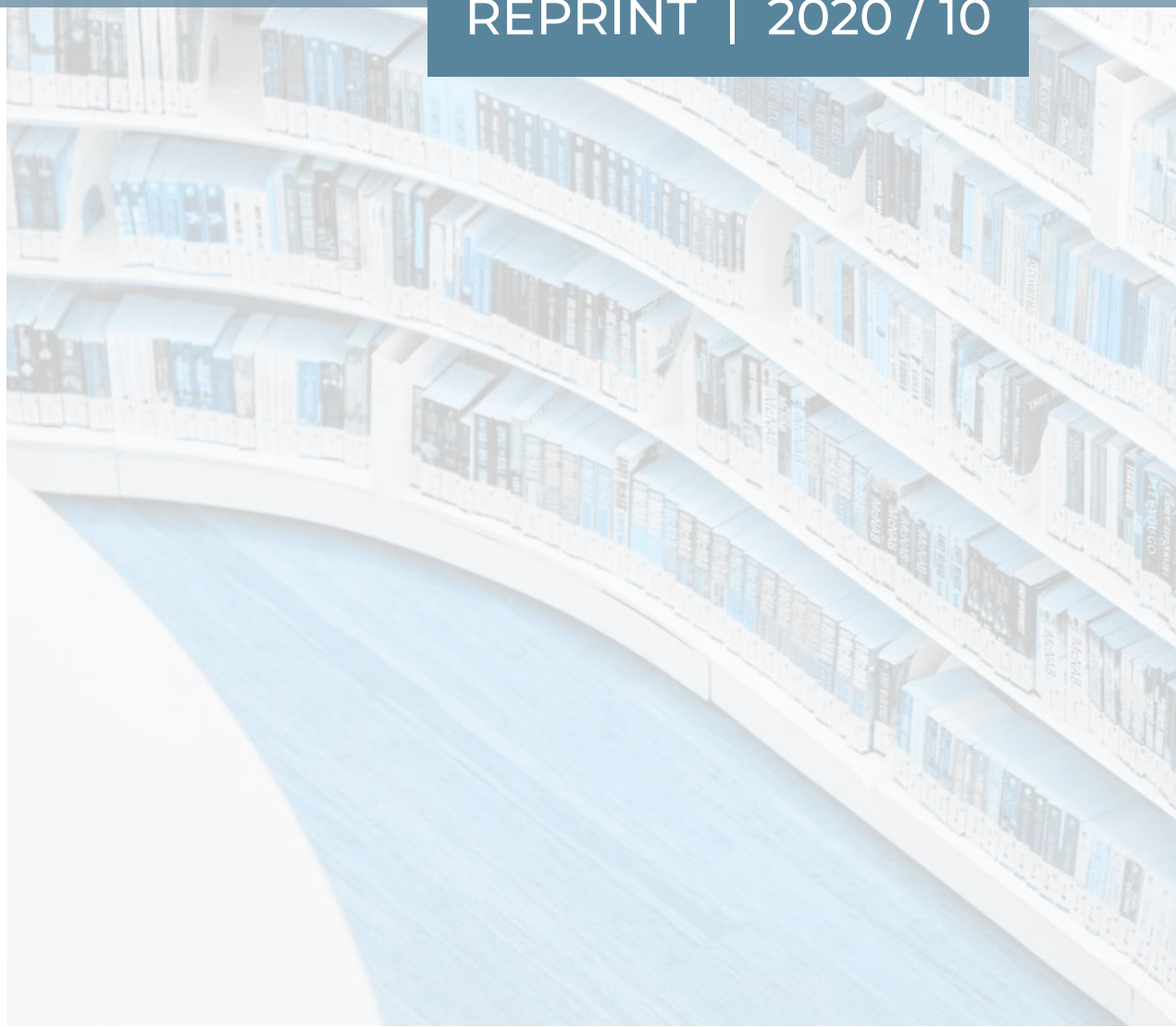
LINEAR CENSORED QUANTILE REGRESSION: A NOVEL MINIMUM-DISTANCE APPROACH

De Backer, M., El Ghouch, A. and
I. Van Keilegom

REPRINT | 2020 / 10



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

Linear Censored Quantile Regression: A Novel Minimum-Distance Approach

Mickaël De Backer* Anouar El Ghouch* Ingrid Van Keilegom†,*

October 11, 2019

Abstract

In this paper, 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 influential 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. This is 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.

Key words: *Beran estimator; bootstrap; double-kernel smoothing; effective dimension reduction; right censoring.*

*Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles. Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. Corresponding e-mail: mickael.debacker@uclouvain.be.

†KU Leuven, Research Centre for Operations Research and Business Statistics. Naamsestraat 69, 3000 Leuven, Belgium.

1 Introduction

Regression analysis is arguably the most common and most powerful statistical tool when it comes to investigating the relationship between certain covariate variables $\boldsymbol{X}$ and a response of interest T. 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. These models stem from the seminal work of Koenker and Bassett (1978) who introduced linear quantile regression as a simple minimization problem through the so-called "check" loss function. Conditional quantiles are particularly convenient when it is sensed that the conditional mean does not properly reflect the impact of the covariates on the dependent variable. They further allow practitioners to perceive a more thorough picture of the conditional distribution of T given $\boldsymbol{X}$. This statement is nicely illustrated in many scientific fields where the study of extremes is relevant, see e.g. the environmental studies of Elsner et al. (2008) and Hirschi et al. (2011) where the impact of changes in the covariates on the upper conditional quantiles is observed to be noticeably stronger than for the center of the conditional distribution. In comparison with their mean regression counterpart, quantile regression models are also praised for their robustness to outliers in the response and their flexibility with respect to the error distribution. A complete and comprehensive introduction to the quantile regression methodology may be found in Koenker (2005).

By expressing quantile regression as the minimization of a conditional expected loss, Koenker and Bassett opened the door to numerous creative parametric, semiparametric and nonparametric modelling techniques applied to quantile regression. In this paper, we choose to restrict ourselves to the estimation of the classical linear regression model. Furthermore, while the proposed methodology could, debatably, solely be devoted to the estimation of a linear model for complete observations, our primary motivation comes from the context of survival analysis where possible right-censoring of the response variable may occur. In this situation, often encountered in many interesting applications, instead of completely observing T, one only witnesses the minimum of it and a censoring variable C. A quantile regression model, through its interpretability, robustness and relaxation of the proportional hazards assumption, provides in these circumstances a valuable complement to the classical Cox regression and accelerated failure time model, as argued in Koenker and Bilias (2001), Koenker and Geling (2001) and Portnoy (2003).

The first study on parametric censored quantile regression can be traced back to the econometric work of Powell (1984; 1986). In the latter work, one assumes that the censoring variable is observable for all observations at hand. In practice however, many applications are confronted to random censoring and to the failure of completely observing C. To account for this specificity, the current literature can essentially be decomposed into three main modelling categories, all boiling down to the initial formulation of Koenker and Bassett in case of no censoring.

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 introduced in mean regression by Koul et al. (1981). This strategy was adapted to the present quantile regression context through different versions and in different linear contexts in Ying et al. (1995), Bang and Tsiatis (2002), Zhou (2006), Shows et al. (2010), Leng and Tong (2013) and Gorfine et al. (2017), among others. These models share the common observation that the expected loss of the unobservable response can easily be recovered from the expected loss of the actually observed response, by only keeping uncensored observations in the procedure and correctly adjusting through the conditional distribution of C given $\boldsymbol{X}$. As a result, for a flexible modelling and without additional assumptions, smoothing of the latter distribution is required. This is turn 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, as it fails to take profit of the robustness of quantile regression in its handling of censored observations in opposition to the second strategy described below.

This second main modelling technique may roughly be rallied under the same idea of proposing an appropriate weighting scheme, although the latter is here defined specifically for quantile regression. The approach is indeed based on the idea that all censored observations are not fundamentally to be treated similarly in this context. For instance, data lying above the quantile function will in fact have the same impact on the estimation of the regression model, regardless of their censoring status. Employing the idea of redistribution-of-mass of Efron (1967), Portnoy (2003) was the first to introduce this consideration into quantile regression through an appropriate weighting scheme, although based on a restrictive global linearity assumption. Relaxing the latter, Wang and Wang (2009) provided a flexible alternative based on the nonparametric estimation of the conditional distribution of T given $\boldsymbol{X}$ in the required weights. Similarly to the ICP technique, this suggests that a flexible modelling here also requires smoothing across the covariates, despite the initial model being parametric. This technique nevertheless engendered a fruitful literature, see e.g. Wang et al. (2013), Wey et al. (2014), Tang and Wang (2015) and Wu and Yin (2016) to name a few.

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 goes back to Lindgren (1997) and suggests to use all observed responses as if they were complete. Underestimation of the quantile function due to censoring is then avoided 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 recently proposed in De Backer et al. (2019), 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, has a broad application range in terms of modelling potential, and is illustrated for the particular case of linear regression. However, similarly to the above-described literature, for flexible modelling 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. Censored quantile regressions are indeed, 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 that requires that not all censoring occurs below the quantile function with probability one. This in turn introduces a natural upper bound to the quantile level one may consider in practice. It then seems natural to wonder if check-function-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 paper is to propose an alternative to the above-mentioned literature on linear censored quantile regression by circumventing the check-based modelling. In fact, our procedure can 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. 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)). The latter are, by construction, limited to small covariate dimensions. This technique was, to the best of our knowledge, never studied for linear regression for complete observations given it induces smoothing for a parametric model. However, 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, 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 practitioners a simple and valuable alternative for robust censored quantile regression estimation.

The rest of the paper is as follows. Section 2 is devoted to the newly proposed estimation procedure. Consistency and asymptotic normality of the estimator are obtained in Section 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 of our work, given these are required for establishing the results of Section 3. An extensive Monte Carlo simulation study is then conducted in Section 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 5 highlights a brief application to real data. Lastly, the proofs of Section 3 are deferred to the Supporting Information.

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 a few generalities on conditional quantile modelling for complete observations that will help us motivate our estimation procedure in the next sections.

2.1 Background on quantile regression modelling for complete observations

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,

$$m_\tau(\boldsymbol{x}) = \inf\{t : F_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \geq \tau\}, \tag{2.1}$$

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

$$m_\tau(\boldsymbol{x}) = \arg\min_a \mathbb{E}\Big[\rho_\tau(T-a)|\boldsymbol{X} = \boldsymbol{x}\Big], \tag{2.2}$$

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. 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 (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 (2.1). Hence, this literature 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 $F_{T|\boldsymbol{X}}$. 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.

2.2 Censored linear regression with low-dimensional covariates

Now, recall that we intend to work in this paper on a linear regression model. That is, we assume that

$$m_\tau(\boldsymbol{X}_i) = \beta_\tau^\mathsf{T} \boldsymbol{X}_i, \tag{2.3}$$

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 $i = 1, \ldots, n$, $F_{T|\boldsymbol{X}}(\beta_\tau^\mathsf{T} \boldsymbol{X}_i|\boldsymbol{X}_i) = \tau$, a natural extension of the inverse-c.d.f. technique is to estimate β_τ by

$$\widehat{\beta}_\tau = \arg\min_\beta \sum_{i=1}^n \left(\widehat{F}(\beta_\tau^\mathsf{T} \boldsymbol{X}_i|\boldsymbol{X}_i) - \tau\right)^2, \tag{2.4}$$

where $\widehat{F}$ is an appropriate estimator of $F_{T|\boldsymbol{X}}$. Note that (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 paper to use the computationally less complicated squared distance as reported in (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

$$\begin{aligned}\widehat{F}^s_{T|\boldsymbol{X}}(t|\boldsymbol{x}) &= \int_{\mathbb{R}} H\left(\frac{t-u}{h_T}\right) \mathrm{d}\widehat{F}_{T|\boldsymbol{X}}(u|\boldsymbol{x}) \\ &= \sum_{i=1}^{n^u} \left(\widehat{F}_{T|\boldsymbol{X}}(Y_{(i)}^u|\boldsymbol{x}) - \widehat{F}_{T|\boldsymbol{X}}(Y_{(i-1)}^u|\boldsymbol{x})\right) H\left(\frac{t - Y_{(i)}^u}{h_T}\right),\end{aligned} \tag{2.5}$$

where h_T is a positive bandwidth parameter, $Y^u_{(0)} = 0$ and $\widehat{F}_{T|\boldsymbol{X}}$ is Beran's local Kaplan-Meier estimator (Beran (1981)), defined as

$$\widehat{F}_{T|\boldsymbol{X}}(t|\boldsymbol{x}) = 1 - \prod_{i=1}^{n} \left\{ 1 - \frac{B_{ni}(\boldsymbol{x})}{\sum_{j=1}^{n} \mathbb{1}(Y_i \leq Y_j) B_{nj}(\boldsymbol{x})} \right\}^{\mathbb{1}(Y_i \leq t,\, \Delta_i = 1)}, \tag{2.6}$$

for a sequence of weights $B_{nj}(\boldsymbol{x})$ adding up to 1. Similarly to Wang and Wang (2009), Leng and Tong (2013) and De Backer et al. (2019), 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. Now, note that $\widehat{F}^s_{T|\boldsymbol{X}}$ may be read in (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, the estimation procedure through (2.4) and (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 (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 (2.4) than a crude step function. 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. The latter emanate from a reduction in 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 5. We nevertheless discuss in Section 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 (2.4). The impact of this strategy is however left for further investigation.

2.3 Censored linear regression for higher-dimensional covariates

When the dimension of the covariates becomes too large to reasonably estimate β_τ using (2.4) and (2.5), similarly to the existing literature on censored quantile regression, we inevitably need to introduce an additional assumption in our model. In this paper, 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

$$T \perp\!\!\!\perp \boldsymbol{X} | (\gamma_{0,1}^{\mathsf{T}} \boldsymbol{X}, \ldots, \gamma_{0,q}^{\mathsf{T}} \boldsymbol{X}), \tag{2.7}$$

where $\perp\!\!\!\perp$ stands for independence, $q < d+1$ for an effective dimension reduction (EDR), and where the γ_0'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. It 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_0's is not implied by (2.7) as such; only the q-dimensional linear subspace of $\mathbb{R}^q$ spanned by the γ_0'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 the base vector of the EDR space 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 γ_0's can be carried out using for instance the methodologies described in Li et al. (1999) or Xia et al. (2010).

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 (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 $(\widehat{\gamma}_{0,1}, \ldots, \widehat{\gamma}_{0,q})$ an estimator of the basis of the EDR space satisfying high-level conditions depicted in Section 3, we propose to estimate β_τ by

$$\widehat{\beta}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \left(\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta^\mathsf{T} \boldsymbol{X}_i | \widehat{\boldsymbol{Z}}_i) - \tau \right)^2, \tag{2.8}$$

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 (2.5) but replacing $\boldsymbol{X}$ with its estimated projections. Lastly, the selection of the value of q, representing the last unknown in (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.

3 Large Sample Properties

We establish in this section both the consistency and asymptotic normality of our proposed estimator $\widehat{\beta}_\tau$ defined in (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 that some purely technical assumptions (in particular assumptions (C5) and (C9) given below) are written considering explicitly the case $q = 1$ in (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 the Supporting Information for sake of brevity. The set of assumptions is then:

(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 $\beta \in \mathcal{B}$, $\inf_{\beta\in\mathcal{B}} \inf_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})} f_{T|\boldsymbol{X}}(\beta^\intercal \boldsymbol{x}|\boldsymbol{x}) > 0$, where $f_{T|\boldsymbol{X}}(\cdot|\boldsymbol{x})$ denotes the conditional density function of T given $\boldsymbol{X} = \boldsymbol{x}$.

(C3) The effective dimension reduction directions $\gamma_{0,j}$, $j = 1, \ldots, q$ belong to a compact subset Ξ of $\mathbb{R}^{d+1}$, and the estimators of the latter satisfy $(\widehat{\gamma}_0 - \gamma_0) = n^{-1} \sum_{i=1}^n \boldsymbol{d}_i + o_{\mathbb{P}}(n^{-1/2})$, where $\boldsymbol{d}_i$ are i.i.d. $q \times (d+1)$-dimensional random vectors with zero mean and positive definite variance-covariance matrix.

(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 $\text{supp}(\boldsymbol{X})$. Define next $\boldsymbol{Z} = \gamma_0^\intercal \boldsymbol{X}$ 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 $F(t, \boldsymbol{z}) :]-\infty, \upsilon] \times \text{supp}(\boldsymbol{Z}) \to [0, M]$ 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 $0 < \eta < 1$.

(i) Define by $\mathcal{F}_F$ the class of distributions $(t, \boldsymbol{x}) \mapsto F_{T|\gamma^\intercal \boldsymbol{X}}(t|\gamma^\intercal \boldsymbol{x})$ 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 $(t, \boldsymbol{x}) \mapsto f_{T|\gamma^\intercal \boldsymbol{X}}(t|\gamma^\intercal \boldsymbol{x})$ 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 $\boldsymbol{x} \in \text{supp}(\boldsymbol{X})$ and for $\beta \in \mathcal{B}$ defined in (C2), the point $\beta^\intercal \boldsymbol{x} \in \mathbb{R}$ lies below υ defined in (C5).

(C8) For some finite constant $C > 0$, the bandwidths $h_{\boldsymbol{X}}$ and h_T are such that $h_{\boldsymbol{X}} \equiv Cn^{-u}$ 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 $L_{Y|\gamma^\intercal \boldsymbol{X}}(t|\gamma^\intercal \boldsymbol{x}) = F_{Y|\gamma^\intercal \boldsymbol{X}}(t|\gamma^\intercal \boldsymbol{x}) f_{\gamma^\intercal \boldsymbol{X}}(\gamma^\intercal \boldsymbol{x})$ where $F_{Y|\gamma^\intercal \boldsymbol{X}}$ is the conditional c.d.f. of Y given $\gamma^\intercal \boldsymbol{X}$, and $f_{\gamma^\intercal \boldsymbol{X}}$ denotes the density of $\gamma^\intercal \boldsymbol{X}$. Then, $\sup_{t\leq\upsilon} \sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})} \sup_{\gamma\in\Xi} \nabla_\gamma L_{Y|\gamma^\intercal \boldsymbol{X}}(t|\gamma^\intercal \boldsymbol{x}) < \infty$, 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 $\gamma_0^\intercal \boldsymbol{X}$. 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 a linear representation of the estimators for the basis of the

EDR space. This namely serves to describe the impact of handling estimated covariate observations on the asymptotic properties of $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$, which, in turn, will impact the asymptotic variance of $\widehat{\beta}_\tau$. Note that this assumption is similar to assumption A4-(ii) in Wang et al. (2013) and may for instance be verified for the estimation procedures of Li et al. (1999) or Xia et al. (2010) under appropriate higher-level conditions. As a remark, we note that for proving only consistency of our proposed estimator $\widehat{\beta}_\tau$, one could consider instead a weaker assumption of $n^{1/2}$-consistency of the estimators for the basis of the EDR space. 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 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 definition of a general class reported in the work of Lopez (2011). It is established in this work that the latter class 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 De Backer et al. (2019), 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 De Backer et al. 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. 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 the Supporting Information.

Theorem 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) and (C9)-(i), for a given quantile level $0 < \tau < 1$ and for $\widehat{\beta}_\tau$ defined in (2.8),*

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

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

The proof of Theorem 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 the Supporting Information under the above-reported assumptions. These 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 consistency, it is observed that only the estimation of $F_{T|\boldsymbol{Z}}$ will influence the asymptotic covariance matrix of our estimator. Hence, only a linear representation for $\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}$ is here required. The latter is again reported in the Supporting Information and extends previous work

of, for instance, Gonzalez-Manteiga and Cadarso-Suarez (1994), Van Keilegom and Veraverbeke (1997) and Du and Akritas (2002).

Theorem 2. *For a given* $0 < \tau < 1$*, under the assumptions of Theorem 1 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 $\Gamma_1 = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top f^2_{T|\boldsymbol{X}}(\beta_\tau^\top \boldsymbol{X}|\boldsymbol{X})]$ *and* $\Sigma = Cov(g_i(\gamma_0^\top \boldsymbol{X}_i))$ *with* $g_i(\cdot)$ *defined in (A.14) in the Supporting Information.*

As a preliminary remark, we point out that Theorem 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 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. This is due to the fact 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 De Backer et al. (2019) to name a few, 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 4.3.

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. We 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 3 will complete the section. All of our simulations are carried out using the statistical computing environment R (R Core Team (2017)).

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 paper, resulting in the following:

ICP: Check-based estimator of Bang and Tsiatis (2002) embodying 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 from (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 (see e.g. section 7.10 of Hastie et al. (2001)) as suggested in Wang and Wang, with candidate bandwidth ranging from 0.05 to 0.5 by 0.05 increments.

AC: 'Adapted-check' estimator of De Backer et al. (2019) for which we apply their proposed MM algorithm 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 (2.4)-(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.

In order to provide a brief remark on the computational burdens for these procedures, one may note that NEW is by construction more computational intensive than ICP and RM. Indeed, the use of a double-kernel approach in NEW involves the evaluation of Beran's estimator $\widehat{F}_{T|\boldsymbol{X}}$ in $n^u \times n$ different points for each iteration of the optimization routine required in (2.4). In contrast, the ICP and RM competitors only require the evaluation of Beran's estimator in n different points at most. On the other hand, it is noted that the AC procedure requires for its part the evaluation of Beran's estimator in n different points at each iteration of the suggested MM algorithm, with the additional computational cost of inverting a $(d+1) \times (d+1)$ matrix at each iteration as well. Hence, comparing the considered estimators based on their current implementation, one logically expects that ICP and RM provide the fastest procedures, while NEW and AC are expected to perform similarly depending on the size $(d+1)$ of the covariate vector and the number of uncensored observations n^u.

Now, 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 using (2.4)-(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.

This more computationally intensive choice 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 $h_{\boldsymbol{X}}$, we employ here a two-step procedure given the computational cost of adding an extra bandwidth; first, a few extra iterations with a very large range for $h_{\boldsymbol{X}}$ serve to determine the order of magnitude in which one is more likely to observe the smallest cross validated prediction error for each DGP, and

second, the results from this first stage are employed to fasten up large scale simulations. 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 (2.4)-(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 De Backer et al., 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). 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 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 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 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 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 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 2.

Table 1 and Figure 1 here.

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 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 paper, 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. To understand the latter result, one should also compare results of $\mathcal{O}_{\rho_\tau}$ and $\mathcal{O}_{NEW}$, as these illustrate that NEW performs here better than $\mathcal{O}_{\rho_\tau}$ simply because $\mathcal{O}_{NEW}$ clearly outperforms $\mathcal{O}_{\rho_\tau}$, not per se because it handles extraordinary well the presence of censoring. 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 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 this DGP. Results for larger sample sizes exhibit the same singular patterns and are therefore omitted for this DGP.

Table 2 and Figure 2 here.

To further illustrate the differences in the results obtained from DGP 1 and 2, we consider in Figure 3 the example of one simulated dataset per DGP for $n = 500$, $\tau = 0.5$ and $p_c = 15\%$. As can be observed, the main difference between the datasets is the heteroscedatsic feature in the data appearing in DGP 2. In DGP 1, the challenge for estimating $F^{-1}_{T|\boldsymbol{X}}$ is thus speculated to be relatively constant across values of $X \sim U(0, 1)$, while it is suspected that in DGP 2 estimation of $F^{-1}_{T|\boldsymbol{X}}$ is in some sense easier for values of the covariate close to 1/2. As it is fairly simple to estimate $F^{-1}_{T|\boldsymbol{X}}$ for most of the observed X_i's in this DGP, one may then speculate that procedures requiring estimation of $F_{T|\boldsymbol{X}}$ will profit from this feature in the data in comparison to their check-based procedures. This might potentially explain the difference between $\mathcal{O}_{\rho_\tau}$ and $\mathcal{O}_{NEW}$, which has direct repercussions when turning to censored data for the comparison between NEW and its competitors.

Figure 3 here.

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.

4.2 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 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) (equations (3.7)-(3.9) in their paper), 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 $\text{RM}_{\hat{q}}$ 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 papers. 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 (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, 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. Similarly 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 3 and Figure 4 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, 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. In particular, $\text{NEW}_{\hat{q}}$ 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 3. The same considerations are reflected through the reported prediction criterion, as can be observed from both Table 3 and the bottom line of Figure 4 where the absolute deviations of all $B = 500$ simulations are depicted instead of only the MAD.

Table 3 and Figure 4 here.

Regarding the considered competitors for censored data, as repeatedly reported in the literature as well, it is further observed in Table 3 that ICP exhibits 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. 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 and Figure 5. 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 DGP to take advantage over AC of its dimension reduction feature. A visual examination of the results reported in Figure 5 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.

Table 4 and Figure 5 here.

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 only the projection $\gamma^\top \boldsymbol{X}$ in our DGP for which one coefficient of γ is equal to 0, and leave out the projection relative to $\beta_\tau^\top \boldsymbol{X}$. 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 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 then advocate for the practical use of an inverse-c.d.f. estimation approach for censored linear quantile regression.

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 paper, 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 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 5 then reports the empirical mean length and empirical coverage probability (in brackets) of the confidence intervals obtained from this simulation study.

Table 5 and Figure 6 here.

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 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 6 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 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.

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 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 paper, 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 our paper, we replace $l_{X,i}$ by $Y_i = \max_{1\leq j\leq 206}(l_{X,j}) - l_{X,i}$, $i = 1, \ldots, 206$, while the variable X will henceforth denote l_{UV} (centered and standardized) for notational convenience. The resulting dataset is plotted in Figure 7.

Figure 7 here.

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, 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 8 the resulting values of the estimated intercept and slope. We focus first on the left part of Figure 8, 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 8, 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.

Figure 8 here.

We now turn to the estimation of the regression parameters for different incremental values of τ, and present the obtained results in Figure 9. Bandwidths are here selected using the same cross validation procedure as in Section 4, while 95% confidence intervals are constructed upon 300 bootstrap samples using the percentile bootstrap illustrated earlier where, 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 9. 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 7 and Figure 9. 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. Finally, 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.

Figure 9 here.

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. Given these results, we note that an interesting line for further investigation would be the development an efficient optimization procedure for the suggested estimator, as the latter is currently more computationally intensive than its competing procedures built upon the check loss function.

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.

Acknowledgments

All authors acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy). M. De Backer and A. El Ghouch further acknowledge financial support from the FSR project IMAQFSR15PROJEL from the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS). I. Van Keilegom also acknowledges support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409).

Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique under convention 2.5020.11.

Address

Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
Corresponding e-mail: mickael.debacker@uclouvain.be.

Supporting Information

Technical proofs of Section 3.

Bibliography

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.

J. Buckley and I. James. Linear regression with censored data. Biometrika, 66(3):429–436, 1979.

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.

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.

M. De Backer, A. El Ghouch, and I. Van Keilegom. An adapted loss function for censored quantile regression. J. Amer. Statist. Assoc., 114(527):1126–1137, 2019.

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.

J. B. Elsner, J. P. Kossin, and T. H. Jagger. The increasing intensity of the strongest tropical cyclones. Nature, 455(7209):92–95, 2008.

R. D. Gill and S. Johansen. A survey of product-integration with a view toward application in survival analysis. The Annals of Statistics, 18(4):1501–1555, 1990.

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.

M. Hirschi, S. I. Seneviratne, V. Alexandrov, F. Boberg, Co. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, and P. Stepanek. Observational evidence for soil-moisture impact on hot extremes in southeastern europe. Nature Geoscience, 4(1):17–21, 2011.

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.

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.

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.

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.

S. Portnoy. Censored regression quantiles. J. Amer. Statist. Assoc., 98(464):1001–1012, 2003.

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 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.

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.

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.

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. Estimation and bootstrap with censored data in fixed design nonparametric regression. Annals of the Institute of Statistical Mathematics, 49(3):467–491, 1997.

C. Vignali, W. Brandt, and D. Schneider. X-ray emission from radio-quiet quasars in the sloan digital sky survey early data release: the α_{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.

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.

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.

L. Zhou. A simple censored median regression estimator. Statistica Sinica, pages 1043–1058, 2006.

$n = 100$		RMSE		MAE		$n = 200$		RMSE		MAE	
p_c	Method	β_0	β_1	β_0	β_1	p_c	Method	β_0	β_1	β_0	β_1
15%	ICP	0.255	0.471	0.179	0.326	15%	ICP	0.193	0.340	0.136	0.220
	RM	0.246	0.462	0.182	0.325		RM	0.194	0.337	0.136	0.232
	AC	0.246	0.460	0.173	0.323		AC	0.194	0.339	0.134	0.223
	NEW	0.239	0.443	0.158	0.292		NEW	0.175	0.309	0.122	0.225
	$\mathcal{O}_{\rho_\tau}$	0.238	0.427	0.170	0.290		$\mathcal{O}_{\rho_\tau}$	0.177	0.311	0.131	0.209
	$\mathcal{O}_{NEW}$	0.221	0.397	0.152	0.267		$\mathcal{O}_{NEW}$	0.165	0.287	0.120	0.202
40%	ICP	0.291	0.557	0.196	0.366	40%	ICP	0.214	0.385	0.150	0.264
	RM	0.280	0.532	0.195	0.342		RM	0.202	0.365	0.139	0.239
	AC	0.285	0.538	0.195	0.350		AC	0.201	0.366	0.140	0.244
	NEW	0.269	0.517	0.179	0.344		NEW	0.193	0.360	0.135	0.258
	$\mathcal{O}_{\rho_\tau}$	0.238	0.427	0.170	0.290		$\mathcal{O}_{\rho_\tau}$	0.177	0.311	0.131	0.209
	$\mathcal{O}_{NEW}$	0.221	0.397	0.152	0.267		$\mathcal{O}_{NEW}$	0.165	0.287	0.120	0.202

Table 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 $\{15\%, 40\%\}$.

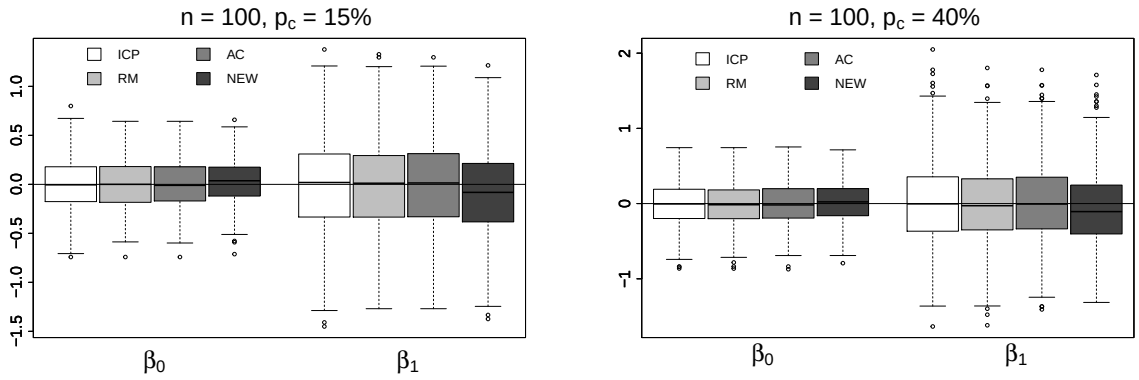


Figure 1: Boxplots relative to DGP 1 and Table 1 (left) for the estimation of β_0 and β_1, both centered (i.e. depicting $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $\tau = 0.5$, $n = 100$ and $p_c \in \{15\%, 40\%\}$.

$p_c = 15\%$		RMSE		MAE	
τ	Method	β_0	β_1	β_0	β_1
0.3	ICP	0.229	0.449	0.152	0.312
	RM	0.214	0.407	0.141	0.286
	AC	0.223	0.424	0.145	0.300
	NEW	0.181	0.309	0.119	0.212
	$\mathcal{O}_{\rho_\tau}$	0.214	0.407	0.146	0.288
	$\mathcal{O}_{NEW}$	0.169	0.309	0.112	0.209
0.5	ICP	0.213	0.439	0.146	0.295
	RM	0.201	0.400	0.131	0.261
	AC	0.208	0.420	0.138	0.262
	NEW	0.142	0.288	0.092	0.200
	$\mathcal{O}_{\rho_\tau}$	0.204	0.402	0.134	0.274
	$\mathcal{O}_{NEW}$	0.145	0.292	0.090	0.193

$p_c = 40\%$		RMSE		MAE	
p_c	Method	β_0	β_1	β_0	β_1
0.3	ICP	0.314	0.552	0.206	0.352
	RM	0.228	0.420	0.154	0.286
	AC	0.229	0.461	0.154	0.313
	NEW	0.205	0.307	0.136	0.212
	$\mathcal{O}_{\rho_\tau}$	0.214	0.407	0.146	0.288
	$\mathcal{O}_{NEW}$	0.169	0.309	0.112	0.209
0.5	ICP	0.282	0.500	0.193	0.350
	RM	0.205	0.396	0.133	0.262
	AC	0.218	0.440	0.148	0.296
	NEW	0.150	0.304	0.104	0.203
	$\mathcal{O}_{\rho_\tau}$	0.204	0.402	0.134	0.274
	$\mathcal{O}_{NEW}$	0.145	0.292	0.090	0.193

Table 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 $\{15\%, 40\%\}$.

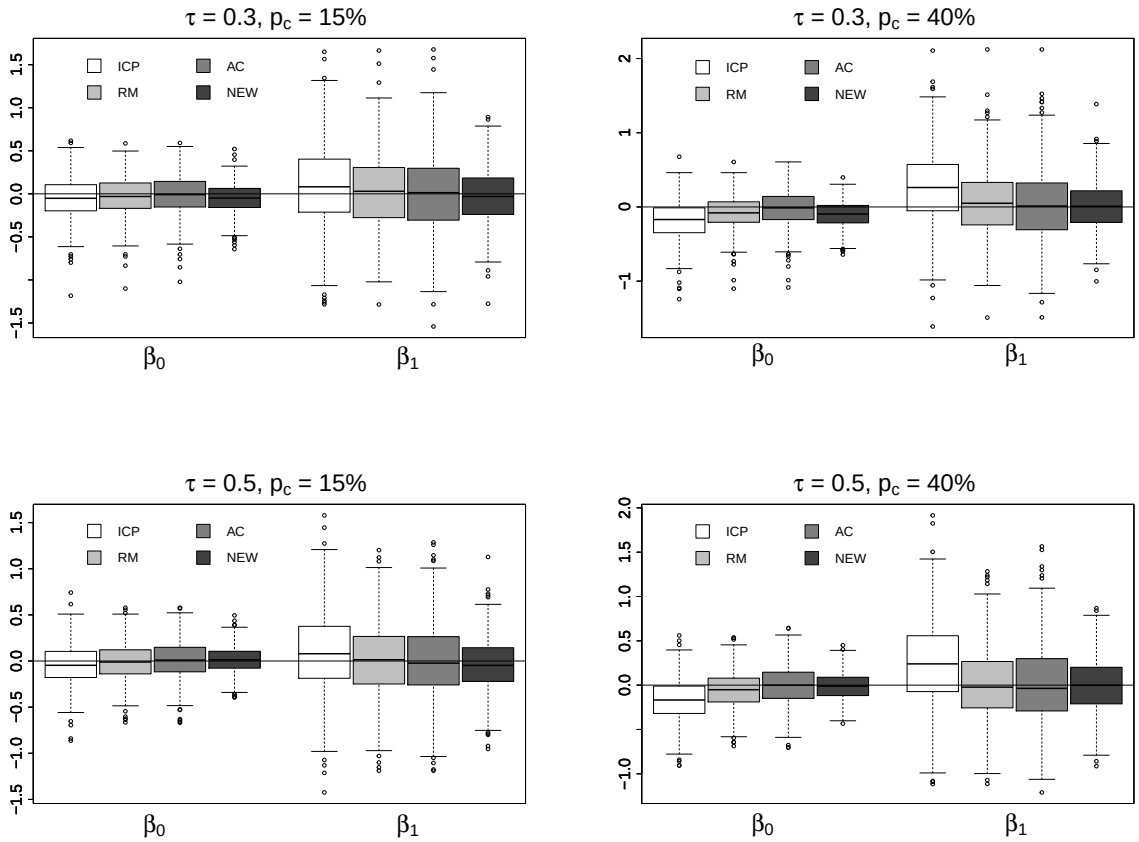


Figure 2: Boxplots relative to DGP 2 and Table 2 for the estimation of β_0 and β_1, both centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $n = 100$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

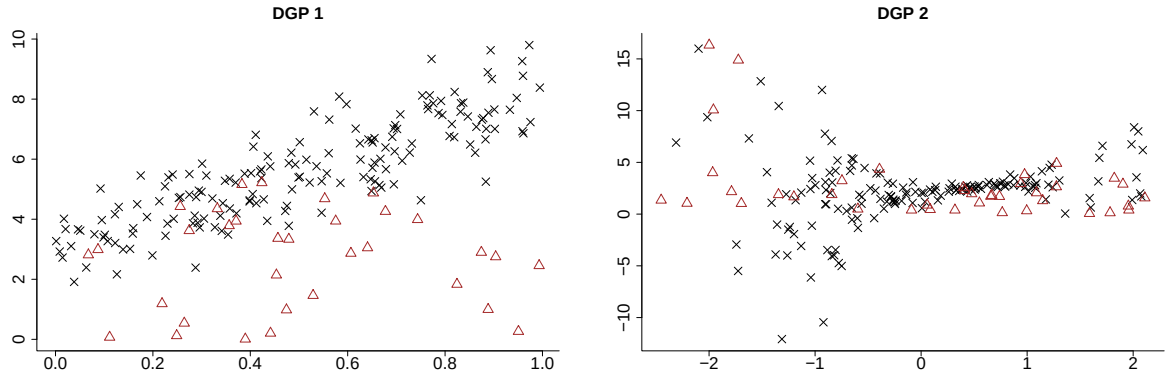


Figure 3: Illustration of one dataset for DGP 1 and 2 with $n = 200$, $\tau = 0.5$ and $p_c = 15\%$. Completely observed times are represented by $\times$, while censored observations are represented by $\triangle$.

p_c	n	Method	RMSE	MAE	MAD
15%	100	ICP	0.502	0.746	0.238
		$\text{RM}_{\hat{q}}$	0.487	0.712	0.232
		AC	0.490	0.711	0.233
		$\text{NEW}_{\hat{q}}$	0.479	0.698	0.228
		NEW_{q}	0.478	0.698	0.228
		$\mathcal{O}_{\rho_\tau}$	0.462	0.717	0.220
		$\mathcal{O}_{NEW}$	0.447	0.682	0.212
	200	ICP	0.346	0.500	0.166
		$\text{RM}_{\hat{q}}$	0.342	0.499	0.165
		AC	0.341	0.489	0.164
		$\text{NEW}_{\hat{q}}$	0.323	0.475	0.155
		NEW_{q}	0.323	0.475	0.155
		$\mathcal{O}_{\rho_\tau}$	0.323	0.485	0.156
		$\mathcal{O}_{NEW}$	0.300	0.439	0.145

p_c	n	Method	RMSE	MAE	MAD
30%	100	ICP	0.565	0.850	0.266
		$\text{RM}_{\hat{q}}$	0.526	0.794	0.252
		AC	0.529	0.791	0.251
		$\text{NEW}_{\hat{q}}$	0.517	0.774	0.246
		NEW_{q}	0.515	0.771	0.245
		$\mathcal{O}_{\rho_\tau}$	0.462	0.717	0.220
		$\mathcal{O}_{NEW}$	0.447	0.682	0.212
	200	ICP	0.391	0.589	0.188
		$\text{RM}_{\hat{q}}$	0.368	0.566	0.177
		AC	0.370	0.563	0.180
		$\text{NEW}_{\hat{q}}$	0.350	0.534	0.168
		NEW_{q}	0.347	0.531	0.166
		$\mathcal{O}_{\rho_\tau}$	0.323	0.485	0.156
		$\mathcal{O}_{NEW}$	0.300	0.439	0.145

Table 3: Simulation results for DGP 3 for both complete and censored data with $\tau = 0.5$ and $n \in \{100, 200\}$.

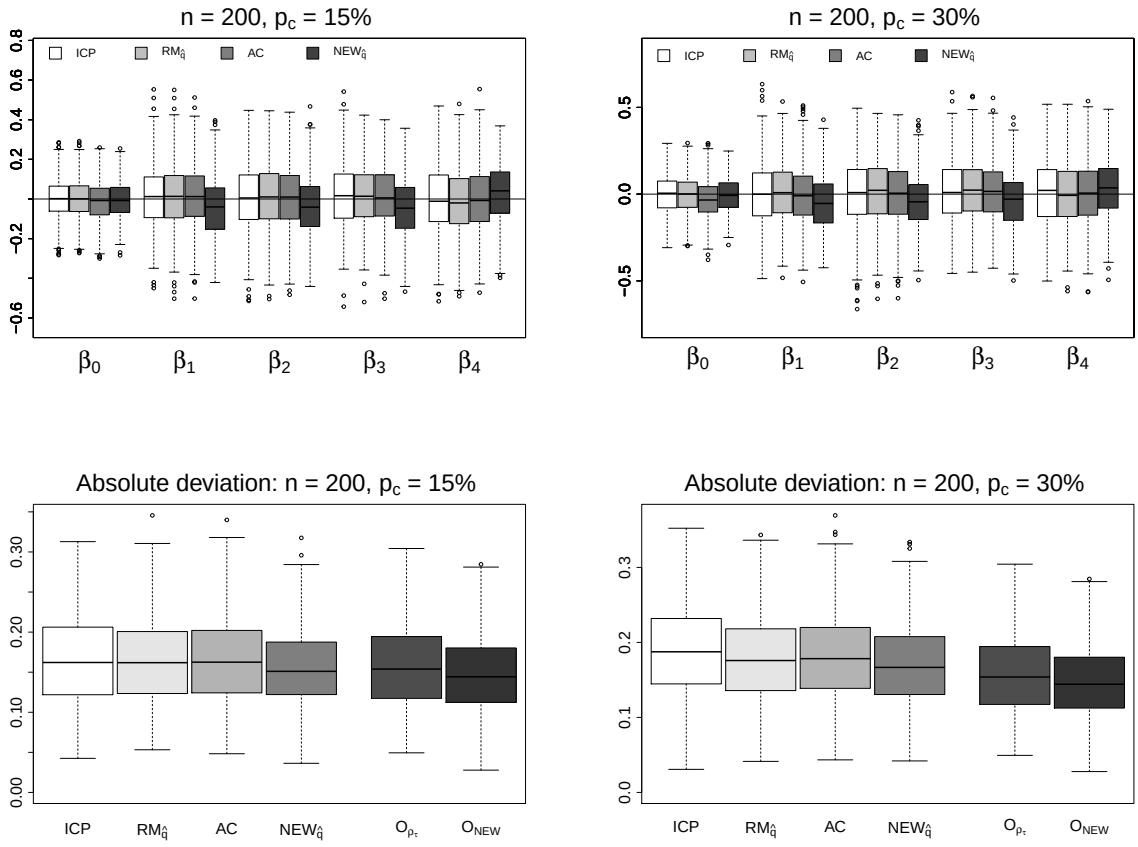


Figure 4: Boxplots relative to DGP 3. The top line reports the estimation of β_τ centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the competitors based on censored data with $\tau = 0.5$, $n = 200$ and $p_c \in \{15\%, 30\%\}$. The bottom line reports for all estimation procedures the absolute deviation results obtained under the same settings.

p_c	τ	Method	RMSE	MAE	MAD
15%	0.3	ICP	0.399	0.609	0.198
		$\text{RM}_{\hat{q}}$	0.401	0.617	0.200
		AC	0.395	0.613	0.197
		$\text{NEW}_{\hat{q}}$	0.381	0.565	0.192
		NEW_{q}	0.365	0.560	0.191
		$\mathcal{O}_{\rho_\tau}$	0.388	0.586	0.192
		$\mathcal{O}_{NEW}$	0.359	0.534	0.183
	0.5	ICP	0.410	0.601	0.200
		$\text{RM}_{\hat{q}}$	0.407	0.597	0.200
		AC	0.406	0.600	0.199
		$\text{NEW}_{\hat{q}}$	0.350	0.511	0.172
		NEW_{q}	0.343	0.507	0.171
		$\mathcal{O}_{\rho_\tau}$	0.387	0.572	0.191
		$\mathcal{O}_{NEW}$	0.341	0.492	0.167

p_c	τ	Method	RMSE	MAE	MAD
30%	0.3	ICP	0.465	0.704	0.231
		$\text{RM}_{\hat{q}}$	0.431	0.625	0.212
		AC	0.440	0.654	0.218
		$\text{NEW}_{\hat{q}}$	0.429	0.621	0.212
		NEW_{q}	0.414	0.618	0.208
		$\mathcal{O}_{\rho_\tau}$	0.388	0.586	0.192
		$\mathcal{O}_{NEW}$	0.359	0.534	0.183
	0.5	ICP	0.506	0.741	0.247
		$\text{RM}_{\hat{q}}$	0.436	0.645	0.212
		AC	0.453	0.667	0.224
		$\text{NEW}_{\hat{q}}$	0.391	0.587	0.200
		NEW_{q}	0.381	0.580	0.192
		$\mathcal{O}_{\rho_\tau}$	0.387	0.572	0.191
		$\mathcal{O}_{NEW}$	0.341	0.492	0.167

Table 4: Simulation results for DGP 4 for both complete and censored data with $n = 200$ and $\tau \in \{0.3, 0.5\}$.

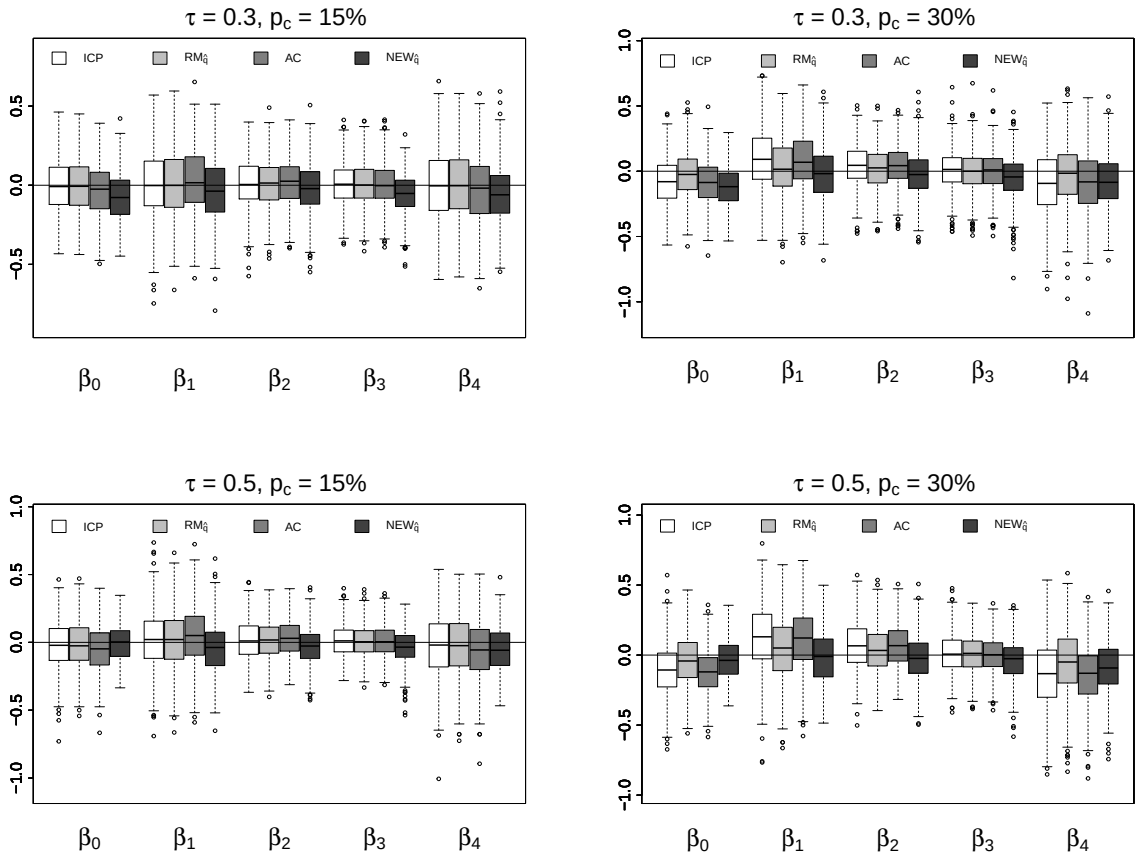


Figure 5: Boxplots relative to DGP 4 and Table 4 for the estimation of β_τ centered (i.e. $\widehat{\beta}_\tau - \beta_\tau$), for the four competitors based on censored data with $n = 200$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

DGP	p_c	τ	RM		AC		NEW	
			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 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 $\tau \in \{0.3, 0.5\}$.

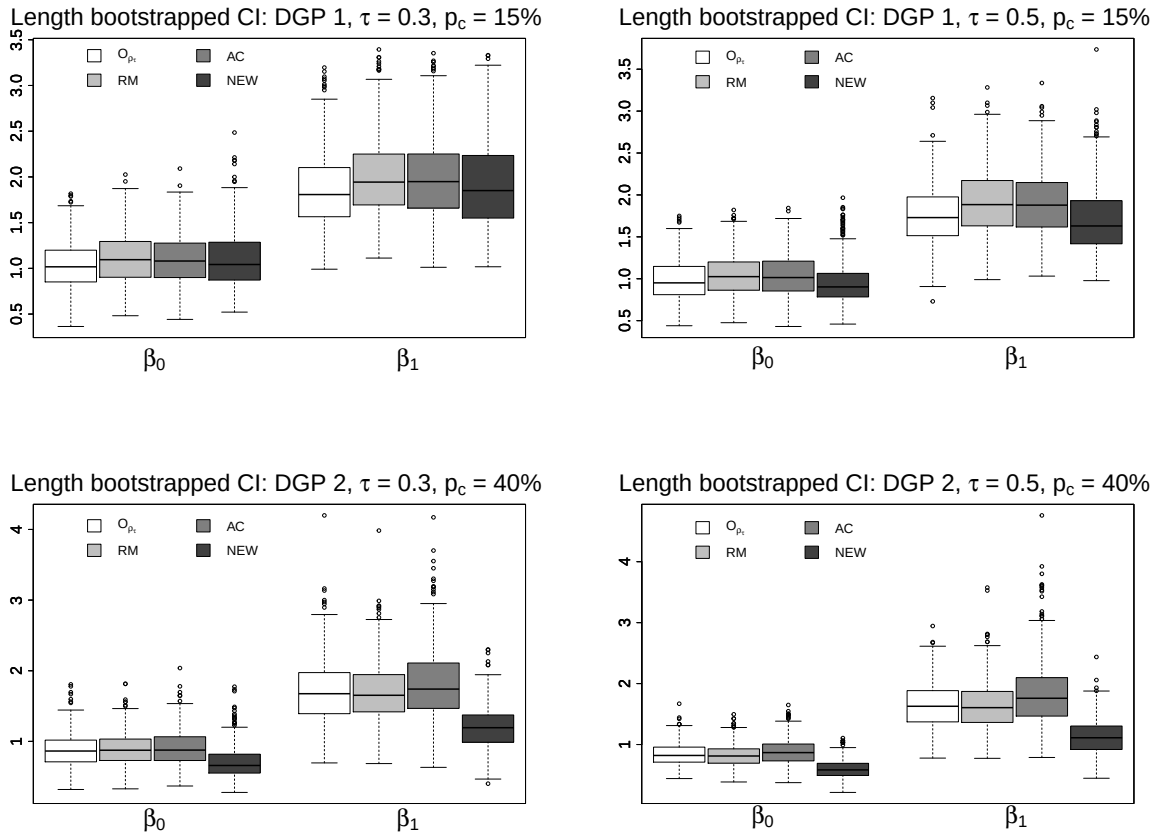


Figure 6: Boxplots of bootstrapped confidence interval length relative to DGP 1 and 2. Sample size is $n = 100$, $\tau \in \{0.3, 0.5\}$ and $p_c \in \{15\%, 40\%\}$.

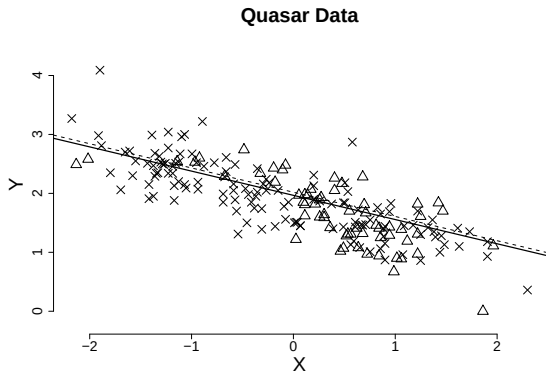


Figure 7: Quasar data of Vignali et al. (2003). Uncensored data points are given by $\times$, while censored observations are represented by $\triangle$. The dotted line represents the Buckley-James mean regression estimation, while the plain line represents the estimated median regression using NEW.

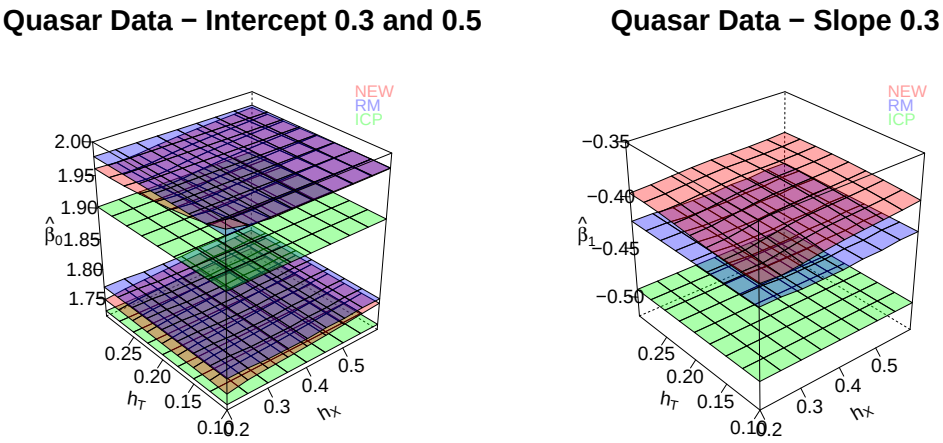


Figure 8: Quasar data, bandwidth sensitivity. The left plot represents estimation of the intercept for $\tau = 0.3$ (bottom) and $\tau = 0.5$ (top) for NEW (red), RM (blue) and ICP (green), for every pair of $(h_T, h_{\boldsymbol{X}})$ considered (RM being independent to h_T and ICP being independent 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 $h_{\boldsymbol{X}}$.

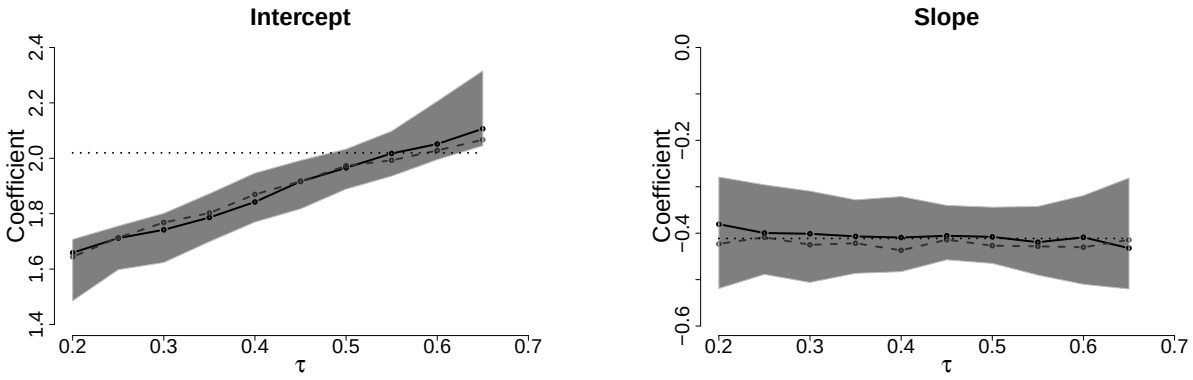


Figure 9: Quasar data. Estimated quantiles coefficients for regressing Y on X. Shaded areas correspond to 95% confidence intervals based on 300 bootstrap samples. Black lines correspond to the estimator NEW, grey lines correspond to RM, and dotted lines correspond to the Buckley-James mean regression estimation.

Supporting Information

We provide in this supporting information the proofs of Section 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 linear representation of the double-kernel version of Beran's estimator with dimension reduction, while the second one reports the uniform consistency only of its corresponding density estimator.

Lemma 1. *Assume conditions (C1), (C3)-(C6), (C8) and (C9)-(i) 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 $F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z}) = \mathbb{P}(Y \leq t, \Delta = 1|\boldsymbol{Z} = \boldsymbol{z})$, we have:*

$$\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \sum_{i=1}^n B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + n^{-1}\sum_{i=1}^n \varphi(Y_i, \Delta_i, t|\boldsymbol{z}) + o_{\mathbb{P}}(n^{-1/2}),$$

where

$$\xi(Y_i, \Delta_i, t|\boldsymbol{z}) = (1 - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}))\left[\int_0^{Y_i \wedge t} \frac{-\mathrm{d}F_{Y,1|\boldsymbol{Z}}(s|\boldsymbol{z})}{\{1 - F_{Y|\boldsymbol{Z}}(s|\boldsymbol{z})\}^2} + \frac{\Delta_i \mathbb{1}(Y_i \leq t)}{1 - F_{Y|\boldsymbol{Z}}(Y_i|\boldsymbol{z})}\right], \tag{A.1a}$$

$$\begin{aligned}
&\varphi(Y_i, \Delta_i, t|\boldsymbol{z}) \\
&= \boldsymbol{d}_i^\intercal \left(\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\nabla_\gamma L_{Y|\gamma_0^\intercal \boldsymbol{X}}(t|\gamma_0^\intercal \boldsymbol{x}) - \frac{L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})}{f^2_{\boldsymbol{Z}}(\boldsymbol{z})}\nabla_\gamma f_{\gamma_0^\intercal \boldsymbol{X}}(\gamma_0^\intercal \boldsymbol{x})\right) \varrho_1\left(F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}), F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right) \\
&+ \boldsymbol{d}_i^\intercal \left(\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\nabla_\gamma L_{Y,1|\gamma_0^\intercal \boldsymbol{X}}(t|\gamma_0^\intercal \boldsymbol{x}) - \frac{L_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})}{f^2_{\boldsymbol{Z}}(\boldsymbol{z})}\nabla_\gamma f_{\gamma_0^\intercal \boldsymbol{X}}(\gamma_0^\intercal \boldsymbol{x})\right) \varrho_2\left(F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}), F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right),
\end{aligned} \tag{A.1b}$$

where $L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) = F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})$, $L_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z}) = F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})$, with $f_{\boldsymbol{Z}}$ denoting the density of $\boldsymbol{Z}$. In expression (A.1b), ϱ_1 (ϱ_2) denotes the partial derivative of ϱ with respect to its first (second) argument, with ϱ defined by

$$\varrho : (F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}), F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})) \mapsto 1 - \prod\left(1 - \frac{\mathrm{d}F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})}{F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})}\right),$$

where $\prod$ is the product-integral operator as defined in Gill and Johansen (1990).

Lemma 2. *Assume the conditions of Lemma 1 hold. Then, for $1 \leq q \leq d+1$, writing $f_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \frac{\partial}{\partial t}F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})$ and $\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) = \frac{\partial}{\partial t}\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$:*

$$\sup_{t\leq\upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - f_{T|\boldsymbol{X}}(t|\boldsymbol{x})| = \sup_{t\leq\upsilon}\sup_{\boldsymbol{x}\in supp(\boldsymbol{X})}|\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})| = o_{\mathbb{P}}(1).$$

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 1.
We start by considering the decomposition

$$\begin{aligned}
\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \{\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} + \{\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} \\
&= (T_1) + (T_2),
\end{aligned}$$

and treat these terms separately. Starting with (T_1), integrating by parts we first have

$$\widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \int \widehat{F}_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z})\widetilde{K}(u)\,\mathrm{d}u. \tag{A.2}$$

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:

$$\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + R_n(t, \boldsymbol{z}), \tag{A.3}$$

where $\sup_{t\leq v}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|R_n(t,\boldsymbol{z})| = O_{\mathbb{P}}\left((\log n)^{3/4}(nh_{\boldsymbol{X}})^{-3/4} + h_{\boldsymbol{X}}^{\nu}\right)$, and ξ is defined in (A.1a). Inserting (A.3) in (A.2), we then have

$$\begin{aligned}\widehat{F}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \int \left\{ F_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z}) + \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t - uh_T|\boldsymbol{z}) + R_n(t - uh_T, \boldsymbol{z}) \right\} \widetilde{K}(u)\mathrm{d}u \\ &= F_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) + \sum_{i=1}^{n} B_{ni}(\boldsymbol{z})\xi(Y_i, \Delta_i, t|\boldsymbol{z}) + O_{\mathbb{P}}\left((\log n/(nh_{\boldsymbol{X}}))^{3/4} + h_{\boldsymbol{X}}^{\nu} + h_T^2\right), \end{aligned} \tag{A.4}$$

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). Note that, under assumption (C8), the term $O_{\mathbb{P}}((\log n/(nh_{\boldsymbol{X}}))^{3/4} + h_{\boldsymbol{X}}^{\nu} + h_T^2)$ is $o_{\mathbb{P}}(n^{-1/2})$.

Now, concerning part (T_2), we first have:

$$\widehat{F}^{s}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = \int \left(\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t - uh_T|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t - uh_T|\boldsymbol{z})\right) \widetilde{K}(u)\,\mathrm{d}u. \tag{A.5}$$

Focusing then on $\widehat{F}_{T|\widehat{\boldsymbol{Z}}}$ and $\widehat{F}_{T|\boldsymbol{Z}}$, we first note that we may rewrite $\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) = \varrho(\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}), \widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}))$ and similarly for $\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})$ (see Gill and Johansen (1990) for more details), where

$$\begin{aligned}\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) &= \sum_{i=1}^{n} \frac{K\left(\frac{\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{Z}}_i}{h_{\boldsymbol{X}}}\right)}{\sum_{k=1}^{n} K\left(\frac{\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{Z}}_k}{h_{\boldsymbol{X}}}\right)} \mathbb{1}(Y_i \leq t), \\ \widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) &= \sum_{i=1}^{n} \frac{K\left(\frac{\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{Z}}_i}{h_{\boldsymbol{X}}}\right)}{\sum_{k=1}^{n} K\left(\frac{\widehat{\boldsymbol{z}} - \widehat{\boldsymbol{Z}}_k}{h_{\boldsymbol{X}}}\right)} \mathbb{1}(Y_i \leq t, \Delta_i = 1).\end{aligned}$$

Then, considering the study of the integrand in (A.5), by Taylor expansion of order one, we have here that

$$\begin{aligned}\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \left(\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})\right) \varrho_1\left(\xi(t|\boldsymbol{z}), \widetilde{\xi}(t|\boldsymbol{z})\right) \\ &\quad + \left(\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right) \varrho_2\left(\xi(t|\boldsymbol{z}), \widetilde{\xi}(t|\boldsymbol{z})\right), \end{aligned} \tag{A.6}$$

for some $\xi(t|\boldsymbol{z})$ between $\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ and $\widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})$, and for some $\widetilde{\xi}(t|\boldsymbol{z})$ between $\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ and $\widehat{F}_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})$. We will now show that we have

$$\begin{aligned}&\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) \\ &\quad = (\widehat{\gamma}_0 - \gamma_0)^{\mathsf{T}} \left(\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \nabla_\gamma L_{Y|\gamma_0^{\mathsf{T}}\boldsymbol{X}}(t|\gamma_0^{\mathsf{T}}\boldsymbol{x}) - \frac{L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})}{f^2_{\boldsymbol{Z}}(\boldsymbol{z})} \nabla_\gamma f_{\gamma_0^{\mathsf{T}}\boldsymbol{X}}(\gamma_0^{\mathsf{T}}\boldsymbol{x}) \right) + o_{\mathbb{P}}(n^{-1/2}), \end{aligned} \tag{A.7a}$$

$$\begin{aligned}&\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z}) \\ &\quad = (\widehat{\gamma}_0 - \gamma_0)^{\mathsf{T}} \left(\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \nabla_\gamma L_{Y,1|\gamma_0^{\mathsf{T}}\boldsymbol{X}}(t|\gamma_0^{\mathsf{T}}\boldsymbol{x}) - \frac{L_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})}{f^2_{\boldsymbol{Z}}(\boldsymbol{z})} \nabla_\gamma f_{\gamma_0^{\mathsf{T}}\boldsymbol{X}}(\gamma_0^{\mathsf{T}}\boldsymbol{x}) \right) + o_{\mathbb{P}}(n^{-1/2}), \end{aligned} \tag{A.7b}$$

where the terms $o_{\mathbb{P}}(n^{-1/2})$ will be observed to hold uniformly in t. Focusing on detailing the result for (A.7a), we first 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 estimating $L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})$. We then rewrite,

$$\begin{aligned}\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \frac{\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})}{\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}})} - \frac{\widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})} \\ &= \frac{1}{\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})}\left(\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})\right) - \frac{\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})}{\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}})\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})}\left(\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}}) - \widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})\right), \quad \text{(A.8)}\end{aligned}$$

from which we focus on the study of $\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ and $\widehat{f}_{\widehat{\boldsymbol{Z}}}$. Developing, we then have

$$\begin{aligned}\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \frac{1}{nh_{\boldsymbol{X}}}\sum_{i=1}^{n}\left\{K\left(\frac{\widehat{\gamma}_0^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right) - K\left(\frac{\gamma_0^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)\right\}\mathbb{1}(Y_i \le t) \\ &= \frac{1}{nh_{\boldsymbol{X}}^2}\sum_{i=1}^{n}K'\left(\frac{\xi^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)(\widehat{\gamma}_0-\gamma_0)^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{X}_i)\mathbb{1}(Y_i \le t) \\ &= (\widehat{\gamma}_0-\gamma_0)^{\mathsf{T}}\nabla_\gamma\,\widehat{L}_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x})\Big|_{\gamma=\xi}, \quad \text{(A.9)}\end{aligned}$$

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), we have that

$$\sup_{t\le\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}\sup_{\gamma\in\Xi}\left|\left|\nabla_\gamma\,\widehat{L}_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x}) - \nabla_\gamma\,L_{Y|\gamma^{\mathsf{T}}\boldsymbol{X}}(t|\gamma^{\mathsf{T}}\boldsymbol{x})\right|\right| = o_{\mathbb{P}}(1).$$

Along with assumption (C9)-(i), we then observe that

$$\sup_{t\le\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})| = O_{\mathbb{P}}(||\widehat{\gamma}_0-\gamma_0||) = O_{\mathbb{P}}(n^{-1/2}),$$

by assumption (C3). A similar development shows that

$$\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}}) - \widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z}) = (\widehat{\gamma}_0-\gamma_0)^{\mathsf{T}}\nabla_\gamma\,\widehat{f}_{\gamma^{\mathsf{T}}\boldsymbol{X}}(\gamma^{\mathsf{T}}\boldsymbol{x})\Big|_{\gamma=\xi},$$

for some ξ between $\widehat{\gamma}_0$ and γ_0. Using the same arguments as above, we then also observe that

$$\sup_{t\le\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}}) - \widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})| = O_{\mathbb{P}}(||\widehat{\gamma}_0-\gamma_0||) = O_{\mathbb{P}}(n^{-1/2}).$$

Returning to (A.8), we then have that

$$\begin{aligned}&\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}) \\ &\qquad= \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\left(\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{L}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})\right) - \frac{L_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})}{f_{\boldsymbol{Z}}^2(\boldsymbol{z})}\left(\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}}) - \widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z})\right) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

from which the result in (A.7a) follows directly from (A.9) and our previous developments on $\widehat{L}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ and $\widehat{f}_{\widehat{\boldsymbol{Z}}}(\widehat{\boldsymbol{z}})$. An exact same development for $\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ instead of $\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}})$ leads to the result in (A.7b) under our assumptions.

Now, returning to (A.5), first note that the results in (A.7a) and (A.7b), along with assumption (C3) and the continuity of ϱ_1 and ϱ_2 (see Gill and Johansen), lead to rewrite (A.6) as

$$\begin{aligned}\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \left(\widehat{F}_{Y|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y|\boldsymbol{Z}}(t|\boldsymbol{z})\right)\varrho_1\left(F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}), F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right) \\ &\quad + \left(\widehat{F}_{Y,1|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right)\varrho_2\left(F_{Y|\boldsymbol{Z}}(t|\boldsymbol{z}), F_{Y,1|\boldsymbol{Z}}(t|\boldsymbol{z})\right) \\ &\quad + o_{\mathbb{P}}(n^{-1/2}). \quad \text{(A.10)}\end{aligned}$$

Then, by inserting the expressions of (A.7a) and (A.7b), along with some standard Taylor expansion and change of variable arguments in (A.5), we have that

$$\begin{aligned}\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= n^{-1}\sum_{i=1}^{n}\varphi(Y_i,\Delta_i,t|\boldsymbol{z}) + O(h_T^2) + o_{\mathbb{P}}(n^{-1/2}) \\ &= n^{-1}\sum_{i=1}^{n}\varphi(Y_i,\Delta_i,t|\boldsymbol{z}) + o_{\mathbb{P}}(n^{-1/2}), \end{aligned} \tag{A.11}$$

provided assumption (C8) holds. Together with (A.4), this last result concludes the proof. □

Proof of Lemma 2. We decompose again

$$\begin{aligned}\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= \{\widehat{f}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} + \{\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{f}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\} \\ &= (T_3) + (T_4).\end{aligned}$$

For the term (T_3), first note that

$$\widehat{f}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) = h_T^{-1}\int \widetilde{K}\left(\frac{t-s}{h_T}\right)\mathrm{d}\widehat{F}_{T|\boldsymbol{Z}}(s|\boldsymbol{z}),$$

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

$$\begin{aligned}(T_{31}) &= h_T^{-1}\int \left\{\widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})\right\}\widetilde{K}'(u)\,\mathrm{d}u \\ &\leq \sup_{t\leq\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - F_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|\,h_T^{-1}\int\left|\widetilde{K}'(u)\right|\mathrm{d}u.\end{aligned}$$

Under the kernel assumption (C4), the latter term will then be observed to be $o_{\mathbb{P}}(1)$ provided $nh_{\boldsymbol{X}}h_T^2 \to \infty$. This will always be verified under assumption (C8), and hence (T_{31}) is $o_{\mathbb{P}}(1)$.

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 (T_3) is $o_{\mathbb{P}}(1)$ under assumption (C8).

It then remains to show the negligibility of (T_4) under our assumptions. For this, using again integration by parts and standard change of variables, we have

$$\begin{aligned}\widehat{f}^s_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{f}^s_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) &= h_T^{-1}\int \left\{\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t-uh_T|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t-uh_T|\boldsymbol{z})\right\}\widetilde{K}'(u)\,\mathrm{d}u \\ &\leq \sup_{t\leq\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}|\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|\,h_T^{-1}\int\left|\widetilde{K}'(u)\right|\mathrm{d}u.\end{aligned}$$

Using the results in the proof of Lemma 1, we then have that $\sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}|\widehat{F}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{F}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|= O_{\mathbb{P}}(n^{-1/2})$, which, together with assumptions (C4) and (C8) leads to

$$\sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}|\widehat{f}^{s}_{T|\widehat{\boldsymbol{Z}}}(t|\widehat{\boldsymbol{z}}) - \widehat{f}^{s}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})|= o_{\mathbb{P}}(1),$$

hereby concluding 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:

$$\begin{aligned} M_n(\beta; F_\gamma, f_\gamma) &= n^{-1}\sum_{i=1}^{n} m(\boldsymbol{X}_i, \beta; F_\gamma, f_\gamma) \\ &= n^{-1}\sum_{i=1}^{n} 2\boldsymbol{X}_i f(\beta^\intercal \boldsymbol{X}_i|\gamma^\intercal \boldsymbol{X}_i)\left(F(\beta^\intercal \boldsymbol{X}_i|\gamma^\intercal \boldsymbol{X}_i) - \tau\right). \end{aligned}$$

Furthermore, we let

$$\begin{aligned} M(\beta; F_\gamma, f_\gamma) &= \mathbb{E}\left[m(\boldsymbol{X}, \beta; F_\gamma, f_\gamma)\right] \\ &= \mathbb{E}\left[2\boldsymbol{X} f(\beta^\intercal \boldsymbol{X}|\gamma^\intercal \boldsymbol{X})\left(F(\beta^\intercal \boldsymbol{X}|\gamma^\intercal \boldsymbol{X}) - \tau\right)\right], \end{aligned}$$

and, conveniently denoting $F_{T|\gamma_0^\intercal \boldsymbol{X}}$ and $f_{T|\gamma_0^\intercal \boldsymbol{X}}$ 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

$$\begin{aligned} d_{\mathcal{F}_F}(F_{\gamma_1}, F^\star_{\gamma_2}) &= \sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}|F(t|\gamma_1^\intercal \boldsymbol{x}) - F^\star(t|\gamma_2^\intercal \boldsymbol{x})| \\ d_{\mathcal{F}_f}(f_{\gamma_1}, f^\star_{\gamma_2}) &= \sup_{t\leq \upsilon}\sup_{\boldsymbol{x}\in\text{supp}(\boldsymbol{X})}|f(t|\gamma_1^\intercal \boldsymbol{x}) - f^\star(t|\gamma_2^\intercal \boldsymbol{x})|, \end{aligned}$$

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 3. *Under assumptions (C1), (C5) and (C7), we have for all positive $\delta_n = o(1)$,*

$$\sup_{||\beta-\beta_\tau||\leq\delta_n}\sup_{d_{\mathcal{F}_F}(F_\gamma,F_0)\leq\delta_n}\sup_{d_{\mathcal{F}_f}(f_\gamma,f_0)\leq\delta_n}||M_n(\beta; F_\gamma, f_\gamma) - M(\beta; F_\gamma, f_\gamma)|| = o_{\mathbb{P}}(n^{-1/2}).$$

Proof of Lemma 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

$$\begin{aligned} &||m(\boldsymbol{x}, \beta_1; F_{\gamma_1}, f_{\gamma_1}) - m(\boldsymbol{x}, \beta_2; F^\star_{\gamma_2}, f^\star_{\gamma_2})|| \\ &\quad= \left|\left|2\boldsymbol{x}\Big(f(\beta_1^\intercal \boldsymbol{x}|\gamma_1^\intercal \boldsymbol{x})\{F(\beta_1^\intercal \boldsymbol{x}|\gamma_1^\intercal \boldsymbol{x}) - \tau\} - f^\star(\beta_2^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x})\{F^\star(\beta_2^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x}) - \tau\}\Big)\right|\right| \\ &\quad\leq K_1\Big(||F(\beta_1^\intercal \boldsymbol{x}|\gamma_1^\intercal \boldsymbol{x}) - F^\star(\beta_1^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x})||+||F^\star(\beta_1^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x}) - F^\star(\beta_2^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x})|| \\ &\qquad\quad+||f(\beta_1^\intercal \boldsymbol{x}|\gamma_1^\intercal \boldsymbol{x}) - f^\star(\beta_1^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x})||+||f^\star(\beta_1^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x}) - f^\star(\beta_2^\intercal \boldsymbol{x}|\gamma_2^\intercal \boldsymbol{x})||\Big) \\ &\quad\leq K_2\Big(d_{\mathcal{F}_F}(F_{\gamma_1}, F^\star_{\gamma_2}) + d_{\mathcal{F}_f}(f_{\gamma_1}, f^\star_{\gamma_2}) + ||\beta_1 - \beta_2||\Big), \end{aligned}$$

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 $N_{[]}(\epsilon, \mathcal{F}_F, ||\cdot||_{\mathcal{F}_F})$ 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 3.

Proof of Theorem 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.. 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 3, note that (1.4) in Chen et al. is straightforwardly satisfied by Lemma 1 under assumption (C8). 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 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

$$\begin{aligned}&\inf_{||\beta-\beta_\tau||>\epsilon} \left|\left|\mathbb{E}\Big[\boldsymbol{X} f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left(F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau\right)\Big]\right|\right| \\ &= \inf_{||\beta-\beta_\tau||>\epsilon} \left|\left|\mathbb{E}\Big[\boldsymbol{X} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\left(F_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}) - F_{T|\boldsymbol{X}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})\right)\Big]\right|\right| \\ &= \inf_{||\beta-\beta_\tau||>\epsilon} \left|\left|\int_{\mathrm{supp}(\boldsymbol{X})} \boldsymbol{x} f_{T|\boldsymbol{X}}(\beta^\mathsf{T}\boldsymbol{x}|\boldsymbol{x}) \int_{\beta_\tau^\mathsf{T}\boldsymbol{x}}^{\beta^\mathsf{T}\boldsymbol{x}} f_{T|\boldsymbol{X}}(t|\boldsymbol{x})\mathrm{d}t\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\right|\right|,\end{aligned}$$

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 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 $\Gamma_1(\beta; s_0)$:

$$\begin{aligned}\Gamma_1(\beta; s_0) &:= \frac{\partial M(\beta; s_0)}{\partial \beta} \\ &= \mathbb{E}\left[2\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\left(f^2_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) + f'_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left\{F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau\right\}\right)\right],\end{aligned}$$

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

$$\begin{aligned}\Gamma_2(\beta; s_0)[s_\gamma - s_0] &:= \lim_{\eta \to 0} \frac{1}{\eta}\left[M(\beta; s_0 + \eta(s_\gamma - s_0)) - M(\beta; s_0)\right] \\ &= 2\,\mathbb{E}\Big[\boldsymbol{X}\Big(f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left\{F(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\} \\ &\qquad + \left\{f(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\}\left\{F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}) - \tau\right\}\Big)\Big].\end{aligned}$$

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

$$\begin{aligned}&||M(\beta; s_\gamma) - M(\beta; s_0) - \Gamma_2(\beta; s_0)[s_\gamma - s_0]|| \\ &\quad = \left|\left|\mathbb{E}\left[2\boldsymbol{X}\left\{f(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - f_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\}\left\{F(\beta^\mathsf{T}\boldsymbol{X}|\gamma^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\}\right]\right|\right| \\ &\quad \leq K_1\, d_{\mathcal{F}_{F_\delta}}(F_\gamma, F_0)\, d_{\mathcal{F}_{f_\delta}}(f_\gamma, f_0) \\ &\quad \leq K_2\, d^2_{\mathcal{S}_\delta}(s_\gamma, s_0), \qquad \text{(A.12)}\end{aligned}$$

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 1 provided assumption (C8) holds. Condition (2.5) is, for its part, verified by Lemma 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 1, we have for $j = 1, \ldots, d+1$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}_j f_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left\{\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\widehat{\gamma}_0^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\}\right] \\
&= \mathbb{E}_{\gamma_0^\mathsf{T}\boldsymbol{X}}\left[(nh_{\boldsymbol{X}})^{-1}\sum_{i=1}^n \frac{K\left(\frac{\gamma_0^\mathsf{T}(\boldsymbol{X}-\boldsymbol{X}_i)}{h_{\boldsymbol{X}}}\right)}{(nh_{\boldsymbol{X}})^{-1}\sum_{k=1}^n K\left(\frac{\gamma_0^\mathsf{T}(\boldsymbol{X}-\boldsymbol{X}_k)}{h_{\boldsymbol{X}}}\right)} g_{1i}(\gamma_0^\mathsf{T}\boldsymbol{X}) + n^{-1}\sum_{i=1}^n g_{2i}(\gamma_0^\mathsf{T}\boldsymbol{X})\right] + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}$$

where the notations $\mathbb{E}_{\boldsymbol{X}}$ and $\mathbb{E}_{\gamma_0^\mathsf{T}\boldsymbol{X}}$ explicitly indicate that the expectations are taken with respect to the distributions of $\boldsymbol{X}$ and $\gamma_0^\mathsf{T}\boldsymbol{X}$, respectively, and where

$$\begin{aligned}
g_{1i}(u) &= \int_{\mathrm{supp}(\boldsymbol{X})} \boldsymbol{x}_j f_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{x}|u)\xi(Y_i, \Delta_i, \beta_\tau^\mathsf{T}\boldsymbol{x}|u) f_{\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}}(\boldsymbol{x}|u)\,\mathrm{d}\boldsymbol{x} \\
g_{2i}(u) &= \int_{\mathrm{supp}(\boldsymbol{X})} \boldsymbol{x}_j f_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{x}|u)\varphi(Y_i, \Delta_i, \beta_\tau^\mathsf{T}\boldsymbol{x}|u) f_{\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X}}(\boldsymbol{x}|u)\,\mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{A.13}$$

Simplifying a bit, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}_j f_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\left\{\widehat{F}^s_{T|\widehat{\boldsymbol{Z}}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\widehat{\gamma}_0^\mathsf{T}\boldsymbol{X}) - F_{T|\boldsymbol{Z}}(\beta_\tau^\mathsf{T}\boldsymbol{X}|\gamma_0^\mathsf{T}\boldsymbol{X})\right\}\right] \\
&\qquad = (nh_{\boldsymbol{X}})^{-1}\sum_{i=1}^n \int_{-\infty}^{+\infty} g_{1i}(\boldsymbol{z}) K\left(\frac{\boldsymbol{z} - \gamma_0^\mathsf{T}\boldsymbol{X}_i}{h_{\boldsymbol{X}}}\right)\mathrm{d}\boldsymbol{z} + n^{-1}\sum_{i=1}^n \mathbb{E}\left[g_{2i}(\gamma_0^\mathsf{T}\boldsymbol{X})\right] + o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

We now write, for $i = 1, \ldots, n$,

$$g_i(u) = g_{1i}(u) + \mathbb{E}\left[g_{2i}(\gamma_0^\mathsf{T}\boldsymbol{X})\right]. \tag{A.14}$$

Using standard change of variables and a Taylor expansion for g_1 under assumption (C9)-(ii), we then have

$$\begin{aligned}
\Gamma_2(\beta_\tau; s_0)[\widehat{s} - s_0] &= 2\,n^{-1}\sum_{i=1}^n g_i(\gamma_0^\mathsf{T}\boldsymbol{X}_i) + O(h_{\boldsymbol{X}}^\nu) + o_{\mathbb{P}}(n^{-1/2}) \\
&= 2\,n^{-1}\sum_{i=1}^n g_i(\gamma_0^\mathsf{T}\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}$$

provided assumptions (C4) and (C8) hold. An application of the central limit theorem under our assumptions then leads to the desired result, hereby concluding the proof. □