

# Global Bahadur representation for nonparametric censored regression quantiles and its applications

---

Efang Kong  
Oliver Linton  
Yingcun Xia

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP33/11

# Global Bahadur Representation for Nonparametric Censored Regression Quantiles and Its Applications

Efang Kong\*

*University of Kent at Canterbury, UK*

Oliver Linton†

*University of Cambridge, UK*

Yingcun Xia‡

*National University of Singapore, Singapore*

**Abstract.** This paper is concerned with the nonparametric estimation of regression quantiles where the response variable is randomly censored. Using results on the strong uniform convergence of U-processes, we derive a global Bahadur representation for the weighted local polynomial estimators, which is sufficiently accurate for many further theoretical analyses including inference. We consider two applications in detail: estimation of the average derivative, and estimation of the component functions in additive quantile regression models.

*Key words:* Bahadur representation; Censored data; Kernel smoothing; Quantile regression; Semiparametric models.

## 1 Introduction

Quantile regression (Koenker and Bassett, 1978), designed originally to render robust estimators against extreme values or outliers among the error terms (Huber, 1973), has attracted tremendous interest in applied work. Equally, censored data regression has always been an important topic in survival analysis, for example, the accelerated failure time

---

\* School of Mathematics, Statistics and Actuarial Science, University of Kent at Canterbury, UK. E-mail: [E.kong@kent.ac.uk](mailto:E.kong@kent.ac.uk).

† Faculty of Economics, University of Cambridge, UK. E-mail: [ob120@cam.ac.uk](mailto:ob120@cam.ac.uk).

‡ Department of Statistics and Applied Probability and RMI, National University of Singapore, Singapore. E-mail: [staxyc@nus.edu.sg](mailto:staxyc@nus.edu.sg).

model, as well as in econometrics, through the well-known Tobit model. A direct consequence of either fixed censoring or random censoring is that it renders the error term to deviate severely from the normal distribution and even worse the conditional moment restrictions of the uncensored model to be violated. Regression quantiles are among the natural choices in analyzing censored data because they may be robust to some censoring, Powell (1984). Most of the existing literature on quantile regressions under censoring adopted a linear/parametric approach; see e.g. Buckley and James (1979), Koul, Susarla and Van Ryzin (1981), Ritov (1990), Ying et al (1995), Honoré, Khan, and Powell (2002), Bang and Tsiatis (2003), and Heuchenne and Van Keilegom (2007a). They assumed that the quantile function belongs to a fixed finite-dimensional space of functions. Under their assumption, the statistical theory of quantile regression for censored data has been well understood and investigated. In this paper, we will focus on the quantile regression models in a nonparametric setting, which imposes no restrictions on the form of the function except for some smoothness properties and likewise we do not restrict the form of the censoring, allowing the censoring distribution to depend in an unknown way on the covariates, so generalizing the setting considered in Honoré, Khan, and Powell (2002).

A small number of estimators exist for nonparametric censored regression models, in most cases focusing on the standard random censoring model. Dabrowska (1992) and Van Keilegom and Veraverbeke (1998) proposed nonparametric censored regression estimators based on quantile methods. Lewbel and Linton (2002) considered the case where the censoring time is a degenerate random variable (i.e., it is constant), extended by Chen, Dahl, and Khan (2005) to allow for heteroscedasticity. Heuchenne and Van Keilegom (2007b, 2008) considered a nonparametric regression model where the error term is independent of the covariates. Linton, Mammen, Nielsen, and Van Keilegom (2011) consider univariate regression models with a variety of censoring schemes and employ estimation methods based on hazard functions.

Bahadur (1966) representations are a useful tool to study the asymptotic properties of estimators especially when the loss function is not smooth, such as in M-estimation and quantile regression. As commented in He and Shao (1996), Bahadur representation approximates the estimator by a sum of independent variables with a smaller-order remainder. As a consequence, many asymptotic properties useful in statistical inference can be derived easily from the Bahadur representation. Under different settings, a number of different Bahadur

representations have been obtained. For example, Carroll (1978) and Martinsek (1989) obtained strong representations for location and regression M-estimators with preliminary scale estimates; Babu (1989) and Pollard (1991) obtained the Bahadur representation for the least absolute deviation regression. Portnoy (1997) obtained the Bahadur representation of quantile smoothing splines. Portnoy (2003) obtained the Bahadur representation for the Cox and censored quantile regression. Chaudhuri (1991a) investigated the pointwise Bahadur representation of nonparametric kernel quantile regression. Kong, Linton and Xia (2009) and Guerre and Sabbah (2009) obtained the uniform Bahadur representation for the quantile local polynomial estimators. Wu (2007) and Zhou and Wu (2010) investigated the Bahadur representation for nonstationary time series data under both parametric and nonparametric settings.

In nonparametric settings, global or uniform asymptotic theory (Bickel, 1972 and Mack and Silverman, 1982) is essential for conducting statistical inference. Because of this, uniform Bahadur representations are more useful than their pointwise counterparts. In this paper, we shall give a global Bahadur representation for nonparametric estimates of censored regression quantiles. We provide two applications of our theory. The first one is the additive model which has been investigated under quantile regression (Linton 2001, De Gooijer and Zerom, 2003; Yu and Lu, 2004) or censored data (Uña Álvarez and Pardiñasa, 2009) separately. However, as far as we know, no one has investigated the model under the combination of the two settings. The second application is to the popular single-index model. Again, this model was investigated under two separate settings. Chaudhuri et al (1991), Wu, Yu and Yu (2010) and Kong and Xia (2011) considered quantile regression of the single-index model, while Lu and Cheng (2007) and Xia, Zhang and Xu (2009) considered the conditional mean regression under random censoring. The global Bahadur representation can also be applied to censored regression quantiles of other semiparametric models; see for example Zhang and Li (2011). Our results are particularly useful for conducting inference about the quantities of interest. The representations we have obtained can directly be used to obtain consistent standard errors in the case where a parametric quantity like the average derivative is of interest or where one wants a pointwise confidence interval for a function like the additive component. They can also be used to obtain uniform confidence bands for such functions, since the detailed probabilistic analysis of the leading terms follows from the well established results for kernel regression and density estimators,

Bickel and Rosenblatt (1973) and Johnston (1982). We remark that recent work of Belloni, Chernozhukov, and Fernández-Val (2011) has provided tools for inference about nonparametric quantile regression based on the series methodology but in the absence of a censoring mechanism.

## 2 The model and estimation method

Suppose that we have real-valued iid observations  $\{(Y_i^*, \mathbf{X}_i), i = 1, \dots, n\}$ , satisfying the model

$$Y_i^* = Q(\mathbf{X}_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where  $Y_i^*$  is the (uncensored and scalar) response variable, while  $\mathbf{X}_i$  is the observed  $p$ -dimensional covariates. Here  $Q(\cdot)$  is an unknown but smooth function, and  $\varepsilon_i$  is the 'error term', which conditional on  $\mathbf{X}$  has a  $\tau$ th quantile equal to zero; i.e.  $Q(\mathbf{X}_i)$  is the conditional  $\tau$ th quantile of  $Y_i^*$  given  $\mathbf{X}_i$ . Or equivalently, through the use of the quantile loss of function, we have

$$Q(\mathbf{X}_i) = \arg \min_a E \rho_\tau(Y_i^* - a | \mathbf{X}_i)$$

where  $\rho_\tau(s) = |s| + (2\tau - 1)s$ . The objective of estimation is the unknown function  $Q(\cdot)$  and its derivatives.

In this paper, we focus on random right censoring; the methodology can be extended to left censoring. Let  $C_i$  denote the censoring variable, with conditional survival function  $G(\cdot | \mathbf{X}_i)$ , i.e. we allow its distribution to depend on  $\mathbf{X}_i$ . In this case, we only observe the triple  $\zeta_i = (Y_i, X_i, d_i)$ , where

$$Y_i = \min\{Y_i^*, C_i\} = \min\{Q(\mathbf{X}_i) + \varepsilon_i, C_i\}, \quad d_i = I\{Y_i^* \leq C_i\}, \quad (2)$$

are the observed (possibly censored) response variable and the censoring indicator, respectively. Equation (1) together with (2) specify a censored quantile regression model, and our main objective is the estimation of  $Q(\cdot)$ , the conditional quantile function of  $Y_i^*$  given  $\mathbf{X}_i$ .

Suppose  $\zeta_i = \{Y_i^*, \mathbf{X}_i, C_i\}, i = 1, \dots, n$ , are i.i.d. random variables, and  $Q(\cdot)$  has partial derivatives up to order  $k$ . For any fixed point  $\mathbf{x} \in R^p$ , the local polynomial estimation of  $Q(\mathbf{x})$  is based the approximation of  $Q(\cdot)$  in the neighborhood of  $\mathbf{x}$  by its  $k$ -order Taylor expansion:

$$Q(\mathbf{X}) \approx Q(\mathbf{x}) + \sum_{1 \leq |u| \leq k} \frac{D^u Q(\mathbf{x})}{\mathbf{u}!} (\mathbf{X} - \mathbf{x})^u, \quad (3)$$

where  $\mathbf{u} = (u_1, \dots, u_p)$  denotes an arbitrary  $p$ -dimensional vector of nonnegative integers,  $[\mathbf{u}] = \sum_{i=1}^p u_i$ ,  $\mathbf{u}! = \prod_{i=1}^p u_i!$ ,  $x^{\mathbf{u}} = \prod_{i=1}^p x_i^{u_i}$  with the convention that  $0^0 = 1$ , and  $D^{\mathbf{u}}$  denotes the differential operator  $\partial^{[\mathbf{u}]} / \partial x_1^{u_1} \dots \partial x_p^{u_p}$ . let  $A = \{\mathbf{u} : [\mathbf{u}] \leq k\}$  and  $n(A) = \#(A)$ . When there is no censoring, the estimates of  $Q(\cdot)$  and its partial derivatives, are obtained by minimizing the function below with respect to  $\mathbf{c} = (c_{\mathbf{u}})_{\mathbf{u} \in A}$ , a vector of length  $n(A)$ ,

$$\sum_{i=1}^n K_{\delta_n}(\mathbf{X}_{i,\mathbf{x}}) \rho_{\tau}\{Y_i - \langle \mathbf{c}, \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \rangle\}, \quad \mathbf{X}_{i,\mathbf{x}} = \mathbf{X}_i - \mathbf{x}, \quad (4)$$

where  $K_{\delta_n}(\cdot) = K(\cdot/\delta_n)$  is some probability density function in  $R^p$  with a smoothing parameter  $\delta_n \rightarrow 0$ ,  $\mathbf{x}(\delta_n, A) = (\mathbf{x}(\delta_n, \mathbf{u}))_{\mathbf{u} \in A}$ , with  $\mathbf{x}(\delta_n, \mathbf{u}) = \delta_n^{-[\mathbf{u}]} \mathbf{x}^{\mathbf{u}}$  defined for any  $\mathbf{x} \in R^p$ , and  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product. Similar ideas have been used in Chaudhuri (1991a, 1991b), and Kong et al (2010).

One of the complications brought about by censoring is that,  $Q(\cdot)$  is the  $\tau$ -quantile of  $Y_i$  iff  $d_i = 1$ . However, the straightforward modification of (4) by restricting the summation to be across those  $i$ 's with  $d_i = 1$  results in a biased estimator. Among many, there are three possible ways to tackle this problem. One is by replacing  $\rho_{\tau}\{Y_i - \langle \mathbf{c}, \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \rangle\}$  with its conditional expectation given  $(Y_i, \mathbf{X}_i, d_i)$ ; see Honoré et al (2002) for its application to the linear quantile regression model. The second is to apply the 'redistribution-of-mass' idea of Efron (1967); see also Portnoy (2003), Peng and Huang (2008), and Wang and Wang (2009). The third strategy which we consider in this paper, is based on the observation that  $E[d_i/G(Y_i|\mathbf{X}_i)] = 1$  (Bang and Tsiatis, 2003), which when plugged into (4) leads to the minimization of

$$\sum_{i=1}^n \frac{d_i}{G(Y_i|\mathbf{X}_i)} \rho_{\tau}\{Y_i - \langle \mathbf{c}, \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \rangle\} K_{\delta_n}(\mathbf{X}_{i,\mathbf{x}}). \quad (5)$$

In practice,  $G(\cdot|\mathbf{X}_i)$  is unknown and has to be estimated. A nonparametric estimator of  $G(\cdot|\mathbf{X}_i)$  is the local Kaplan-Meier estimator  $\hat{G}_n(\cdot|\mathbf{X}_i)$  (Gonzalez-Manteigaa and Cadarso-Suarez, 1994) defined as

$$\hat{G}_n(t|\mathbf{x}) = \prod_{j=1}^n \left\{ 1 - \frac{B_{nj}(\mathbf{x})}{\sum_{k=1}^n I(Y_k \geq Y_j) B_{nk}(\mathbf{x})} \right\}^{\beta_j(t)}, \quad (6)$$

where  $\beta_j(t) = I(Y_j \leq t, d_j = 0)$ , and  $B_{nk}(\mathbf{x})$ ,  $k = 1, \dots, n$  is a sequence of nonnegative weights adding up to 1. We adopt this idea of local Kaplan-Meier estimator, but with a slightly different choice for  $B_{nk}(\cdot)$ : the local polynomial 'equivalent kernel/weight'; see, Fan

and Gijbels (1996) and Masry (1996) for more details. Specifically, for some positive integer  $\kappa_1$ , define

$$B_{nk}(\mathbf{x}) = \mathbf{e}_1^\top \tilde{\Sigma}_n^{-1}(\mathbf{x}) \tilde{B}_{nk}(\mathbf{x}), \quad \tilde{B}_{nk}(\mathbf{x}) = \mathbf{X}_{k\mathbf{x}}(h_n, A_1) \tilde{K}_{h_n}(\mathbf{X}_{k\mathbf{x}}), \quad (7)$$

where  $A_1 = \{\mathbf{u} : [\mathbf{u}] \leq \kappa_1\}$ ,  $\mathbf{e}_1$  is the  $n(A_1) \times 1$  vector  $(1, 0, \dots, 0)^\top$ ,

$$\tilde{\Sigma}_n(\mathbf{x}) = \frac{1}{n} \sum_{k=1}^n \tilde{K}_{h_n}(\mathbf{X}_{k\mathbf{x}}) \mathbf{X}_{k\mathbf{x}}(h_n, A_1) \mathbf{X}_{k\mathbf{x}}(h_n, A_1)^\top,$$

$\tilde{K}(\cdot) : R^p \rightarrow R^+$  is yet another kernel density function, and  $h_n \in R^+$  is the associated smoothing parameter, not necessarily identical to the  $K(\cdot)$  and  $\delta_n$  used above.

Substituting  $\hat{G}_n(\cdot|\cdot)$  for  $G(\cdot|\cdot)$  in (8), we propose to estimate  $\{D^{\mathbf{u}}Q(\mathbf{x}) : [\mathbf{u}] \in A\}$  by

$$\mathbf{c}_n(\mathbf{x}) = (c_{n,\mathbf{u}}(\mathbf{x}))_{\mathbf{u} \in A} \stackrel{def}{=} \arg \min_{\mathbf{c}} \sum_{i=1}^n \frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)} \rho_\tau\{Y_i - \langle \mathbf{c}, \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \rangle\} K_{\delta_n}(\mathbf{X}_{i,\mathbf{x}}). \quad (8)$$

Since  $0 < \tau < 1$ ,  $\rho_\tau(s)$  goes to infinity as  $|s| \rightarrow \infty$ . Thus the minima of (8) always exists.

The reason to use a weight function  $B_{nk}(\mathbf{x})$  as in (7), instead of the commonly used Nadaraya-Waston's type weights such as in Wang and Wang (2009), is that the latter is no longer up for the job of yielding a bias which is 'negligible relative to variance' in multivariate settings; whereas the 'local polynomial equivalent kernel', though a bit more complicated, renders a bias of order  $h_n^{\kappa_1+1}$ , which can be arbitrarily small for large value of  $\kappa_1$ .

A minor complication resulted from using weight (7) is that the corresponding K-M estimator (6) is not necessarily a proper survival function as  $B_{nk}(\cdot)$  could be negative. However, this shouldn't cause much concern, as firstly, the result proved in Gonzalez-Manteigaa and Cadarso-Suarez (1994) on the local K-M estimator (6) doesn't rely on the positivity of  $B_{nk}(\cdot)$ . Secondly, in practice, a simple truncation can always be applied to ensure  $0 \leq \hat{G}_n(\cdot|\mathbf{x}) \leq 1$ ; see Spierdijk (2008) for a similar observation.

### 3 Notations and Assumptions

Let  $D$  be an open convex set in  $R^p$  and for  $s_0 = l + \gamma$ , with non-negative integer  $l$  and  $0 < \gamma \leq 1$ , we say a function  $m(\cdot) : R^p \rightarrow R$  has the order of smoothness  $s_0$  on  $D$ , denoted by  $m(\cdot) \in H_{s_0}(D)$ , if its partial derivatives up to order  $l$  exists and there exists a constant  $C > 0$ , such that

$$|D^{\mathbf{u}}m(\mathbf{x}_1) - D^{\mathbf{u}}m(\mathbf{x}_2)| \leq C|\mathbf{x}_1 - \mathbf{x}_2|^\gamma, \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in D \text{ and } [\mathbf{u}] = l.$$

Note that  $|\cdot|$  in this paper stands for the supnorm, i.e. for  $\mathbf{x} = (x_1, \dots, x_p)^\top \in R^p$ ,  $|\mathbf{x}| = \max_{1 \leq i \leq p} |x_i|$ .

For any  $\mathbf{t} \in [-1, 1]^p$ , denote by  $\mathbf{t}(A)$  the vector of length  $n(A)$  with elements  $(\mathbf{t}^{\mathbf{u}})_{\mathbf{u} \in A}$ . Let  $\Sigma(A)$  be the  $n(A) \times n(A)$  matrix

$$\Sigma(A) = \int_{[-1, 1]^p} \mathbf{t}(A) \mathbf{t}(A)^\top dt.$$

$\mathbf{t}(A_1)$  and matrix  $\Sigma(A_1)$  are similarly defined. It is assumed throughout this paper that both matrices,  $\Sigma(A)$  and  $\Sigma(A_1)$  are invertible.

Let  $f(\cdot)$  be the marginal probability density function of  $\mathbf{X}_i$ . For any  $\mathbf{x} \in R^p$ , denote by  $g(\cdot|\mathbf{x})$ ,  $f_0(\cdot|\mathbf{x})$ , and  $f_\varepsilon(\cdot|\mathbf{x})$ , the probability density functions of  $C_i$ ,  $Y_i^*$  and  $\varepsilon_i$  conditional on  $\mathbf{X}_i = \mathbf{x}$ . Let  $F_0(t|\mathbf{x}) = P(Y_i^* \leq t|X_i = \mathbf{x})$ ,  $F_\varepsilon(t|\mathbf{x}) = P(\varepsilon_i \leq t|X_i = \mathbf{x})$ .

We make the following assumptions:

[A1]  $f(\mathbf{x})$  is positive on a compact set  $D \subset R^p$  and  $f \in H_{s_1}(D)$ .

[A2] The quantile function  $Q(\cdot)$  has the order of smoothness  $s_2$ , i.e.  $Q(\cdot) \in H_{s_2}(D)$ .

[A3]  $f_\varepsilon(t|\mathbf{x})$ , considered as a function of  $\mathbf{x}$  belongs to  $H_{s_3}(D)$  for all  $t$  in a neighborhood of zero.  $f_\varepsilon(0|\mathbf{x})$  is positive for all  $\mathbf{x} \in D$ , and its first derivative with respect to  $t$  exists continuously for values of  $t$  in a neighborhood of zero for all  $\mathbf{x} \in D$ .

[A4] The censoring variable  $\{C_i\}$  is conditionally independent of  $\varepsilon_i$  given  $\mathbf{X}_i$ ; and for any  $\mathbf{x} \in R^p$ , there exists some finite  $\pi_0$ , which might depend on  $\mathbf{x}$ , such that  $G(\tau_0|\mathbf{x}) = 0$  and  $\inf_{\mathbf{x}} P(C_i = \pi_0|\mathbf{x}) > 0$ .

[A5] The functions  $f_0(0|\mathbf{x})$  and  $g(t|\mathbf{x})$  are uniformly bounded both in  $t$  and in  $\mathbf{x}$ . Both belong to  $H_{s_4}(D)$  and their  $\kappa_1(= [s_4])$ -order partial derivatives with respect to  $\mathbf{x}$  belong to  $H_{s_4}(D)$ , uniformly in  $t$ .

[A6] The kernel function  $\tilde{K}(\cdot)$  is a probability density function on  $R^p$  with a compact support. It is symmetric and Lipschitz continuous of order 1 with finite variance.

[A7] The bandwidth  $h_n$  is chosen such that  $nh_n^{2s_4+p}/\log n \rightarrow \infty$ ,  $nh_n^{3p}/\log n \rightarrow 0$ ,  $nh_n^{p+4}/\log n < \infty$ .

[A8] The relationship between the two smoothing parameters  $\delta_n$  and  $h_n$  is such that  $nh_n^{2p}/(\delta_n^p \log n) \rightarrow \infty$ .

**Remark** [A1]-[A3] are standard regularity conditions assumed in the context of local polynomial smoothing; see also Chaudhuri et al (1997). Especially, [A2] implies that, if  $|\mathbf{X} - \mathbf{x}| \leq \delta_n$ , then the error resulted from approximating  $Q(\mathbf{X})$  by the  $k$ -order ( $k = [s_2]$ )

Taylor expansion

$$Q_n(\mathbf{X}, \mathbf{x}) = \sum_{\mathbf{u} \in A} c_{n,\mathbf{u}}(\mathbf{x}) [(\mathbf{X} - \mathbf{x})/\delta_n]^{\mathbf{u}}$$

is of order  $O(\delta_n^{s_2})$ , uniformly over  $\mathbf{x} \in D$  and  $\{\mathbf{X} : |\mathbf{X} - \mathbf{x}| \leq \delta_n\}$ . [A4] implies a upper bound on the censoring values and the positive mass on the upper boundary of their support. This guarantees that  $d_i/\hat{G}_n(Y_i|\mathbf{X}_i)$  is uniformly finite in large samples; this condition can be satisfied by artificially censoring all observations at some point  $\pi_0(\leq \max_i Y_i)$ . [A5]-[A7] are imposed such that the local K-M estimator  $\hat{G}_n(\cdot|\mathbf{X}_i)$  admits the almost sure representation in terms of a sum of independent random variables. Note that compared to that in Gonzalez-Manteigaa and Cadarso-Suarez (1994) or Wang and Wang (2009), which focuses on univariate covariate, [A6] is stronger, which is necessary to ensure that the bias of the K-M estimator is negligible relative to the stochastic terms.

To facilitate the subsequent discussion on application, we will focus on the estimation of  $\mathbf{c}_n(\mathbf{x}) = (c_{n,\mathbf{u}}(\mathbf{x}))_{\mathbf{u} \in A}$  with  $c_{n,\mathbf{u}}(\mathbf{x}) = \delta_n^{[\mathbf{u}]} D^{\mathbf{u}} Q(\mathbf{x})/\mathbf{u}!$  with  $\mathbf{x} = \mathbf{X}_j$ ,  $j = 1, \dots, n$ . Note that for simplification purposes, we choose  $K(\cdot)$  to be the uniform density on  $[-1, 1]^p$ :  $\mathcal{U}[-1, 1]^p$ . Results obtained in this paper hold for the use of any symmetric probability density functions with a compact support. We will derive uniform convergence rate and the Bahadur type representation of  $\hat{\mathbf{c}}_n(\mathbf{X}_j)$  defined as

$$\hat{\mathbf{c}}_n(\mathbf{X}_j) = \arg \min_{\mathbf{c}} \sum_{i \in S_n(\mathbf{x})} \frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)} \rho_{\tau}\{Y_i - \langle \mathbf{c}, \mathbf{X}_{i,j}(\delta_n, A) \rangle\}, \quad \mathbf{X}_{i,j} = \mathbf{X}_i - \mathbf{X}_j \quad (9)$$

where  $S_n(\mathbf{X}_j)$  is the index set defined as

$$S_n(\mathbf{X}_j) = \{i : 1 \leq i \leq n, i \neq j, |\mathbf{X}_{ij}| \leq \delta_n\}, \quad N_n(\mathbf{X}_j) = \#(S_n(\mathbf{X}_j)),$$

#### 4 Convergence rate and asymptotic representation

Our first result concerns the almost sure representation of the local K-M estimator  $\hat{G}_n(\cdot|.)$ :

**Lemma 4.1** *Under [A4]-[A7], we have with probability one,*

$$\sup_{\mathbf{x}} \sup_t |\hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x})| = O\left(\left(\frac{\log n}{nh_n^p}\right)^{1/2}\right) \quad (10)$$

$$\begin{aligned} \hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x}) &= \mathbf{e}_1^\top \Sigma^{-1}(A_1)G(t|\mathbf{x}) \frac{1}{n} \sum_{j=1}^n \tilde{B}_{nj}(\mathbf{x}) \xi(Y_j, d_j, t, \mathbf{x}) \\ &\quad + O\left(\left(\frac{\log n}{nh_n^p}\right)^{3/4}\right) \end{aligned} \quad (11)$$

uniformly in  $\mathbf{x}$  as well as in  $t$ , where, for  $j = 1, \dots, n$ ,

$$\begin{aligned}\xi(Y_j, d_j, t, \mathbf{x}) &= \frac{I\{Y_j \leq t, d_j = 0\}}{h(Y_j|\mathbf{x})} - \int_0^{\min(Y_j, t)} \frac{d\Gamma(s|\mathbf{x})}{h(s|\mathbf{x})} \\ h(t|\mathbf{x}) &= 1 - P(Y_j \leq t|\mathbf{x}) = G(t|\mathbf{x})(1 - F(t|\mathbf{x})) \\ \Gamma(t|\mathbf{x}) &= -\ln(G(t|\mathbf{x})) = -\int_0^t \frac{dG(t|\mathbf{x})}{G(t|\mathbf{x})}.\end{aligned}$$

The next Theorem gives the almost sure convergence rate of  $\hat{\mathbf{c}}_n(\mathbf{X}_j)$  uniform in  $j = 1, \dots, n$  :

**Theorem 4.2** *Suppose [A4]-[A8] and assume that  $f_{\mathbf{x}}(\cdot)$  is positive and continuous. Then under [A2] and [A3] with  $s_2 > 0$ ,  $s_3 > 1$ , (3) holds with  $k = [s_2]$ , and that the bandwidth  $\delta_n$  in the definition (9) is such that*

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2s_2 + d} < \kappa.$$

then we have with probability one,

$$\sup_{1 \leq j \leq n} |\hat{\mathbf{c}}_n(\mathbf{X}_j) - \mathbf{c}_n(\mathbf{X}_j)| = O[(\log n/n^{1-\kappa p})^{1/2}]. \quad (12)$$

**Remark** Note that conditions in Theorem 4.2 could be weakened; see Chaudhuri (1991b). The uniformity can be extended to cover the whole compact set  $\mathcal{D}$ , which can be easily verified as follows. Cover  $\mathcal{D}$  with  $J_n^p$  number of cubes side length  $2\delta_n$  and let  $S_{n,r}$ ,  $1 \leq r \leq J_n^p$ , be a typical such cube with center at  $\mathbf{x}_{n,r}$ . Once  $\hat{\mathbf{c}}_n(\mathbf{x}_{n,r})$  are obtained through minimizing (9) with  $\mathbf{x}_{n,r}$  in place of  $\mathbf{X}_j$ , the estimates of  $\mathbf{c}_n(\mathbf{x})$  for any fixed point  $\mathbf{x}$  with  $S_{n,r}$  is defined as

$$\hat{\mathbf{c}}_{n,\mathbf{u}}(\mathbf{x}) = \delta_n^{[\mathbf{u}]} / \mathbf{u}! D^{\mathbf{u}}[\langle \hat{\mathbf{c}}_n(\mathbf{x}_{n,r}), (\mathbf{x}_{n,r} - \mathbf{x})(\delta_n, A) \rangle], \quad \text{if } \mathbf{x} \in S_{n,r} \quad (13)$$

where the differential operator  $D^{\mathbf{u}}$  is with respect to  $\mathbf{x}$ . Under [A2], the problem thus translates into a problem of establishing the uniform convergence rate for  $\hat{\mathbf{c}}_n(\mathbf{x}_{n,r})$ , which can be proved in exactly the same way as (12), by applying

Regarding the strong uniform Bahadur type representation of  $\hat{\mathbf{c}}_n(\cdot)$ , we have

**Theorem 4.3** *Suppose conditions of Theorem 4.2 hold and*

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2(s_2 + p)} < \kappa < \frac{1}{p},$$

we have

$$\begin{aligned}\hat{\mathbf{c}}_n(\mathbf{X}_j) - \mathbf{c}_n(\mathbf{X}_j) &= \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) \left[ \tau - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} \right] \\ &\quad - \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \sum_{k=1}^n \tilde{Q}(\zeta_j, \zeta_k) \Sigma^{-1}(A_1) \mathbf{e}_1 + R_n(\mathbf{X}_j),\end{aligned}\quad (14)$$

where

$$\begin{aligned}\Sigma_n^{-1}(\mathbf{x}) &= E_i \left\{ f_{\varepsilon|\mathbf{X}}(0|\mathbf{X}_i) \mathbf{X}_{i\mathbf{x}}(\delta_n, A) \mathbf{X}_{i\mathbf{x}}^\top(\delta_n, A) | \mathbf{X}_i \in S_n(\mathbf{x}) \right\} \\ \tilde{Q}(\zeta_j, \zeta_k) &= \int \mathbf{X}_{ij}(\delta_n, A) \mathbf{X}_{ki}^\top(h_n, A_1) E_{Y_i|\mathbf{X}_i} [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \xi(Y_k, d_k, Y_i, \mathbf{X}_i) d\mathbf{X}_i\end{aligned}$$

and

$$\max_{1 \leq j \leq n} |R_n(\mathbf{X}_j)| = O(n^{-3(1-\kappa p)/4} [\log n]^{3/4}), \quad a.e.$$

### Remark

Compared to the result in Chaudhuri et al (1997) (Lemma 4.1), the extra term in (14) can be interpreted as the 'correction term' resulted from the preliminary estimation of the survival function  $G(\cdot)$ . Similar observation has been made by Honoré et al (2002) for linear quantile regression under censoring. The uniformity results again can easily strengthened to be over any  $n^\iota$  of  $\mathbf{x}$ 's:  $\mathbf{x}_1, \dots, \mathbf{x}_{n^\iota}$  with  $\iota$  being finite. Similarly, if [A2] and [A3], which concern functions (random variables) varying with  $\tau$ , are satisfied uniformly in  $\tau$  as well in addition to in  $\mathbf{x}$  and (or)  $t$ , then the uniformity results in Theorem 4.3 can be easily extended to cover estimation at different quantile levels  $\tau_1, \dots, \tau_{n^\iota}$  with finite  $\iota$ .

Under conditions of Theorem 4.3, we have

$$\max_{1 \leq j \leq n} \delta_n^{-1} |R_n(\mathbf{X}_j)| = o(n^{-1/2}) \quad a.e. \quad (15)$$

provided that

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2(s_2 + p)} < \kappa < \frac{1}{4 + 3p}, \quad (16)$$

and

$$(ii) \quad \max_{1 \leq j \leq n} |R_n(\mathbf{X}_j)| = o(n^{-1/2}) \quad a.e. \quad (17)$$

provided that

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2(s_2 + p)} < \kappa < \frac{1}{3p}. \quad (18)$$

(15) will be used for the average derivative estimator, and (17) will be used for the additive quantile regression model.

## 5 Applications

In this section, we will demonstrate how the results in Theorem 4.3 can be used to obtain the asymptotic properties of a class of estimators through two examples

### 5.1 The Average Derivative Estimator

Define the average gradient vector

$$\beta = (\beta_1, \dots, \beta_p)^\top = E(\nabla Q(\mathbf{X})),$$

which gives a concise summary of quantile specific regression effects, i.e. the average change in the quantile of the response as the  $i$ th covariate is perturbed while the other covariates are held fixed. This parameter has been of great interest in econometrics following Härdle and Stoker (1989). See for example Chaudhuri et al (1997) who extended this theory to quantile regression case. Here we study the estimation of  $\beta$  in the presence of censoring using the average derivative method.

Let  $\nabla \hat{Q}(X_j)$  be the nonparametric estimator of the gradient of the conditional quantile  $Q(\mathbf{x})$  at  $\mathbf{x} = \mathbf{X}_j$ , defined in (8), i.e.

$$\nabla \hat{Q}(X_j) = (\hat{c}_{n,\mathbf{u}}(\mathbf{X}_j))_{[\mathbf{u}]=1}.$$

The average derivative estimator of  $\beta$  is thus defined as

$$\hat{\beta} = \frac{1}{n} \sum_{j=1}^n \nabla \hat{Q}(X_j). \quad (19)$$

To establish the asymptotics of  $\hat{\beta}$ , we assume that [A1]-[A3] hold with  $s_1 = s_3 = 1 + \gamma$  ( $\gamma > 0$ ),  $s_2 > 3 + 3p/2$  and (3) holds with  $k = [s_2]$ , then according to Theorem 4.3, under (16) for any  $\mathbf{a} = (a_1, \dots, a_p)^\top \in R^p$ , we have

$$\begin{aligned} & \mathbf{a}^\top (\hat{\beta} - \beta) \\ = & \mathbf{a}^\top \left[ \frac{1}{n} \sum_{j=1}^n \nabla Q(\mathbf{X}_j) - \beta \right] + o(n^{-1/2}) \\ & + \mathbf{A}^\top \frac{1}{n\delta_n} \sum_{j=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) \left[ \tau - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} \right] \quad (20) \end{aligned}$$

$$+ \mathbf{A}^\top \frac{1}{n^2} \sum_{j,k=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \tilde{Q}(\zeta_j, \zeta_k) \mathbf{e}_1^\top \Sigma^{-1}(A_1) \quad (21)$$

where  $\mathbf{A}$  is an  $n(A) \times 1$  factor, defined as

$$\mathbf{A} = (0, \mathbf{a}^\top, \mathbf{0})^\top$$

Firstly note that the term (20) has been shown (Chaudhuri et al 1997, Theorem 2.1) to have the following asymptotic form

$$\frac{1}{n} \sum_{j=1}^n \frac{d_j}{G(Y_j|\mathbf{X}_j)} [\tau - I\{\varepsilon_j \leq 0\}] \frac{\nabla f(\mathbf{X}_j)}{f_{\varepsilon, \mathbf{x}}(0, \mathbf{X}_j)} + o_p(n^{-1/2}). \quad (22)$$

Observe next that, the smallest eigenvalue of  $\Sigma_n(\mathbf{x})$  is bounded away from zero uniformly over  $\mathbf{x} \in \mathcal{D}$  and that the term inside the square bracket of (21) is uniformly bounded, following from [A5]. Therefore, the term in (21) is essentially a U-statistic plus an asymptotically negligible term, i.e.

$$\frac{\mathbf{A}^\top}{n^2} \left[ \sum_{j,k=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \tilde{Q}(\zeta_j, \zeta_k) \right] \Sigma^{-1}(A_1) \mathbf{e}_1 = \mathbf{A}^\top \frac{U_n}{n(n-1)} \Sigma^{-1}(A_1) \mathbf{e}_1 + o_p(n^{-1/2}),$$

where  $U_n = \sum_{1 \leq j < k \leq n} \xi_n(\mathbf{Z}_j, \mathbf{Z}_k)$ ,  $\mathbf{Z}_j = (\mathbf{X}_j, Y_j)$ ,  $\xi_n(\mathbf{Z}_j, \mathbf{Z}_k) = \eta_n(\mathbf{Z}_j, \mathbf{Z}_k) + \eta_n(\mathbf{Z}_k, \mathbf{Z}_j)$ , and

$$\eta_n(\mathbf{Z}_j, \mathbf{Z}_k) = \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} \tilde{Q}(\zeta_j, \zeta_k).$$

To analyze  $U_n$ , first note that  $E[\xi_n(\mathbf{Z}_j, \mathbf{Z}_k)] = E[\eta_n(\mathbf{Z}_k, \mathbf{Z}_j)] = 0$ . Consider the Hoeffding decomposition of  $U_n$  (see, e.g., Serfling (1980)) and define the projection of  $U_n$  as

$$P_n = (n-1) \sum_{k=1}^n g_n(\mathbf{Z}_k),$$

with  $g_n(\mathbf{Z}_k) = E_j[\xi_n(\mathbf{Z}_k, \mathbf{Z}_j)] = E_j[\eta_n(\mathbf{Z}_k, \mathbf{Z}_j)]$ . We thus have, through arguments similar to that in Chaudhuri et al (1997), that

$$\begin{aligned} E(U_n - P_n)^2 &= \frac{n(n-1)}{2} \{E[\xi_n^2(\mathbf{Z}_k, \mathbf{Z}_j)] - 2E[g_n^2(\mathbf{Z}_k)]\} \\ &\leq \frac{n(n-1)}{2} E[\xi_n^2(\mathbf{Z}_k, \mathbf{Z}_j)] = O(n^2/\delta_n^2) \end{aligned}$$

whence  $U_n = P_n + o_p(n^{3/2})$ . We move on to study  $g_n(\cdot)$ . Noting (32) and (28), i.e.

$$\begin{aligned} \sup_{\mathbf{x} \in R^p} \left| \frac{N_n(\mathbf{x})}{n} - \delta_n^p f(\mathbf{x}) \right| &= o(1) \quad a.e. \\ \Sigma_n^{-1}(\mathbf{x}) &= f_{\varepsilon|\mathbf{X}}^{-1}(0|\mathbf{x}) \Sigma^{-1}(A) + \delta_n \frac{\Sigma^{-1}(A) \sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{x}) \Sigma^{-1}(A)}{f_{\varepsilon|\mathbf{X}}^2(0|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})} + O(\delta_n^{s_3} + \delta_n^2), \end{aligned}$$

we have

$$\delta_n^p \frac{\Sigma_n^{-1}(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} = \frac{1}{n} \frac{\Sigma^{-1}(A)}{f_{\varepsilon, \mathbf{X}}(0, \mathbf{X}_j)} + \frac{\delta_n}{n} \frac{\Sigma^{-1}(A) \sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{X}_j) \Sigma^{-1}(A)}{f_{\varepsilon, \mathbf{X}}^2(0, \mathbf{X}_j)} + O(\delta_n^{s_3} + \delta_n^2),$$

uniformly in  $j = 1, \dots, n$ . Therefore,

$$\begin{aligned} \frac{1}{n^2} P_n &= \frac{1}{n} \sum_{k=1}^n g_n(\mathbf{Z}_k) \\ &= \frac{\Sigma^{-1}(A)}{n \delta_n^{p+1}} \sum_{k=1}^n E_j \left[ f_{\varepsilon, \mathbf{X}}^{-1}(0, \mathbf{X}_j) \tilde{Q}(\zeta_j, \zeta_k) \right] \\ &\quad + \frac{1}{n \delta_n^p} \Sigma^{-1}(A) \left[ \sum_{k=1}^n \int \frac{\sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{X}_j)}{f_{\varepsilon, \mathbf{X}}^2(0, \mathbf{X}_j)} \Sigma^{-1}(A) \tilde{Q}(\zeta_j, \zeta_k) f_{\mathbf{X}}(\mathbf{X}_j) d\mathbf{X}_j \right] \\ &\quad + o_p(n^{-1/2}) \end{aligned} \tag{23}$$

The two leading terms in (23) turn out to be of order  $o_p(n^{-1/2})$ . We only deal with the first term to illustrate. Note that

$$\begin{aligned} &E_j \left\{ f_{\varepsilon, \mathbf{X}}^{-1}(0, \mathbf{X}_j) \tilde{Q}(\zeta_j, \zeta_k) \right\} \\ &= \int \mathbf{X}_{ij}(\delta_n, A) \mathbf{X}_{ki}^\top(h_n, A_1) E_{Y_i | \mathbf{X}_i} [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \xi(Y_k, d_k, Y_i, \mathbf{X}_i) \frac{f(\mathbf{X}_j)}{f_{\varepsilon | \mathbf{X}}(0 | \mathbf{X}_j)} d\mathbf{X}_i d\mathbf{X}_j. \end{aligned}$$

We assume that if regarded as a function of  $(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$ ,

$$E\{[I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \xi(Y_k, d_k, Y_i, \mathbf{X}_i) | \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k\},$$

is continuous with respect of all its arguments, then based on the fact that  $\int \mathbf{t}(A_2) d\mathbf{t} = (1, 0, \dots, 0)$ ,  $\Sigma(A)^{-1} \int \mathbf{t}(A) d\mathbf{t} = (1, 0, \dots, 0)$ , and  $E_k[\xi(Y_k, d_k, Y_i, \mathbf{X}_i)] = 0$ , we have

$$\mathbf{A}^\top \frac{\Sigma^{-1}(A)}{n \delta_n^{p+1}} \sum_{k=1}^n E_j \left[ f_{\varepsilon, \mathbf{X}}^{-1}(0, \mathbf{X}_j) \tilde{Q}(\zeta_j, \zeta_k) \right] \Sigma^{-1}(A_1) \mathbf{e}_1 = o(n^{-1/2}) \text{ a.e.}$$

Hence,

$$\hat{\beta} - \beta = \frac{1}{n} \sum_{j=1}^n \nabla Q(\mathbf{X}_j) - \beta + \frac{1}{n} \sum_{j=1}^n \frac{d_j}{G(Y_j | \mathbf{X}_j)} [\tau - I\{\varepsilon_j \leq 0\}] \frac{\nabla f(\mathbf{X}_j)}{f_{\varepsilon, \mathbf{X}}(0, \mathbf{X}_j)} + o_p(n^{-1/2}).$$

This is exactly the same as that obtained in Chaudhuri et al (1997, Theorem 2.1) in the case of no censoring. Empirical interpretation for this could be that the 'averaging' in the construction of  $\hat{\beta}$  together with the 'polynomial smoothing' used in the local K-M estimator (6) have canceled out the correction term in Theorem 4.3 resulted from the preliminary estimation of the survival function  $G(\cdot)$ .

## 5.2 ADDITIVE QUANTILE REGRESSION MODEL UNDER RANDOM CENSORING

In this section, we apply our main result to derive estimators of the additive quantile-regression model again under random censoring. Specifically we assume an additive structure for the function  $Q(\cdot)$  in model (2), i.e.

$$Q(\mathbf{x}) = Q(x_1, \dots, x_p) = c + Q_1(x_1) + \dots + Q_p(x_p), \quad (24)$$

where  $c$  is an unknown constant and  $Q_k(\cdot)$ ,  $k = 1, \dots, p$  are unknown functions which have been normalized such that  $EQ_k(x_k) = 0$ ,  $k = 1, \dots, p$ . For previous work on additive quantile regression model, see Linton (2001) Yu and Lu (2004) and Horowitz and Lee (2005). Now to estimate the component functions in (24),  $Q_1(\cdot)$  say, we consider the marginal integration method; this involves estimating function  $Q(\cdot)$  and then integrating it over certain directions. Partition  $\mathbf{x}$  as  $(x_1, \mathbf{x}_2)$  where  $x_1$  is the one dimensional direction of interest and  $\mathbf{x}_2$  is a  $p - 1$  dimensional nuisance direction. Accordingly, partition  $\mathbf{X}_i = (X_{i1}, \mathbf{X}_{i2})$ . Define the functional

$$\phi_1(x_1) = \int Q(x_1, \mathbf{x}_2) f_2(\mathbf{x}_2) d\mathbf{x}_2, \quad (25)$$

where  $f_2(\mathbf{x}_2)$  is the joint probability density of  $\mathbf{X}_{2i}$ . Under the additive structure (24), the difference between  $\phi_1(\cdot)$  and  $Q_1(\cdot)$  is a constant. Replace  $Q(\cdot)$  in (25) with  $\hat{\mathbf{c}}_{n1}(x_1, \mathbf{x}_2)$ , the first element of  $\hat{\mathbf{c}}_n(x_1, \mathbf{x}_2)$  defined in (9), with  $\mathbf{X}_j$  replaced by  $(x_1, \mathbf{x}_2)$ , and  $\phi_1(x_1)$  can thus be estimated by the sample version of (25):

$$\phi_{n1}(x_1) = n^{-1} \sum_{j=1}^n \hat{\mathbf{c}}_{n1}(x_1, \mathbf{X}_{j2}).$$

In the case of mean regression, Linton and Härdle (1996) and Hengartner and Sperlich (2005) suggested that for  $\phi_{n1}(\cdot)$  to be asymptotically normal, bandwidth used for the direction of interest  $X_1$  should be different from those for the  $p - 1$  nuisance directions. However, for simplification purposes, we assume that the same bandwidth is used for all directions.

Let  $\mathbf{X}_j^* = (x_1, \mathbf{X}_{j2})$  and  $\mathbf{X}_{ij}^* = \mathbf{X}_i - \mathbf{X}_j^*$ . According to Theorem 4.3, we have

$$\begin{aligned} \hat{\mathbf{c}}_{n1}(\mathbf{X}_j^*) - \mathbf{c}_{n1}(\mathbf{X}_j^*) &= \hat{\mathbf{e}}_1^\top \frac{\Sigma_n^{-1}(\mathbf{X}_j^*)}{N_n(\mathbf{X}_j^*)} \sum_{i \in S_n(\mathbf{X}_j^*)} \frac{d_i}{G(Y_i | \mathbf{X}_i)} \mathbf{X}_{ij}^* (\delta_n, A) \left[ \tau - I \left\{ Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j^*) \right\} \right] \\ &\quad - \hat{\mathbf{e}}_1^\top \frac{1}{n} \sum_{k=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j^*)}{N_n(\mathbf{X}_j^*)} \tilde{Q}(\zeta_j^*, \zeta_k) \mathbf{e}_1^\top \Sigma^{-1}(A_1) + R_n(\mathbf{X}_j^*), \end{aligned}$$

where  $\tilde{\mathbf{e}}_1 = (1, 0, \dots, 0)^\top$  is a  $n(A) \times 1$  vector, and  $\tilde{Q}(\zeta_j^*, \zeta_k)$  is defined similarly to  $\tilde{Q}(\zeta_j^*, \zeta_k)$  with  $\mathbf{X}_j^*$  replacing  $\mathbf{X}_j$  and  $\mathbf{X}_{ij}^*$  replacing  $\mathbf{X}_{ij}$ , which together with the additive structure (24) assumed for  $Q(\cdot)$ , leads to

$$\begin{aligned} \phi_{n1}(x_1) &= \phi_1(x_1) + \frac{1}{n} \sum_{j=1}^n \mathbf{Q}_2(\mathbf{X}_{j2}) + o(n^{-1/2}) \\ &+ \tilde{\mathbf{e}}_1^\top \frac{1}{n} \sum_{j=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j^*)}{N_n(\mathbf{X}_j^*)} \sum_{i \in S_n(\mathbf{X}_j^*)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}^*(\delta_n, A) \left[ \tau - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j^*)\} \right] \\ &- \tilde{\mathbf{e}}_1^\top \frac{1}{n^2} \sum_{j,k} \frac{\Sigma_n^{-1}(\mathbf{X}_j^*)}{N_n(\mathbf{X}_j^*)} \tilde{Q}(\zeta_j^*, \zeta_k) \mathbf{e}_1^\top \Sigma^{-1}(A_1) \end{aligned} \quad (26)$$

$$(27)$$

where  $\mathbf{Q}_2(x_2, \dots, x_p) = Q_1(x_2) + \dots + Q_p(x_p)$ . Note  $\phi_{n1}(x_1)$  is, by definition the average of  $n$ - sub-vectors of  $\hat{\mathbf{c}}_n(\cdot)$ , The only difference being that for  $\phi_{n1}(x_1)$ , the average is taken along the  $p - 1$  nuisance directions, while for ADE (19), the average is taken along all  $p$  directions. Therefore, as in the case of ADE (19), the leading term in (27) is negligible (similar to (21)), and methodologies in Chaudhuri et al (1997) can be used to tackle term (26), as shown in the case of ADE. Specifically, we have

$$\begin{aligned} &\frac{1}{n} \sum_{j=1}^n \frac{\Sigma_n^{-1}(\mathbf{X}_j^*)}{N_n(\mathbf{X}_j^*)} \sum_{i \in S_n(\mathbf{X}_j^*)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}^*(\delta_n, A) \left[ \tau - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j^*)\} \right] \\ &= \frac{\Sigma^{-1}(A)}{n\delta_n} \sum_{i=1}^n \frac{d_i I\{|X_{i1} - x| \leq \delta_n\}}{G(Y_i|\mathbf{X}_i)} \left[ \tau - I\{\varepsilon_i \leq 0\} \right] \left[ \frac{f_2(\mathbf{X}_{i2})}{f_{\varepsilon, \mathbf{X}}(0, x_1, \mathbf{X}_{i2})} \right] \times \\ &\int_{[0,1]^{\otimes(p-1)}} [\delta_n^{-1}(X_{i1} - x), \nu](A) d\nu + o_p(\delta_n^{-1} n^{-1/2}), \end{aligned}$$

where  $X_{i1}$  stands for the first element of  $\mathbf{X}_i$ . This together with the fact that (27) is of order  $O_p(n^{-1/2})$ , leads to

$$\begin{aligned} \phi_{n1}(x_1) &= \phi_1(x_1) + \tilde{\mathbf{e}}_1^\top \frac{\Sigma^{-1}(A)}{n\delta_n} \sum_{i=1}^n \frac{d_i I\{|X_{i1} - x| \leq \delta_n\}}{G(Y_i|\mathbf{X}_i)} \left[ \tau - I\{\varepsilon_i \leq 0\} \right] \left[ \frac{f_2(\mathbf{X}_{i2})}{f_{\varepsilon, \mathbf{X}}(0, x_1, \mathbf{X}_{i2})} \right] \times \\ &\int_{[0,1]^{\otimes(p-1)}} [\delta_n^{-1}(X_{i1} - x), \nu](A) d\nu + o_p\left(\frac{1}{n^{1/2}\delta_n}\right). \end{aligned}$$

Define  $\mathbf{b}(A) = \int_{[0,1]^{\otimes p}} \mathbf{t}(A) dt$ . Asymptotic normality for  $\phi_{n1}(\cdot)$  can thus be established, with mean zero and covariance equal to  $n^{-1} \delta_n^{-1} \tilde{\mathbf{e}}_1^\top \Sigma^{-1}(A) \mathbf{b}(A) \mathbf{b}(A)^\top \Sigma^{-1}(A) \tilde{\mathbf{e}}_1$  multiplied by

$$\int \frac{[\tau - I\{\varepsilon_i \leq 0\}]^2 f_2^2(\mathbf{X}_2)}{G\{Q(x_1, \mathbf{X}_2) + \varepsilon | \mathbf{X} = (x_1, \mathbf{X}_2)\} f_{\varepsilon, \mathbf{X}}(0, x_1, \mathbf{X}_{i2})} d\varepsilon d\mathbf{X}_2$$

To conduct pointwise inference one only needs to estimate the unknown quantities in the asymptotic variance, which is straightforward. For uniform confidence bands, one can proceed as Johnston (1982).

## 6 Discussion

In this paper, we have obtained the Bahadur representation for the local polynomial estimator of a nonparametric quantile regression function. The weighting scheme suggested by Bang and Tsiatis (2002) is adopted to deal with the presence of random censoring. Two examples have been provided to demonstrate the usefulness of the results in establishing the asymptotic properties of estimators. We nevertheless point out that due to the nature of this weighting scheme, information contained in the censored observations is largely lost. It is therefore worthwhile examining other weighting schemes which makes more efficient use of the data, such as those by Portnoy (2003), Peng and Huang (2008) and Wang and Wang (2009). Or to replace  $\rho_\tau(\cdot)$  in (8) in with  $E[\rho_\tau(\cdot)|Y_i, d_i \mathbf{X}_i]$  as considered in Honoré et al (2002) for linear quantile regression under independent censoring. Study and comparison of these alternative methods in the context of nonparametric censored quantile regression will be part of our ongoing research on this subject. The presence of initial consistent estimators with a linear expansion greatly facilitates this work.

## Appendix

**Proposition 6.1** *If  $\delta_n \approx n^{-\kappa}$ , with  $0 < \kappa < 1/p$ , there exists another pair of positive constants  $K_1 < K_2$ , such that  $Pr(\liminf E_n) = 1$ , where*

$$E_n = \{K_1 n^{1-\kappa p} \leq N_n(\mathbf{X}_j) \leq K_2 n^{1-\kappa p}, \text{ for all } j = 1, \dots, n\}.$$

which can be strengthened as

$$\sup_{\mathbf{x} \in R^p} \left| \frac{N_n(\mathbf{x})}{n} - \delta_n^p f(\mathbf{x}) \right| = o(1) \quad a.e. \quad (28)$$

Similarly, we have under [A2] and [A3],

$$\sup_{\mathbf{x} \in \mathcal{D}} \left| \Sigma_n(\mathbf{x}) - f_{\varepsilon|\mathbf{X}}(0|\mathbf{x})\Sigma(A) - \frac{\delta_n}{f_{\mathbf{X}}(\mathbf{x})} \sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{x}) \right| = O(\delta_n^{s_3}), \quad a.e. \quad (29)$$

where  $f_{\varepsilon, \mathbf{X}}$  denotes the joint probability density function of  $(\varepsilon, \mathbf{x})$ ,  $f_{\varepsilon, \mathbf{X}}^{(i)}$ ,  $i = 1, \dots, p$ , its first order partial derivatives, and for each  $1 \leq l \leq p$ ,  $\Sigma_l^*$  is the corresponding  $n(A) \times n(A)$

matrix with a typical entry

$$\sigma_{\mathbf{u}, \mathbf{v}, \mathbf{e}_k} = \int_{[-1, 1]^{\otimes p}} \mathbf{t}^{\mathbf{u} + \mathbf{v} + \mathbf{e}_k} d\mathbf{t}$$

with  $\mathbf{e}_k$  being the  $k$ th column of the  $p \times p$  identity matrix; and under [A1] and [A6] ,

$$\sup_{\mathbf{x} \in \mathcal{D}} |\tilde{\Sigma}_n(\mathbf{x}) - f(\mathbf{x})\Sigma(A_1)| = O\left((nh_n^p/\log n)^{-1/2} + h_n\right) \text{ a.e.} \quad (30)$$

The proof follows directly from application of GlivenkoCantelli Theorem. Using the von Neumann expansion for the inverse matrix, we further have

$$\Sigma_n^{-1}(\mathbf{x}) = f_{\varepsilon|\mathbf{X}}^{-1}(0|\mathbf{x})\Sigma^{-1}(A) + \delta_n \frac{\Sigma^{-1}(A) \sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{x})\Sigma^{-1}(A)}{f_{\varepsilon|\mathbf{X}}^2(0|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})} + O(\delta_n^{s_3} + \delta_n^2), \quad (31)$$

$$\Sigma_n^{-1}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\varepsilon|\mathbf{X}}(0|\mathbf{x})}\Sigma^{-1}(A) + \delta_n \frac{\Sigma^{-1}(A) \sum_{l=1}^p \Sigma_l^* f_{\varepsilon, \mathbf{X}}^{(l)}(0, \mathbf{x})\Sigma^{-1}(A)}{f_{\varepsilon|\mathbf{X}}^2(0|\mathbf{x})} + O(\delta_n^{s_3} + \delta_n^2). \quad (32)$$

**Proof of Lemma 4.1** This follows directly from (30), Theorem 2.1 and Theorem 2.3 of Gonzalez-Manteigaa and Cadarso-Suarez (1994). Note that the fact that the weight  $\tilde{B}_{nj}(\cdot)$  might be negative does not affect the validity of the proof.  $\blacksquare$

We now list a few facts used in the proof. For any  $\mathbf{x} \in \mathcal{D}$ , let  $\omega_{\delta_n}(\cdot, \mathbf{x})$  denote the conditional density of the vector  $\delta_n^{-1}(\mathbf{X} - \mathbf{x})$ , given that  $|\mathbf{X} - \mathbf{x}| \leq \delta_n$ .

[F1] Then under [A1],  $\omega_{\delta_n}(\mathbf{t}, \mathbf{x})$  converges uniformly both in  $\mathbf{t}$  and  $\mathbf{x}$ , to the uniform density on  $[-1, 1]^p$ .

The proof of [F1] is straightforward; also see Chaudhuri(1991b). We now move on to derive the explicit form of  $\hat{\mathbf{c}}_n(\mathbf{x})$ ,  $\mathbf{x} = \mathbf{X}_j$ ,  $1 \leq j \leq n$ .

Let  $\mathbf{T} = \{i : 1 \leq i \leq n, d_i = 1\}$ ,  $DX_n(\mathbf{x})$  be the matrix with rows given by the vectors  $\{\mathbf{X}_{i\mathbf{x}}(\delta_n, A), i \in S_n(\mathbf{x}) \cap \mathbf{T}\}$ , and  $VY_n(\mathbf{x})$  be the corresponding column vector with components  $\{Y_i, i \in S_n(\mathbf{x}) \cap \mathbf{T}\}$ . For any subset  $\mathbf{h} \subset S_n(\mathbf{x}) \cap \mathbf{T}$ , such that  $\sharp(\mathbf{h}) = n(A)$ , denote by  $DX_n(\mathbf{x}, \mathbf{h})$ , the corresponding  $n(A) \times n(A)$  matrix with rows  $\{\mathbf{X}_{i\mathbf{x}}(\delta_n, A), i \in \mathbf{h}\}$ , and by  $VY_n(\mathbf{x}, \mathbf{h})$ , the  $n(A)$  dimensional column vector  $\{Y_i, i \in \mathbf{h}\}$ . Define

$$H_n(\mathbf{x}) = \{\mathbf{h} : \mathbf{h} \subset S_n(\mathbf{x}) \cap \mathbf{T}, \sharp(\mathbf{h}) = n(A), DX_n(\mathbf{x}, \mathbf{h}) \text{ has full rank}\}$$

The following two facts will play a crucial role in the proofs of the Theorems.

[F2] If  $DX_n(\mathbf{x})$  has rank  $n(A)$ , then there is a subset  $\mathbf{h} \in H_n(\mathbf{x})$ , such that (9) has at least one minima of the form

$$\hat{\mathbf{c}}_n(\mathbf{x}) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1} V Y_n(\mathbf{x}, \mathbf{h}).$$

[F3] For the  $\mathbf{h}$  specified in [F2],  $L_n(\mathbf{x}, \mathbf{h}) \in [\tau - 1, \tau]^{n(A)}$  which stands for the  $n(A)$ -dimensional interval in  $R^{n(A)}$ , where

$$\begin{aligned} L_n(\mathbf{x}, \mathbf{h}) &= \sum_{i \in \bar{\mathbf{h}}} d_i \left[ \frac{1}{2} - \frac{1}{2} \text{sign} \left\{ Y_i - \left\langle \mathbf{X}_{i\mathbf{x}}(\delta_n, A), \hat{\mathbf{c}}_n(\mathbf{x}) \right\rangle \right\} - \tau \right] \\ &\quad \times \left\{ \hat{G}_n(Y_i | \mathbf{X}_i) \right\}^{-1} \mathbf{X}_{i\mathbf{x}}(\delta_n, A) \left[ W_n(h) DX_n(\mathbf{x}, \mathbf{h}) \right]^{-1}, \end{aligned}$$

where  $\bar{\mathbf{h}} = S_n(\mathbf{x}) \setminus \mathbf{h}$  denotes its complement in  $S_n(\mathbf{x})$ ,  $\text{sign}(a)$  is  $+1, 0$ , or  $-1$  depending on whether  $x$  is positive, zero or negative, and  $W_n(\mathbf{h})$  is the diagonal matrix with elements  $\{\hat{G}_n(Y_i | \mathbf{X}_i), i \in \mathbf{h}\}$ . Moreover,  $\hat{\mathbf{c}}_n(\mathbf{x})$  is the unique minima of (9) iff  $L_n(\mathbf{x}, \mathbf{h}) \in (\tau - 1, \tau)^{n(A)}$ .

**Remark** Noticing the linearity of the loss function  $\rho_\tau(\cdot)$ , [F2] and [F3] can be proved in exactly the same manner as Theorem 3.1 and 3.3 in Koenker and Bassett (1978); see Chaudhuri (1991a) for parallel results. Note that the form of  $\hat{\mathbf{c}}_n(\mathbf{x})$  specified in [F2] is free from the K-M estimator  $\hat{G}_n(\cdot)$ , and appears to be identical to the minimizer of,

$$\min_{\mathbf{c}} \sum_{i \in S_n(\mathbf{x})} \rho_\tau\{Y_i - P_n(\mathbf{c}, \mathbf{x}, \mathbf{X}_i)\},$$

which is another version of (9) with equal weights. They are, however, distinct, since the subsets  $\mathbf{h}$  they are related to are usually different. This can be seen from the [F3], the necessary and sufficient condition  $\mathbf{h}$  has to satisfy, which does involves  $\hat{G}_n(\cdot)$ , and thus is different from Fact 6.4 in Chaudhuri (1991a). For illustration purposes, consider a simple example, where we have only two observations  $\{Y_1, Y_2\}$ , with  $Y_1 < Y_2$ , then the solution set to the minimization problem  $\min_y \{|Y_1 - y| + |y - Y_2|\}$  with equal weights is  $[Y_1, Y_2]$ . However, the weighted minimization problem  $\min_y \{a_1 |Y_1 - y| + a_2 |y - Y_2|\}$  for some positive  $a_1 \neq a_2$ , has a unique solution,  $Y_1$ , if  $a_1 > a_2$ , and  $Y_2$ , if  $a_1 < a_2$ . Therefore, the two solutions sets may overlap, but they usually do not coincide.

Under [A2], we have for any  $\mathbf{x} \in D$ ,  $k = [s_2]$ , all sufficiently large  $n$ , and any bounded  $\mathbf{t} \in [-1, 1]^{\otimes p}$ ,  $Q(\mathbf{x} + \mathbf{t}\delta_n)$  can be approximated by the  $k$ -th order Taylor polynomial

$$Q_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x}) = \sum_{\mathbf{u} \in A} c_{n,\mathbf{u}}(\mathbf{x}) \mathbf{t}^{\mathbf{u}} = \langle \mathbf{c}_n(\mathbf{x}), \mathbf{t} \rangle, \quad (33)$$

and the remainder  $r(\mathbf{t}\delta_n, \mathbf{x}) = Q(\mathbf{x} + \mathbf{t}\delta_n) - Q_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x})$  satisfies

$$|r(\mathbf{t}\delta_n, \mathbf{x})| \leq C(|\mathbf{t}|\delta_n)^{s_2}, \quad (34)$$

uniformly over  $\mathbf{t} \in [-1, 1]^{\otimes p}$  and  $\mathbf{x} \in D$ . Define

$$\hat{Q}_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x}) = \langle \hat{\mathbf{c}}_n(\mathbf{x}), \mathbf{t} \rangle. \quad (35)$$

**Proof of Theorem 4.2.** For any positive constant  $K_1$  and a generic  $\mathbf{x} \in R^p$ , which stands for any one of  $\mathbf{X}_j, j = 1, \dots, n$ , let  $U_n$  be the event defined as

$$U_n(\mathbf{x}) = \{|\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})| \geq K_1[n\delta_n^p/\log n]^{-1/2}\}.$$

According to the Borel-Cantelli lemma, the assertion in Theorem 4.2 will follow, if there exists some  $K_1 > 0$ , such that

$$\sum_n nP(U_n(\mathbf{x})) < \infty.$$

To obtain an uniform upper bound for  $P(U_n(\mathbf{x}))$ , for any given vector  $\Delta_n \in R^{n(A)}$ , set

$$Z_{ni}(\mathbf{x}) = \left[ \frac{1}{2} - \frac{1}{2} \text{sign} \left\{ \varepsilon_i - \langle \mathbf{X}_{i\mathbf{x}}(\delta_n, A), \Delta_n \rangle + r_n(\mathbf{X}_{i\mathbf{x}}, \mathbf{x}) \right\} - \tau \right] \mathbf{X}_{i\mathbf{x}}(\delta_n, A), \quad (36)$$

where  $r_n(\mathbf{X}_{i\mathbf{x}}, \mathbf{x})$  is the remainder from the Taylor expansion (33). Using results (10) on the strong uniform consistency of K-M estimator, i.e.

$$\sup_{\mathbf{x}} \sup_{t \leq \tau(\mathbf{x})} |\hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x})| = O\left(\left(\frac{\log n}{nh_n^p}\right)^{1/2}\right)$$

we have  $W_n(\mathbf{h}) = W(\mathbf{h}) + o(1)$  a.e., where  $W(\mathbf{h})$  is the diagonal matrix with elements  $\{G(Y_i|\mathbf{X}_i), i \in \mathbf{h}\}$ . Consequently, the assertion in [F3] that  $L_n(\mathbf{x}, \mathbf{h}) \in (\tau - 1, \tau)^{d+1}$  implies that there exists some constant  $\phi_1 > 0$ , which depends on  $n(A)$ , such that  $|L_{n1}(\mathbf{x}, \mathbf{h}) + L_{n2}(\mathbf{x}, \mathbf{h})| \leq \phi_1$ , where

$$\begin{aligned} L_{n1}(\mathbf{x}, \mathbf{h}) &= \sum_{i \in \mathbf{h}} \{G(Y_i|\mathbf{X}_i)\}^{-1} Z_{ni}(\mathbf{x}) d_i, \\ L_{n2}(\mathbf{x}, \mathbf{h}) &= \sum_{i \in \mathbf{h}} \frac{G(Y_i|\mathbf{X}_i) - \hat{G}_n(Y_i|\mathbf{X}_i)}{G(Y_i|\mathbf{X}_i) \hat{G}_n(Y_i|\mathbf{X}_i)} Z_{ni}(\mathbf{x}) d_i, \end{aligned}$$

where  $Z_{ni}(\mathbf{x})$  is defined as in (36) with  $\Delta_n = \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})$ . As  $E[d_i|\mathbf{X}_i, Y_i] = G(Y_i|\mathbf{X}_i)$ , we have in parallel to FACT 6.5 in Chaudhuri (1991a) that there exists positive constants  $\epsilon_1^*, \epsilon_2^*, c_5^*$  and  $M_2^*$ , such that

$$\left| E \left[ Z_{ni}(\mathbf{x}) d_i / G(Y_i|\mathbf{X}_i) \right] \right| \geq \min\{\epsilon_1^*, c_5^* |\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})|\},$$

whenever  $|r_n(\mathbf{X}_{i\mathbf{x}}, \mathbf{x})| \leq \epsilon_2^*$  and  $|\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})| \geq M_3^* |r_n(\mathbf{X}_{i\mathbf{x}}, \mathbf{x})|$ , where  $M_3^* \geq M_2^*$ . Therefore, if event  $U_n$  is true, i.e.  $|\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})| \geq K_1 [n\delta_n^p / \log n]^{-1/2}$ , for some positive  $K_1$ , we have from  $r_n(\mathbf{X}_{i\mathbf{x}}, \mathbf{x}) = O(|\delta_n|^{s_3}) = o([n\delta_n^p / \log n]^{-1/2})$ , for  $\kappa > 1/(2s_3 + d)$ , that there exists some constant  $c_5 > 0$ , such that

$$\left| E \left[ Z_{ni}(\mathbf{x}) d_i / G(Y_i | \mathbf{X}_i) \right] \right| \geq c_5 [n\delta_n^p / \log n]^{-1/2}.$$

This combined with the facts that  $|L_{n1}(\mathbf{x}, \mathbf{h}) + L_{n2}(\mathbf{x}, \mathbf{h})| \leq \phi_1$ , for some  $\phi_1 > 0$ ,  $\sharp(\bar{\mathbf{h}}) = O(n\delta_n^p)$  (Proposition 6.1), and  $\sup_{\mathbf{x} \in \mathcal{D}} L_{n2}(\mathbf{x}, \mathbf{h}) = O\{(n\delta_n^p \log n)^{1/2}\}$  a.e. which follows from (10), leads to the conclusion that there exists some  $K_1^* > 0$ , such that  $U_n(\mathbf{x})$  is contained in the event

$$\left\{ \text{for some } \mathbf{h} \in H_n(\mathbf{x}), \left| \sum_{i \in \bar{\mathbf{h}}} \{Z_{ni}(\mathbf{x}) d_i / G(Y_i | \mathbf{X}_i) - E[Z_{ni}(\mathbf{x}) d_i / G(Y_i | \mathbf{X}_i)]\} \right| \geq K_1^* [n\delta_n^p \log n]^{1/2}, \right. \\ \left. \text{with } \Delta_n = \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x}), \hat{\mathbf{c}}_n(\mathbf{x}) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1} V Y_n(\mathbf{x}, \mathbf{h}), \text{ and } |\Delta_n| \geq K_1 [n\delta_n^p / \log n]^{-1/2} \right\}.$$

Apply Bernstein's inequality to  $\sum_{i \in \bar{\mathbf{h}}} Z_{ni}(\mathbf{x}) d_i / G(Y_i | \mathbf{X}_i)$ , we have by noting that  $\sharp(H_n(\mathbf{x})) = O\{(n\delta_n^p)^{n(A)}\}$ , and that  $Z_{ni}(\mathbf{x}) d_i / G(Y_i | \mathbf{X}_i)$  is bounded, there exist constants  $c_6 > 0, c_7 > 0$  and an integer  $N_1 > 0$ , such that

$$P(U_n(\mathbf{x})) \leq c_6 (n\delta_n^p)^{n(A)} \exp(-c_7 \log n) \quad (37)$$

uniformly in  $\mathbf{x} = \mathbf{X}_1, \dots, \mathbf{X}_n$ . By letting  $K_1$ , thus  $K_1^*$  sufficiently large, we indeed have  $\sum_n nP(U_n(\mathbf{x})) < \infty$ . ■

**Proof of Theorem 4.3.** Again, here the generic  $\mathbf{x} \in R^p$  should be interpreted as any of the  $\mathbf{X}_j, j = 1, \dots, n$ . The proof consists of the following steps

*Step 1:* Define

$$\begin{aligned}
\tilde{H}_n(\mathbf{x}, \delta_n, \mathbf{c}_n(\mathbf{x})) &= \int_{[-1,1]^p} F_\varepsilon\{\langle \mathbf{c}_n(\mathbf{x}), \mathbf{t}(A) \rangle - Q(\mathbf{x} + \mathbf{t}\delta_n)\} \mathbf{t}(A) \omega_{\delta_n}(\mathbf{t}, \mathbf{x}) d\mathbf{t} \\
&= \int_{[-1,1]^p} F_\varepsilon\{Q_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x}) - Q(\mathbf{x} + \mathbf{t}\delta_n)\} \mathbf{t}(A) \omega_{\delta_n}(\mathbf{t}, \mathbf{x}) d\mathbf{t} \\
&= \int_{[-1,1]^p} F_\varepsilon\{r(\mathbf{t}\delta_n, \mathbf{x})\} \mathbf{t}(A) \omega_{\delta_n}(\mathbf{t}, \mathbf{x}) d\mathbf{t} \\
\tilde{H}_n(\mathbf{x}, \delta_n, \hat{\mathbf{c}}_n(\mathbf{x})) &= \int_{[-1,1]^p} F_\varepsilon\{\langle \hat{\mathbf{c}}_n(\mathbf{x}), \mathbf{t}(A) \rangle - Q(\mathbf{x} + \mathbf{t}\delta_n)\} \mathbf{t}(A) \omega_{\delta_n}(\mathbf{t}, \mathbf{x}) d\mathbf{t} \\
&= \int_{[-1,1]^p} F_\varepsilon\{\hat{Q}_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x}) - Q(\mathbf{x} + \mathbf{t}\delta_n)\} \mathbf{t}(A) \omega_{\delta_n}(\mathbf{t}, \mathbf{x}) d\mathbf{t},
\end{aligned}$$

and

$$R_n^{(1)}(\mathbf{x}) = \tilde{H}_n(\mathbf{x}, \delta_n, \hat{\mathbf{c}}_n(\mathbf{x})) - \tilde{H}_n(\mathbf{x}, \delta_n, \mathbf{c}_n(\mathbf{x})) - \Sigma_n(\mathbf{x})[\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})].$$

Then as shown in *Step 1* on page 773 of Chaudhuri (1991a), that by Theorem 4.2, [A1] and [A2] we have

$$\sup_j |R_n^{(1)}(\mathbf{X}_j)| = O\{[n^{(1-\kappa p)} / \log n]^{-3/4}\} \quad (38)$$

almost surely.

*Step 2:* Define the  $n(A)$ -dimensional random vector  $\chi_n(\mathbf{x})$  as

$$\begin{aligned}
\chi_n(\mathbf{x}) &= \sum_{i \in S_n(\mathbf{x})} \left[ \frac{d_i}{G(Y_i | \mathbf{X}_i)} \mathbf{X}_{ix}(\delta_n, A) I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{x})\} - \tilde{H}_n(\mathbf{x}, \delta_n, \hat{\mathbf{c}}_n(\mathbf{x})) \right] \\
&\quad - \sum_{i \in S_n(\mathbf{x})} \left[ \frac{d_i}{G(Y_i | \mathbf{X}_i)} \mathbf{X}_{ix}(\delta_n, A) I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{x})\} - \tilde{H}_n(\mathbf{x}, \delta_n, \mathbf{c}_n(\mathbf{x})) \right],
\end{aligned}$$

and for some constant  $K_3 > 0$ , the corresponding event

$$W_n(\mathbf{x}) = \left\{ |\chi_n(\mathbf{x})| \geq K_3 [\log n]^{3/4} n^{(1-\kappa p)/4} \right\}$$

Also for  $\mathbf{h} \in H_n(\mathbf{x})$ , and large enough  $n$ , define

$$\begin{aligned}
\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}) &= [DX_n(\mathbf{x}, \mathbf{h})]^{-1} VY_n(\mathbf{x}, \mathbf{h}), \quad \hat{Q}_n^{\mathbf{h}}(\mathbf{X}_i, \mathbf{x}) = \langle \hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}), \mathbf{X}_{ix}(\delta_n, A) \rangle, \\
\chi_n^{\mathbf{h}}(\mathbf{x}) &= \sum_{i \in \bar{\mathbf{h}}} \left[ \frac{d_i}{G(Y_i | \mathbf{X}_i)} \mathbf{X}_{ix}(\delta_n, A) I\{Y_i \leq \hat{Q}_n^{\mathbf{h}}(\mathbf{X}_i, \mathbf{x})\} - \tilde{H}_n(\mathbf{x}, \delta_n, \hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x})) \right] \\
&\quad - \sum_{i \in \bar{\mathbf{h}}} \left[ \frac{d_i}{G(Y_i | \mathbf{X}_i)} \mathbf{X}_{ix}(\delta_n, A) I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{x})\} - \tilde{H}_n(\mathbf{x}, \delta_n, \mathbf{c}_n(\mathbf{x})) \right].
\end{aligned}$$

Then in view of definition of the events  $A_n$  (i.e. unique solution),  $U_n(\mathbf{x})$  and [F2], the event  $W_n(\mathbf{x}) \cap A_n \cap \overline{U_n(\mathbf{x})}$  is contained in the event

$$\left\{ \begin{array}{l} \text{for some } \mathbf{h} \in H_n(\mathbf{x}), |\chi_n^{\mathbf{h}}(\mathbf{x})| \geq K_4 [\log n]^{3/4} n^{(1-\kappa p)/4} \\ \text{and } |\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})| \leq K_1 [n^{(1-\kappa p)} / \log n]^{-1/2} \end{array} \right\} \cap A_n$$

for large enough  $n$ , where  $K_4 = K_3/2$  and we have implicitly used the fact that  $[\log n]^{3/4} n^{(1-\kappa p)/4} \rightarrow \infty$  and that  $\sharp(\mathbf{h}) = p$ . As argued in Chaudhuri (1991a), given the set  $S_n(\mathbf{x})$ ,  $\mathbf{h} \in H_n$ , and the set of  $\{(\mathbf{X}_i, Y_i) : i \in \mathbf{h}\}$ , the terms in the sum defining  $\chi_n^{\mathbf{h}}(\mathbf{x})$  are IID with mean 0, and variance-covariance matrix with Euclidean norm of the same order as  $|\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})|$ , which is  $O([n^{(1-\kappa p)} / \log n]^{-1/2})$ . This result follows from the fact that the presence of the indicator function  $I(\cdot)$  in the definition of  $\chi_n^{\mathbf{h}}(\mathbf{x})$  causes the terms in the sums acts in a similar way as a random vector with Binomial components. As  $G(\cdot)$  is abounded away from zero, an application of the Bernstein's inequality to the sum defining  $\chi_n^{\mathbf{h}}(\mathbf{x})$  yields a result similar to (37), i.e. there exist constant  $c_8 > 0, c_9 > 0$ , such that

$$P(W_n(\mathbf{x}) \cap A_n \cap \overline{U_n(\mathbf{x})}) \leq c_8 n^{(1-\kappa p)n(A)} \exp(-c_9 \log n) = o(n^{-2}),$$

by choosing  $K_3$ , hence  $c_9$  sufficiently large. Therefore, we have

$$\sup_j |\chi_n^{\mathbf{h}}(\mathbf{X}_j)| = O([\log n]^{3/4} n^{(1-\kappa p)/4}) \quad (39)$$

*Step 3:* Combining (38) and (39), we have

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \\ &= \frac{1}{N_n(\mathbf{X}_j)} \chi_n^{\mathbf{h}}(\mathbf{X}_j) + \tilde{H}_n(\mathbf{X}_j, \delta_n, \hat{\mathbf{c}}_n(\mathbf{X}_j)) - \tilde{H}_n(\mathbf{X}_j, \delta_n, \mathbf{c}_n(\mathbf{X}_j)) \\ & \quad - \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \\ &= O([n^{(1-\kappa p)} / \log n]^{-3/4}) + \Sigma_n(\mathbf{X}_j) [\hat{\mathbf{c}}_n(\mathbf{X}_j) - \mathbf{c}_n(\mathbf{X}_j)] \\ & \quad + \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \quad (40) \\ & \quad + \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} d_i \left[ \frac{1}{G(Y_i|\mathbf{X}_i)} - \frac{1}{\hat{G}_n(Y_i|\mathbf{X}_i)} \right] \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \end{aligned}$$

uniformly for  $\mathbf{x} = \mathbf{X}_j$ ,  $j = 1, \dots, n$ . Note that according to [F3], term (40) is of order  $O(n^{\kappa p - 1})$ , and is thus negligible. The rest of the proof is left as Lemma 6.2.  $\blacksquare$

**Lemma 6.2** Let  $\gamma_n = \log n/n^{1-\kappa p}$ . Then, with probability 1,

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\}}{G(Y_i|\mathbf{X}_i)\hat{G}_n(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \\ &= \frac{1}{N_n(\mathbf{X}_j)} \sum_{k=1}^n E_i \mathcal{F}_{\mathbf{X}_j}(\zeta_i, \zeta_k) + O([n^{1-\kappa p}/\log n]^{-3/4}) + o(n^{-1/2}) \end{aligned}$$

uniformly in  $j = 1, \dots, n$ , where

$$\begin{aligned} \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) &= \frac{d_i I\{\mathbf{X}_i \in S_n(\mathbf{x})\} \mathbf{X}_{i\mathbf{x}}(\delta_n, A)}{f_{\mathbf{X}}(\mathbf{X}_i) G(Y_i|\mathbf{X}_i)} [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{x})\} - \tau] \\ &\quad \times \mathbf{e}_1^\top \Sigma^{-1}(A_1) \tilde{B}_{h_n}(\mathbf{X}_{ki}) q(Y_k, Y_i, \mathbf{X}_i), \end{aligned}$$

and  $E_i(\cdot)$  stands for expectation taken with respect to the joint distribution of  $(\mathbf{X}_i, Y_i)$  with the other argument held fixed.

**Proof.** The proof consists of the following steps

*Step 1* According to (10), [A6], and the facts that  $G(\cdot)$  is bounded below from zero, and  $|\mathbf{X}_{ij}(\delta_n, A)| \leq 1$  for all  $i \in S_n(\mathbf{X}_j)$ , it is obvious that

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\}}{G(Y_i|\mathbf{X}_i)\hat{G}_n(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \\ &= \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i|\mathbf{X}_i)} [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\} \\ &\quad + o(n^{-1/2}), \end{aligned} \tag{41}$$

*Step 2:* For the leading term in (41), replace  $I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\}$  with  $I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\}$ , the resulting remainder, as shown in Lemma 6.3, is of order  $O(\gamma_n^{3/4})$ , i.e.

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i|\mathbf{X}_i)} [I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\} \\ &= \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i|\mathbf{X}_i)} [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\} \\ &\quad + O(\gamma_n^{3/4}) \end{aligned}$$

uniformly in  $j = 1, \dots, n$ .

*Step 3:* Using the result in Lemma 4.1, under [A7], with probability one,

$$\hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x}) = \frac{\varepsilon_1^\top \Sigma^{-1}(\mathbf{x})}{n f_X(\mathbf{x})} \sum_{j=1}^n \tilde{K}_{h_n}(\mathbf{X}_{j\mathbf{x}}) \xi(Y_j, d_j, t, \mathbf{x}) + o(n^{-1/2}),$$

uniformly in  $t$  and  $\mathbf{x}$ , we have

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i | \mathbf{X}_i)} [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \{\hat{G}_n(Y_i | \mathbf{X}_i) - G(Y_i | \mathbf{X}_i)\} \\ &= \frac{1}{nN_n(\mathbf{X}_j)} \sum_{i,k=1}^n \mathcal{F}_{\mathbf{X}_j}(\zeta_i, \zeta_k) + o(n^{-1/2}), \end{aligned} \quad (42)$$

uniformly in  $\mathbf{X}_j$ ,  $j = 1, \dots, n$ .

*Step 4:* Built on *Step 3*, it remains to show that

$$\frac{1}{nN_n(\mathbf{x})} \sum_{i,k=1}^n \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) = \frac{1}{N_n(\mathbf{x})} \sum_{k=1}^n E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)] + o([\log n/N_n]^\alpha + n^{-1/2}),$$

uniformly in  $\mathbf{x} \in \mathcal{D}$ .

Firstly, from the definition of  $\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)$  and noting that the 'own observation' terms are asymptotically negligible, we know the leading term on the right hand side of (43) can be written a U-statistic plus an asymptotically negligible term:

$$\frac{1}{n(n-1)} \sum_{i \neq k}^n \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) = \frac{1}{2n(n-1)} \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k), \quad (43)$$

where  $\mathcal{H}_{\mathbf{x}}(\cdot, \cdot)$  is a symmetric function defined as

$$\mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) = \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) + \mathcal{F}_{\mathbf{x}}(\zeta_k, \zeta_i).$$

Consider the Hoeffding decomposition of  $\mathcal{H}_{\mathbf{x}}(\cdot, \cdot)$

$$\mathcal{H}_{\mathbf{x}}^0(\zeta_i, \zeta_k) = \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) - E_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) - E_k \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) + E \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k),$$

where  $E_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k)$  standing for taking expectation w.r.t  $\zeta_i$  with  $\zeta_k$  held fixed. Since

$$\begin{aligned} E_k \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) &= E_k[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) + \mathcal{F}_{\mathbf{x}}(\zeta_k, \zeta_i)] = E_k[\mathcal{F}_{\mathbf{x}}(\zeta_k, \zeta_i)] \\ E_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) &= E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)], \quad E \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) = 0, \end{aligned}$$

We thus have

$$\begin{aligned} \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) &= \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}^0(\zeta_i, \zeta_k) + \sum_{i \neq k} E_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) + \sum_{i \neq k} E_k \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) - \sum_{i \neq k} E \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) \\ &= 2(n-1) \sum_{k=1}^n E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)] + \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}^0(\zeta_i, \zeta_k) \end{aligned} \quad (44)$$

For the third term, to apply Proposition 4 in Arcones (1995), we need to verify that the class of functions  $\{\mathcal{H}_{\mathbf{x}}^0(\cdot, \cdot) : \mathbf{x} \in \mathcal{D}\}$  is Euclidean with constant envelope, referred to as the uniformly bounded VC subgraph class in Arcones (1995). This is because, first of all, the class of functions  $\{\mathcal{F}_{\mathbf{x}}(\cdot, \cdot) : \mathbf{x} \in \mathcal{D}\}$  is uniformly bounded (CONDITION 7). Secondly, as  $\Sigma^{-1}\tilde{B}_{h_n}(\mathbf{X}_{ki})q(Y_k, Y_i, \mathbf{X}_i)$  is independent of  $\mathbf{x}$ , we note from Lemma 2.14 (i) and (ii) in Pakes and Pollard (1989) that it suffices to show the Euclidean property for the two classes (a)  $(I\{\mathbf{X}_i \in \mathcal{S}_n(\mathbf{x})\}\mathbf{X}_{ix}(\delta_n, A) : \mathbf{x} \in \mathcal{D})$ , (b)  $(I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{x})\} : \mathbf{x} \in \mathcal{D})$ . This is indeed true for the envelope  $F \equiv 1$ , following directly from Lemma 22(ii) in Nolan and Pollard (1987) as  $I(\cdot)$  is of bounded variation. Therefore, according to Proposition 4 in Arcones (1995), there exists some constant  $c_0 > 0$ , such that for any  $\epsilon > 0$  and  $1 > \alpha > 0$ ,

$$Pr\left\{\max_{\mathbf{x} \in \mathcal{D}} \left| \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}^0(\zeta_i, \zeta_k) \right| \geq \epsilon N_n n (\log n / N_n)^\alpha \right\} < 2 \exp(-c_0 N_n^{1-\alpha} \log^\alpha n) = o(n^{-2}).$$

By the Borel-Cantelli lemma, we have with probability one,

$$\max_{\mathbf{x} \in \mathcal{D}} \frac{1}{nN_n} \left| \sum_{i \neq k} \mathcal{H}_{\mathbf{x}}^0(\zeta_i, \zeta_k) \right| = o([\log n / N_n]^\alpha), \text{ for any } \alpha < 1.$$

This together with (43) and (44) leads to

$$\begin{aligned} & \frac{1}{nN_n(\mathbf{x})} \sum_{i,k=1}^n \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) \\ &= \frac{1}{n} \sum_{k=1}^n E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)] + o([\log n / N_n]^\alpha + n^{-1/2}) \end{aligned}$$

almost surely, for any  $\alpha < 1$  with the  $o(\cdot)$  uniform in  $\mathbf{x} \in \mathcal{D}$ . Moreover, noting (30), it is straightforward to check that  $E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)]$  does coincide with the second leading term in Theorem 4.3. This completes the proof.  $\blacksquare$

**Lemma 6.3** *Under conditions assumed for Theorem 4.3, we have*

$$\begin{aligned} & \sum_{i \in \mathcal{S}_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i | \mathbf{X}_i)} \left[ I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} \right] \{ \hat{G}_n(Y_i | \mathbf{X}_i) - G(Y_i | \mathbf{X}_i) \} \\ &= O(n^{1-\kappa p} \gamma_n^{3/4}), \end{aligned} \tag{45}$$

uniformly in  $j = 1, \dots, n$ .

**Proof** Based on (10), and the facts that  $|\mathbf{X}_{ij}(\delta_n, A)| \leq 1$ ,  $G(\cdot)$  is bounded away from zero, the term in (45) is bounded by  $O\{(nh_n^p/\log n)^{1/2}\}$  multiplied by

$$\begin{aligned} & \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \left| \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i | \mathbf{X}_i)} I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} \right| \\ & \leq \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} d_i |I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\} - I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\}| \\ & \leq \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} I\{\varepsilon_i \in I_{ni}(\mathbf{X}_j)\} \leq \frac{1}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} I\{\varepsilon_i \in D_n\}, \end{aligned}$$

where

$$I_{ni}(\mathbf{x}) = \left[ r_n(\mathbf{X}_{ix}, \mathbf{x}) - |\langle \mathbf{X}_{ix}(\delta_n, A), \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x}) \rangle|, r_n(\mathbf{X}_{ix}, \mathbf{x}) + |\langle \mathbf{X}_{ix}(\delta_n, A), \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x}) \rangle| \right],$$

and  $D_n = [-K_1 \gamma_n^{1/2}, K_1 \gamma_n^{1/2}]$ , for some  $K_1 > 0$ , and the last equality follows from (34), Theorem 4.2 and the fact that  $\delta_n^{s_3} = o(\gamma_n^{1/2})$ .

As  $E I\{\varepsilon_i \in D_n\} = O(\gamma_n^{1/2}) = o\{(nh_n^p/\log n)^{1/2} \gamma_n^{3/4}\}$ , obviously (45) will follow if we can show that

$$\sup_j \sum_{i \in S_n(\mathbf{X}_j)} \left[ I\{\varepsilon_i \in D_n\} - E[I\{\varepsilon_i \in D_n\}] \right] = O\{(nh_n^p \log n)^{1/2} \gamma_n^{-1/4}\}. \quad (46)$$

To this aim, for any positive constant  $K_2$ , and  $\mathbf{x} \in R^p$ , define

$$U_n(\mathbf{x}) = \left\{ \sum_{i \in S_n(\mathbf{x})} I\{\varepsilon_i \in D_n\} - E[I\{\varepsilon_i \in D_n\}] \geq K_2 \{(nh_n^p \log n)^{1/2} \gamma_n^{-1/4}\} \right\}.$$

Applying Bernstein's inequality, we have

$$P(U_n(\mathbf{x})) \leq 2 \exp\left\{ - \frac{K_2^2 n h_n^p \log n / \gamma_n^{1/2}}{4n^{1-\kappa p} \gamma_n^{1/2} + 2K_2 (n h_n^p \log n)^{1/2} / \gamma_n^{1/4}} \right\} = o(n^{-2}),$$

i.e.  $\sum n P(U_n(\mathbf{x})) < \infty$ , which, according to the Borel-Cantelli lemma, leads to (46).  $\blacksquare$

## REFERENCES

- Arcones, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statistics and Probability Letters* **22**, 239-247.
- Babu, G. j. (1989). Strong representations for LAD estimators in linear models. *Probability Theory and Related Fields* **83**, 547-558.

- Bahadur, R.R. (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics* **37**, 577-580.
- Buckley, J. and I. James (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.
- Bang, H. and Tsiatis, A. A. (2002) Median regression with censored cost Data. *Biometrics* **58**, 643-649.
- Belloni, A., Chernozhukov, V. and Fernández-Val, I. (2011) Conditional quantile processes based on series or many regressors. CEMMAP working paper cwp1911.
- Chaudhuri, P. (1991a). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics* **19**, 760-777.
- Chaudhuri, P. (1991b). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* **39**, 246-269.
- Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics* **25**, 715-744.
- Chen, S. Dahl, G.B. and Khan, S. (2005) Nonparametric identification and estimation of a censored location-regression model. *Journal of the American Statistical Association* **100** 212-221.
- Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics* **14**, 181-192.
- Dabrowska, D. M. (1992). Nonparametric quantile regression with censored data. *Sankhyā Ser. A* **54**, 252-259.
- De Gooijer, J. G. and D. Zerom (2003). On additive conditional quantiles with high dimensional covariates. *Journal of the American Statistical Association* **98**, 135-146.
- de Uña-Álvarez, J. and Roca-Pardiñasa, J. (2009). Additive models in censored regression. *Computational Statistics and Data Analysis* **53**, 3490-3501.
- Efron, B. (1967). The two-sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics, IV*, eds. Le Cam, L. and Neyman, J. Prentice-Hall: New York, 831-853.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall: London.
- Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics* **4**, 65-78.

- Guerre, E., and Sabbah, C (2009) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function, September 2009, 33p. Forthcoming, *Econometric Theory*. <http://arxiv.org/abs/1105.5038>
- Härdle, W. and T.M. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of American Statistical Association* **84**, 986-995.
- He, X. and Shao, Q. M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics* **24**, 2608-2630.
- Heuchenne, C. and I. Van Keilegom (2007a). Nonlinear regression with censored data. *Technometrics* **49**, 34-44.
- Heuchenne, C. and I. Van Keilegom (2007b). Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis* **98**, 1558-1582.
- Heuchenne, C. and I. Van Keilegom (2008). Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics* (to appear).
- Huber, P. J. (1981). *Robust Statistics*. Wiley: New York.
- Honoré, B., Khan, S. and Powell, J. (2002) Quantile regression under random censoring. *Journal of Econometrics* **109**, 67-105.
- Johnston, G. (1982) Probabilities of maximal deviations of nonparametric regression function estimates. *Journal of Multivariate Analysis* **12**, 402-414.
- Jones, L. K. (1987). On a conjecture of Huber concerning the convergence of projection pursuit regression. *The Annals of Statistics* **15**, 880-882.
- Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. John and Wiley: New York.
- Kim, M. O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics* **35**, 92-108.
- Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387-421.
- Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.
- Koenker, R. and Biliyas, Y. (2001). Quantile regression for duration data: a reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics* **26**, 199-220.

- Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* **26**, 1529-1564.
- Koul, H., Susarla, V. and J. Van Ryzin (1981). Regression analysis with randomly right censored data. *The Annals of Statistics* **42**, 1276-1288.
- Lewbel, A. and Linton, O. (2002). Nonparametric censored and truncated regression. *Econometrica* **70**, 765-779.
- Linton, O. (2001). Estimating additive nonparametric models by partial Lq Norm: The Curse of Fractionality. *Econometric Theory* **17**, 1037-1050.
- Linton, O. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika* **83**, 529-540.
- Linton, O., Mammen, E., Nielsen, J. P. and Van Keilegom, I. (2011). Nonparametric regression with filtered Data. *Bernoulli* Forthcoming
- Lo, S. H. and Singh, K. (1983). The product-limit estimator and the bootstrap: some asymptotic representations *Probability Theory and Related Fields* **71**, 455-465.
- Lu X. and Cheng, T. L. (2007). Randomly censored partially linear single-index models. *Journal of Multivariate Analysis* **98**, 1895-1922
- Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields* **61**, 405-415.
- Masry, E. (1996) Multivariate local polynomial regression for time series: uniform strong convergence and rates. *Journal of Time Series Analysis* **17**, 571-599.
- Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *The Annals of Statistics* **15**, 780-799.
- Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.
- Patilea, V. (2011). Inference in conditional moment restrictions models in the presence of censoring on the dependent variable. *Manuscript*.
- Portnoy, S. (1997). Local asymptotics for quantile smoothing splines. *The Annals of Statistics* **25**, 414-434.
- Portnoy, S. (2003) Censored regression quantiles. *Journal of the American Statistical Association* **98**, 1001-1012.
- Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25** 303-325.
- Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 27-41.

- Spierdijk, L. (2008). Nonparametric conditional hazard rate estimation: a local linear approach. *Computational Statistics and Data Analysis* **52**, 2419-2434.
- Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *The Annals of Statistics* **27**, 1745-1784.
- Van Keilegom, I. and N. Veraverbeke (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *Journal of Statistical Planning and Inference* **69**, 115-131.
- Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **104**, 1117-1128.
- Wang, Y., He, S., Zhu, L. and Yuen, K. C. (2007). Asymptotics for a censored generalized linear model with unknown link function. *Probability Theory and Its Related Fields* **138**, 235-267.
- Wu, T. Z., Yu, K. and Yu, Y. (2010) Single-index quantile regression. *Journal of Multivariate Analysis* **101**, 1607-1621.
- Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178-184.
- Xia, Y., Xu, J. and Zhang, D. (2010) Dimension reduction and semi-parametric estimation of survival Models. *Journal of the American Statistical Association*. **105**, 278-290.
- Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics* **31**, 333-346.
- Zhang, W. and Li, J. (2010) A semiparametric threshold model for censored longitudinal data analysis. *Journal of the American Statistical Association* (to appear)