. **J. Amer. Statist. Assoc.**, 53(282):457–481, 1958.

G. Kauermann, C. Schellhase, and D. Ruppert. Flexible copula density estimation with penalized hierarchical B-splines. **Scandinavian Journal of Statistics**, 40(4):685–705, 2013.

J.-M. Kim, Y.-S. Jung, E.A. Sungur, K.-H. Han, C. Park, and I. Sohn. A copula method for modeling directional dependence of genes. **BMC bioinformatics**, 9(1):225, 2008.

R. Koenker. **Quantile Regression**. Cambridge Univ. Press., 2005.

R. Koenker and G. Jr. Bassett. Regression quantiles. **Econometrica**, 46(1):33–50, 1978.

R. Koenker and Y. Bilias. Quantile regression for duration data: A reappraisal of the Pensylvania employment bonus experiments. **Empirical Economics**, 26: 199–220, 2001.

R. Koenker and O. Geling. Reappraising medfly longevity: a quantile regression survival analysis. **J. Amer. Statist. Assoc.**, 96:458–468, 2001.

H. Koul, V. Susarla, and J. Van Ryzin. Regression analysis with randomly right-censored data. **The Annals of Statistics**, pages 1276–1288, 1981.

D. Kraus and C. Czado. D-vine copula based quantile regression. **Computational Statistics & Data Analysis**, 110:1–18, 2017.

B. Langholz and L. Goldstein. Risk set sampling in epidemiologic cohort studies. **Statistical Science**, pages 35–53, 1996.

E. Leconte, S. Poiraud-Casanova, and C. Thomas-Agnan. Smooth conditional distribution function and quantiles under random censorship. **Lifetime Data Analysis**, 8(3):229–246, 2002.

C. Leng and X. Tong. A quantile regression estimator for censored data. **Bernoulli**, 19(1):344–361, 2013.

K.-C. Li. Sliced inverse regression for dimension reduction. **J. Amer. Statist. Assoc.**, 86(414):316–327, 1991.

K.-C. Li, J.-L. Wang, and C.-H. Chen. Dimension reduction for censored regression data. **The Annals of Statistics**, 27(1):1–23, 1999.

Q. Li and J. S. Racine. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. **Journal of Business & Economic Statistics**, 26(4):423–434, 2008.

Q. Li and J. S. Racine. Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. **Journal of Econometrics**, 201(1):72–94, 2017.

Q. Li, J. Lin, and J. S. Racine. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. **Journal of Business & Economic Statistics**, 31(1):57–65, 2013.

A. Lindgren. Quantile regression with censored data using generalized L-1 minimization. **Computational Statistics & Data Analysis**, 23(4):509–524, 1997.

S.-H. Lo and K. Singh. The product-limit estimator and the bootstrap: some asymptotic representations. **Probability Theory and Related Fields**, 71(3): 455–465, 1986.

O. Lopez. Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. **Communications in Statistics-Theory and Methods**, 40(15):2639–2660, 2011.

A. López-Cheda, R. Cao, M.A. Jácome, and I. Van Keilegom. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. **Computational Statistics & Data Analysis**, 105:144–165, 2017.

J. H. Lubin, J. D. Boice, C. Edling, R. W. Hornung, G. R. Howe, E. Kunz, R. A. Kusiak, H. I. Morrison, E. P. Radford, J. M. Samet, et al. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. **Journal of the National Cancer Institute**, 87(11):817–827, 1995.

Y. Ma and G. Yin. Censored quantile regression with covariate measurement errors. **Statistica Sinica**, pages 949–971, 2011.

C. Martins-Filho and F. Yao. A note on the use of V and U statistics in nonparametric models of regression. **Annals of the Institute of Statistical Mathematics**, 58(2):389–406, 2006.

R.G. Miller. **Survival analysis**. John Wiley & Sons, New York, 1981.

D.F. Moore. **Applied Survival Analysis Using R**. Springer, New York, 2016.

T. Nagler. Kernel methods for vine copula estimation. Masterarbeit, Technische Universität München, Aug 2014.

T. Nagler and C. Czado. Evading the curse of dimensionality in multivariate kernel density estimation with simplified vines. **Journal of Multivariate Analysis**, 151:69–89, 2016.

T. Nagler and T. Vatter. Solving estimating equations with copulas. **arXiv preprint arXiv:1801.10576**, 2018.

R. Nelsen. **An Introduction to Copulas.** Springer, New York, 2006.

H. Noh, A. El Ghouch, and T. Bouezmarni. Copula-based regression estimation and inference. **J. Amer. Statist. Assoc.**, 108:676–688, 2013.

H. Noh, A. El Ghouch, and I. Van Keilegom. Semiparametric conditional quantile estimation through copula-based multivariate models. **Journal of Business & Economic Statistics**, 33(2):167–178, 2015.

D.H. Oh and A. Patton. Modelling dependence in high dimensions with factor copulas. **Manuscript, Duke University**, 2012.

M. Omelka, I. Gijbels, and N. Veraverbeke. Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. **The Annals of Statistics**, 37 (5B):3023–3058, 2009.

J.C. Pardo-Fernández, I. Van Keilegom, and W. González-Manteiga. Goodness-of-fit tests for parametric models in censored regression. **Canadian Journal of Statistics**, 35(2):249–264, 2007.

L. Peng and Y. Huang. Survival analysis with quantile regression models. **J. Amer. Statist. Assoc.**, 103(482):637–649, 2008.

S. Portnoy. Censored regression quantiles. **J. Amer. Statist. Assoc.**, 98(464): 1001–1012, 2003.

S. Portnoy and R. Koenker. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. **Statistical Science**, 12(4):279–300, 1997.

J.L. Powell. Least absolute deviations estimation for the censored regression model. **Journal of Econometrics**, 25(3):303–325, 1984.

J.L. Powell. Censored regression quantiles. **Journal of Econometrics**, 32:143–155, 1986.

R.L. Prentice. Linear rank tests with right censored data. **Biometrika**, 65(1): 167–179, 1978.

R Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL http://www.R-project.org/.

R-D. Reiss. Nonparametric estimation of smooth distribution functions. **Scandinavian Journal of Statistics**, pages 116–119, 1981.

G. Salvadori and C. De Michele. On the use of copulas in hydrology: theory and practice. **Journal of Hydrologic Engineering**, 12(4):369–380, 2007.

A. Sancetta and S. Satchell. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. **Econometric theory**, 20(3): 535–562, 2004.

N. Schallhorn, D. Kraus, T. Nagler, and C. Czado. D-vine quantile regression with discrete variables. **arXiv preprint arXiv:1705.08310**, 2017.

J. Segers. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. **Bernoulli**, 18(3):764–782, 2012.

R.J. Serfling. **Approximation Theorems of Mathematical Statistics**. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series. Wiley, 1980. ISBN 9780471024033.

X. Shen, Y. Zhu, and L. Song. Linear B-spline copulas with applications to nonparametric estimation of copulas. **Computational Statistics & Data Analysis**, 52(7):3806–3819, 2008.

J. H. Shih and T. A. Louis. Inferences on the association parameter in copula models for bivariate survival data. **Biometrics**, pages 1384–1399, 1995.

J. H. Shows, W. Lu, and H. H. Zhang. Sparse estimation and inference for censored median regression. **Journal of Statistical Planning and Inference**, 140(7): 1903–1917, 2010.

B. Silverman. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. **The Annals of Statistics**, pages 177–184, 1978.

M. Sklar. Fonctions de répartition à n dimensions et leurs marges. **Publ. Inst. Stitst. Univ. Paris**, 8:229–231, 1959.

V. Spokoiny, W. Wang, and W. K. Härdle. Local quantile regression. **Journal of Statistical Planning and Inference**, 143(7):1109–1129, 2013.

J. Stöber, H. Joe, and C. Czado. Simplified pair copula constructions - limitations and extensions. **Journal of Multivariate Analysis**, 119(0):101–118, 2013.

C. J. Stone. Consistent nonparametric regression. **The Annals of Statistics**, pages 595–620, 1977.

W. Stute. The oscillation behavior of empirical processes. **The Annals of Probability**, pages 86–107, 1982.

Y. Tang and H. Wang. Penalized regression across multiple quantiles under random censoring. **Journal of Multivariate Analysis**, 141:132–146, 2015.

T.M. Therneau and P.M. Grambsch. **Modeling Survival Data: Extending the Cox Model**. Springer, New York, 2000.

R. Tibshirani. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society. Series B (Methodological)**, pages 267–288, 1996.

H. Tsukahara. Semiparametric estimation in copula models. **Canadian Journal of Statistics**, 33(3):357–375, 2005.

B.W. Turnbull. Nonparametric estimation of a survivorship function with doubly censored data. **J. Amer. Statist. Assoc.**, 69(345):169–173, 1974.

A. W. Van der Vaart and J. A. Wellner. **Weak Convergence and Empirical Processes**. Springer, 1996.

I. Van Keilegom and M. G. Akritas. Transfer of tail information in censored regression models. **The Annals of Statistics**, pages 1745–1784, 1999.

I. Van Keilegom and N. Veraverbeke. Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. **Communications in Statistics-Theory and Methods**, 25(10):2251–2265, 1996.

I. Van Keilegom and N. Veraverbeke. Estimation and bootstrap with censored data in fixed design nonparametric regression. **Annals of the Institute of Statistical Mathematics**, 49(3):467–491, 1997a.

I. Van Keilegom and N. Veraverbeke. Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. **Communications in Statistics-Theory and Methods**, 26(4):853–869, 1997b.

C. Vignali, W. Brandt, and D. Schneider. X-ray emission from radio-quiet quasars in the sloan digital sky survey early data release: the $\alpha_{\rm ox}$ dependence upon ultraviolet luminosity. **The Astronomical Journal**, 125(2):433–443, 2003.

H. J. Wang, J. Zhou, and Y. Li. Variable selection for censored quantile regresion. **Statistica Sinica**, 23(1):145–167, 2013.

H.J. Wang and L. Wang. Locally weighted censored quantile regression. **J. Amer. Statist. Assoc.**, 104:1117–1128, 2009.

A. Wey, L. Wang, and K. Rudser. Censored quantile regression with recursive partitioning-based weights. **Biostatistics**, 15(1):170–181, 2014.

T. Wu, K. Yu, and Y. Yu. Single-index quantile regression. **Journal of Multivariate Analysis**, 101(7):1607–1621, 2010.

Y. Wu and G. Yin. Multiple imputation for cure rate quantile regression with censored data. **Biometrics**, 2016.

Y. Xia, D. Zhang, and J. Xu. Dimension reduction and semiparametric estimation of survival models. **J. Amer. Statist. Assoc.**, 105(489):278–290, 2010.

J. Xu and Y. Peng. Nonparametric cure rate estimation with covariates. **Canadian Journal of Statistics**, 42(1):1–17, 2014.

Z. Ying, S.H. Jung, and L.J. Wei. Survival analysis with median regression models. **J. Amer. Statist. Assoc.**, 90:178–184, 1995.

K. Yu and MC Jones. Local linear quantile regression. **J. Amer. Statist. Assoc.**, 93(441):228–237, 1998.

K. Yu, Z. Lu, and J. Stander. Quantile regression: applications and current research areas. **Journal of the Royal Statistical Society: Series D (The Statistician)**, 52(3):331–350, 2003.

L. Zhou. A simple censored median regression estimator. **Statistica Sinica**, pages 1043–1058, 2006.

M. Zhou. M-estimation in censored linear models. **Biometrika**, 79(4):837–841, 1992.

L. Zhu, M. Huang, and R. Li. Semiparametric quantile regression with high-dimensional covariates. **Statistica Sinica**, 22(4):1379–1401, 2012.

H. Zou. The adaptive lasso and its oracle properties. **J. Amer. Statist. Assoc.**, 101(476):1418–1429, 2006.