Ann Inst Stat Math (2012) 64:765–790
DOI 10.1007/s10463-011-0335-8

Empirical likelihood for conditional quantile with left-truncated and dependent data

Han-Ying Liang · Jacobo de Uña-Álvarez

Received: 30 July 2009 / Revised: 23 March 2011 / Published online: 17 July 2011

Abstract In this paper, we employ the method of empirical likelihood to construct confidence intervals for a conditional quantile in the presence and absence of auxiliary information, respectively, for the left-truncation model. It is proved that the empirical likelihood ratio admits a limiting chi-square distribution with one degree of freedom when the lifetime observations with multivariate covariates form a stationary α-mixing sequence. For the problem of testing a hypothesis on the conditional quantile, it is shown that the asymptotic power of the test statistic based on the empirical likelihood ratio with the auxiliary information is larger than that of the one based on the standard empirical likelihood ratio. The finite sample performance of the empirical likelihood confidence intervals in the presence and absence of auxiliary information is investigated through simulations.

Keywords Empirical likelihood · Conditional quantile · Truncated data · α-mixing · Auxiliary information

1 Introduction

Let Y be a response variable with continuous distribution function (df) $\tilde{F}(\cdot)$ and let $\mathbf{X}$ be a random vector of covariates taking its values in $\mathbb{R}^d\,(d \geq 1)$ with joint density $l(\cdot)$.

H.-Y. Liang (✉)
Department of Mathematics, Tongji University,
Shanghai 200092, People's Republic of China
e-mail: hyliang83@yahoo.com

H.-Y. Liang · J. de Uña-Álvarez
Department of Statistics and OR, Facultad de Ciencias Económicas y Empresariales,
Universidad de Vigo, Campus Lagoas-Marcosende, Vigo 36310, Spain
e-mail: jacobo@uvigo.es

Throughout the paper, $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. For any $\mathbf{x}$, the conditional df of Y given $\mathbf{X} = \mathbf{x}$, is $F(y|\mathbf{x}) = \mathbb{E}[I(Y \leq y)|\mathbf{X} = \mathbf{x}]$, which can be written as

$$F(y|\mathbf{x}) = \int_{-\infty}^{y} f(\mathbf{x}, t)\, \mathrm{d}t / l(\mathbf{x}) \overset{\text{def}}{:=} \frac{F_1(\mathbf{x}, y)}{l(\mathbf{x})}, \tag{1}$$

where $f(\cdot, \cdot)$ is the probability density function of $(\mathbf{X}, Y)$ (assumed to exist), and $l(\cdot)$ is assumed to be positive at $\mathbf{x}$. In the context of regression, it is of interest to estimate $F(y|\mathbf{x})$ and/or the pertaining quantile function $\theta_q = \inf\{y : F(y|\mathbf{x}) \geq q\}$ for $q \in (0, 1)$. Indeed, it is well known that the conditional quantile functions (especially the conditional median function) can give a good description of the data (cf. Chaudhuri et al. 1997), because of their robustness to heavy-tailed error distributions and outliers. Many authors have considered this problem under random sampling; for example Mehra et al. (1991), Fan et al. (1994), and Xiang (1996).

In practice, the response variable Y may be subject to random censoring and/or truncation. This is the case, for example, in regression models with a lifetime as response variable. Under right-censoring, Dabrowska (1992) established a Bahadur-type representation of the kernel quantile estimator; see also Van Keilegom and Veraverbeke (1998) for the fixed design regression framework or Iglesias-Pérez (2003) for the inclusion of left-truncation. Furthermore, for the censored setup, Xiang (1995) obtained the deficiency of the sample quantile estimator with respect to a kernel estimator by using the coverage probability. Ould-Saïd (2006) constructed a kernel estimator of the conditional quantile under censoring, and established its strong uniform convergence rate; Liang and de Uña-Álvarez (2011) proved the strong uniform convergence and asymptotic normality of this estimator under dependence assumptions.

In this paper we are interested in the random left truncation model. Left-truncated data occur in astronomy, economics, epidemiology and biometry; see Woodroofe (1985), Feigelson and Babu (1992) and He and Yang (1994). Recently, Ould-Saïd and Tatachak (2007) constructed a new kernel estimator of the conditional density for the left-truncation model in the independent data setting. But dependent data may arise in applications; for example, when sampling clustered lifetimes (family members, or repeated measurements), see Cai et al. (2000).

There is some literature devoted to conditional df and conditional quantile estimation under dependence. To mention some examples, Cai (2002) investigated the asymptotic normality and the weak convergence of a weighted Nadaraya–Watson conditional df and quantile estimator for α-mixing time series. Honda (2000) dealt with α-mixing processes and proved the uniform convergence and asymptotic normality of an estimate of θ_q using the local polynomial fitting method. Ferraty et al. (2005) considered quantile regression under dependence when the conditioning variable is infinite dimensional. Non-parametric conditional median predictors for time series based on the double kernel method and the constant kernel method were proposed by Gannoun et al. (2003). A nice extension of the conditional quantile process theory to set-indexed processes under strong mixing was established in Polonik and Yao (2002). In addition, Zhou and Liang (2000) reported asymptotic analysis of a kernel conditional median estimator for dependent data. Lecoutre and Ould-Saïd (1995) provided the uniform strong consistency of a kernel-type estimator of the conditional df under censoring and strong mixing conditions. However, to the best of our knowledge, empirical likelihood confidence intervals for the conditional quantile with truncated and dependent data have not been investigated so far.

In some instances, some auxiliary information about the conditional distribution function is available in the sense that there exist $\kappa(\kappa \geq 1)$ functions $g_1(y), \ldots, g_\kappa(y)$ such that

$$\Psi(\mathbf{x}) = \mathbb{E}(g(Y)|\mathbf{X} = \mathbf{x}) = 0, \tag{2}$$

where $g(y) = (g_1(y), \ldots, g_\kappa(y))^\tau$ is an κ-dimensional vector. This model is of interest in many circumstances where some partial information about the conditional distribution of the sample is known. For example, for given $\mathbf{x}$, if the conditional mean $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ is known, we have (2) by taking $g(y) = y - m(\mathbf{x})$; if the conditional distribution is symmetric about a known constant y_0 for a given $\mathbf{x}$, then (2) holds for $g(y) = I(y \geq y_0) - \frac{1}{2}$; finally, if one knows that a proportion p_0 of the responses falls in a given interval (a, b), i.e., $\mathbb{P}(a < Y < b|\mathbf{X} = \mathbf{x}) = p_0$, then one can take $g(y) = I(a < y < b) - p_0$.

The empirical likelihood (EL) was introduced by Owen (1988, 1990) for a mean vector for i.i.d. observations, and has many advantages over normal approximation-based methods and the bootstrap for constructing confidence intervals (see Hall 1992; Hall and La Scala 1990). For example, the EL confidence intervals do not have a predetermined shape, whereas confidence intervals based on the asymptotic normality of an estimator have a symmetry implied by asymptotic normality; on the basis of the coverage probability, the EL method is a competitive and favorable method which outperforms the Wald-type method and can overcome the under-coverage probability problem for small sample size (see Zhou and Li 2008); one of the nice features of the EL method particularly appreciated in censored data analysis is that one can construct confidence intervals without estimating the variance of the statistic; moreover, it has better performance than the traditional normal approximation (Wald) method (see Zhao 2011). Furthermore, the EL confidence intervals respect the range of the parameter: if the parameter is positive, then the confidence interval contains no negative values. Another preferred characteristic is that the EL confidence interval is transformation respecting. For complete data setting, the EL methods have been studied extensively by many authors. We refer the reader to Chen and Hall (1993) for confidence intervals of a quantile; further investigations have been carried out by Zhou and Jing (2003). Quantile estimation in the context of survey sampling is considered in Chen and Wu (2002), while confidence intervals of quantiles for weakly dependent data were constructed by Chen and Wong (2009). Again, an attractive feature of the EL is that it can be used to make sharper inferences when some auxiliary information is available (see Chen and Qin 1993; Zhang 1997; Qin and Wu 2001).

In this paper, we propose using the EL for the construction of confidence intervals for θ_q in the presence and absence of auxiliary information (2), respectively, for the left-truncation model. It is proved that the EL ratio admits a limiting chi-square distribution with one degree of freedom when the lifetime observations with multivariate covariates form a stationary α-mixing sequence. Hypothesis test for θ_q shows that the asymptotic power of the test statistic based on the EL ratio with the auxiliary information (2) is larger than that based on the standard empirical likelihood ratio. In addition, we investigate through simulations the finite sample performance of the empirical likelihood confidence intervals in the presence and absence of auxiliary information.

Recall that a sequence $\{\xi_i, i \geq 1\}$ is said to be α-mixing if the α-mixing coefficient

$$\alpha(m) := \sup_{k\geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{m+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $m \to \infty$, where $\mathcal{F}_l^m$ denotes the σ-algebra generated by $\xi_l, \xi_{l+1}, \ldots, \xi_m$ with $l \leq m$. Among various mixing conditions used in the literature, α-mixing is reasonably weak and is known to be fulfilled for many stochastic processes including many time series models. Withers (1981) derived the conditions under which a linear process is α-mixing. In fact, under very mild assumptions, linear autoregressive and more generally bilinear time series models are strongly mixing with mixing coefficients decaying exponentially, i.e., $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$. See Doukhan (1994), p. 99, for more details. We only say that the α-mixing has been used in applications with clustered survival data; for instance, Cai and Kim (2003).

The rest of the paper is organized as follows. Section 2 introduces the empirical likelihood ratio for the conditional quantile in the presence and absence of auxiliary information, respectively. Main results are formulated in Sect. 3. A simulation study is presented in Sect. 4. Section 5 lists some preliminary lemmas, which are used in the proof of the main results. The proofs of the main results are given in Sect. 6. In Sect. 7, we collect some known results, which are used in the proof of the preliminary lemmas. The proofs of the preliminary lemmas are deferred to Sect. 8.

2 Estimator

In order to formalize things, let $\{(\mathbf{X}_k, Y_k, T_k), 1 \leq k \leq N\}$ be a sequence of random vectors distributed as $(\mathbf{X}, Y, T)$, where T is the truncation variable. For the components of $(\mathbf{X}, Y, T)$, in addition to the assumptions and notation for $\mathbf{X}$ and Y made at the beginning of the Sect. 1, we assume throughout that T and $(\mathbf{X}, Y)$ are independent, and that T has continuous df G. Let $F(\cdot, \cdot)$ be the joint df of the random vector $(\mathbf{X}, Y) \in \mathbb{R}^{d+1}$. Without loss of generality, we assume that Y and T are both nonnegative random variables, as usual in survival analysis. In the random left-truncation model, the lifetime Y_i is interfered by the truncation random variable T_i in such a way that both Y_i and T_i are observable only when $Y_i \geq T_i$, whereas neither is observed if $Y_i < T_i$ for $i = 1, \ldots, N$, where the N is the potential sample size. Due to the occurrence of truncation, the N is unknown, and n (the size of the actually observed sample) is random with $n \leq N$. Let $\mu = \mathbb{P}(Y \geq T)$ be the probability that the random variable Y is observable. Since $\mu = 0$ implies that no data can be observed, we suppose throughout the paper that $\mu > 0$. Note that the N is unknown and the n is known (although random); hence, our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the conditional probability P with respect to the actually observed n-sample instead. Furthermore, $\mathbb{E}$ and E will denote the expectation operators under $\mathbb{P}$ and P, respectively. In the sequel, the observed sample $\{(\mathbf{X}_i, Y_i, T_i), 1 \leq i \leq n\}$ is assumed to be a stationary α-mixing sequence.

Note that $C(y) = \mathbb{P}(T \leq y \leq Y | Y \geq T) = \mu^{-1} G(y)[1 - \tilde{F}(y)]$. Then the empirical estimator of $C(y)$ is defined by $C_n(y) = n^{-1} \sum_{i=1}^n I(T_i \leq y \leq Y_i)$. Following the idea of Lynden-Bell (1971), the non-parametric maximum likelihood estimators of the dfs $\tilde{F}$ and G are given by

$$1 - \tilde{F}_n(y) = \prod_{i:Y_i \leq y} \left(1 - \frac{1}{nC_n(Y_i)}\right) \quad \text{and} \quad G_n(y) = \prod_{i:T_i > y} \left(1 - \frac{1}{nC_n(T_i)}\right).$$

The estimator of μ is defined (cf. He and Yang 1998) by $\mu_n = G_n(y)[1 - \tilde{F}_n(y-)]C_n^{-1}(y)$, where $\tilde{F}_n(y-)$ denotes the left-limit of $\tilde{F}_n$ at y.

For any df W, let $a_W = \inf\{y : W(y) > 0\}$ and $b_W = \sup\{y : W(y) < 1\}$ be its two endpoints. Since T is independent of $(\mathbf{X}, Y)$, the conditional joint distribution of $(\mathbf{X}, Y, T)$ is given by

$$\begin{aligned} H^*(\mathbf{x}, y, t) &= P(\mathbf{X} \leq \mathbf{x}, Y \leq y, T \leq t) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}, Y \leq y, T \leq t | Y \geq T) \\ &= \mu^{-1} \int_{\mathbf{s} \leq \mathbf{x}} \int_{a_G \leq v \leq y} G(v \wedge t) F(\mathrm{d}\mathbf{s}, \mathrm{d}v). \end{aligned}$$

Taking $t = +\infty$, the observed pair $(\mathbf{X}, Y)$ then has the following df $F^*(\cdot, \cdot)$: $F^*(\mathbf{x}, y) = H^*(\mathbf{x}, y, \infty) = \mu^{-1} \int_{\mathbf{s} \leq \mathbf{x}} \int_{a_G \leq v \leq y} G(v) F(\mathrm{d}\mathbf{s}, \mathrm{d}v)$, which yields that

$$F(d\mathbf{x}, dy) = [\mu^{-1} G(y)]^{-1} F^*(\mathrm{d}\mathbf{x}, \mathrm{d}y) \quad \text{for} \quad y > a_G. \tag{3}$$

2.1 Standard empirical likelihood ratio

Note that, by using (A2)(i) and (A7) in Sect. 3, from (3) we have

$$\begin{aligned} &\frac{\mu}{h_n^d} E\left\{K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) G^{-1}(Y_i) \left(I(Y_i \leq \theta_q) - q\right)\right\} \\ &= \int_{\mathbb{R}^d} K(\mathbf{s}) l(\mathbf{x} - h_n \mathbf{s}) \left[\mathbb{E}\left(I(Y \leq \theta_q) | \mathbf{X} = \mathbf{x} - h_n \mathbf{s}\right) - q\right] \mathrm{d}\mathbf{s} \to 0, \end{aligned}$$

where K is some kernel function on $\mathbb{R}^d$, $(h_n)_{n \geq 1} \searrow 0$ as $n \nearrow \infty$.
This motivates the introduction of the empirical likelihood function $\bar{H} = \prod_{i=1}^n p_i$, where $p_1, \ldots, p_n$ are subject to the restrictions:

$$p_i \geq 0, \ \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) \left(I(Y_i \leq \theta_q) - q\right) = 0. \tag{4}$$

The maximum of $\bar{H}$ can be found via Lagrange multipliers. It may be shown that $\bar{H}_{\max} = \prod_{i=1}^n \bar{p}_i$, where

$$\bar{p}_i = \frac{1}{n} \frac{1}{1 + \lambda(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(I(Y_i \le \theta_q) - q)}, \quad i = 1, \ldots, n,$$

and $\lambda(\theta_q)$ is the solution of the following equation:

$$\sum_{i=1}^n \frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(I(Y_i \le \theta_q) - q)}{1 + \lambda(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(I(Y_i \le \theta_q) - q)} = 0.$$

We propose the following empirical log-likelihood ratio function at θ_q:

$$\bar{l}_n(\theta_q) = 2\sum_{i=1}^n \log\left[1 + \lambda(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(I(Y_i \le \theta_q) - q)\right].$$

2.2 Empirical likelihood ratio with auxiliary information

By using (3), (2) is equivalent to

$$E(g(Y)G^{-1}(Y)|\mathbf{X} = \mathbf{x}) = 0. \tag{5}$$

To make use of (2), i.e. (5), we introduce the empirical likelihood function $\tilde{H} = \prod_{i=1}^n p_i$, where $p_1, \ldots, p_n$ are subject to the restrictions:

$$p_i \ge 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i) = 0. \tag{6}$$

It may be shown that $\tilde{H}_{\max} = \prod_{i=1}^n \tilde{p}_i$, where

$$\tilde{p}_i = \frac{1}{n} \frac{1}{1 + \eta_1^\tau K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i)}, \quad i = 1, \ldots, n,$$

and η_1 is the solution of equation $\sum_{i=1}^n \frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i)}{1+\eta_1^\tau K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i)} = 0$. Meanwhile, we consider the empirical likelihood function $\hat{H} = \prod_{i=1}^n p_i$, where $p_1, \ldots, p_n$ are subject to the restrictions (4) and (6). It may be shown that $\hat{H}_{\max} = \prod_{i=1}^n \hat{p}_i$, where

$$\hat{p}_i = \frac{1}{n} \frac{1}{1 + \eta_2^\tau(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(g^\tau(Y_i), I(Y_i \le \theta_q) - q)^\tau}, \quad i = 1, \ldots, n,$$

and $\eta_2(\theta_q)$ is the solution of the following equation:

$$\sum_{i=1}^{n} \frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(g^\tau(Y_i), I(Y_i \le \theta_q) - q)^\tau}{1 + \eta_2^\tau(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(g^\tau(Y_i), I(Y_i \le \theta_q) - q)^\tau} = 0. \tag{7}$$

The empirical log-likelihood ratio function at θ_q with the auxiliary information (2) is defined by

$$\begin{aligned}
\hat{l}_n(\theta_q) &= -2 \log \prod_{i=1}^{n} \frac{\hat{p}_i}{\tilde{p}_i} \\
&= 2 \sum_{i=1}^{n} \log \left[1 + \eta_2^\tau(\theta_q) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) \left(g^\tau(Y_i), I(Y_i \le \theta_q) - q\right)^\tau \right] \\
&\quad -2 \sum_{i=1}^{n} \log \left[1 + \eta_1^\tau K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i) \right].
\end{aligned}$$

3 Main results

In the sequel, let C and c denote generic finite positive constants, whose values are unimportant and may change from line to line. Let $C(l)$ represent the set of continuity points of function l. The norm of a $n_1 \times n_2$ matrix $A = (a_{ij})_{n_1 \times n_2}$ is defined by $\|A\| = (\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} a_{ij}^2)^{1/2}$ for $n_1, n_2 \ge 1$. All limits are taken as the sample size n tends to ∞, unless specified otherwise. Let $U(\mathbf{x})$ represent a neighborhood of $\mathbf{x}$. Put $u(\mathbf{x}) = \mathbb{E}\{G^{-1}(Y)|\mathbf{X} = \mathbf{x}\}$, $v(\mathbf{x}) = \mathbb{E}\{I(Y \le \theta_q) G^{-1}(Y)|\mathbf{X} = \mathbf{x}\}$, $V(\mathbf{x}) = \mathbb{E}\{g(Y) g^\tau(Y) G^{-1}(Y)|\mathbf{X} = \mathbf{x}\}$ and $W(\mathbf{x}) = \mathbb{E}\{g(Y) G^{-1}(Y)[I(Y \le \theta_q) - q]|\mathbf{X} = \mathbf{x}\}$.

In order to formulate the main results, we need the following assumptions.

(A0) $a_G < a_{\tilde{F}}$ and $b_G < b_{\tilde{F}}$.

(A1) For all integers $j \ge 1$, the joint density $l_j^*(\cdot, \cdot)$ of $\mathbf{X}_1$ and $\mathbf{X}_{j+1}$ w.r.t. P exists on $\mathbb{R}^d \times \mathbb{R}^d$ and satisfies $l_j^*(\mathbf{s}, \mathbf{t}) \le C$ for $(\mathbf{s}, \mathbf{t}) \in U(\mathbf{x}) \times U(\mathbf{x})$.

(A2) (i) The kernel $K(\cdot)$ is a bounded function with compact support on $\mathbb{R}^d$;
(ii) $\int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1$; (iii) $\int_{\mathbb{R}^d} x_1^{i_1} \ldots x_d^{i_d} K(\mathbf{x}) d\mathbf{x} = 0$ for non-negative integers $i_1, \ldots, i_d$ with $i_1 + \cdots + i_d = 1$.

(A3) The sequence $\alpha(n)$ satisfies that
(i) there exist positive integers $\eta := \eta_n$ such that $\eta = o((nh_n^d)^{1/2})$ and $\lim_{n\to\infty} (nh_n^{-d})^{1/2} \alpha(\eta) = 0$;
(ii) there exist $r > 2$ and $\delta > 1 - 2/r$ such that $\sum_{l=1}^{\infty} l^\delta [\alpha(l)]^{1-2/r} < \infty$.

(A4) $n^{r-2} h_n^{rd} \ge c_0 > 0$, $n^{-1} h_n^{-d(1+4/r)} = O(1)$ and $nh_n^{d+4} \to 0$, where r is the same as in (A3).

(A5) For $1 \le k, l \le \kappa$, $j \ge 1$ and $(\mathbf{s}, \mathbf{t}) \in U(\mathbf{x}) \times U(\mathbf{x})$,
(i) $E(|g_k(Y_1) g_l(Y_1) g_k(Y_{1+j}) g_l(Y_{1+j})||\mathbf{X}_1 = \mathbf{s}, \mathbf{X}_{1+j} = \mathbf{t}) < \infty$, $E(|g_k(Y_1) g_k(Y_{1+j})||\mathbf{X}_1 = \mathbf{s}, \mathbf{X}_{1+j} = \mathbf{t}) < \infty$;

(ii) $E(|g_k(Y_1)g_l(Y_{1+j})||\mathbf{X}_1 = \mathbf{s}, \mathbf{X}_{1+j} = \mathbf{t}) < \infty$, $E(|g_k(Y_1)||\mathbf{X}_1 = \mathbf{s}, \mathbf{X}_{1+j} = \mathbf{t}) < \infty$ and $E(|g_k(Y_{1+j})||\mathbf{X}_1 = \mathbf{s}, \mathbf{X}_{1+j} = \mathbf{t}) < \infty$.

(A6) For $1 \le k, l \le \kappa$, $\mathbb{E}(|g_k(Y)g_l(Y)|^r|\mathbf{X} = \mathbf{s}) < \infty$ and $\mathbb{E}(|g_k(Y)|^r|\mathbf{X} = \mathbf{s}) < \infty$ for $\mathbf{s} \in U(\mathbf{x})$, where r is the same as in (A3).

(A7) (i) The second partial derivative of $\Psi(\mathbf{s})$ and $l(\mathbf{s})$ is bounded in $U(\mathbf{x})$;
(ii) The second partial derivative with respect to $\mathbf{s}$ of $F(\theta_q|\mathbf{s}))$ and $l(\mathbf{s})$ is bounded in $U(\mathbf{x})$.

(A8) (i) $u(\mathbf{s})$ and $v(\mathbf{s})$ are continuous at $\mathbf{x}$;
(ii) $V(\mathbf{s})$ is continuous at $\mathbf{x}$ and $V(\mathbf{x})$ is positive definite;
(iii) $W(\mathbf{s})$ is continuous at $\mathbf{x}$.

(A9) The matrix Σ_3 in Sect. 5 is positive definite.

(B1) For all integers $j \ge 1$, the joint density $l_j^*(\cdot,\cdot,\cdot,\cdot)$ of $(\mathbf{X}_1, \mathbf{X}_{j+1}, Y_1, Y_{j+1})$ w.r.t. P exists on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$ and satisfies $l_j^*(\mathbf{s}, \mathbf{t}, y_1, y_2) \le C$ for $(\mathbf{s}, \mathbf{t}, y_1, y_2) \in U(\mathbf{x}) \times U(\mathbf{x}) \times U(\theta_q) \times U(\theta_q)$.

(B2) $f(\mathbf{s}, y)$ is continuous at $(\mathbf{x}, \theta_q)$ and $f(\mathbf{x}, \theta_q) > 0$.

Remark 1 (a) Condition $a_G < a_{\tilde{F}}$ in (A0) implies $G(Y) \ge G(a_{\tilde{F}}) > 0$, which ensures $G_n(Y_i) \ne 0$ eventually, so the given empirical log-likelihood ratio functions are well defined for large n. Assumptions (A1), (A5) and (B1) are mainly technical, which are employed to simplify the calculations of covariances in the proof of Theorems 1 and 2 below, these assumptions are redundant for the independent setting.

(b) Assumptions (A3) and (A4) imply restrictions when choosing the bandwidth. In order to illustrate the practical implications of these assumptions, choose $h_n = cn^{-1/(d+4)}(logn)^{-1}$ so $nh_n^{d+4} \to 0$ is satisfied. Then, it is easily seen that this bandwidth satisfies the two former conditions in (A4), provided that $r > \max(4, 2 + d/2)$. Furthermore, assume $\alpha(n) = O(n^{-\lambda})$ for some λ ; then, (A3) automatically holds if λ is large enough, specifically $\lambda > \max(2d + 1, r(d+1)/(r-2))$ (note that λ can be arbitrarily large if $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$). We see that, in essence, (A3) imposes a relationship between the degree of dependence in the data and the dimension of the covariates, so high dimensional data could not be handled under strong dependence (i.e. small λ).

Theorem 1 *Suppose that (A0)–(A4), (A7)(ii) and (A8)(i) are satisfied. Let* $\alpha(n) = O(n^{-\gamma})$ *for some* $\gamma \ge [r(r+2)]/[2(r-2)]$. *Then* $\bar{l}_n(\theta_q) \xrightarrow{\mathcal{D}} \chi_1^2$, *further* $\bar{l}_n(\theta_n) \xrightarrow{\mathcal{D}} \chi_1^2(\Delta^2)$ *for constant* Δ *if (B1) and (B2) are satisfied, where* $\theta_n = \theta_q + (nh_n^d)^{-1/2}\frac{\{[(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})]\mu l(\mathbf{x})\int_{\mathbb{R}^d} K^2(\mathbf{s})d\mathbf{s}\}^{1/2}}{f(\mathbf{x},\theta_q)}\Delta$.

Remark 2 Together with the empirical likelihood ratio test, a Wald-type test statistic for the conditional quantile may be introduced by using the preliminary Lemmas in Sects. 5 and 6 below (specifically, Lemmas 5.3(d) and 7.4). However, as discussed in the Sect. 1, the EL approach avoids the problem of variance estimation while it performs better when the sample size is small. For these reasons we do not further develop the Wald approach here.

Theorem 2 *Suppose that (A0)–(A9) are satisfied and that* $\alpha(n) = O(n^{-\gamma})$ *for some* $\gamma \geq [r(r+2)]/[2(r-2)]$. *Then* $\hat{l}_n(\theta_q) \xrightarrow{\mathcal{D}} \chi_1^2$, *further* $\hat{l}_n(\theta_n) \xrightarrow{\mathcal{D}} \chi_1^2(\rho^2)$ *for constant* ρ *if (B1) and (B2) are satisfied, where* θ_n *is the same as in Theorem 1, and* $\rho^2 = \frac{(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})}{(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})-W^\tau(\mathbf{x})V^{-1}(\mathbf{x})W(\mathbf{x})}\Delta^2$.

Remark 3 (a) Choosing $0 < \delta < 1$, let c_δ satisfy $P(\chi_1^2 > c_\delta) = \delta$. Then, both of $\bar{I}_\delta = \{\theta | \bar{l}_n(\theta) \leq c_\delta\}$ and $\hat{I}_\delta = \{\theta | \hat{l}_n(\theta) \leq c_\delta\}$ are level $1-\delta$ asymptotic confidence intervals of θ_q. In view of Theorems 1–2, both of $\bar{I}_\delta$ and $\hat{I}_\delta$ have the correct asymptotic coverage probability $1-\delta$, and $\hat{I}_\delta$ reduces to $\bar{I}_\delta$ in the absence of the auxiliary information (2).

(b) Consider the hypothesis test: $H_0 : \theta_q = \theta_0$ versus $H_A : \theta_q = \theta_n$, where θ_n is defined as in Theorem 1 with θ_0 in the place of θ_q. In view of Theorems 1–2, $\{(\mathbf{X}_i, Y_i, T_i), 1 \leq i \leq n | \bar{l}_n(\theta_0) > c_\delta\}$ and $\{(\mathbf{X}_i, Y_i, T_i), 1 \leq i \leq n | \hat{l}_n(\theta_0) > c_\delta\}$ are level δ asymptotic rejective intervals of H_0. Furthermore, from Theorems 1–2, the power of the test is asymptotically, respectively

$$\lim_{n\to\infty} P_{\theta_n}(\bar{l}_n(\theta_0) > c_\delta) = \lim_{n\to\infty} P_{\theta_n}(\bar{l}_n(\theta_n - \Omega_n\Delta) > c_\delta) = P(\chi_1^2(\Delta^2) > c_\delta),$$
$$\lim_{n\to\infty} P_{\theta_n}(\hat{l}_n(\theta_0) > c_\delta) = \lim_{n\to\infty} P_{\theta_n}(\hat{l}_n(\theta_n - \Omega_n\Delta) > c_\delta) = P(\chi_1^2(\rho^2) > c_\delta),$$

where $\Omega_n = (nh_n^d)^{-1/2}\frac{\{[(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})]\mu l(\mathbf{x})\int_{\mathbb{R}^d} K^2(\mathbf{s})d\mathbf{s}\}^{1/2}}{f(\mathbf{x},\theta_q)}$. Using the fact that the non-central chi-squared distribution is stochastically increasing in its non-centrality parameter for any given degrees of freedom, we have $P(\chi_1^2(\rho^2) > c_\delta) \geq P(\chi_1^2(\Delta^2) > c_\delta)$ since $\rho^2 \geq \Delta^2$. In particular, if $W(\mathbf{x}) \neq 0$, then $\rho^2 > \Delta^2$, and hence $P(\chi_1^2(\rho^2) > c_\delta) > P(\chi_1^2(\Delta^2) > c_\delta)$. These facts indicate that the empirical likelihood ratio test with auxiliary information (2) has powers at least as large as the standard empirical likelihood ratio test which does not utilize auxiliary information (2), i.e., the asymptotic power of the test statistic based on the empirical likelihood ratio with the auxiliary information (2) is larger than the one based on the standard empirical likelihood ratio.

4 Simulation study

In this section, we carry out a simulation study to investigate the finite sample performance of the empirical likelihood confidence intervals for the quantile θ_q in the presence and absence of auxiliary information. We consider the cases $d = 1$ and $d = 2$, these are, a real valued covariate and a two-dimensional covariate, respectively.

(a) First, we consider the simulation in the one-dimensional case ($d = 1$). In order to obtain a sequence $\{X_i, Y_i, T_i\}$ fulfilling the α-mixing property after truncation, we generate the observed data as follows.

(1) Drawing of (X_1, Y_1, T_1):
Step 1 Draw $e_1 \sim N(0, 1)$, and take $X_1 = e_1$;
Step 2 Compute Y_1 from the model $Y_1 = 1.5X_1^2 + \sin(\pi X_1) + 0.6\epsilon_1$, where $\epsilon_1 \sim N(0, 1)$;

Step 3 Draw $T_1 \sim N(\beta, 1)$, where β is adapted in order to get different values of μ. If $Y_1 < T_1$, we reject the datum and go back to Step 2, do this until $Y_1 \geq T_1$.

(2) Drawing of (X_2, Y_2, T_2):

Step 4 Draw X_2 from the AR(1) models $X_2 = \rho X_1 + e_2$, where $e_2 \sim N(0, 1)$ and $|\rho| < 1$ is some constant;

Step 5 Compute Y_2 from the model $Y_2 = 1.5X_2^2 + \sin(\pi X_2) + 0.6\epsilon_2$, where $\epsilon_2 \sim N(0, 1)$;

Step 6 Draw $T_2 \sim N(\beta, 1)$. If $Y_2 < T_2$, we reject the datum and go back to Step 5, do this until $Y_2 \geq T_2$.

By repeating step (2) above, we generate the observed data $(X_i, Y_i, T_i), i = 1, \ldots, n$, where $X_i = \rho X_{i-1} + e_i, e_i \sim N(0, 1), Y_i = 1.5X_i^2 + \sin(\pi X_i) + 0.6\epsilon_i$ where ϵ_i are i.i.d. random variables with distribution $N(0, 1)$ and $T_i \sim N(\beta, 1)$, and everything is distributed conditionally on $Y_i \geq T_i$. Besides, the α-mixing property of the observable X_i is immediately transferred to the (X_i, Y_i, T_i). Obviously, the regression function is $m(x) = \mathbb{E}(Y|X = x) = 1.5x^2 + \sin(\pi x)$. Assume that $m(1) = 1.5$ is known. For illustration of the proposed empirical likelihood method, we use this as auxiliary information, by considering function $g(y) = y - 1.5$. The target is θ_q with $q = 0.5$ (i.e. the conditional median). Note that the true value of θ_q at $x = 1$ is 1.5. The sample size is $n = 300$. The Gaussian kernel for $K(\cdot)$, and the bandwidth $h_n = 0.3$, are used in the computations.

In Table 1 we report the coverage probabilities and average lengths of 95% confidence intervals for $\theta_q = 1.5$ at point $x = 1$, constructed from the empirical likelihood method with $(\hat{I}_\delta)$ and without $(\bar{I}_\delta)$ the auxiliary information, along $M = 1{,}000$ Monte Carlo replications. We take $\rho = 0.3$ and three different values of the truncation parameter: $\mu \approx 0.7, 0.8$ and 0.9 (which results in approximately 30, 20 and 10% of truncation, respectively). At the same time, we also report the achieved coverage probabilities of $\bar{I}_\delta$ and $\hat{I}_\delta$ for $\theta_q = 1.4, 1.3, 1.2$. These coverage probabilities are identical to one minus the power of the corresponding empirical likelihood ratio two-sided tests for the null hypothesis $H_0 : \theta_q = \theta_0$ at level $\delta = 0.05$.

From Table 1, it can be seen that (i) the average interval lengths of $\hat{I}_\delta$ are smaller than those of $\bar{I}_\delta$, i.e., the empirical likelihood confidence interval with the auxiliary information improves the standard empirical likelihood confidence interval in the sense of achieving a certain amount of reduction in interval width; (ii) for values of θ_q other than $\theta_q = 1.5$, the achieved coverage probabilities of $\hat{I}_\delta$ are always lower than those of $\bar{I}_\delta$, reflecting that the empirical likelihood ratio test with the auxiliary information is more powerful than the standard empirical likelihood ratio test; and (iii) both empirical likelihood methods behave a bit better under lighter truncation proportions.

To study the influence of the dependence of the observations, we consider different degrees of dependence; specifically, we take $\rho = 0.2, 0.4, 0.6$ and 0.8 in model (a) above. We report in Table 2 the coverage probabilities and average lengths of 95% confidence intervals for the conditional median $\theta_q = 1.5$ at $x = 1$ in the presence and absence of auxiliary information along $M = 1{,}000$ Monte Carlo replications, with $\mu \approx 0.9$. Table 2 shows that (i) as the dependence of the observations increases (i.e., as the value of ρ increases), the attained coverages apart more from the nominal 95%,

Table 1 Average lengths (AL) and coverage probabilities of 95% confidence intervals (CI) for $\theta_q = 1.5$ (true) and other θ_q values (wrong) along 1,000 trials for simulated model (a) and several truncation proportions $(1-\mu)$: empirical likelihood with $(\hat{I}_\delta)$ and without $(\bar{I}_\delta)$ auxiliary information

μ	CI	AL	$\theta_q = 1.5$	$\theta_q = 1.4$	$\theta_q = 1.3$	$\theta_q = 1.2$
0.7	$\bar{I}_\delta$	0.3569	0.8890	0.6840	0.3890	0.0810
	$\hat{I}_\delta$	0.2142	0.9030	0.6390	0.2100	0.0030
0.8	$\bar{I}_\delta$	0.3492	0.9120	0.6890	0.3030	0.0390
	$\hat{I}_\delta$	0.2118	0.9190	0.5880	0.0830	0.0020
0.9	$\bar{I}_\delta$	0.3163	0.9250	0.6860	0.2430	0.0340
	$\hat{I}_\delta$	0.1941	0.9290	0.6510	0.1450	0.0010

Table 2 Average lengths (AL) and coverage probabilities (CP) of 95% confidence intervals (CI) for $\theta_q = 1.5$ along 1,000 trials for simulated model (a) and several dependence degrees ρ: empirical likelihood with $(\hat{I}_\delta)$ and without $(\bar{I}_\delta)$ auxiliary information

CI	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
	AL	CP	AL	CP	AL	CP	AL	CP
$\bar{I}_\delta$	0.3319	0.9130	0.3309	0.9080	0.3601	0.8940	0.3944	0.8590
$\hat{I}_\delta$	0.2004	0.9250	0.2074	0.9220	0.2156	0.9180	0.2340	0.8720

while the intervals get wider; and (ii) for each fixed ρ, the average interval lengths of $\hat{I}_\delta$ are smaller than those of $\bar{I}_\delta$, and the coverage probabilities of $\hat{I}_\delta$ are closer to the nominal than those of $\bar{I}_\delta$. This is in accordance with the results displayed in Table 1.

(b) Now, we consider the simulation in the case $d = 2$. Using the same method as in (a) above [step (1) and repetition of step (2)], we generate the observed data $(\mathbf{X}_i, Y_i, T_i)$, $i = 1, \ldots, n$, where $\mathbf{X}_i = (X_{1,i}, X_{2,i})$, from the following model:

$$
\begin{aligned}
Y_i &= 1.5X_{1,i}^2 + \sin(\pi X_{2,i}) + 0.72\epsilon_i, \;\; X_{1,i} = 0.1X_{1,i-1} + 0.73e_{1,i}, \\
X_{2,i} &= 0.4X_{2,i-1} + 0.8e_{2,i},
\end{aligned}
$$

where $e_{1,i}$, $e_{2,i}$ and ϵ_i are i.i.d. random variables with distribution $N(0,1)$ and $T_i \sim N(\beta, 1)$, and everything is distributed conditionally on $Y_i \geq T_i$. From Lemma 2 of Cai (2001), it follows that $\{\mathbf{X}_i\}$ is a sequence of α-mixing random variables. As before, the α-mixing property of the $\mathbf{X}_i$ is immediately transferred to the $(\mathbf{X}_i, Y_i, T_i)$. The true regression function is given by $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}) = 1.5x_1^2 + \sin(\pi x_2)$ at $\mathbf{x} = (x_1, x_2)$, and the conditional median coincides with $m(\mathbf{x})$. Assume that $m(1,1) = 1.5$ is known. Hence, for the proposed empirical likelihood method, we introduce the auxiliary information through the function$g(y) = y - 1.5$. The product kernel $K(x_1)K(x_2)$ with a Gaussian $K(\cdot)$ is used. We consider $n = 300$ as sample size and the bandwidth $h_n = 0.33$. Similarly as in Table 1 for the one-dimensional case, in Table 3 the coverage proportions and average lengths of the empirical likelihood 95% confidence intervals (with and without the auxiliary information) for $q = 0.5$ and $\theta_q = 1.5$ [which is true value of θ_q at $\mathbf{x} = (1, 1)$], as well as for $\theta_q = 1.8, 2.1, 2.4$,

Table 3 Average lengths (AL) and coverage probabilities of 95% confidence intervals (CI) for $\theta_q = 1.5$ (true) and other θ_q values (wrong) along 1,000 trials for simulated model (b) and several truncation proportions $(1-\mu)$: empirical likelihood with $(\hat{I}_\delta)$ and without $(\bar{I}_\delta)$ auxiliary information

μ	CI	AL	$\theta_q = 1.5$	$\theta_q = 1.8$	$\theta_q = 2.1$	$\theta_q = 2.4$
0.7	$\bar{I}_\delta$	1.1645	0.8460	0.5770	0.2710	0.0730
	$\hat{I}_\delta$	0.6712	0.8960	0.5550	0.1120	0.0060
0.8	$\bar{I}_\delta$	1.1872	0.8860	0.6460	0.3380	0.1020
	$\hat{I}_\delta$	0.6639	0.9100	0.5420	0.1150	0.0050
0.9	$\bar{I}_\delta$	1.1848	0.9020	0.6750	0.3510	0.1180
	$\hat{I}_\delta$	0.6656	0.9150	0.5630	0.0980	0.0040

are reported. As above, we take three different values for the truncation parameter ($\mu \approx 0.7$, 0.8 and 0.9) and 1,000 Monte Carlo replications.

From Table 3, it can be seen that the relative performance of the two empirical likelihood confidence intervals $\hat{I}_\delta$ and $\bar{I}_\delta$ for $d = 2$ is similar to that corresponding to the case $d = 1$. We can also get similar conclusions about the influence of the truncation proportion and of the chosen value for θ_q. Indeed, the numerical results in Table 3 suggest that the introduction of the auxiliary information in the construction of the confidence interval has more relevance in the two-dimensional case, both in terms of interval length reduction and statistical power.

5 Preliminary lemmas

Let $w_{1in}(\theta) = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(I(Y_i \le \theta) - q)$, $w_{1i}(\theta) = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-1}(Y_i)$ $(I(Y_i \le \theta) - q)$,

$$w_{2in} = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i) g(Y_i), \quad w_{2i} = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-1}(Y_i) g(Y_i),$$

$w_{3in}(\theta) = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G_n^{-1}(Y_i)(g^\tau(Y_i), I(Y_i \le \theta) - q)^\tau$, and $w_{3i}(\theta) = K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)$ $G^{-1}(Y_i)(g^\tau(Y_i), I(Y_i \le \theta) - q)^\tau$. Then $w_{lin}(\theta) = w_{li}(\theta)\left[1 + \frac{G(Y_i) - G_n(Y_i)}{G_n(Y_i)}\right]$ $(l = 1, 2, 3)$. Set

$$\Sigma_1 = [(1-2q)v(\mathbf{x}) + q^2 u(\mathbf{x})]\mu l(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{s})d\mathbf{s}, \ \Sigma_2 = \mu l(\mathbf{x}) V(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{s})d\mathbf{s}$$

and $\Sigma_3 = \begin{pmatrix} V(\mathbf{x}) & W(\mathbf{x}) \\ W^\tau(\mathbf{x}) & (1-2q)v(\mathbf{x}) + q^2 u(\mathbf{x}) \end{pmatrix} \mu l(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{s})d\mathbf{s}$.

Lemma 1 *Suppose that (A0)–(A9) are satisfied. Let $\alpha(n) = O(n^{-\gamma})$ for some $\gamma \ge [r(r+2)]/[2(r-2)]$. Set $\Theta_n = \frac{\mu}{nh_n^d} \sum_{i=1}^n w_{3i}(\theta_q) w_{3i}^\tau(\theta_q)$. Then*

(a) $\Theta_n = \Sigma_3/\mu + o_p(1)$; (b) $\eta_2(\theta_q) = O_p((nh_n^d)^{-1/2})$;

(c) $\eta_2(\theta_q)=\Theta_n^{-1}\cdot\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)+o_p((nh_n^d)^{-1/2})$; (d) $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)\xrightarrow{\mathcal{D}} N(0,\Sigma_3)$.

Further, let $\theta_n=\theta_q+(nh_n^d)^{-1/2}c_1$ *for some* $c_1>0$. *Assume that (B1) and (B2) are satisfied, then (a)–(c) still hold for* $\theta_q=\theta_n$, *and* $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_n)\xrightarrow{\mathcal{D}} N(\phi,\Sigma_3)$ *with* $\phi=(0^\tau, c_1 f(\mathbf{x},\theta_q))^\tau$.

Lemma 2 *Suppose that (A0)–(A6), (A7)(i) and (A8)(ii) are satisfied. Let* $\alpha(n)=O(n^{-\gamma})$ *for some* $\gamma\ge[r(r+2)]/[2(r-2)]$, *then*

(a) $\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{2i}w_{2i}^\tau=\Sigma_2/\mu+o_p(1)$; (b) $\eta_1=O_p((nh_n^d)^{-1/2})$;

(c) $\eta_1=\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{2i}w_{2i}^\tau\right)^{-1}\cdot\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{2i}+o_p((nh_n^d)^{-1/2})$;

(d) $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{2i}\xrightarrow{\mathcal{D}} N(0,\Sigma_2)$.

Lemma 3 *Suppose that (A0)–(A4), (A7)(ii) and (A8)(i) are satisfied. Let*$\alpha(n)=O(n^{-\gamma})$ *for some* $\gamma\ge r^2/(r-2)$, *then*

(a) $\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{1i}^2(\theta_q)=\Sigma_1/\mu+o_p(1)$; (b) $\lambda(\theta_q)=O_p((nh_n^d)^{-1/2})$;

(c) $\lambda(\theta_q)=\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{1i}^2(\theta_q)\right)^{-1}\cdot\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{1i}(\theta_q)+O_p((nh_n^d)^{-1})$;

(d) $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_q)\xrightarrow{\mathcal{D}} N(0,\Sigma_1)$.

Further, let $\theta_n=\theta_q+(nh_n^d)^{-1/2}c_1$ *for some* $c_1>0$. *Assume that (B1) and (B2) are satisfied, then (a)-(c) still hold for* $\theta_q=\theta_n$, *and* $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_n)\xrightarrow{\mathcal{D}} N(c_1 f(\mathbf{x},\theta_q),\Sigma_1)$.

Lemma 4 (Srivastava and Khatri 1979, p. 64, Corollary 2.11.2) *Let* $A_{p\times p}$ *be a symmetric matrix with Rank*$(A)=k$, $X\sim N(\mu_0,\Sigma_{p\times p})$, *where* Σ *is a positive definite matrix and matrix* ΣA *is idempotent. Then* $X^\tau AX\sim\chi_k^2(\mu_0^\tau A\mu_0)$.

6 Proof of main results

Proof of Theorem 1 First we prove $\bar{l}_n(\theta_q)\xrightarrow{\mathcal{D}}\chi_1^2$. By Taylor's expansion we have

$$\begin{aligned}\bar{l}_n(\theta_q)&=2\lambda(\theta_q)\sum_{i=1}^n w_{1i}(\theta_q)-\lambda^2(\theta_q)\sum_{i=1}^n w_{1i}^2(\theta_q)\\&\quad+2\lambda(\theta_q)\sum_{i=1}^n w_{1i}(\theta_q)\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\\&\quad-\lambda^2(\theta_q)\sum_{i=1}^n w_{1i}^2(\theta_q)\left[2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right]\\&\quad+O\left(|\lambda(\theta_q)|^3\max_{1\le i\le n}|w_{1in}(\theta_q)|\sum_{i=1}^n w_{1i}^2(\theta_q)\left(1+\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right).\qquad(8)\end{aligned}$$

Since $E|w_{1i}(\theta_q)|^r < \infty$ from (A2)(i), $\max_{1\le i\le n}|w_{1i}(\theta_q)| = o(n^{1/r})$ $a.s.$ from the proof of Lemma 3 in Owen (1990). In view of (3), from (A2)(i) and (A7)(ii) we have $\frac{1}{nh_n^d}\sum_{i=1}^n E|w_{1i}(\theta_q)| = O(1)$, which implies $\frac{1}{nh_n^d}\sum_{i=1}^n |w_{1i}(\theta_q)| = O_p(1)$. Hence, in view of Lemmas 3 and 8 we obtain that

$$\begin{aligned}&\left|\lambda(\theta_q)\sum_{i=1}^n w_{1i}(\theta_q)\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right|\\ &\quad\le \frac{|\lambda(\theta_q)|\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}{G(a_{\tilde F})-\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}\sum_{i=1}^n|w_{1i}(\theta_q)| = o_p(1), \end{aligned} \tag{9}$$

$$\left|\lambda^2(\theta_q)\sum_{i=1}^n w_{1i}^2(\theta_q)\left[2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right]\right| = o_p(1), \tag{10}$$

$$\begin{aligned}&|\lambda(\theta_q)|^3\max_{1\le i\le n}|w_{1in}(\theta_q)|\sum_{i=1}^n w_{1i}^2(\theta_q)\left(1+\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\\ &\quad= o_p((n^{r-2}h_n^{rd})^{-1/2r}) = o_p(1). \end{aligned} \tag{11}$$

Therefore, using (a) and (d) in Lemmas 3, (8)–(11) yield that

$$\bar l_n(\theta_q) = \left(\frac{\mu^2}{nh_n^d}\sum_{i=1}^n w_{1i}^2(\theta_q)\right)^{-1}\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_q)\right)^2 + o_p(1)\xrightarrow{\mathcal{D}}\chi_1^2.$$

Next we prove $\bar l_n(\theta_n)\xrightarrow{\mathcal{D}}\chi_1^2(\Delta^2)$. Lemma 3 shows that $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_n)\xrightarrow{\mathcal{D}} N(c_1 f(\mathbf{x},\theta_q),\Sigma_1)$, where $c_1 = \frac{\{[(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})]\mu l(\mathbf{x})\int_{\mathbb{R}^d}K^2(\mathbf{s})d\mathbf{s}\}^{1/2}}{f(\mathbf{x},\theta_q)}\Delta$. Note that $(c_1 f(\mathbf{x},\theta_q))^2\Sigma_1^{-1} = \Delta^2$. So, by using Lemma 4, similarly to the arguments as in (8)–(11) one can get

$$\bar l_n(\theta_n) = \left(\frac{\mu^2}{nh_n^d}\sum_{i=1}^n w_{1i}^2(\theta_n)\right)^{-1}\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_n)\right)^2 + o_p(1)\xrightarrow{\mathcal{D}}\chi_1^2(\Delta^2).$$

□

Proof of Theorem 2 First we prove $\hat l_n(\theta_q)\xrightarrow{\mathcal{D}}\chi_1^2$. By Taylor's expansion we have

$$\begin{aligned}\hat l_n(\theta_q) &= 2\sum_{i=1}^n\log[1+\eta_2^\tau(\theta_q)w_{3in}(\theta_q)] - 2\sum_{i=1}^n\log[1+\eta_1^\tau w_{2in}]\\ &= \left\{2\eta_2^\tau(\theta_q)\sum_{i=1}^n w_{3i}(\theta_q) - \eta_2^\tau(\theta_q)\left(\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)\right)\eta_2(\theta_q)\right\}\end{aligned}$$

$$
\begin{aligned}
&-\left\{2\eta_1^\tau \sum_{i=1}^n w_{2i} - \eta_1^\tau\left(\sum_{i=1}^n w_{2i}w_{2i}^\tau\right)\eta_1\right\}\\
&+2\eta_2^\tau(\theta_q)\sum_{i=1}^n w_{3i}(\theta_q)\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)} + O\left(\sum_{i=1}^n |\eta_2^\tau(\theta_q)w_{3in}(\theta_q)|^3\right)\\
&-\eta_2^\tau(\theta_q)\left[\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)\left(2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right.\right.\\
&\left.\left.+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right)\right]\eta_2(\theta_q) - 2\eta_1^\tau\sum_{i=1}^n w_{2i}\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\\
&+O\left(\sum_{i=1}^n|\eta_1^\tau w_{2in}|^3\right)+\eta_1^\tau\left[\sum_{i=1}^n w_{2i}w_{2i}^\tau\left(2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right.\right.\\
&\left.\left.+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right)\right]\eta_1. \qquad (12)
\end{aligned}
$$

From (3), (A2)(i) and (A6) we have $\frac{\mu}{nh_n^d}\sum_{i=1}^n E\|w_{3i}(\theta_q)\| = O(1)$, then $\frac{\mu}{nh_n^d}\sum_{i=1}^n \|w_{3i}(\theta_q)\| = O_p(1)$. Similarly, $\frac{\mu}{nh_n^d}\sum_{i=1}^n \|w_{3i}(\theta_q)\|^2 = O_p(1)$, $\frac{\mu}{nh_n^d}\sum_{i=1}^n \|w_{2i}\| = O_p(1)$ and $\frac{\mu}{nh_n^d}\sum_{i=1}^n \|w_{2i}\|^2 = O_p(1)$. Note that $w_{3in}(\theta_q) = w_{3i}(\theta_q)[1 + (G(Y_i)-G_n(Y_i))/G_n(Y_i)]$, so from Lemmas 1–2 and 8 it follows that

$$
\begin{aligned}
&\left|\eta_2^\tau(\theta_q)\sum_{i=1}^n w_{3i}(\theta_q)\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right|\\
&\quad\le \frac{\|\eta_2(\theta_q)\|\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}{G(a_{\tilde F})-\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}\sum_{i=1}^n\|w_{3i}(\theta_q)\| = o_p(1), \qquad (13)
\end{aligned}
$$

and by using (23) in Sect. 8

$$
\begin{aligned}
&\sum_{i=1}^n|\eta_2^\tau(\theta_q)w_{3in}(\theta_q)|^3 = o_p(1),\quad \left|\eta_1^\tau\sum_{i=1}^n w_{2i}\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right| = o_p(1),\\
&\qquad\qquad \sum_{i=1}^n|\eta_1^\tau w_{2in}|^3 = o_p(1), \qquad (14)
\end{aligned}
$$

$$
\begin{aligned}
&\left|\eta_2^\tau(\theta_q)\left[\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)\left(2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right)\right]\eta_2(\theta_q)\right|\\
&\quad = o_p(1), \qquad (15)
\end{aligned}
$$

$$\left|\eta_1^\tau\left[\sum_{i=1}^n w_{2i}w_{2i}^\tau\left(2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right)\right]\eta_1\right| = o_p(1). \tag{16}$$

Therefore, on applying Lemmas 1, 2, from (12)–(16) we have

$$\begin{aligned}\hat{l}_n(\theta_q) &= \left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)\right)^\tau\left(\frac{\mu^2}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)\right)^{-1}\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)\right)\\ &\quad-\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{2i}\right)^\tau\left(\frac{\mu^2}{nh_n^d}\sum_{i=1}^n w_{2i}w_{2i}^\tau\right)^{-1}\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{2i}\right)+o_p(1)\\ &= \left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)\right)^\tau\Gamma\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)\right)+o_p(1),\end{aligned}$$

where $\Gamma = \left(\Sigma_3^{-1}-\begin{pmatrix}\Sigma_2^{-1} & 0\\ 0 & 0\end{pmatrix}\right)$. Note that $\Gamma\Sigma_3$ is an idempotent matrix, and $Rank(\Gamma) = Rank(\Gamma\Sigma_3) = 1$. Therefore $\hat{l}_n(\theta_q) \xrightarrow{\mathcal{D}} \chi_1^2$ by Lemmas 1 and 4.

Next we verify $\hat{l}_n(\theta_n) \xrightarrow{\mathcal{D}} \chi_1^2(\rho^2)$. Since $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_n) \xrightarrow{\mathcal{D}} N(\phi, \Sigma_3)$ with $\phi = (0^\tau, c_1 f(\mathbf{x}, \theta_q))^\tau$ and $\phi^\tau\Gamma\phi = \frac{(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})}{(1-2q)v(\mathbf{x})+q^2u(\mathbf{x})-W^\tau(\mathbf{x})V^{-1}(\mathbf{x})W(\mathbf{x})}\Delta^2 = \rho^2$, on applying Lemmas 1 and 4, from $\hat{l}_n(\theta_n) = (\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q))^\tau\Gamma(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_q)) + o_p(1)$, we obtain that $\hat{l}_n(\theta_n) \xrightarrow{\mathcal{D}} \chi_1^2(\rho^2)$. □

Acknowledgments The authors are grateful to the editor, associate editor, and anonymous referee for their helpful comments. This research was supported by the Grants MTM2008-03129 of the Spanish Ministry of Science and Innovation and Project 10PXIB300068PR of the Xunta de Galicia, Spain, and also by the National Natural Science Foundation of China (10871146).

7 Appendix A

Lemma 5 (Volkonskii and Rozanov 1959) *Let $V_1, \ldots, V_m$ be α-mixing random variables measurable with respect to the σ-algebra $\mathcal{F}_{i_1}^{j_1}, \ldots, \mathcal{F}_{i_m}^{j_m}$, respectively, with $1 \le i_1 < j_1 < \cdots < j_m \le n$, $i_{l+1} - j_l \ge w \ge 1$ and $|V_j| \le 1$ for $l, j = 1, 2, \ldots, m$. Then $|E(\prod_{j=1}^m V_j) - \prod_{j=1}^m EV_j| \le 16(m-1)\alpha(w)$, where $\mathcal{F}_a^b = \sigma\{V_i, a \le i \le b\}$ and $\alpha(w)$ is the mixing coefficient.*

Lemma 6 (Hall and Heyde 1980, Corollary A.2, p. 278) *Suppose that X and Y are random variables such that $E|X|^p < \infty$, $E|Y|^q < \infty$, where $p, q > 1$, $p^{-1}+q^{-1} < 1$. Then*

$$|EXY - EXEY| \le 8\|X\|_p\|Y\|_q\left\{\sup_{A\in\sigma(X), B\in\sigma(Y)}|P(A\cap B)-P(A)P(B)|\right\}^{1-p^{-1}-q^{-1}}.$$

Lemma 7 (Shao and Yu 1996, Theorem 4.1) *Let* $2 < p < q \leq \infty, 2 < \lambda \leq q$ *and* $\{X_n, n \geq 1\}$ *be an* α*-mixing sequence of random variables with* $EX_n = 0$ *and the mixing coefficients* $\{\hat{\alpha}(j)\}$*. Assume that* $\hat{\alpha}(n) \leq Cn^{-\gamma}$ *for some* $C > 0$ *and* $\gamma > 0$*. Then there exists* $Q = Q(p, q, \lambda, \gamma, C) < \infty$ *such that* $E|\sum_{i=1}^n X_i|^p \leq Qn^{p/2} \max_{1\leq i\leq n} \|X_i\|_q^p$ *if* $\gamma \geq pq/[2(q-p)]$.

Lemma 8 (Liang et al. 2011) *Suppose that* $\alpha(n) = O(n^{-\gamma})$ *for some* $\gamma > 3$*. Then, under (A0) we have* $\sup_{y\geq a_{\tilde{F}}} |G_n(y) - G(y)| = O_p(n^{-1/2})$.

8 Appendix B

Proof of Lemma 1 (a) It is easy to see that

$$\Theta_n = \frac{\mu}{nh_n^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-2}(Y_i)\begin{pmatrix} g(Y_i)g^\tau(Y_i) & g(Y_i)(I(Y_i\leq\theta_q)-q) \\ g^\tau(Y_i)(I(Y_i\leq\theta_q)-q) & (I(Y_i\leq\theta_q)-q)^2 \end{pmatrix}.$$

It is easy to verify that $E\Theta_n = \frac{\mu}{nh_n^d}\sum_{i=1}^n E(w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)) = \Sigma_3/\mu + o(1)$. Then, according to $X_n = EX_n + O_p(\sqrt{\mathrm{Var}X_n})$, to prove (a) we need only to prove that $\mathrm{Var}(\Theta_n) \to 0$, further it suffices to show that for $1 \leq k, l \leq \kappa$

$$P_n := \mathrm{Var}\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-2}(Y_i)g_k(Y_i)g_l(Y_i)\right) = O((nh_n^d)^{-1}) \to 0; \tag{17}$$

$$\mathrm{Var}\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-2}(Y_i)g_k(Y_i)(I(Y_i\leq\theta_q)-q)\right) = O((nh_n^d)^{-1}) \to 0; \tag{18}$$

$$\mathrm{Var}\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-2}(Y_i)(I(Y_i\leq\theta_q)-q)^2\right) = O((nh_n^d)^{-1}) \to 0. \tag{19}$$

We prove only (17), the proofs of (18) and (19) are similar. Write

$$\begin{aligned} P_n &= \left(\frac{\mu}{nh_n^d}\right)^2 \left\{\sum_{i=1}^n \mathrm{Var}\left(K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{g_k(Y_i)g_l(Y_i)}{G^2(Y_i)}\right)\right. \\ &\quad \left. + \sum_{i\neq j}\mathrm{Cov}\left(K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{g_k(Y_i)g_l(Y_i)}{G^2(Y_i)}, K^2\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\frac{g_k(Y_j)g_l(Y_j)}{G^2(Y_i)}\right)\right\} \\ &:= P_{1n} + P_{2n}. \end{aligned}$$

From (3), (A2)(i) and (A6) it is easy to verify that $P_{1n} = O((nh_n^d)^{-1})$. Let $\xi_i = \frac{\mu}{h_n^d} K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \frac{g_k(Y_i)g_l(Y_i)}{G^2(Y_i)}$. For $i < j$, applying (A1), (A2)(i) and (A5)–(A6) we have

$$|\mathrm{Cov}(\xi_i, \xi_j)| = \Bigg| \frac{\mu^2}{h_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K^2\left(\frac{\mathbf{x}-\mathbf{s}}{h_n}\right) K^2\left(\frac{\mathbf{x}-\mathbf{t}}{h_n}\right)$$
$$\times E\left(\frac{g_k(Y_i)g_l(Y_i)g_k(Y_j)g_l(Y_j)}{G^2(Y_i)G^2(Y_j)} \Bigg| \mathbf{X}_i = \mathbf{s}, \mathbf{X}_j = t\right) l^*_{j-i}(\mathbf{s}, \mathbf{t}) \mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}$$
$$- \frac{\mu^2}{h_n^{2d}} \left(\int_{\mathbb{R}^d} K^2\left(\frac{\mathbf{x}-\mathbf{s}}{h_n}\right) \mathbb{E}\left(\frac{g_k(Y)g_l(Y)}{G(Y)} |\mathbf{X} = \mathbf{s}\right) l(\mathbf{s}) \mathrm{d}\mathbf{s}\right)^2 \Bigg| = O(1). \tag{20}$$

On the other hand, from Lemma 6 it follows that $|\mathrm{Cov}(\xi_i, \xi_j)| \le C[\alpha(j-i)]^{1-2/r}(E|\xi_i|^r)^{2/r}$ and $E|\xi_i|^r = O(h_n^{-d(r-1)})$. Then from (A3)(ii) we have

$$P_{2n} = \frac{1}{n^2}\left\{\sum_{|i-j|\le[h_n^{-d}]} + \sum_{|i-j|>[h_n^{-d}]}\right\} \mathrm{Cov}(\xi_i, \xi_j) = O((nh_n^d)^{-1}).$$

(b) Write $\eta_2(\theta_q) = \lambda\beta_2(\theta_q)$, where $\lambda \ge 0$ and $\|\beta_2(\theta_q)\| = 1$. From (7) we find

$$0 \ge \frac{\lambda\beta_2^\tau(\theta_q)\Theta_n\beta(\theta_q)}{1+\lambda\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3in}(\theta_q)|} - \left|\beta_2^\tau(\theta_q)\cdot\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)\right|$$
$$-\frac{\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}{G(a_{\tilde F})-\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}\left(\frac{\mu}{nh_n^d}\sum_{i=1}^n|\beta_2^\tau(\theta_q)w_{3i}(\theta_q)|\right.$$
$$\left.+\frac{2\lambda\beta_2^\tau(\theta_q)\Theta_n\beta(\theta_q)}{1+\lambda\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3in}(\theta_q)|}\right)-\left(\frac{\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}{G(a_{\tilde F})-\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}\right)^2$$
$$\frac{\lambda\beta_2^\tau(\theta_q)\Theta_n\beta(\theta_q)}{1+\lambda\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3in}(\theta_q)|}. \tag{21}$$

Note that Σ_3 is positive definite by (A9). (a) implies $\beta_2^\tau(\theta_q)\Theta_n\beta_2(\theta_q) \ge t_0 + o_p(1)$ where t_0 is the smallest eigenvalue of Σ_3/μ, (d) follows $\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q) = O_p((nh_n^d)^{-1/2})$, and it is easy to verify that $\frac{\mu}{nh_n^d}\sum_{i=1}^n|\beta_2^\tau(\theta_q)w_{3i}(\theta_q)| = O_p(1)$. Therefore, from (21) and Lemma 8 we obtain that

$$\frac{\lambda}{1+\lambda\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3in}(\theta_q)|} = O_p((nh_n^d)^{-1/2}). \tag{22}$$

Since $E|\beta_2^\tau(\theta_q)w_{3i}(\theta_q)|^r < \infty$ from (A2)(i) and (A6), $\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3i}(\theta_q)| = o(n^{1/r})$ $a.s.$ from the proof of Lemma 3 in Owen (1990). Hence, from Lemma 8 we have

$$\max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3in}(\theta_q)| \le \max_{1\le i\le n}|\beta_2^\tau(\theta_q)w_{3i}(\theta_q)|$$
$$\times\left(1+\frac{\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}{G(a_{\tilde F})-\sup_{y\ge a_{\tilde F}}|G_n(y)-G(y)|}\right)=o_p(n^{1/r}).$$

Therefore, $n^{r-2}h_n^{rd} \ge c_0 > 0$ and (22) yield that $\lambda = O_p((nh_n^d)^{-1/2}), \eta_2(\theta_q) = O_p((nh_n^d)^{-1/2})$, and

$$\max_{1\le i\le n}|\eta_2^\tau(\theta_q)w_{3in}(\theta_q)| = O_p((nh_n^d)^{-1/2})o_p(n^{1/r}) = o_p(1). \tag{23}$$

(c) From (7) we write

$$\begin{aligned}0 &= \frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q) + \frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)} - \Theta_n\eta_2(\theta_q)\\
&\quad -\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}(\theta_q)^\tau\eta_2(\theta_q)\left[2\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}+\left(\frac{G(Y_i)-G_n(Y_i)}{G_n(Y_i)}\right)^2\right]\\
&\quad +\frac{\mu}{nh_n^d}\sum_{i=1}^n\frac{w_{3in}(\theta_q)(\eta_2^\tau(\theta_q)w_{3in}(\theta_q))^2}{1+\eta_2^\tau(\theta_q)w_{3in}(\theta_q)}.\end{aligned} \tag{24}$$

Using (A6) and (A2)(i) it follows that $\frac{\mu}{nh_n^d}\sum_{i=1}^n\|w_{3i}(\theta_q)\|^2 = O_p(1)$ and $\frac{\mu}{nh_n^d}\sum_{i=1}^n\|w_{3i}(\theta_q)\| = O_p(1)$. Hence, in view of $\max_{1\le i\le n}\|w_{3i}(\theta_q)\| = o(n^{1/r})$ $a.s.$, (b) and (23) we have

$$\frac{\mu}{nh_n^d}\sum_{i=1}^n\|w_{3i}(\theta_q)w_{3i}(\theta_q)^\tau\eta_2(\theta_q)\| \le \frac{\mu}{nh_n^d}\sum_{i=1}^n\|w_{3i}(\theta_q)\|^2\|\eta_2^\tau(\theta_q)\|$$
$$= O_p((nh_n^d)^{1/2})\left\|\frac{\mu}{nh_n^d}\sum_{i=1}^n\frac{w_{3in}(\theta_q)(\eta_2^\tau(\theta_q)w_{3in}(\theta_q))^2}{1+\eta_2^\tau(\theta_q)w_{3in}(\theta_q)}\right\| = o_p((nh_n^d)^{-1/2}).$$

Therefore, on applying Lemma 8, from (24) it follows that $\eta_2(\theta_q) = \Theta_n^{-1}\cdot\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q) + o_p((nh_n^d)^{-1/2})$.

(d) It suffices to show that $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n[w_{3i}(\theta_q)-Ew_{3i}(\theta_q)] \xrightarrow{\mathcal{D}} N(0,\Sigma_3)$, $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n Ew_{3i}(\theta_q)\to 0$. In order to prove $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n Ew_{3i}(\theta_q)\to 0$, we need only to verify that

$$I_{1n}(\mathbf{x}) := \frac{\mu}{\sqrt{nh_n^d}} \sum_{i=1}^{n} E\left(K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) G^{-1}(Y_i)(I(Y_i \le \theta) - q)\right) \to 0$$

and $I_{2n}(\mathbf{x}) := \frac{\mu}{\sqrt{nh_n^d}} \sum_{i=1}^{n} E(K(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}) G^{-1}(Y_i) g(Y_i)) \to 0$. By using (3) and (2), from (A2)(iii) and (A7) we have $I_{1n}(\mathbf{x}) = (nh_n^d)^{1/2} \int_{\mathbb{R}^d} K(\mathbf{s}) l(\mathbf{x} - h_n\mathbf{s})[\mathbb{E}(I(Y \le \theta_q)|\mathbf{X} = \mathbf{x} - h_n\mathbf{s}) - q]d\mathbf{s} = O(nh_n^{d+4})^{1/2} \to 0$. Similarly, $I_{2n}(\mathbf{x}) \to 0$.

Next we prove $\frac{\mu}{\sqrt{nh_n^d}} \sum_{i=1}^{n} [w_{3i}(\theta_q) - Ew_{3i}(\theta_q)] \xrightarrow{\mathcal{D}} N(0, \Sigma_3)$, it is sufficient to prove that for $\mathbf{a} = (a_1, \ldots, a_\kappa, a_{\kappa+1})^\tau := (\tilde{\mathbf{a}}^\tau, a_{\kappa+1})^\tau \ne 0$, $\frac{\mu \mathbf{a}^\tau}{\sqrt{nh_n^d}} \sum_{i=1}^{n} [w_{3i}(\theta_q) - Ew_{3i}(\theta_q)] \xrightarrow{\mathcal{D}} N(0, \mathbf{a}^\tau \Sigma_3 \mathbf{a})$.

Note that (A3)(i) implies that there exists a sequence of positive integers $\delta_n \to \infty$ such that $\delta_n \eta = o((nh_n^d)^{1/2})$, $\delta_n (nh_n^{-d})^{1/2} \alpha(\eta) \to 0$. Let $\pi := \pi_n = [\frac{n}{\beta+\eta}]$, $\beta := \beta_n = [(nh_n^d)^{1/2}/\delta_n]$. Then

$$\eta/\beta \to 0, \quad \pi\alpha(\eta) \to 0, \quad \pi\eta/n \to 0, \quad \beta/n \to 0, \quad \beta/(nh_n^d)^{1/2} \to 0. \tag{25}$$

Next partition the set $\{1, 2, \ldots, n\}$ into $2\pi_n + 1$ subsets with large blocks of size $\beta := \beta_n$ and small blocks of size $\eta := \eta_n$. Let y_{mn}, y'_{mn} and y''_{wn} be defined as follows:

$$y_{mn} = \sum_{i=k_m}^{k_m+\beta-1} \frac{\mu \mathbf{a}^\tau}{\sqrt{h_n^d}} (w_{3i}(\theta_q) - Ew_{3i}(\theta_q)),$$

$$y'_{mn} = \sum_{j=l_m}^{l_m+\eta-1} \frac{\mu \mathbf{a}^\tau}{\sqrt{h_n^d}} (w_{3i}(\theta_q) - Ew_{3i}(\theta_q)),$$

where $y''_{\pi n} = \sum_{k=\pi(\beta+\eta)+1}^{n} \frac{\mu \mathbf{a}^\tau}{\sqrt{h_n^d}} (w_{3i}(\theta_q) - Ew_{3i}(\theta_q))$, $k_m = (m-1)(\beta+\eta)+1$, $l_m = (m-1)(\beta+\eta)+\beta+1$, $m = 1, \ldots, \pi$. Then

$$\begin{aligned} \frac{\mu \mathbf{a}^\tau}{\sqrt{nh_n^d}} \sum_{i=1}^{n} [w_{3i}(\theta_q) - Ew_{3i}(\theta_q)] &= n^{-1/2} \left\{ \sum_{m=1}^{\pi} y_{mn} + \sum_{m=1}^{\pi} y'_{mn} + y''_{\pi n} \right\} \\ &:= n^{-1/2} \left\{ S'_n + S''_n + S'''_n \right\}. \end{aligned}$$

Hence, it suffices to show that

$$n^{-1} E(S''_n)^2 \to 0, \quad n^{-1} E(S'''_n)^2 \to 0, \quad \mathrm{Var}(n^{-1/2} S'_n) \to \mathbf{a}^\tau \Sigma_3 \mathbf{a}, \tag{26}$$

$$\left| E\exp\left(it \sum_{m=1}^{\pi} n^{-1/2} y_{mn} \right) - \prod_{m=1}^{\pi} E\exp(itn^{-1/2} y_{mn}) \right| \to 0, \tag{27}$$

$$g_n(\epsilon) = \frac{1}{n} \sum_{m=1}^{\pi} Ey_{mn}^2 I(|y_{mn}| > \epsilon\sqrt{n}) \to 0 \quad \forall \epsilon > 0. \tag{28}$$

We first establish (26). Note that

$$\begin{aligned}\frac{1}{n}E(S_n'')^2 &= \frac{1}{n}\sum_{m=1}^{\pi}\sum_{i=l_m}^{l_m+\eta-1}\mathrm{Var}\left(\mu h_n^{-d/2}\mathbf{a}^\tau w_{3i}(\theta_q)\right)+\frac{2}{n}\sum_{1\le i<j\le\pi}\mathrm{Cov}(y'_{in},y'_{jn})\\ &\quad+\frac{2}{n}\sum_{m=1}^{\pi}\sum_{l_m\le i<j\le l_m+\eta-1}\mathrm{Cov}\left(\mu h_n^{-d/2}\mathbf{a}^\tau w_{3i}(\theta_q),\mu h_n^{-d/2}\mathbf{a}^\tau w_{3j}(\theta_q)\right)\\ &:= J_{1n}(\mathbf{x})+J_{2n}(\mathbf{x})+J_{3n}(\mathbf{x}).\end{aligned}$$

It is not difficult to verify that

$$\mathrm{Var}\left(\mu h_n^{-d/2}\mathbf{a}^\tau w_{3i}(\theta_q)\right)=\mathbf{a}^\tau\Sigma_3\mathbf{a}+o(1), \tag{29}$$

which yields that $J_{1n}(\mathbf{x}) = O(\pi q/n) = o(1)$ from (25). Since both of $J_{2n}(\mathbf{x})$ and $J_{3n}(\mathbf{x})$ are bounded by $\frac{2}{n}\sum_{1\le i<j\le n}|\mathrm{Cov}(\mu h_n^{-d/2}\mathbf{a}^\tau w_{3i}(\theta_q),\mu h_n^{-d/2}\mathbf{a}^\tau w_{3j}(\theta_q))|$, to prove $|J_{2n}(\mathbf{x})| = o(1)$ and $|J_{3n}(\mathbf{x})| = o(1)$, it suffices to show that

$$\frac{1}{nh_n^d}\sum_{1\le i<j\le n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q),\mathbf{a}^\tau w_{3j}(\theta_q))|\to 0. \tag{30}$$

Next, let c_n (specified below) be a sequence of integers such that $c_n\to\infty$ and $c_nh_n^d\to 0$. Write

$$\begin{aligned}&\frac{1}{nh_n^d}\sum_{1\le i<j\le n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q),\mathbf{a}^\tau w_{3j}(\theta_q))|\\ &\quad=\frac{1}{nh_n^d}\sum_{0<j-i\le c_n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q),\mathbf{a}^\tau w_{3j}(\theta_q))|\\ &\qquad+\frac{1}{nh_n^d}\sum_{j-i>c_n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q),\mathbf{a}^\tau w_{3j}(\theta_q))|.\end{aligned} \tag{31}$$

Note that $\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q),\mathbf{a}^\tau w_{3j}(\theta_q)) = \mathbf{a}^\tau\left(E(w_{3i}(\theta_q)w_{3j}^\tau(\theta_q))\right)\mathbf{a}-\left(E(\mathbf{a}^\tau w_{3i}(\theta_q))\right)^2$ and

$$\begin{aligned}w_{3i}(\theta_q)w_{3j}^\tau(\theta_q)&=\frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)}{G(Y_i)G(Y_j)}\\ &\quad\times\begin{pmatrix} g(Y_i)g^\tau(Y_j) & g(Y_i)(I(Y_j\le\theta_q)-q)\\ g^\tau(Y_j)(I(Y_i\le\theta_q)-q) & (I(Y_i\le\theta_q)-q)(I(Y_j\le\theta_q)-q)\end{pmatrix}.\end{aligned}$$

Similarly to the arguments as in (20), for $1 \le k, l \le \kappa$ we have

$$E\left|\frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)}{G(Y_i)G(Y_j)}g_k(Y_i)g_l(Y_j)\right| = O(h_n^{2d}),$$

$$E\left|\frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)}{G(Y_i)G(Y_j)}g_k(Y_i)(I(Y_j \le \theta_q) - q)\right| = O(h_n^{2d}),$$

$$E\left|\frac{K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)}{G(Y_i)G(Y_j)}(I(Y_i \le \theta_q) - q)(I(Y_j \le \theta_q) - q)\right| = O(h_n^{2d}).$$

Then from $[E(\mathbf{a}^\tau w_{3i}(\theta_q))]^2 = O(h_n^{2d})$ we have $|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q), \mathbf{a}^\tau w_{3j}(\theta_q))| = O(h_n^{2d})$, further

$$\frac{1}{nh_n^d}\sum_{0<j-i\le c_n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q), \mathbf{a}^\tau w_{3j}(\theta_q))| = O(c_n h_n^d) \to 0. \tag{32}$$

On the other hand, from Lemma 6 we have $|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q), \mathbf{a}^\tau w_{3j}(\theta_q))| \le C[\alpha(j - i)]^{1-2/r}(E|\mathbf{a}^\tau w_{3i}(\theta_q)|^r)^{2/r}$ and $E|\mathbf{a}^\tau w_{3i}(\theta_q)|^r = O(h_n^d)$. Therefore, choose $c_n = h_n^{-d(1-2/r)/\delta}$ and in view of (A3)(ii) we obtain that $\frac{1}{nh_n^d}\sum_{j-i>c_n}|\mathrm{Cov}(\mathbf{a}^\tau w_{3i}(\theta_q), \mathbf{a}^\tau w_{3j}(\theta_q))| \le Cc_n^{-\delta}h_n^{-d(1-2/r)}\sum_{l=c_n}^{\infty} l^\delta[\alpha(l)]^{1-2/r} \to 0$, which, together with (31) and (32), yields (30).

Similarly, from (29), (30) and $\pi\beta/n \to 1$ it is easy to show that $n^{-1}E(S_n''')^2 \to 0$ and $\mathrm{Var}(n^{-1/2}S_n') \to \mathbf{a}^\tau\Sigma_3\mathbf{a}$. As to (27), according to Lemma 5 we have

$$\left|E\exp\left(it\sum_{m=1}^{\pi}n^{-1/2}y_{mn}\right) - \prod_{m=1}^{\pi}E\exp\left(itn^{-1/2}y_{mn}\right)\right| \le 16\pi\alpha(\eta + 1),$$

which tends to zero by (25).

Finally, we establish (28). In Lemma 7, taking $p = 1 + r/2, \lambda = q = r$, then $\gamma \ge pq/[2(q - p)] = [r(r + 2)]/[2(r - 2)]$, and from $E|\mathbf{a}^\tau w_{3i}(\theta_q)|^r = O(h_n^d)$ we have

$$\begin{aligned} &Ey_{mn}^2 I(|y_{mn}| > \epsilon\sqrt{n}) \\ &\quad \le \epsilon^{1-r/2}n^{-(r-2)/4}E|y_{mn}|^{1+r/2} \le Cn^{-(r-2)/4}\beta^{(2+r)/4}h_n^{d(4-r^2)/4r}. \end{aligned}$$

Therefore, (A4) and $\delta_n \to \infty$ yield that $g_n(\epsilon) \le C\delta_n^{-(r-2)/4}(n^{-1}h_n^{-d(1+4/r)})^{(r-2)/8} \to 0$.

(e) We first prove that

$$\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-1}(Y_i)I(\theta_q<Y_i\le\theta_n)=c_1f(\mathbf{x},\theta_q)+o_p(1). \tag{33}$$

Put $\xi_i=K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-1}(Y_i)I(\theta_q<Y_i\le\theta_n)$. The proof follows the line as in (a). In view of $\mathbf{x}\in C(l)$, (A2)(ii) and (B2), from (3) we have

$$\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^{n}E\xi_i=c_1f(\mathbf{x},\theta_q)+o(1). \tag{34}$$

Note that $\frac{\mu^2}{nh_n^d}\sum_{i=1}^n\mathrm{Var}(\xi_i)=O((nh_n^d)^{-1/2})$ and $|\mathrm{Cov}(\xi_i,\xi_j)|\le C\min\{h_n^d/n,\ h_n^d/n)^{1/r}[\alpha(j-i)]^{1-2/r}\}$. Then, from (A3) (ii) it follows that

$$\begin{aligned}&\mathrm{Var}\left(\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^{n}\xi_i\right)\\&=\frac{\mu^2}{nh_n^d}\left\{\sum_{i=1}^{n}\mathrm{Var}(\xi_i)+\left(\sum_{0<j-i\le[(n/h_n^d)^{1/2}]}+\sum_{j-i>[(n/h_n^d)^{1/2}]}\right)\mathrm{Cov}(\xi_i,\xi_j)\right\}\\&=O((nh_n^d)^{-1/2}).\end{aligned} \tag{35}$$

Therefore, (34) and (35) yield (33). Note that

$$\sum_{i=1}^{n}w_{3i}(\theta_n)=\sum_{i=1}^{n}w_{3i}(\theta_q)+\left(0^\tau,\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-1}(Y_i)I(\theta_q<Y_i\le\theta_n)\right)^\tau.$$

Therefore, from (d) and (33) it follows that $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{3i}(\theta_n)\xrightarrow{\mathcal{D}}N(\phi,\Sigma_3)$ and

$$\frac{\mu}{nh_n^d}\sum_{i=1}^{n}w_{3i}(\theta_n)=\frac{\mu}{nh_n^d}\sum_{i=1}^{n}w_{3i}(\theta_q)+O_p((nh_n^d)^{-1/2}). \tag{36}$$

Next we verify (a) when $\theta_q=\theta_n$. Obviously

$$\begin{aligned}&w_{3i}(\theta_n)w_{3i}^\tau(\theta_n)=w_{3i}(\theta_q)w_{3i}^\tau(\theta_q)+K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-2}(Y_i)\\&\quad\begin{pmatrix}0 & g(Y_i)I(\theta_q<Y_i\le\theta_n)\\ g^\tau(Y_i)I(\theta_q<Y_i\le\theta_n) & I(\theta_q<Y_i\le\theta_n)[1+2(I(Y_i\le\theta_q)-q)]\end{pmatrix}.\end{aligned} \tag{37}$$

Note that for $1 \le k \le \kappa$, by using (A2)(i) and (A6) it follows that

$$\frac{\mu}{nh_n^d}\sum_{i=1}^n E\left|K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-2}(Y_i)g_k(Y_i)I(\theta_q < Y_i \le \theta_n)\right| \to 0; \tag{38}$$

$$\frac{\mu}{nh_n^d}\sum_{i=1}^n E\left|K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)G^{-2}(Y_i)I(\theta_q < Y_i \le \theta_n)[1+2(I(Y_i \le \theta_q)-q)]\right| \to 0. \tag{39}$$

Therefore, in view of (a), from (37) to (39) it yields that

$$\frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_n)w_{3i}^\tau(\theta_n) = \frac{\mu}{nh_n^d}\sum_{i=1}^n w_{3i}(\theta_q)w_{3i}^\tau(\theta_q) + o_p(1) = \Sigma_3/\mu + o_p(1). \tag{40}$$

By applying (36) and (40), similarly to the evaluate in (b) and (c) above, the conclusion in (b) and (c) remains true for $\theta_q = \theta_n$. □

Proof of Lemma 2 Following the line as in the proof of Lemma 1, Lemma 2 can be proved. □

Proof of Lemma 3 Results (a)–(d) in Lemma 3 are obtained from Lemma 2 for the particular choice $g(y) = I(y \le \theta_q) - q$. Then, we prove only the last statement of the Lemma, that is, $\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_n) \xrightarrow{\mathcal{D}} N(c_1 f(\mathbf{x}, \theta_q), \Sigma_1)$. From (33) we have

$$\frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_n) = \frac{\mu}{\sqrt{nh_n^d}}\sum_{i=1}^n w_{1i}(\theta_q) + c_1 f(\mathbf{x}, \theta_q) + o_p(1),$$

and then the result follows by (d) in Lemma 3. □

References

Cai, Z. W. (2001). Estimating a distribution function for censored time series data. *Journal of Multivariate Analysis, 78*, 299–318.

Cai, Z. W. (2002). Regression quantiles for time series. *Econometric Theory, 18*, 169–192.

Cai, J., Kim, J. (2003). Nonparametric quantile estimation with correlated failure time data. *Lifetime Data Analysis, 9*, 357–371.

Cai, T., Wei, L. J., Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika, 87*, 867–878.

Chaudhuri, P., Doksum, K., Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics, 25*, 715–744.

Chen, J., Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika, 80*, 107–116.

Chen, J., Wu, C. (2002). Estimation of distribution function and quantiles using the model calibrated pseudo empirical likelihood method. *Statistica Sinica, 12*, 1223–1239.

Chen, S. X., Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics, 21*, 1166–1181.

Chen, S. X., Wong, C. M. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statistica Sinica, 19*, 71–81.

Dabrowska, D. (1992). Nonparametric quantile regression with censored data. *Sankhyà, 54*, 252–259.

Doukhan, P. (1994). *Mixing: Properties and examples, lecture notes in statistics* (Vol. 85). Berlin: Springer.

Fan, J., Hu, T. C., Truong, Y. K. (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics, 21*, 433–446.

Feigelson, E. D., Babu, G. J. (1992). *Statistical challenges in modern astronomy*. New York: Springer.

Ferraty, F., Rabhi, A., Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. *Sankyhà, 67*, 378–398.

Gannoun, A., Saracco, J., Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *Journal of Statistical Planning and Inference, 117*, 207–223.

Hall, P. (1992). *The bootstrap and edgeworth expansion*. New York: Springer.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.

Hall, P., La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review, 58*, 109–127.

He, S., Yang, G. (1994). Estimating a lifetime distribution under different sampling plan. In S. S. Gupta, J. O. Berger (eds.), *Statistical decision theory and related topics* (Vol. 5, pp. 73–85). NewYork: Springer.

He, S., Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics, 26*, 1011–1027.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Annals of the Institute of Statistical Mathematics, 52*, 459–470.

Iglesias-Pérez, M. C. (2003). Strong representation of a conditional quantile function estimator with truncated and censored data. *Statistics & Probability Letters, 65*, 79–91.

Lecoutre, J. P., Ould-Saïd, E. (1995). Convergence of the conditional Kaplan-Meier estimate under strong mixing. *Journal of Statistical Planning and Inference, 44*, 359–369.

Liang, H. Y., Uña-Álvarez, J. (2011). Asymptotic properties of conditional quantile estimator for censored dependent observations. *Annals of the Institute of Statistical Mathematics, 63*, 267–289.

Liang, H. Y., de Uña-Álvarez, J., Iglesias-Pérez, M. C. (2011). Local polynomial estimation of a conditional mean function with dependent truncated data. *Test*, in Press.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society, 155*, 95–118.

Mehra, K. L., Rao, M. S., Upadrasta, S. P. (1991). A smooth conditional quantile estimator and related applications of conditional empirical processes. *Journal of Multivariate Analysis, 37*, 151–179.

Ould-Saïd, E. (2006). A strong uniform convergence rate of kernel conditional quantile estimator under random censorship. *Statistics & Probability Letters, 76*, 579–586.

Ould-Saïd, E., Tatachak, A. (2007). Asymptotic properties of the kernel estimator of the conditional mode for the left truncated model. *Comptes Rendus de l'Académie Des Sciences I Mathematics, 344*, 651–656.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika, 75*, 237–249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *The Annals of Statistics, 18*, 90–120.

Polonik, W., Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *Journal of Multivariate Analysis, 80*, 234–255.

Qin, Y. S., Wu, Y. (2001). An estimator of a conditional quantile in the presence of auxiliary information. *Journal of Statistical Planning and Inference, 99*, 59–70.

Shao,Q., Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability, 24*, 2098–2127.

Srivastava, M. S., Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. New York: North-Holland.

Van Keilegom, I., Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *Journal of Statistical Planning and Inference, 69*, 115–131.

Volkonskii, V. A., Rozanov, Y. A. (1959). Some limit theorems for random functions. *Theory of Probability and its Applications, 4*, 178–197.

Withers, C. S. (1981). Conditions for linear processes to be strong mixing. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete, 57*, 477–480.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics, 13*, 163–177.

Xiang, X. (1995). Deficiency of samples quantile estimator with respect to kernel estimator for censored data. *The Annals of Statistics, 23*, 836–854.

Xiang, X. (1996). A kernel estimator of a conditional quantile. *Journal of Multivariate Analysis, 59*, 206–216.

Zhang, B. (1997). Empirical likelihood confidence intervals for M-functionals in the presence of auxiliary information. *Statistics & Probability Letters, 32*, 87–97.

Zhao, Y. (2011). Empirical likelihood inference for the accelerated failure time model. *Statistics & Probability Letters, 81*, 603–610.

Zhou, M., Li, G. (2008). Empirical likelihood analysis of the Buckley-James estimator. *Journal of Multivariate Analysis, 99*, 649–664.

Zhou, W., Jing, B.-Y. (2003). Adjusted empirical likelihood method for quantiles. *Annals of the Institute of Statistical Mathematics, 55*, 689–703.

Zhou, Y., Liang, H. (2000). Asymptotic normality for L1 norm kernel estimator of conditional median under α-mixing dependence. *Journal of Multivariate Analysis, 73*, 136–154.