Econometric Theory, **29**, 2013, 941–968.
doi:10.1017/S0266466612000813

GLOBAL BAHADUR REPRESENTATION FOR NONPARAMETRIC CENSORED REGRESSION QUANTILES AND ITS APPLICATIONS

EFANG KONG
University of Kent at Canterbury

OLIVER LINTON
University of Cambridge

YINGCUN XIA
Nanjing University and National University of Singapore

This paper is concerned with the nonparametric estimation of regression quantiles of a response variable that is randomly censored. Using results on the strong uniform convergence rate of U-processes, we derive a global Bahadur representation for a class of locally weighted polynomial estimators, which is sufficiently accurate for many further theoretical analyses including inference. Implications of our results are demonstrated through the study of the asymptotic properties of the average derivative estimator of the average gradient vector and the estimator of the component functions in censored additive quantile regression models.

1. INTRODUCTION

Quantile regression (Koenker and Bassett, 1978), originally designed to render estimators robust against extreme values or outliers among the error terms (Huber, 1981), has since attracted tremendous interest both in theoretical statistics and in applied area; see Koenker (2005) and Koenker and Bilias (2001) for a comprehensive literature review. Equally, regression problems based on censored data have always been an important topic in survival analysis, e.g., the accelerated failure time model, as well as in labor economics (Buchinsky, 1994) and econometrics, such as the well-known Tobit model. A direct consequence of censoring is that it causes the error term to deviate from the normal distribution, and the conditional moment restrictions of the uncensored model might be violated. Regression quantiles are among the natural choices for analyzing censored data, as they are more

The authors thank Professor Arthur Lewbel (co-editor) and three referees for helpful comments. Xia's work is partially supported by a grant of NUS R-155-000-121-112. Address correspondence to Efang Kong, School of Mathematics, Statistics and Actuarial Science, University of Kent at Canterbury, UK; e-mail: E.kong@kent.ac.uk.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

"resilient" against such distortions. Most of the existing studies in quantile regression with censoring data adopted a parametric approach; see, e.g., Buckley and James (1979), Koul, Susarla, and Van Ryzin (1981), Ritov (1990), Ying, Jung, and Wei (1995), Honoré, Khan, and Powell (2002), Bang and Tsiatis (2002), and Heuchenne and Van Keilegom (2007b). In this paper we focus on nonparametric estimation of regression quantiles, i.e., no preassumption is made on the form of the conditional quantile function, besides that it should satisfy a certain degree of smoothness. We also relax the requirements imposed on the censoring scheme, allowing the (usually unknown) distribution function of the censoring variable to be dependent on the covariates.

A small number of estimators exist for nonparametric censored quantile 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 of fixed censoring, extended by Chen, Dahl, and Khan (2005) to allow for heteroskedasticity, while Heuchenne and Van Keilegom (2007a, 2008) examined 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) representation is 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 noted in He and Shao (1996), Bahadur representation approximates the estimator by a sum of independent variables with a smaller-order remainder. Consequently, 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) derived the 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) studied the Bahadur representation of quantile smoothing splines, and Portnoy (2003) studied the Bahadur representation for the Cox and censored quantile regression. Chaudhuri (1991b) investigated the pointwise Bahadur representation of nonparametric kernel quantile regression. In nonparametric settings, global or uniform asymptotic theory (Bickel and Rosenblatt, 1973; Mack and Silverman, 1982) is essential for conducting statistical inference. Because of this, uniform Bahadur representations are more useful than their pointwise counterparts. Kong, Linton, and Xia (2010) and Guerre and Sabbah (2012) obtained the uniform Bahadur representation for the quantile local polynomial estimators. Wu (2005) and Zhou and Wu (2009) investigated the Bahadur representation for nonstationary time series data under both parametric and nonparametric settings.

In this paper we introduce a nonparametric local polynomial estimator of the quantile regression function and its derivatives, derived from minimizing a locally

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

weighted objective function where the weights are estimated from the data to adjust for the presence of censoring. Under some regularity conditions, we will obtain the global Bahadur representation for these estimators and show that it can achieve the optimal rate of convergence (Stone, 1980). The implications of the presence of random censoring are twofold. First, it complicates the estimation, especially when the censoring distribution can depend on the covariates, as we allow.[1] Second, the derivation of the Bahadur representation is much more involved, as the behavior of the estimated, thus random, weights needs to be taken into account.

In Section 5 we present two examples to showcase the implications of our results. The first concerns the estimation of the average gradient vector. Again, this has been studied under two separate settings: Chaudhuri, Doksum, and Samarov (1997), Wu, Yu, and Yu (2010), and Kong and Xia (2012) in quantile regression, while Lu and Cheng (2007) and Xia, Zhang, and Xu (2009) for censored mean regression. The second example focuses on the additive model, which has been used to model the regression quantiles (Linton, 2001; De Gooijer and Zerom, 2003; Yu and Lu, 2004) or regression mean but with censored data (de Uña Álvarez and Roca-Pardiñasa, 2009). Yet no one has investigated the use of the additive model for estimating regression quantiles under censoring.

Through these two examples, it will become clear that our results are particularly useful for conducting inference about a variety of quantities of interest. The representations we have obtained can be directly used to obtain consistent standard errors in the case where a parametric quantity like the average gradient vector 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; Johnston, 1982). We remark that the recent work of Belloni, Chernozhukov, and Fernández-Val (2011) has provided tools for inference about nonparametric quantile regression based on the series methodology, but this is done in the absence of censoring.

2. THE MODEL AND ITS ESTIMATION

Suppose $\{(T_i, \mathbf{X}_i), \iota = 1, \ldots, n\}$ are independent and identically distributed (i.i.d.) observations generated according to

$$T_i = Q(\mathbf{X}_i) + \varepsilon_i, \qquad 1 \leq i \leq n, \tag{1}$$

where T_i is the observed value of the univariate dependent variable T, while $\mathbf{X}_i$ is the observed value of the p-dimensional covariates $\mathbf{X}$. Here, $Q(.)$ is an unknown but smooth function, and ε_i is the "error term," which conditional on $\mathbf{X}$ has an τth quantile equal to zero. In other words, $Q(\mathbf{X}_i)$ is the τth quantile of T_i conditional

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

on $\mathbf{X}_i$. Or equivalently, through the use of the quantile loss of function, it is defined as

$$Q(\mathbf{X}_i) = \arg\min_a \mathbb{E}\rho_\tau(T_i - a|\mathbf{X}_i)$$

where $\rho_\tau(s) = |s| + (2\tau - 1)s$.

In this paper we consider the case where T_i is not necessarily observable. Instead, it is subject to random right censoring; the methodology can be easily adapted for left censoring. Let C_i denote the censoring variable, with conditional survival function $G(.|\mathbf{X}_i)$ given $\mathbf{X}_i$; i.e., C_i is not required to be independent of $\mathbf{X}_i$. The observations are made on the triple $\zeta_i = (Y_i, X_i, d_i)$, where

$$Y_i = \min\{T_i, C_i\} = \min\{Q(\mathbf{X}_i) + \varepsilon_i, C_i\}, \qquad d_i = I\{T_i \le C_i\}, \tag{2}$$

are, respectively, the observed (possibly censored) value of the response variable and the censoring indicator. Jointly (1) and (2) specify a censored quantile regression model, and our main objective is the estimation of $Q(.)$ and its partial derivatives, assuming that $Q(.)$ is smooth enough to have partial derivatives up to order k.

For any fixed point $\mathbf{x} \in R^p$, the local polynomial estimation of $Q(\mathbf{x})$ and its partial derivative is based on the fact that $Q(.)$ can be approximated by its k-order Taylor expansion in the neighborhood of $\mathbf{x}$,

$$Q(\mathbf{X}) \approx Q(\mathbf{x}) + \sum_{1 \le [u] \le k} \frac{D^{\mathbf{u}} Q(\mathbf{x})}{\mathbf{u}!}(\mathbf{X} - \mathbf{x})^{\mathbf{u}}, \tag{3}$$

where $\mathbf{u} = (u_1, \ldots, u_p)$ denotes a generic 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 $\partial^{[\mathbf{u}]}/\partial x_1^{u_1} \cdots \partial x_p^{u_p}$. Let $A = \{\mathbf{u} : [\mathbf{u}] \le k\}$ and $n(A) = \sharp(A)$, the cardinality of A.

We start with the ideal scenario, where $\mathbf{Z}_i = \{T_i, \mathbf{X}_i, C_i\}, i = 1, \ldots, n$, are directly observable. The estimates of $Q(.)$ and its partial derivatives can be obtained by minimizing the function below with respect to $\mathbf{c} = (c_{\mathbf{u}})_{\mathbf{u} \in A} \in R^{n(A)}$, a vector of length $n(A)$,

$$\sum_{i=1}^n \rho_\tau \left\{ T_i - \mathbf{c}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \right\} I\{|\mathbf{X}_{i,\mathbf{x}}| \le \delta_n\}, \qquad \mathbf{X}_{i,\mathbf{x}} = \mathbf{X}_i - \mathbf{x}, \tag{4}$$

where $\mathbf{X}_{i,\mathbf{x}} = \mathbf{X}_i - \mathbf{x}$, $\delta_n \to 0$, as $n \to \infty$, is a smoothing parameter, $|.|$ stands for the uniform norm, and for any $\mathbf{x} \in R^p$, $\mathbf{x}(\delta_n, A) = (\mathbf{x}(\delta_n, \mathbf{u}))_{\mathbf{u} \in A}$, with $\mathbf{x}(\delta_n, \mathbf{u}) = \delta_n^{-[\mathbf{u}]} \mathbf{x}^{\mathbf{u}}$. The use of the uniform kernel $I(|.| \le \delta_n)$ is for simplification purposes only (Chaudhuri 1991a, 1991b); $I(.)$ can be replaced by a general multivariate kernel function, e.g., $K_{\delta_n}(.) = K(./\delta_n)$, where $K(.)$ is some probability density

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

function in R^p with a compact support. Results presented in this paper are still valid under such generalization.

The problem with simply submitting Y_i for T_i in (4) is that $Q(,\mathbf{X}_i)$ may not be τth quantile of Y_i unless $Y_i = T_i$, or equivalently, $d_i = 1$. One may then be tempted to restrict the summation in (4) over those i's such that $d_i = 1$; however, this will result in a biased estimator. There are currently three possible ways to deal with the presence of censoring. One is by replacing $\rho_\tau\left\{T_i - \mathbf{c}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A)\right\}$ 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 with C_i assumed to be independent of both $\mathbf{X}_i$ and T_i. 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) for applications of this idea in linear quantile regression. The third strategy, the one adopted in this paper, is based on the observation that, if T_i and Y_i are independent conditional on $\mathbf{X}_i$, then $\mathbb{E}[d_i / G(Y_i|\mathbf{X}_i)|\mathbf{X}_i, T_i] = \mathbb{E}[d_i / G(T_i|\mathbf{X}_i)|\mathbf{X}_i, T_i] = 1$, and, consequently,

$$\mathbb{E}[d_i / G(Y_i|\mathbf{X}_i)\rho_\tau\{Y_i - a)\}] = \mathbb{E}[\rho_\tau\{T_i - a)\}].$$

See also Bang and Tsiatis (2002). Incorporating this observation with (4) implies that we should instead minimize the target function

$$\sum_{i=1}^{n} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \rho_\tau \left\{ Y_i - \mathbf{c}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \right\} I\{|\mathbf{X}_{i,\mathbf{x}}| \le \delta_n\}.$$

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

$$\hat{G}_n(t|\mathbf{x}) = \prod_{j=1}^{n} \left\{ 1 - \frac{B_{nj}(\mathbf{x})}{\sum_{k=1}^{n} I(Y_k \ge Y_j) B_{nk}(\mathbf{x})} \right\}^{\beta_j(t)}, \tag{5}$$

where $\beta_j(t) = I(Y_j \le t, d_j = 0)$, and $B_{nk}(\mathbf{x}),\ k = 1, \ldots, n$ is a sequence of non-negative weights adding up to 1. We choose $B_{nk}(.)$ to be the local polynomial "equivalent kernel/weight"; see Fan and Gijbels (1996) and Masry (1996) for more details. Specifically, assuming that $G(.|\mathbf{x})$ is smooth enough to have derivatives up to order κ_1, define

$$B_{nk}(\mathbf{x}) = \mathbf{e}_1^\top \left[\tilde{\Sigma}_n(\mathbf{x})\right]^{-1} \mathbf{X}_{k,\mathbf{x}}(h_n, A_1) I\{|\mathbf{X}_{k,\mathbf{x}}| \le h_n\}, \tag{6}$$

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

where $\mathbf{e}_1$ from now on stands for a column vector $(1, 0, \ldots, 0)^\top$ of length clear from the context, $A_1 = \{\mathbf{u} : [\mathbf{u}] \le \kappa_1\}$, and $h_n \in R^+$ is yet another smoothing

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

parameter, possibly different from δ_n used above. Again, the uniform kernel $I\{|.| \le h_n\}$, could be replaced by any aforementioned appropriate multivariate kernel.

Substituting $\hat{G}_n(.|.)$ for $G(.|.)$ in (7), we propose to estimate $\{D^{\mathbf{u}}Q(\mathbf{x}) : [\mathbf{u}] \in A\}$ by $\hat{\mathbf{c}}_n(\mathbf{x}) = (\hat{c}_{n,\mathbf{u}}(\mathbf{x}))_{\mathbf{u}\in A}$, the minima of

$$\stackrel{def}{=} \arg\min_{\mathbf{c}} \sum_{i=1}^{n} \frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)} \rho_\tau \left\{ Y_i - \mathbf{c}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \right\} I\left\{ |\mathbf{X}_{i,\mathbf{x}}| \le \delta_n \right\}. \tag{7}$$

Since $0 < \tau < 1$, $\rho_\tau(s)$ goes to infinity as $|s| \to \infty$. Thus the minima of (7) always exists.

Instead of the commonly used Nadaraya-Watson weight (Wang and Wang, 2009) or Gasser-Müller's type weight (Gonzalez-Manteiga and Cadarso-Suarez, 1994), the reason that we opt for a weight as described in (6) is so we can have a K-M estimator with bias of order $O(h_n^{\kappa_1+1})$, which is "negligible relative to variance" for large κ_1.

A minor inconvenience from using the "local polynomial weight" is that the corresponding K-M estimator (5) is not necessarily a proper survival function, as $B_{nk}(.)$ could be negative. However, this shouldn't cause much concern. On one hand, the almost sure representation of the local K-M estimator (5) does not rely on $B_{nk}(.)$ being positive (Gonzalez-Manteiga and Cadarso-Suarez, 1994). On the other hand, in practice, a simple truncation can always be applied to ensure $0 \le \hat{G}_n(.|\mathbf{x}) \le 1$; see Spierdijk (2008) for a similar observation.

3. NOTATIONS AND ASSUMPTIONS

Let $\mathcal{D}$ be an open convex set in R^p and for $s_0 = l + \gamma$, with nonnegative integer l and $0 < \gamma \le 1$, we say a function $m(.) : R^p \to R$ has the order of smoothness s_0 on $\mathcal{D}$, denoted by $m(.) \in H_{s_0}(\mathcal{D})$, if it is differentiable up to order l and there exists a constant $C > 0$, such that

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

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

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

Here, $\mathbf{t}(A_1)$ 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 $f(.)$ be the marginal probability density function of $\mathbf{X}_i$. For any $\mathbf{x} \in R^p$, denote by $g(.|\mathbf{x})$, $f_0(.|\mathbf{x})$, and $f_\varepsilon(.|\mathbf{x})$ the probability density functions of C_i, T_i, and ε_i conditional on $\mathbf{X}_i = \mathbf{x}$. Let

$$F_0(t|\mathbf{x}) = \Pr(T_i \le t | X_i = \mathbf{x}), \qquad F_\varepsilon(t|\mathbf{x}) = \Pr(\varepsilon_i \le t | X_i = \mathbf{x})$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

We assume the following conditions hold throughout the paper unless stated otherwise.

Assumption A1. There exists an open convex set $\mathcal{D}$, such that $f(.)$ is positive on $\mathcal{D}$ and $f(.) \in H_{s_1}(\mathcal{D})$, for some $s_1 > 0$.

Assumption A2. The conditional quantile function $Q(.) \in H_{s_2}(\mathcal{D})$ for some $s_2 > 0$.

Assumption A3. Assume $f_\varepsilon(t|\mathbf{x})$, when seen as a function of $\mathbf{x}$, belongs to $H_{s_3}(\mathcal{D})$ for some $s_3 > 0$, uniformly in t in a neighborhood of zero. Moreover, $f_\varepsilon(0|\mathbf{x})$ is bounded away from zero uniformly in $\mathbf{x} \in \mathcal{D}$, and its first-order derivative with respect to t exists and is continuous in a neighborhood of zero for all $\mathbf{x} \in \mathcal{D}$.

Assumption A4. The censoring variable $\{C_i\}$ is conditionally independent of ε_i given $\mathbf{X}_i$; and for any $\mathbf{x} \in \mathcal{D}$, there exists some finite π_0, which might depend on $\mathbf{x}$, such that $G(\pi_0|\mathbf{x}) = 0$ and $\inf_{\mathbf{x}} P(C_i = \pi_0|\mathbf{x}) > 0$.

Assumption A5. Functions $f_0(t|\mathbf{x})$ and $g(t|\mathbf{x})$, when seen as functions of $\mathbf{x}$, both belong to $H_{s_4}(\mathcal{D})$ for some $s_4 > 0$, uniformly in t.

Assumption A6. The bandwidth h_n in the local K-M estimator is chosen such that $nh_n^{2s_4+p}/\log n \to 0$, $nh_n^{3p}/\log n \to 0$, $nh_n^{p+4}/\log n < \infty$;

Assumption A7. The relationship between the two smoothing parameters δ_n and h_n is such that $\delta_n = o(h_n)$ and $nh_n^{2p}/(\delta_n^p \log n) \to \infty$.

Remark. Assumptions A1–A3 are standard assumptions of local polynomial estimation in quantile regression; see also Chaudhuri et al. (1997). Among them, Assumption A2 implies that, if $|\mathbf{X} - \mathbf{x}| \le \delta_n$, then the error resulted from approximating $Q(\mathbf{X})$ by the $[s_2]$-order 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 $\{(\mathbf{X}, \mathbf{x}) : \mathbf{X}, \mathbf{x} \in \mathcal{D},\ |\mathbf{X} - \mathbf{x}| \le \delta_n\}$. Assumption A4 suggests that T_i and C_i need not be independent if this is a result of their mutual dependency on $\mathbf{X}_i$. Also, given $\mathbf{X}_i$, there is a positive mass on the upper boundary of the support of the censoring variable. This guarantees that $d_i/\hat{G}_n(Y_i|\mathbf{X}_i)$ is uniformly finite in large samples; this condition can always be met by artificially censoring all observations at some point $\pi_0 (\le \max_i Y_i)$. Assumptions A5–A6 are imposed such that the local K-M estimator $\hat{G}_n(.|\mathbf{X}_i)$ admits the almost sure representation in terms of a sum of independent errors. Note that Assumption A6 is stronger compared to its counterpart in Gonzalez-Manteiga and Cadarso-Suarez (1994) or Wang and Wang (2009), which focused on univariate $\mathbf{X}$. This is so such that the bias of the K-M estimator is negligible relative to its variance. Assumption A7 is imposed such that the

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

higher-order remainder term in the almost-sure representation of $\hat{G}_n(.|.)$ is negligible.

To facilitate the subsequent discussion in Section 5, 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 $\mathbf{x} = \mathbf{X}_j,\ j = 1,\ldots,n$. We will derive the uniform convergence rate and the Bahadur-type representation of $\hat{\mathbf{c}}_n(\mathbf{X}_j)$, the estimator of $\mathbf{c}_n(\mathbf{X}_j)$,

$$\hat{\mathbf{c}}_n(\mathbf{X}_j) = \arg\min_{\mathbf{c}} \sum_{i\in S_n(\mathbf{X}_j)} \frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)} \rho_\tau\left\{Y_i - \mathbf{c}^\top \mathbf{X}_{ij}(\delta_n, A)\right\}, \qquad \mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j \tag{8}$$

where the index set $S_n(\mathbf{X}_j)$ and its cardinality are defined as

$$S_n(\mathbf{X}_j) = \{i : 1 \le i \le n, i \ne j, |\mathbf{X}_{ij}| \le \delta_n\}, \qquad N_n(\mathbf{X}_j) = \sharp(S_n(\mathbf{X}_j)).$$

The reason for us to consider the above leave-one-out version is the same as given in Chaudhuri et al. (1997), i.e., to simplify the conditioning arguments used at various stages of the proofs. The non-leave-one-out estimator 7 is asymptotically first-order equivalent to its leave-one-out counterpart.

4. CONVERGENCE RATE AND ASYMPTOTIC REPRESENTATION

Our first result concerns the almost sure representation of the local K-M estimator $\hat{G}_n(.|.)$. For $j = 1,\ldots,n$, define

$$\xi(Y_j, d_j, t, \mathbf{x}) = \frac{I\{Y_j \le t, d_j = 0\}}{h(Y_j|\mathbf{x})} - \int_0^{\min(Y_j,t)} \frac{d\Gamma(s|\mathbf{x})}{h(s|\mathbf{x})},$$

where $h(t|\mathbf{x}) = 1 - \Pr(Y_j \le t|\mathbf{x}) = G(t|\mathbf{x})(1 - F_0(t|\mathbf{x}))$, $\Gamma(t|\mathbf{x}) = -\ln(G(t|\mathbf{x})) = -\int_0^t \frac{dG(t|\mathbf{x})}{G(t|\mathbf{x})}$, and $\tilde{H}(t|\mathbf{x}) = \Pr\{Y_j \le t, d_j = 0|\mathbf{x}\} = -\int_0^t (1 - F(s|\mathbf{x}))dG(s|\mathbf{x})$.

LEMMA 4.1. *Suppose that Assumptions A4–A7 hold and* $\kappa_1 = [s_4]$. *Then with probability one,*

$$\sup_{\mathbf{x}\in\mathcal{D}} \sup_t |\hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x})| = O\left(\left(\frac{\log n}{nh_n^p}\right)^{1/2}\right) \tag{9}$$

$$\hat{G}_n(t|\mathbf{x}) - G(t|\mathbf{x}) = G(t|\mathbf{x}) \frac{1}{\tilde{N}_n(\mathbf{x})} \sum_{k\in\tilde{S}_n(\mathbf{x})} [\mathbf{X}_{k\mathbf{x}}(h_n, A_1)]^\top \xi(Y_k, d_k, t, \mathbf{x}) \tag{10}$$

$$\times [f(\mathbf{x})\Sigma(A_1)]^{-1}\mathbf{e}_1 + O\left(\left(\frac{\log n}{nh_n^p}\right)^{3/4}\right)$$

uniformly in $\mathbf{x} \in \mathcal{D}$ *as well as in* t, *where*

$$\tilde{S}_n(\mathbf{x}) = \{i : 1 \le i \le n,\ |\mathbf{X}_{i,\mathbf{x}}| \le h_n\}, \qquad \tilde{N}_n(\mathbf{x}) = \sharp\left(\tilde{S}_n(\mathbf{x})\right).$$

The next theorem gives the strong uniform convergence rate of $\hat{\mathbf{c}}_n(\mathbf{X}_j)$, $j = 1,\ldots,n$.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

THEOREM 4.2. *Suppose Assumptions A1–A7 hold with* $s_1 > 0, s_2 > 0, s_3 > 0$. *Let* $k = [s_2]$, *and the bandwidth* δ_n *is chosen such that*

$$\delta_n \propto n^{-\kappa}, \quad \text{for some } \frac{1}{2s_2 + p} \leq \kappa < \frac{1}{p}.$$

Then we have, with probability one,

$$\sup_{1 \leq j \leq n} |\hat{\mathbf{c}}_n(\mathbf{X}_j) - \mathbf{c}_n(\mathbf{X}_j)| = O\left(\left\{n^{1-\kappa p}/\log n\right\}^{-1/2}\right). \tag{11}$$

Remark. The optimal rate of convergence of Stone (1980) can be achieved by setting $\kappa = 1/(2s_2 + p)$.

We take this moment to briefly describe how uniformity in Theorem 4.2 can be extended to cover the whole compact set $\mathcal{D}$. Cover $\mathcal{D}$ with $J_n^p = O(n^{kp})$ number of cubes side length $2\delta_n$, and let $S_{n,r}$ be a typical such cube with center at $\mathbf{x}_{n,r}$, $1 \leq r \leq J_n^p$. Obtain estimates of $\hat{\mathbf{c}}_n(\mathbf{x}_{n,r})$ through minimizing (3) with $\mathbf{x}_{n,r}$ substituted for $\mathbf{X}_j$. For any $\mathbf{x} \in S_{n,r}$, construct the estimates of $\mathbf{c}_n(\mathbf{x})$ as

$$\hat{c}_{n,\mathbf{u}}(\mathbf{x}) = \delta_n^{[\mathbf{u}]}/\mathbf{u}! D^{\mathbf{u}} \left[\left\{\hat{\mathbf{c}}_n(\mathbf{x}_{n,r})\right\}^\top (\mathbf{x}_{n,r} - \mathbf{x})(\delta_n, A)\right],$$

where the differential operator $D^{\mathbf{u}}$ is with respect to $\mathbf{x}$. Under Assumption A2, uniformity over $\mathcal{D}$ thus reduces to uniformity over $\mathbf{x}_{n,r}$, which can be proved in exactly the same way as (11), by appealing to the Borel-Cantelli lemma.

For any $\mathbf{x} \in \mathcal{D}$, define

$$\Sigma_n(\mathbf{x}) = \mathbb{E}_i \left[f_{\varepsilon|\mathbf{X}}(0|\mathbf{X}_i) \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \mathbf{X}_{i,\mathbf{x}}^\top(\delta_n, A) | \mathbf{X}_i \in S_n(\mathbf{x}) \right],$$

$$T_n(\zeta_j, \zeta_k) = \mathbb{E}_i \Big[\mathbf{X}_{ij}(\delta_n, A) \mathbf{X}_{ki}^\top(h_n, A_1) [I\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\} - \tau] \times \xi(Y_k, d_k, Y_i, \mathbf{X}_i)/f(\mathbf{X}_i) | \mathbf{X}_i \in S_n(\mathbf{X}_k) \Big],$$

where $\mathbb{E}_i(.)$ stands for expectation taken with respect to the distribution of $(\mathbf{X}_i, Y_i)$.

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

THEOREM 4.3. *Suppose conditions in Theorem 4.2 hold with* $s_3 > 1/2$, *and the bandwidth* δ_n *is chosen such that*

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

Then we have

$$\begin{aligned} &\hat{\mathbf{c}}_n(\mathbf{X}_j) - \mathbf{c}_n(\mathbf{X}_j) \\ &\quad = \frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)} \sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}(\delta_n, A) \left[\tau - I\left\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\right\}\right] \\ &\qquad - \frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)} \sum_{k=1}^{n} T_n(\zeta_j, \zeta_k) [\Sigma(A_1)]^{-1} \mathbf{e}_1 + R_n(\mathbf{X}_j), \end{aligned} \tag{12}$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

where with probability one,

$$\max_{1\le j\le n}|R_n(\mathbf{X}_j)| = O\left(\left\{n^{1-\kappa p}/\log n\right\}^{-3/4}\right).$$

Remark 1. In the absence of censoring, Chaudhuri (1991b, Thm. 3.3) derived a local Bahadur representation. Our results differ from his in two aspects. First, the factor $d_i/G(Y_i|\mathbf{X}_i)$ rectifies the "bias" caused by censoring. Second, the plugging-in of the preliminary estimator $\hat{G}_n(.|.)$ of the survival function $G(.|.)$ leads to the second term in (12), which is nonexistent in Chaudhuri (1991b). Similar observation has been made by Honoré et al. (2002) for linear quantile regression under censoring.

Remark 2. For any finite $\iota > 0$, let $0 < \tau_1 < \tau_2 < \cdots < \tau_{n^\iota} < 1$ stand for n^ι different quantile levels. If Assumptions A2 and A3 are also satisfied uniformly in $\tau = \tau_1, \ldots, \tau_{n^\iota}$, the above uniformity results can be easily extended to cover estimators associated with these quantile levels.

Remark 3. The above Bahadur representation is valid for the "optimal" nonparametric function estimation bandwidth $\delta_n \propto n^{-1/(2s_2+p)}$. Yet, in many applications "undersmoothing," i.e., bandwidth smaller than optimal, is often necessary to obtain the desired asymptotic results. This is so that the bias of the estimators, of order $O(\delta_n^{s_2})$, is $o\left(n^{-1/2}\right)$. This implies that we need $\kappa > 1/(2s_2)$ at least. Similar observations have been made in Chaudhuri et al. (1997), where readers can find references containing examples of undersmoothing.

5. APPLICATIONS

In this section we showcase through two examples the implications of Theorem 4.3 in establishing asymptotic properties of a class of estimators. For the remainder of this paper, a.s. stands for almost surely.

First note that under the conditions in Theorem 4.3, we have

$$\max_{1\le j\le n}\delta_n^{-1}|R_n(\mathbf{X}_j)| = o\left(n^{-1/2}\right) \quad \text{a.s.}$$

provided that

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2s_2} \le \kappa < \frac{1}{4+3p}; \tag{13}$$

and

$$\max_{1\le j\le n}|R_n(\mathbf{X}_j)| = o\left(n^{-1/2}\right) \quad \text{a.s.},$$

provided that

$$\delta_n \propto n^{-\kappa}, \quad \text{with} \quad \frac{1}{2s_2} \le \kappa < \frac{1}{3p}. \tag{14}$$

Note that (15) will be used in Section 5.1 and (14) in Section 5.2.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

5.1. The Censored Average Derivative Estimator

Let $\nabla Q(\mathbf{X}) = \partial Q(\mathbf{X})/\partial \mathbf{X}$ denote the gradient vector of $Q(.)$. Then the average gradient vector

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

gives a concise summary of quantile-specific regression effects; i.e., the average change in the quantile of the response as the ith covariate is perturbed, while the other covariates are held fixed. This parameter has been of great interest in econometrics following the work of Härdle and Stoker (1989). Here we study the estimation of β in the presence of censoring using the average derivative method.

Let $\nabla \hat{Q}(X_j)$ be the nonparametric estimator of $\nabla Q(\mathbf{X})$ derived from (3); i.e.,

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

and consequently we can construct an estimate of β as

$$\hat{\beta} = \frac{1}{n}\sum_{j=1}^{n} \nabla \hat{Q}(X_j). \tag{15}$$

We refer to (15) as the censored average derivative estimator (c-ADE). To establish the asymptotic property of $\hat{\beta}$, we assume that conditions in Theorem 4.3 hold with $s_1 = s_3 = s_4 = 1+\gamma$, for some $\gamma > 0$, and δ_n is chosen such that (13) holds for some $s_2 > 3p/2+2$.

According to Theorem 4.3, for any $\mathbf{b} = (b_1, \ldots, b_p)^\top \in R^p$, we have

$$\mathbf{b}^\top(\hat{\beta}-\beta) = \mathbf{b}^\top\left[\frac{1}{n}\sum_{j=1}^{n}\nabla Q(\mathbf{X}_j) - \beta\right] + o\left(n^{-1/2}\right) + \mathbf{B}^\top \frac{1}{n\delta_n}\sum_{j=1}^{n}\frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)}$$

$$\times \sum_{i\in S_n(\mathbf{X}_j)} \frac{d_i}{G(Y_i|\mathbf{X}_i)}\mathbf{X}_{ij}(\delta_n, A)\left[\tau - I\left\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{X}_j)\right\}\right] \tag{16}$$

$$+\mathbf{B}^\top \frac{1}{n\delta_n}\sum_{j,k=1}^{n}\frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)} T_n(\zeta_j, \zeta_k)[\Sigma(A_1)]^{-1}\mathbf{e}_1 \quad \text{a.s.,} \tag{17}$$

where $\mathbf{B}$ is an $n(A)\times 1$ vector, defined as $\mathbf{B} = (0, \mathbf{b}^\top, \mathbf{0})^\top$. First, note that following similar lines as in Chaudhuri et al. (1997, pp. 736–739), we can show that the term (16) is asymptotically equivalent to

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

We now move on to study term (17). Define

$$\tilde{T}_n(\zeta_j, \zeta_k) = \mathbb{E}_i\Big[\mathbf{X}_{ij}(\delta_n, A)\mathbf{X}_{ki}^\top(h_n, A_1)[I\{Y_i \le Q(\mathbf{X}_i)\} - \tau]$$

$$\times \xi(Y_k, d_k, Y_i, \mathbf{X}_i)/f(\mathbf{X}_i)|\mathbf{X}_i \in S_n(\mathbf{X}_k)\Big].$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Then it is easy to see that the error resulting from replacing $T_n(\zeta_j, \zeta_k)$ in (17) with $\tilde{T}_n(\zeta_j, \zeta_k)$ is of order $o_p\left(n^{-1/2}\right)$. Now let

$$\eta_n(\mathbf{Z}_j, \mathbf{Z}_k) = \mathbf{B}^\top [N_n(\mathbf{X}_j)\Sigma_n(\mathbf{X}_j)]^{-1}\tilde{T}_n(\zeta_j, \zeta_k)[\Sigma(A_1)]^{-1}\mathbf{e}_1,$$

$$\xi_n(\mathbf{Z}_j, \mathbf{Z}_k) = \eta_n(\mathbf{Z}_j, \mathbf{Z}_k) + \eta_n(\mathbf{Z}_k, \mathbf{Z}_j), \qquad U_n = \sum_{1 \le j < k \le n} \xi_n(\mathbf{Z}_j, \mathbf{Z}_k).$$

Then we have

$$\mathbf{B}^\top \frac{1}{n\delta_n}\left[\sum_{j,k=1}^{n} \frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)}\tilde{T}_n(\zeta_j, \zeta_k)\right][\Sigma(A_1)]^{-1}\mathbf{e}_1 = \frac{1}{n\delta_n}U_n + o\left(n^{-1/2}\right) \quad \text{a.s.},$$

as long as $\kappa < 1/(2p-2)$, where we have implicitly used (A.1) in the Appendix and the fact that the smallest eigenvalue of $\Sigma_n(\mathbf{X}_j)$ is bounded away from zero uniformly in j and k.

To analyze U_n, first note that $\mathbb{E}[\xi_n(\mathbf{Z}_j, \mathbf{Z}_k)] = \mathbb{E}[\eta_n(\mathbf{Z}_k, \mathbf{Z}_j)] = 0$. 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 $g_n(\mathbf{Z}_k) = \mathbb{E}_j[\xi_n(\mathbf{Z}_j, \mathbf{Z}_k)] = \mathbb{E}_j[\eta_n(\mathbf{Z}_j, \mathbf{Z}_k)]$. We thus have, through arguments similar to that in Chaudhuri et al. (1997), that

$$\begin{aligned}\mathbb{E}(U_n - P_n)^2 &= \frac{n(n-1)}{2}\left\{\mathbb{E}\left[\xi_n^2(\mathbf{Z}_k, \mathbf{Z}_j)\right] - 2\mathbb{E}\left[g_n^2(\mathbf{Z}_k)\right]\right\} \\ &\le \frac{n(n-1)}{2}\mathbb{E}\left[\xi_n^2\left(\mathbf{Z}_k, \mathbf{Z}_j\right)\right] = O\left(\delta_n^{-p}\right) = o\left(n\delta_n^2\right), \end{aligned} \tag{18}$$

as long as $\kappa < 1/(p+2)$. We move on to study $g_n(.)$. To this aim, we need to use facts (A.5) and (A.1) given in the Appendix, to get that with probability one,

$$\begin{aligned}&\delta_n^p \frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)} \\ &\quad = \frac{1}{n}\frac{[\Sigma(A)]^{-1}}{f_{\varepsilon,\mathbf{X}}(0, \mathbf{X}_j)} + \frac{\delta_n}{n}\frac{[\Sigma(A)]^{-1}\sum_{l=1}^{p}\Sigma_l^* f_{\varepsilon,\mathbf{X}}^{(l)}(0, \mathbf{X}_j)[\Sigma(A)]^{-1}}{f_{\varepsilon,\mathbf{X}}^2(0, \mathbf{X}_j)} + O\left(\delta_n^{s_3} + \delta_n^2\right),\end{aligned}$$

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

$$\begin{aligned}\frac{1}{n\delta_n}P_n &= \frac{1}{\delta_n}\sum_{k=1}^{n} g_n(\mathbf{Z}_k) + o\left(n^{-1/2}\right) \\ &= \mathbf{B}^\top \frac{[\Sigma(A)]^{-1}}{n\delta_n^{p+1}}\sum_{k=1}^{n}\mathbb{E}_j\left[f_{\varepsilon,\mathbf{X}}^{-1}(0, \mathbf{X}_j)\tilde{T}_n(\zeta_j, \zeta_k)\right] \\ &\quad + \frac{1}{n\delta_n^p}[\Sigma(A)]^{-1}\sum_{k=1}^{n}\mathbb{E}_j\left[\frac{\sum_{l=1}^{p}\Sigma_l^* f_{\varepsilon,\mathbf{X}}^{(l)}(0, \mathbf{X}_j)}{f_{\varepsilon,\mathbf{X}}^2(0, \mathbf{X}_j)}\tilde{T}_n(\zeta_j, \zeta_k) f(\mathbf{X}_j)\right] \\ &\quad + o_p\left(n^{-1/2}\right). \end{aligned} \tag{19}$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

The handling of the two leading terms in (19) are very similar. We only deal with the first term to illustrate. Let

$$M(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k) = \mathbb{E}\{[I\{Y_i \le Q(\mathbf{X}_i)\} - \tau]\xi(Y_k, d_k, Y_i, \mathbf{X}_i)|\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k\}\frac{f(\mathbf{X}_j)}{f_{\varepsilon|\mathbf{X}}(0|\mathbf{X}_j)}. \tag{20}$$

First note that in view of Assumptions A1–A4, it is easy to see that $M(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k) \in H_{1+\gamma}(\mathcal{D})$. Moreover, $M(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k)$ itself and its first-order partial derivatives with respect to $\mathbf{X}_i$ and $\mathbf{X}_j$ all have mean zero. Therefore, the first term in (19) can be expressed as

$$\begin{aligned}
&\mathbf{B}^\top\left[\delta_n^{p+1}\Sigma(A)\right]^{-1}\mathbb{E}_{ij}\left[\mathbf{X}_{ij}(\delta_n, A)\mathbf{X}_{ki}^\top(h_n, A_1)M(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k)\Big||\mathbf{X}_{ki}| \le h_n\right][\Sigma(A_1)]^{-1}\mathbf{e}_1 \\
&\quad = \mathbb{E}_i\Big[\mathbf{B}^\top[\delta_n\Sigma(A)]^{-1}\Big\{\int_{\mathbf{t}\in[-1,1]^d} M(\mathbf{X}_i, \mathbf{X}_i + \delta_n\mathbf{t}, \mathbf{Z}_k)\mathbf{t}(A)d\mathbf{t}\Big\} \\
&\qquad\quad \times [\mathbf{X}_{ki}(h_n, A_1)]^\top\Big|\mathbf{X}_{ki}| \le h_n\Big][\Sigma(A_1)]^{-1}\mathbf{e}_1 \\
&\quad = \mathbb{E}_i\Big[\Big\{\mathbf{b}^\top M_1(\mathbf{X}_i, \mathbf{Z}_k) + \delta_n^\gamma W_{n1}(\mathbf{X}_i, \mathbf{Z}_k)\Big\}[\mathbf{X}_{ki}(h_n, A_1)]^\top\Big|\mathbf{X}_{ki}| \le h_n\Big][\Sigma(A_1)]^{-1}\mathbf{e}_1 \\
&\quad = \mathbf{b}^\top M_1(\mathbf{Z}_k) + h_n W_{n2}(\mathbf{Z}_k) + \delta_n^\gamma W_{n3}(\mathbf{Z}_k),
\end{aligned} \tag{21}$$

where $W_{n1}(.)$, $W_{n2}(.)$, and $W_{n2}(.)$ stand for various zero-mean uniformly bounded random terms, and

$$M_1(\mathbf{X}_i, \mathbf{Z}_k) = \frac{\partial M(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Z}_k)}{\partial \mathbf{X}_j}|_{\mathbf{X}_j=\mathbf{X}_i}, \qquad M_1(\mathbf{Z}_k) \stackrel{def}{=} M_1(\mathbf{X}_k, \mathbf{Z}_k). \tag{22}$$

Note that for the second equality in (21), we have implicitly used the facts

$$\mathbf{B}^\top[\Sigma(A)]^{-1}\int \mathbf{t}(A)[\mathbf{t}(A)]^\top d\mathbf{t} = \mathbf{B}^\top,$$

$$\mathbf{B}^\top[\Sigma(A)]^{-1}\int \mathbf{t}(A)d\mathbf{t} = 0, \qquad \mathbf{e}_1^\top[\Sigma(A_1)]^{-1}\int \mathbf{t}(A_1)d\mathbf{t} = 1.$$

In a similar manner, the second term in (19) can be shown to be of order $o_p\left(n^{-1/2}\right)$. Gathering results (17), (18), (19), and (21), we have

$$\hat{\beta} - \beta = \frac{1}{n}\sum_{k=1}^n \nabla Q(\mathbf{X}_k) - \beta + \frac{1}{n}\sum_{k=1}^n\left[\frac{d_k\{\tau - I(\varepsilon_k \le 0)\}}{G(Y_k|\mathbf{X}_k)}\frac{\nabla f(\mathbf{X}_k)}{f_{\varepsilon,\mathbf{x}}(0, \mathbf{X}_k)} + M_1(\mathbf{Z}_k)\right] + o_p\left(\frac{1}{n^{1/2}}\right). \tag{23}$$

This would be exactly the same as the one obtained in Chaudhuri et al. (1997, Thm. 2.1) if not for the factor $d_k/G(Y_k|\mathbf{X}_k)$ and the presence of $M_1(\mathbf{Z}_k)$. The former reflects the presence of censoring, while the latter reflects the impact of plugging in the K-M estimator. If $C_i = \infty$, then $d_k/G(Y_k|\mathbf{X}_k) \equiv 1$ and

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

$M_1(\mathbf{Z}_k) \equiv 0$, for all $k = 1, \ldots, n$. In other words, the results we derived here do coincide with those in Chaudhuri et al. (1997) in the absence of censoring.

Asymptotic normality of $\hat{\beta}$ is straightforward from (23), since the three terms $\nabla Q(\mathbf{X}_k) - \beta$, $\dfrac{d_k\{\tau - I(\varepsilon_k \le 0)\}}{G(Y_k|\mathbf{X}_k)} \dfrac{\nabla f(\mathbf{X}_k)}{f_{\varepsilon,\mathbf{x}}(0,\mathbf{X}_k)}$, and $M_1(\mathbf{Z}_k)$ are all zero-meaned with finite variances. While it is relatively easy to work out the variances of the first two,

$$\mathrm{Var}(\nabla Q(\mathbf{X}_k)), \qquad \mathbb{E}\left[\frac{\mathrm{Var}\{I(\varepsilon_k \le 0)|\mathbf{X}_k\}}{G(Y_k|\mathbf{X}_k)} \frac{\nabla f(\mathbf{X}_k)\{\nabla f(\mathbf{X}_k)\}^\top}{\{f_{\varepsilon,\mathbf{x}}(0,\mathbf{X}_k)\}^2}\right],$$

the variance of $M_1(\mathbf{Z}_k)$ and its covariances with the other two terms take too complicated a form to be included here.

5.2. The Additive Quantile Regression Model

Suppose the regression quantile function $Q(.)$ in model (2) admits an "additive" form; i.e.,

$$Q(\mathbf{x}) = Q(x_1, \ldots, x_p) = c + Q_1(x_1) + \cdots + Q_p(x_p), \tag{24}$$

where c is an unknown constant, and $Q_k(.)$, $k = 1, \ldots, p$, are unknown functions that have been normalized such that $\mathbb{E}[Q_k(X_k)] = 0$, $k = 1, \ldots, p$. For previous work on additive regression model, see Linton (2001), Yu and Lu (2004), and Horowitz and Lee (2005). To estimate the component functions in (24), $Q_1(.)$ say, we consider the marginal integration method, which involves first estimating $Q(.)$ and then integrating it over certain directions. Partition $\mathbf{x}$ as $\mathbf{x} = (x_1, \mathbf{x}_2)$, where x_1 is the one-dimensional direction of interest and $\mathbf{x}_2$ is the $p-1$ dimensional nuisance direction. Accordingly, partition $\mathbf{X}_j = (X_{j1}, \mathbf{X}_{j2})$. Define the functional

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

where $f_2(\mathbf{x}_2)$ is the joint probability density of $\mathbf{X}_{i2}$. Under the additive structure (24), $\phi_1(.) = c + Q_1(.)$. Therefore, estimation of $Q_1(.)$ transfers into that of $\phi_1(.)$. Based on (25), an estimate $\phi_{n1}(x_1)$ of $\phi_1(x_1)$ can be obtained by replacing $f_2(.)$ in (25) with the empirical distribution function of $\mathbf{X}_{j2}$, and $Q(.)$ with its local polynomial estimate $\hat{\mathbf{c}}_{n1}(x_1, \mathbf{X}_{j2})$, the first element of $\hat{\mathbf{c}}_n(x_1, \mathbf{X}_{j2})$; i.e.,

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

In the context of mean regression, Linton and Härdle (1996) and Hengartner and Sperlich (2005) suggested that for $\phi_{n1}(.)$ to be asymptotically normal, the bandwidth used for the direction of interest x_1 should be different from those for the $p-1$ nuisance directions. However, for ease of expression, we assume that the same bandwidth is used for all directions.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Let $\mathbf{X}_j^* = (x_1, \mathbf{X}_{j2})$ and $\mathbf{X}_{ij}^* = \mathbf{X}_i - \mathbf{X}_j^*$. It follows from Theorem 4.3 that

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

where $T_n(\zeta_j^*, \zeta_k)$ is defined similarly to $T_n(\zeta_j^*, \zeta_k)$, 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 $Q(.)$ leads to

$$\begin{aligned}
\phi_{n1}(x_1) &= \phi_1(x_1) + \frac{1}{n}\sum_{j=1}^{n} \mathbf{Q}_2(\mathbf{X}_{j2}) + o\left(n^{-1/2}\right) \\
&\quad + \mathbf{e}_1^\top \frac{1}{n} \sum_{j=1}^{n} \frac{\left[\Sigma_n\left(\mathbf{X}_j^*\right)\right]^{-1}}{N_n\left(\mathbf{X}_j^*\right)} \sum_{i \in S_n\left(\mathbf{X}_j^*\right)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}^*(\delta_n, A)\left[\tau - I\left\{Y_i \le Q_n\left(\mathbf{X}_i, \mathbf{X}_j^*\right)\right\}\right] \qquad \textbf{(26)} \\
&\quad - \mathbf{e}_1^\top \frac{1}{n^2} \sum_{j,k} \frac{\left[\Sigma_n\left(\mathbf{X}_j^*\right)\right]^{-1}}{N_n\left(\mathbf{X}_j^*\right)} T_n\left(\zeta_j^*, \zeta_k\right) \mathbf{e}_1^\top [\Sigma(A_1)]^{-1}, \qquad \textbf{(27)}
\end{aligned}$$

where $\mathbf{Q}_2(x_2, \ldots, x_p) = Q_2(x_2) + \cdots + Q_p(x_p)$. Note that $\phi_{n1}(x_1)$ is, by definition, the average of n subvectors of $\hat{\mathbf{c}}_n(.)$, with the average taken along the $p-1$ nuisance directions, while for ADE (15), the average is taken along all p directions. Therefore, as in the case of ADE, we can conclude that the term (27), like (17), is negligible compared to others, and the dealing of term (26) is similar to that of (17). Specifically, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n} \frac{\left[\Sigma_n\left(\mathbf{X}_j^*\right)\right]^{-1}}{N_n\left(\mathbf{X}_j^*\right)} \sum_{i \in S_n\left(\mathbf{X}_j^*\right)} \frac{d_i}{G(Y_i|\mathbf{X}_i)} \mathbf{X}_{ij}^*(\delta_n, A)\left[\tau - I\left\{Y_i \le Q_n\left(\mathbf{X}_i, \mathbf{X}_j^*\right)\right\}\right] \\
&\quad = \frac{[\Sigma(A)]^{-1}}{n\delta_n} \sum_{i=1}^{n} \frac{d_i I\{|X_{i1} - x| \le \delta_n\}}{G(Y_i|\mathbf{X}_i)} [\tau - I\{\varepsilon_i \le 0\}] \left[\frac{f_2(\mathbf{X}_{i2})}{f_{\varepsilon,\mathbf{X}}(0, x_1, \mathbf{X}_{j2})}\right] \\
&\qquad \times \int_{[0,1]^{p-1}} \left[\delta_n^{-1}(X_{i1} - x), v\right](A) dv + o_p\left(\delta_n^{-1} n^{-1/2}\right),
\end{aligned}$$

where X_{j1} stands for the first element of $\mathbf{X}_j$. Therefore,

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

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Write $\mathbf{b}(A) = \int_{[0,1]^{\otimes p}} \mathbf{t}(A)d\mathbf{t}$. Asymptotic normality for $(n\delta_n)^{1/2}(\phi_{n1}(.) - \phi_1(x_1))$ can thus be established, with mean zero and covariance equal to $\mathbf{e}_1^\top[\Sigma(A)]^{-1}\mathbf{b}(\mathbf{A})\mathbf{b}(\mathbf{A})^\top[\Sigma(A)]^{-1}\mathbf{e}_1$ multiplied by

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

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

6. NUMERICAL STUDIES

In this section we carry out a small-scale simulation study to investigate the finite-sample performance of the c-ADE estimator (15). For comparison, we also include in the study the naive average derivative estimator (n-ADE), i.e., the ADE using only uncensored data as if they were not subject to any censoring.

Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \varepsilon$, and ϵ be i.i.d. $N(0,1)$ random variables. The response variable T and censoring variable C are such that

$$T = \exp(\beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_2\mathbf{x}_3) + \varepsilon, \qquad C = \exp(\mathbf{x}_2 + c_0 + \epsilon), \tag{28}$$

with $\beta_1 = -1, \beta_2 = 0, \beta_3 = 1$, and some constant c_0, which dictate the probability that $C < T$, i.e., the censoring rate. Specifically, $c_0 = -1, -0.5, 0, 1, 1.5, 2, 3$, and ∞ correspond to censoring rates of about 64%, 57%, 50%, 33%, 24%, 17%, 7%, and 0%, respectively.

Model (28) is a single-index model at all quantile levels, with $\theta_0 = (\beta_1,\beta_2,\beta_3)^\top/(\beta_1^2+\beta_2^2+\beta_3^2)^{1/2} = (-1,0,1)^\top/\sqrt{2}$. For any standardized estimator $\hat{\theta}$ of θ_0, define the estimation error as

$$err = \left(1 - |\theta_0^\top\hat{\theta}|\right)^{1/2}.$$

The averaged estimation errors based on 200 replications are shown in Figure 1. It is immediately evident that the higher the censoring rate, the bigger the improvement c-ADE is over n-ADE. Also, as sample size increases, the superiority of c-ADE over n-ADE becomes more pronounced.

7. DISCUSSION

Kong et al. (2010) derived the Bahadur representation for the local polynomial estimator of a nonparametric M-(quantile) regression function for complete data. This paper considers the same problem, but for censored data, thus building a bridge joining Bahadur representation for nonparametric quantile regression and

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

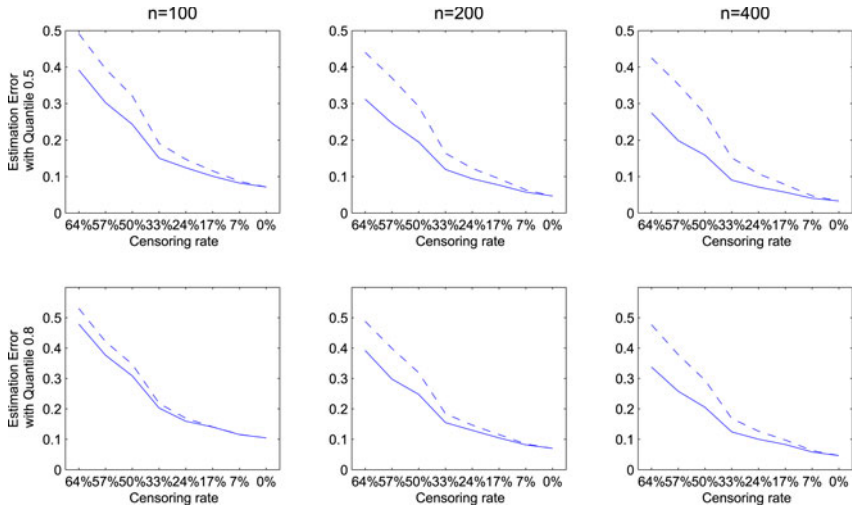


FIGURE 1. The average estimation errors for different combinations of sample size n, quantile level, and censoring rate. The dashed line represents the averaged estimation error of n-ADE; the solid line represents the averaged estimation error of c-ADE.

(random) censoring. Two examples have been provided to demonstrate the usefulness of the results in establishing the asymptotic properties of estimators commonly used in statistical inference. One of the examples showcases that the result obtained in Chaudhuri et al. (1997) is in fact a special case of ours.

NOTE

1. Under fixed censoring, quantile regression of a certain order may be consistent, but this is not the case under random censoring.

REFERENCES

Arcones, M.A. (1995) A Bernstein-type inequality for U-statistics and U-processes. *Statistics and Probability Letters* 22, 239–247.

Babu, G. (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.

Bang, H. & A.A. Tsiatis (2002) Median regression with censored cost data. *Biometrics* 58, 643–649.

Belloni, A., V. Chernozhukov, & I. Fernández-Val (2011) Conditional Quantile Processes Based on Series or Many Regressors. CEMMAP working paper cwp1911.

Bickel, P. & M. Rosenblatt (1973) On some global measures of the deviation of density function estimators. *Annals of Statistics* 1, 1071–1095.

Buchinsky, M. (1994) Changes in the U.S. wage structure 1963-1987: Application of quantile regression. *Econometrica* 62, 405–458.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Buckley, J. & I. James (1979) Linear regression with censored data. *Biometrika* 66, 429–436.
Carroll, R.J. (1978) On almost sure expansions for M-estimates. *Annals of Statistics* 6, 314–318.
Chaudhuri, P. (1991a) Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.
Chaudhuri, P. (1991b) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760–777.
Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.
Chen, S., G.B. Dahl, & S. Khan (2005) Nonparametric identification and estimation of a censored location-regression model. *Journal of the American Statistical Association* 100, 212–221.
Dabrowska, D.M. (1992) Nonparametric quantile regression with censored data. *Sankhyā Series A* 54, 252–259.
De Gooijer, J.G. & D. Zerom (2003) On additive conditional quantiles with high dimensional covariates. *Journal of the American Statistical Association* 98, 135–146.
de Uña-Álvarez, J. & J. Roca-Pardiñasa (2009) Additive models in censored regression. *Computational Statistics and Data Analysis* 53, 3490–3501.
Efron, B. (1967) The two-sample problem with censored data. In L. Le Cam & J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics, IV*, pp. 831–853. Prentice-Hall.
Fan, J. & I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*. Chapman and Hall.
Gonzalez-Manteiga, W. & C. Cadarso-Suarez (1994) Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics* 4, 65–78.
Guerre, E. & C. Sabbah (2012) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* 28, 87–129.
Härdle, W. & T.M. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of American Statistical Association* 84, 986–995.
He, X. & Q.M. Shao (1996) A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Annals of Statistics* 24, 2608–2630.
Hengartner, N. & S. Sperlich (2005) Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis* 95, 246–272.
Heuchenne, C. & I. Van Keilegom (2007a) Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis* 98, 1558–1582.
Heuchenne, C. & I. Van Keilegom (2007b) Nonlinear regression with censored data. *Technometrics* 49, 34–44.
Heuchenne, C. & I. Van Keilegom (2008) Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics* 62, 439–463.
Honoré, B., S. Khan, & J. Powell (2002) Quantile regression under random censoring. *Journal of Econometrics* 109, 67–105.
Horowitz, J. & S. Lee (2005) Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* 100, 1238–1249.
Huber, P.J. (1981) *Robust Statistics*. Wiley.
Johnston, G. (1982) Probabilities of maximal deviations of nonparametric regression function estimates. *Journal of Multivariate Analysis* 12, 402–414.
Koenker, R. (2005) *Quantile Regression*. Cambridge University Press.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.
Koenker, R. & Y. Bilias (2001) Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics* 26, 199–220.
Kong, E., O. Linton, & Y. Xia (2010) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* 26, 1529–1564.
Kong, E. & Y. Xia (2012) A single-index quantile regression model and its estimation. *Econometric Theory* 28, 730–768.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Koul, H., V. Susarla, & J. Van Ryzin (1981) Regression analysis with randomly right censored data. *Annals of Statistics* 42, 1276–1288.
Lewbel, A. & O. Linton (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. & W. Härdle (1996) Estimation of additive regression models with known links. *Biometrika* 83, 529–540.
Linton, O., E. Mammen, J.P. Nielsen, & I. Van Keilegom (2011) Nonparametric regression with filtered data. *Bernoulli* 17, 60–87.
Lu, X. & T.L. Cheng (2007) Randomly censored partially linear single-index models. *Journal of Multivariate Analysis* 98, 1895–1922.
Mack, Y.P. & B.W. Silverman (1982) Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields* 61, 405–415.
Martinsek, A. (1989) Almost Sure Expansion for M-estimators and S-estimators in Regression. Technical Report 25, Dept. Statistics, University of Illinois.
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. & D. Pollard (1987) U-processes: Rates of convergence. *Annals of Statistics* 15, 780–799.
Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
Peng, L. & Y. Huang (2008) Survival analysis with quantile regression models. *Journal of the American Statistical Association* 103, 637–649.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
Portnoy, S. (1997) Local asymptotics for quantile smoothing splines. *Annals of Statistics* 25, 414–434.
Portnoy, S. (2003) Censored regression quantiles. *Journal of the American Statistical Association* 98, 1001–1012.
Ritov, Y. (1990) Estimation in a linear regression model with censored data. *Annals of Statistics*. 18, 27–41.
Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley.
Spierdijk, L. (2008) Nonparametric conditional hazard rate estimation: A local linear approach. *Computational Statistics and Data Analysis* 52, 2419–2434.
Stone, C.J. (1980) Optimal rates of convergence for nonparametric estimators. *Annals of Statistics* 8, 1348–1360.
Van Keilegom, I. & 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. & L. Wang (2009) Locally weighted censored quantile regression. *Journal of the American Statistical Association* 104, 1117–1128.
Wang, Y., S. He, L. Zhu, & K.C. Yuen (2007) Asymptotics for a censored generalized linear model with unknown link function. *Probability Theory and Its Related Fields* 138, 235–267.
Wu, T.Z., K. Yu, & Y. Yu (2010) Single-index quantile regression. *Journal of Multivariate Analysis* 101, 1607–1621.
Wu, W.B. (2005) On the Bahadur representation of sample quantiles for dependent sequences. *Annals of Statistics* 33, 1934–1963.
Xia, Y., D. Zhang, & J. Xu (2010) Dimension reduction and semi-parametric estimation of survival models. *Journal of the American Statistical Association* 105, 278–290.
Ying, Z., S.H. Jung, & L.J. Wei (1995) Survival analysis with median regression models. *Journal of the American Statistical Association* 90, 178–184.
Yu, K. & Z. Lu (2004) Local linear additive quantile regression. *Scandinavian Journal of Statistics* 31, 333–346.
Zhou, Z. & W. Wu (2009) Local linear quantile estimation for non-stationary time series. *Annals of Statistics* 37, 2696–2729.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

APPENDIX

PROPOSITION A.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 $Pr(\liminf F_n) = 1$, where*

$$F_n = \left\{ K_1 n^{1-\kappa p} \le N_n(\mathbf{X}_j) \le K_2 n^{1-\kappa p}, \quad \text{for all } j = 1, \ldots, n \right\}.$$

which can be strengthened as

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

Similarly, we have under Assumptions A2 and A3,

$$\sup_{\mathbf{x} \in \mathcal{D}} \left| \Sigma_n(\mathbf{x}) - f_{\varepsilon|\mathbf{X}}(0|\mathbf{x}) \Sigma(A) - \frac{\delta_n}{f(\mathbf{x})} \sum_{l=1}^{p} \Sigma_l^* f_{\varepsilon,\mathbf{X}}^{(l)}(0, \mathbf{x}) \right| = O\left(\delta_n^{s_3}\right), \quad a.s., \tag{A.2}$$

where $f_{\varepsilon,\mathbf{X}}$ denotes the joint probability density function of $(\varepsilon, \mathbf{x})$, $f_{\varepsilon,\mathbf{X}}^{(i)}$, $i = 1, \ldots, p$, its first-order partial derivatives, and for each $1 \le l \le p$, Σ_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 kth column of the $p \times p$ identity matrix; and under Assumptions A1 and A6,

$$\sup_{\mathbf{x} \in \mathcal{D}} |\tilde{\Sigma}_n(\mathbf{x}) - f(\mathbf{x}) \Sigma(A_1)| = O\left(\left(n h_n^p / \log n \right)^{-1/2} + h_n \right) \quad a.s. \tag{A.3}$$

The proof follows directly from application of the Glivenko–Cantelli theorem. Using the von Neumann expansion for the inverse matrix, we further have

$$[\Sigma_n(\mathbf{x})]^{-1} = f_{\varepsilon|\mathbf{X}}^{-1}(0|\mathbf{x})[\Sigma(A)]^{-1} + \delta_n \frac{[\Sigma(A)]^{-1} \sum_{l=1}^{p} \Sigma_l^* f_{\varepsilon,\mathbf{X}}^{(l)}(0, \mathbf{x})[\Sigma(A)]^{-1}}{f_{\varepsilon|\mathbf{X}}^2(0|\mathbf{x}) f(\mathbf{x})} + O\left(\delta_n^{s_3} + \delta_n^2\right), \tag{A.4}$$

$$[\Sigma_n(\mathbf{x})]^{-1} f(\mathbf{x}) = \frac{f(\mathbf{x})}{f_{\varepsilon|\mathbf{X}}(0|\mathbf{x})}[\Sigma(A)]^{-1} + \delta_n \frac{[\Sigma(A)]^{-1} \sum_{l=1}^{p} \Sigma_l^* f_{\varepsilon,\mathbf{X}}^{(l)}(0, \mathbf{x})[\Sigma(A)]^{-1}}{f_{\varepsilon|\mathbf{X}}^2(0|\mathbf{x})} + O\left(\delta_n^{s_3} + \delta_n^2\right). \tag{A.5}$$

Proof of Lemma 4.1. This follows directly from (A.3), Theorem 2.1, and Theorem 2.3 of Gonzalez-Manteiga and Cadarso-Suarez (1994). Note that the fact that the weight $\tilde{B}_{nj}(.)$ might be negative does not affect the validity of the proof. ■

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

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

Fact F1. Then under Assumption 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 Fact F1 is straightforward; also see Chaudhuri (1991a). We now move on to derive the matrix form of $\hat{\mathbf{c}}_n(\mathbf{x})$, $\mathbf{x} = \mathbf{X}_j,\ 1 \le j \le n$, as in Chaudhuri (1991a,1991b).

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

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

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

Fact F2. If $DX_n(\mathbf{x})$ has rank $n(A)$, then there is a subset $\mathbf{h} \in H_n(\mathbf{x})$, such that (7) has at least one minima of the form

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

Fact F3. For the $\mathbf{h}$ specified in Fact 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

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

where $\bar{\mathbf{h}} = S_n(\mathbf{x}) \backslash \mathbf{h}$ denotes its complement in $S_n(\mathbf{x})$, $\mathrm{sign}(a)$ 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 (7) iff $L_n(\mathbf{x}, \mathbf{h}) \in (\tau - 1, \tau)^{n(A)}$.

Remark. Noticing the linearity of the loss function $\rho_\tau(.)$, Facts F2 and F3 can be proved in exactly the same manner as Theorems 3.1 and 3.3 in Koenker and Bassett (1978); see Chaudhuri (1991b) for parallel results. Note that the form of $\hat{\mathbf{c}}_n(\mathbf{x})$ specified in Fact F2 is free from the K-M estimator $\hat{G}_n(.)$, 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 (7) 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 Fact F3, the necessary and sufficient condition $\mathbf{h}$ has to satisfy, which does involves $\hat{G}_n(.)$, and thus is different from Fact 6.4 in Chaudhuri (1991b). 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 $a_1 \ne a_2$, 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.

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Under Assumption A2, we have for any $\mathbf{x} \in \mathcal{D}$, $k = [s_2]$, 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 kth order Taylor polynomial

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

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

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

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

$$\hat{Q}_n(\mathbf{x} + \mathbf{t}\delta_n, \mathbf{x}) = \mathbf{t}^\top \hat{\mathbf{c}}_n(\mathbf{x}).$$

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 $\mathbf{X}_j$, $j = 1, \ldots, n$, let U_n be the event defined as

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

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

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

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

$$Z_{ni}(\mathbf{x}) = \left[\frac{1}{2} - \frac{1}{2} \operatorname{sign} \left\{ \varepsilon_i - \Delta_n^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) + r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x}) \right\} - \tau \right] \mathbf{X}_{i,\mathbf{x}}(\delta_n, A), \tag{A.6}$$

where $r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x})$ is the remainder defined above. Using results (9) on the strong uniform consistency of K-M estimator, i.e.,

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

we have $W_n(\mathbf{h}) = W(\mathbf{h}) + o(1)$ a.s., where $W(\mathbf{h})$ is the diagonal matrix with elements $\{G(Y_i|\mathbf{X}_i), i \in \mathbf{h}\}$. Consequently, the assertion in Fact 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 $|L_{n1}(\mathbf{x}, \mathbf{h}) + L_{n2}(\mathbf{x}, \mathbf{h})| \le \phi_1$, where

$$L_{n1}(\mathbf{x}, \mathbf{h}) = \sum_{i \in \mathbf{h}} \{G(Y_i|\mathbf{X}_i)\}^{-1} Z_{ni}(\mathbf{x}) d_i,$$

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

where $Z_{ni}(\mathbf{x})$ is defined as in (A.6) with $\Delta_n = \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})$. First note that as a result of (9), (A.1), and assumptions that $\delta_n = o(h_n)$, we have $\sup_{\mathbf{x} \in \mathcal{D}} L_{n2}(\mathbf{x}, \mathbf{h}) = o\{(n\delta_n^p \log n)^{1/2}\}$ a.s. Therefore, from $|L_{n1}(\mathbf{x}, \mathbf{h}) + L_{n2}(\mathbf{x}, \mathbf{h})| \le \phi_1$, we have

$$L_{n1}(\mathbf{x}, \mathbf{h}) = o\left\{ (n\delta_n^p \log n)^{1/2} \right\} \quad \text{a.s.} \tag{A.7}$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Next, as $\mathbb{E}[d_i|\mathbf{X}_i, Y_i] = G(Y_i|\mathbf{X}_i)$, Proposition 6.1 in Chaudhuri (1991b) states that there exist positive constants $\epsilon_1^*, \epsilon_2^*, c_5^*$, and M_2^*, such that

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

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 $r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x}) = O(|\delta_n|^{s_2}) = o([n\delta_n^p/\log n]^{-1/2})$, for $\kappa \geq 1/(2s_3 + d)$, that there exists some constant $c_5 > 0$, such that

$$\left|\mathbb{E}\left[Z_{ni}(\mathbf{x})d_i/G(Y_i|\mathbf{X}_i)\right]\right| \geq c_5\left[n\delta_n^p/\log n\right]^{-1/2}. \tag{A.8}$$

Now (A.7) and (A.8) jointly imply that there exists some $K_1^* > 0$, such that $U_n(\mathbf{x})$ is contained in the event

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

$$\left.\text{with } \Delta_n = \hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x}), \hat{\mathbf{c}}_n(\mathbf{x}) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1}VY_n(\mathbf{x}, \mathbf{h}), \text{and } |\Delta_n| \geq K_1[n\delta_n^p/\log n]^{-1/2}\right\}.$$

Applying 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

$$P(U_n(\mathbf{x})) \leq c_6(n\delta_n^p)^{n(A)}\exp(-c_7\log n) \tag{A.9}$$

uniformly in $\mathbf{x} = \mathbf{X}_1, \ldots, \mathbf{X}_n$. 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 $\mathbf{X}_j$, $j = 1, \ldots, n$. The proof consists of the following steps.

Step 1. Define

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

and

$$R_n^{(1)}(\mathbf{x}) = \tilde{H}_n(\mathbf{x}, \delta_n, \hat{\mathbf{c}}_n(\mathbf{x})) - \tilde{H}_n(\mathbf{x}, \delta_n, \mathbf{c}_n(\mathbf{x})) - \Sigma_n(\mathbf{x})[\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})]. \tag{A.10}$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Then, as shown in Step 1 of Chaudhuri (1991b, p. 773), by Theorem, 4.2 and Assumptions A1 and A2, we have

$$\sup_j |R_n^{(1)}(\mathbf{X}_j)| = O\left\{[n^{(1-\kappa p)}/\log n]^{-(1+s_3)/2}\right\} \tag{A.11}$$

almost surely, which is of order $O\{[n^{(1-\kappa p)}/\log n]^{-3/4}\}$, if $s_3 \geq 1/2$.

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

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

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

$$\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}) = [DX_n(\mathbf{x},\mathbf{h})]^{-1}\,VY_n(\mathbf{x},\mathbf{h}), \qquad \hat{Q}_n^{\mathbf{h}}(\mathbf{X}_i,\mathbf{x}) = \left\{\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x})\right\}^{\top}\mathbf{X}_{i,\mathbf{x}}(\delta_n, A),$$

$$\chi_n^{\mathbf{h}}(\mathbf{x}) = \sum_{i\in\bar{\mathbf{h}}}\left[\frac{d_i}{G(Y_i|\mathbf{X}_i)}\mathbf{X}_{i,\mathbf{x}}(\delta_n, A)I\{Y_i \leq \hat{Q}_n^{\mathbf{h}}(\mathbf{X}_i,\mathbf{x})\} - \tilde{H}_n(\mathbf{x},\delta_n,\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}))\right]$$
$$- \sum_{i\in\bar{\mathbf{h}}}\left[\frac{d_i}{G(Y_i|\mathbf{X}_i)}\mathbf{X}_{i,\mathbf{x}}(\delta_n, A)I\{Y_i \leq Q_n(\mathbf{X}_i,\mathbf{x})\} - \tilde{H}_n(\mathbf{x},\delta_n,\mathbf{c}_n(\mathbf{x}))\right].$$

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

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

for large enough n, where $K_4 = K_3/2$, and we have implicitly used the fact that $[\log n]^{3/4}n^{(1-\kappa p)/4} \to \infty$ and that $\sharp(\mathbf{h}) = p$. As argued in Chaudhuri (1991b), 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 i.i.d. 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 $I(.)$ in the definition of $\chi_n^{\mathbf{h}}(\mathbf{x})$ causes the terms in the sums to act in a similar way as a random vector with binomial components. As $G(.)$ 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 (A.9); i.e., there exist constants $c_8 > 0, c_9 > 0$ such that

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

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

$$\sup_j |\chi_n^{\mathbf{h}}(\mathbf{X}_j)| = O\left([\log n],^{3/4}\, n^{(1-\kappa p)/4}\right). \tag{A.12}$$

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Step 3. Combining (A.10), (A.11) and (A.12), we have

$$\begin{aligned}
&\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i}{G(Y_i|\mathbf{X}_i)}\mathbf{X}_{ij}(\delta_n,A)[I\{Y_i\le Q_n(\mathbf{X}_i,\mathbf{X}_j)\}-\tau]\\
&\quad=\frac{1}{N_n(\mathbf{X}_j)}\chi_n^{\mathbf{h}}(\mathbf{X}_j)+\tilde{H}_n(\mathbf{X}_j,\delta_n,\hat{\mathbf{c}}_n(\mathbf{X}_j))-\tilde{H}_n\big(\mathbf{X}_j,\delta_n,\mathbf{c}_n(\mathbf{X}_j)\big)\\
&\qquad-\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i}{G(Y_i|\mathbf{X}_i)}\mathbf{X}_{ij}(\delta_n,A)\big[I\big\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j)\big\}-\tau\big]\\
&\quad=\Sigma_n(\mathbf{X}_j)[\hat{\mathbf{c}}_n(\mathbf{X}_j)-\mathbf{c}_n(\mathbf{X}_j)]+O\left\{\left[n^{(1-\kappa p)}/\log n\right]^{-3/4}\right\}\\
&\qquad-\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i}{\hat{G}_n(Y_i|\mathbf{X}_i)}\mathbf{X}_{ij}(\delta_n,A)\left[I\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j)\}-\tau\right]\\
&\qquad-\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}d_i\left[\frac{1}{G(Y_i|\mathbf{X}_i)}-\frac{1}{\hat{G}_n(Y_i|\mathbf{X}_i)}\right]\\
&\qquad\times\mathbf{X}_{ij}(\delta_n,A)\left[I\left\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j)\right\}-\tau\right] \qquad \textbf{(A.13)}
\end{aligned}$$

uniformly for $\mathbf{x}=\mathbf{X}_j,\ \ j=1,\dots,n$. Note that according to Fact F3, the second term on the right-hand side of (A.13) is of order $O(n^{\kappa p-1})=o\{[n^{(1-\kappa p)}/\log n]^{-3/4}\}$. It remains to show that the last terms equal

$$\frac{[\Sigma_n(\mathbf{X}_j)]^{-1}}{N_n(\mathbf{X}_j)}\sum_{k=1}^{n}T_n(\zeta_j,\zeta_k)\Sigma^{-1}(A_1)\mathbf{e}_1+R_n(\mathbf{X}_j)+O\left\{\left[n^{(1-\kappa p)}/\log n\right]^{-3/4}\right\}.$$

This is accomplished by the arguments immediately below and Lemma A.2. ■

Define

$$\begin{aligned}
\mathcal{F}_{\mathbf{x}}(\zeta_i,\zeta_k)=&\frac{d_iI\{\mathbf{X}_i\in S_n(\mathbf{x})\}\mathbf{X}_{i,\mathbf{x}}(\delta_n,A)}{f(\mathbf{X}_i)G(Y_i|\mathbf{X}_i)}[I\{Y_i\le Q_n(\mathbf{X}_i,\mathbf{x})\}-\tau]\\
&\times[\mathbf{X}_{ki}(h_n,A_1)]^{\top}\xi(Y_k,d_k,Y_i,\mathbf{X}_i)I\{\mathbf{X}_k\in S_n(\mathbf{X}_i)\}.
\end{aligned}$$

Now $\mathbb{E}_i[\mathcal{F}_{\mathbf{x}}(\zeta_i,\zeta_k)]$ stands for expectation taken with respect to the joint distribution of $(\mathbf{X}_i,Y_i)$ with the other argument held fixed. Then based on (A.3), Theorem 4.3 would follow if we can show that

LEMMA A.2. *With probability one,*

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

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

Proof. Let $\gamma_n=\log n/(n\delta_n^p)$. The proof consists of the following steps.

Step 1. According to (9), Assumption A7, and the facts that $G(.)$ is bounded below from zero and $|\mathbf{X}_{ij}(\delta_n,A)|\le 1$ for all $i\in S_n(\mathbf{X}_j)$, it is obvious that

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

$$\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i\{\hat{G}_n(Y_i|\mathbf{X}_i)-G(Y_i|\mathbf{X}_i)\}}{G(Y_i|\mathbf{X}_i)\hat{G}_n(Y_i|\mathbf{X}_i)}\mathbf{X}_{ij}(\delta_n,A)\Big[I\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j)\}-\tau\Big]$$
$$=\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i\mathbf{X}_{ij}(\delta_n,A)\}}{G^2(Y_i|\mathbf{X}_i)}\Big[I\big\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j)\big\}-\tau\Big]$$
$$\times\big\{\hat{G}_n(Y_i|\mathbf{X}_i)-G(Y_i|\mathbf{X}_i)\big\}+O\left(\gamma_n^{3/4}\right), \tag{A.14}$$

Step 2. Replacing $I\{Y_i\le\hat{Q}_n(\mathbf{X}_i,\mathbf{X}_j\}$ above with $I\{Y_i\le Q_n(\mathbf{X}_i,\mathbf{X}_j)\}$, we have

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

uniformly in $j=1,\ldots,n$, the proof of which is left in Lemma A.3.

Step 3. Using the result in Lemma 4.1, under Assumption A7, with probability one,

$$\hat{G}_n(t|\mathbf{x})-G(t|\mathbf{x})=\frac{G(t|\mathbf{x})}{nh_n^p f(\mathbf{x})}\sum_k I\{\mathbf{X}_k\in S_n(\mathbf{x})\}\xi(Y_k,d_k,t,\mathbf{x})[\mathbf{X}_{k,\mathbf{x}}(h_n,A_1)]^\top$$
$$\times[\Sigma(A_1)]^{-1}\mathbf{e}_1+O\left(\gamma_n^{3/4}\right)$$

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

$$\frac{1}{N_n(\mathbf{X}_j)}\sum_{i\in S_n(\mathbf{X}_j)}\frac{d_i\mathbf{X}_{ij}(\delta_n,A)}{G^2(Y_i|\mathbf{X}_i)}[I\{Y_i\le Q_n(\mathbf{X}_i,\mathbf{X}_j)\}-\tau]\big\{\hat{G}_n(Y_i|\mathbf{X}_i)-G(Y_i|\mathbf{X}_i)\big\}$$
$$=\frac{1}{nh_n^p N_n(\mathbf{X}_j)}\sum_{i,k=1}^n\mathcal{F}_{\mathbf{X}_j}(\zeta_i,\zeta_k)[\Sigma(A_1)]^{-1}\mathbf{e}_1+O\left(\gamma_n^{3/4}\right) \tag{A.15}$$

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

Step 4. We will show that with probability one,

$$\frac{1}{nh_n^p N_n(\mathbf{x})}\sum_{i,k=1}^n\mathcal{F}_{\mathbf{x}}(\zeta_i,\zeta_k)=\frac{1}{h_n^p N_n(\mathbf{x})}\sum_{k=1}^n E_i[\mathcal{F}_{\mathbf{x}}(\zeta_i,\zeta_k)]+o\left(\big[\log n/N_n(\mathbf{x})\big]^\alpha\right)$$

uniformly in $\mathbf{x}\in\mathcal{D}$, for any $\alpha<1$.

First, 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 (A.16) can be written as a U-statistic plus an asymptotically negligible term,

$$\frac{1}{n(n-1)}\sum_{i\ne k}\mathcal{F}_{\mathbf{x}}(\zeta_i,\zeta_k)=\frac{1}{2n(n-1)}\sum_{i\ne k}\mathcal{H}_{\mathbf{x}}(\zeta_i,\zeta_k), \tag{A.16}$$

where $\mathcal{H}_{\mathbf{x}}(.,.)$ 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 $\mathcal{H}_{\mathbf{x}}(.,.)$ defined as

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

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

where $\mathbb{E}_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k)$ stands for taking expectation w.r.t ζ_i with ζ_k held fixed. Since

$$\mathbb{E}_k \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) = \mathbb{E}_k[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) + \mathcal{F}_{\mathbf{x}}(\zeta_k, \zeta_i)] = \mathbb{E}_k[\mathcal{F}_{\mathbf{x}}(\zeta_k, \zeta_i)],$$
$$\mathbb{E}_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) = \mathbb{E}_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)], \qquad \mathbb{E}\mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) = 0.$$

We thus have

$$\begin{aligned}\sum_{i \neq k} \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) &= \sum_{i \neq k} \mathcal{H}^0_{\mathbf{x}}(\zeta_i, \zeta_k) + \sum_{i \neq k} \mathbb{E}_i \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) + \sum_{i \neq k} \mathbb{E}_k \mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) \\ &\quad - \sum_{i \neq k} \mathbb{E}\mathcal{H}_{\mathbf{x}}(\zeta_i, \zeta_k) \\ &= 2(n-1) \sum_{k=1}^{n} \mathbb{E}_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)] + \sum_{i \neq k} \mathcal{H}^0_{\mathbf{x}}(\zeta_i, \zeta_k). \end{aligned} \tag{A.17}$$

For the third term, to apply Proposition 4 in Arcones (1995), we need to verify that the class of functions $\{\mathcal{H}^0_{\mathbf{x}}(.,.) : \mathbf{x} \in \mathcal{D}\}$ is Euclidean with constant envelope, referred to as the uniformly bounded VC subgraph class in Arcones. We argue as follows. First, the class of functions $\{\mathcal{F}_{\mathbf{x}}(.,.) : \mathbf{x} \in \mathcal{D}\}$ is uniformly bounded. Second, as $\Sigma^{-1}(A_1)\tilde{B}_{h_n}(\mathbf{X}_{ki})q(Y_k, Y_i, \mathbf{X}_i)$ 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) $(I\{\mathbf{X}_i \in S_n(\mathbf{x})\}\mathbf{X}_{i,\mathbf{x}}(\delta_n, A) : \mathbf{x} \in \mathcal{D})$, (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 $I(.)$ is of bounded variation.

Therefore, according to Proposition 4 in Arcones (1995), noting that $|\mathcal{H}^0_{\mathbf{x}}| = O\left(\delta_n^p h_n^p\right)$, there exists some constant $c_0 > 0$, such that for any $\epsilon > 0$ and $1 > \alpha > 0$,

$$\Pr\left\{\max_{\mathbf{x} \in \mathcal{D}} \left| \sum_{i \neq k} \mathcal{H}^0_{\mathbf{x}}(\zeta_i, \zeta_k) \right| \geq \epsilon n^2 \delta_n^p h_n^p (\gamma_n)^{3/4} \right\} < 2\exp\left(-c_0 n^2 (\gamma_n)^{3/4}\right) = o\left(n^{-2}\right).$$

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

$$\max_{\mathbf{x} \in \mathcal{D}} \frac{1}{n N_n(\mathbf{x})} \left| \sum_{i \neq k} \mathcal{H}^0_{\mathbf{x}}(\zeta_i, \zeta_k) \right| = o\left(\gamma_n^{3/4}\right), \quad \text{for any } \alpha < 1.$$

This, together with (A.16) and (A.17), leads to

$$\begin{aligned}\frac{1}{n N_n(\mathbf{x})} \sum_{i,k=1}^{n} \mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k) &= \frac{1}{N_n(\mathbf{x})} \sum_{k=1}^{n} \mathbb{E}_i[\mathcal{F}_{\mathbf{x}}(\zeta_i, \zeta_k)] \\ &\quad + o\left\{(n\delta_n^p / \log n)^{-3/4}\right\} \quad \text{a.s.},\end{aligned}$$

for any $\alpha < 1$, where the $o(.)$ uniform in $\mathbf{x} \in \mathcal{D}$. This completes the proof. ■

LEMMA A.3. *Under conditions in Theorem 4.3, we have*

$$\sum_{i \in S_n(\mathbf{X}_j)} \frac{d_i \mathbf{X}_{ij}(\delta_n, A)}{G^2(Y_i|\mathbf{X}_i)} \left[I\left\{Y_i \leq \hat{Q}_n(\mathbf{X}_i, \mathbf{X}_j)\right\} - I\left\{Y_i \leq Q_n(\mathbf{X}_i, \mathbf{X}_j)\right\} \right]$$
$$\times \left\{\hat{G}_n(Y_i|\mathbf{X}_i) - G(Y_i|\mathbf{X}_i)\right\} = O\left(n^{1-\kappa p} \gamma_n^{3/4}\right). \tag{A.18}$$

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

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press

Proof. Based on (9) and the fact that $|\mathbf{X}_{ij}(\delta_n, A)| \le 1$, $G(.)$ is bounded away from zero, the term in (A.18) is bounded by $O\{(nh_n^p/\log n)^{-1/2}\}$ multiplied by

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

where

$$\begin{aligned}
I_{ni}(\mathbf{x}) = \Big[& r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x}) - |\{\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})\}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A)|, r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x}) \\
& + \left| \left\{\hat{\mathbf{c}}_n(\mathbf{x}) - \mathbf{c}_n(\mathbf{x})\right\}^\top \mathbf{X}_{i,\mathbf{x}}(\delta_n, A) \right| \Big],
\end{aligned}$$

and $D_n = \left[-K_1 \gamma_n^{1/2}, K_1 \gamma_n^{1/2}\right]$, for some $K_1 > 0$, and the last equality follows from Theorem 4.2 and the fact that $|r_n(\mathbf{X}_{i,\mathbf{x}}, \mathbf{x})| = O(\delta_n^{s_2}) = o\left(\gamma_n^{1/2}\right)$.

Since $nh_n^{2p}/(\delta_n^p \log n) \to \infty$ (Assumption A7), $\mathbb{E}I\{\varepsilon_i \in \mathcal{D}_n\} = O\left(\gamma_n^{1/2}\right) = o\left\{(nh_n^p/\log n)^{1/2} \gamma_n^{3/4}\right\}$, obviously (A.18) will follow if we can show that

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

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

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

Applying Bernstein's inequality to $U_n(\mathbf{x})$, we have

$$\Pr\left[U_n(\mathbf{x})) \le 2 \exp\left\{ -\frac{K_2^2 nh_n^p \log n/\gamma_n^{1/2}}{4K_2 n^{1-\kappa p} \gamma_n^{1/2} + 2K_2 \left(nh_n^p \log n\right)^{1/2}/\gamma_n^{1/4}} \right\} \right] = o\left(n^{-2}\right);$$

i.e., $\sum n P(U_n(\mathbf{x})) < \infty$, for sufficiently large K_2. This, according to the Borel-Cantelli lemma, leads to (A.18). ■

https://doi.org/10.1017/S0266466612000813 Published online by Cambridge University Press