Ann Inst Stat Math (2012) 64:255–273
DOI 10.1007/s10463-010-0310-9

Asymptotic normality of Powell's kernel estimator

Kengo Kato

Received: 18 May 2009 / Revised: 22 April 2010 / Published online: 14 October 2010

Abstract We establish asymptotic normality of Powell's kernel estimator for the asymptotic covariance matrix of the quantile regression estimator for both i.i.d. and weakly dependent data. As an application, we derive the optimal bandwidth that minimizes the approximate mean squared error of the kernel estimator. We also derive the corresponding results to censored quantile regression.

Keywords Asymptotic normality · Bandwidth selection · Censored quantile regression · Density estimation · Kernel method · Quantile regression

1 Introduction

This paper establishes asymptotic normality of Powell's (1991) kernel estimator for the asymptotic covariance matrix of the quantile regression estimator. Let us first introduce a quantile regression model. Let $(Y_i, \mathbf{X}_i)(i = 1, 2, \ldots, n)$ be i.i.d. observations from $(Y, \mathbf{X})$ where Y is a response variable and $\mathbf{X}$ is a d-dimensional covariate vector. The τth ($\tau \in (0, 1)$) conditional linear quantile regression model is defined as

$$Q_Y(\tau|\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_0(\tau), \tag{1}$$

where $Q_Y(\tau|\mathbf{X}) = \inf\{y : \mathrm{P}(Y \le y|\mathbf{X}) \ge \tau\}$ is the τth conditional quantile function of Y given $\mathbf{X}$. Koenker and Bassett (1978) propose the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau)$ for $\boldsymbol{\beta}_0(\tau)$ which minimizes the objective function $\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}_i'\boldsymbol{\beta})$, where $\rho_\tau(u) = \{\tau - I(u \le 0)\}u$ is called the check function ($I(\cdot)$ denotes the indicator function). It is well known that,

K. Kato (✉)
Department of Mathematics, Graduate School of Science, Hiroshima University,
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
e-mail: kkato@hiroshima-u.ac.jp

under suitable regularity conditions, $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau)$ satisfies consistency and asymptotic normality (see Koenker 2005, Chapter 4). Let $f(y|\mathbf{x})$ denote the conditional density of Y given $\mathbf{X} = \mathbf{x}$. Then the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau) - \boldsymbol{\beta}_0(\tau))$ is given by $\mathbf{J}^{-1}(\tau)\boldsymbol{\Sigma}(\tau)\mathbf{J}^{-1}(\tau)$, where $\mathbf{J}(\tau) = \mathrm{E}[f(\mathbf{X}'\boldsymbol{\beta}_0(\tau)|\mathbf{X})\mathbf{X}\mathbf{X}']$ and $\Sigma(\tau) = \tau(1-\tau)\mathrm{E}[\mathbf{X}\mathbf{X}']$. The estimation of the matrix $\Sigma(\tau)$ is straightforward. However, the matrix $\mathbf{J}(\tau)$ involves the conditional density and we should care for the estimation of $\mathbf{J}(\tau)$. Section 3.4 of Koenker (2005) introduces two approaches to the estimation of the matrix $\mathbf{J}(\tau)$. The first one, suggested by Hendricks and Koenker (1992), is a natural extension of the scalar sparsity estimation by Siddiqui (1960). On the other hand, Powell (1991) proposes the kernel estimator

$$\hat{\mathbf{J}}_{\mathrm{P}}(\tau) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}(\tau)}{h}\right)\mathbf{X}_i\mathbf{X}_i',$$

where $\hat{\boldsymbol{\beta}}(\tau)$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0(\tau)$ (usually, we take $\hat{\boldsymbol{\beta}}(\tau) = \hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau)$) and $K(\cdot)$ is the uniform kernel $K(u) = 2^{-1}I(|u| \leq 1)$. Powell (1991) shows that $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ is consistent under some regularity conditions. Especially, he imposes the condition on the bandwidth h that $h \to 0$ and $nh^2 \to \infty$. The recent study by Angrist et al. (2006) shows that $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ is uniformly consistent over a closed interval of τ even when the model is misspecified. However, to the author's knowledge, there is no literature that rigorously studies the asymptotic distribution of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ itself.

This paper establishes asymptotic normality of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ under the conditions that the conditional density is twice continuously differentiable and that the bandwidth h is such that $h \to 0$ and $(n^{1/2}h)/\log n \to \infty$. The condition on the bandwidth is close to the one required for proving consistency of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$. As an application, we evaluate the approximate mean squared error (AMSE) of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ and derive the optimal h that minimizes the AMSE, which is another contribution of this paper. Since the kernel estimator contains the estimated parameter in the sum, the direct calculation of the mean squared error (MSE) is infeasible. So the evaluation of the MSE is not a trivial task. We also derive the corresponding results to censored quantile regression. It should be noted that Powell originally suggests to use the kernel method to estimate the asymptotic covariance matrix for censored quantile regression (see Powell 1984, 1986). In addition, we extend the results to weakly dependent data.

We now review the literature related to this paper. Koul (1992) discusses the uniform convergence of the kernel estimator of the error density in a linear model based on the weak convergence results of the residual empirical processes. Chai et al. (1991), Chai and Li (1993) and Li (1995) show several important asymptotic results for the kernel estimation of the error density in a linear model with fixed design when using the least squares method and the least absolute deviation method to estimate the coefficients. Especially, the latter two papers show asymptotic normality of the histogram estimator (namely, the estimator using the uniform kernel) of the error density. Unfortunately, the proof of Lemma 4 in Chai and Li (1993), which is a key to their asymptotic normality results, is incorrect. See the remark after the proof of Lemma 1 below. Besides, the differences of the present paper from theirs are as follows: (i) Chai and Li treat the estimation of the scalar unconditional error density and the present paper treats the estimation of the matrix that involves the conditional density. This difference affects the bandwidth selection. See Sect. 3. (ii) Chai and Li impose the stringent condition that the covariate vectors are bounded over all observations. The present paper removes this condition. (iii) Chai and Li only treat independent data, while the present paper treats both i.i.d. and weakly dependent data. After submitting the paper, I found that Hall and Horowitz (1990) address a similar issue to ours; however, they do not directly analyze Powell's estimator that the present paper handles, and their proof strategy is different from ours.

The estimation of the innovation density in parametric time series models is studied by Robinson (1987), Liebscher (1999), Müller et al. (2005) and Schick and Wefelmeyer (2007). Among them, Liebscher (1999) establishes asymptotic normality of the residual-based kernel estimator of the innovation density for a nonlinear autoregressive model. He assumes that the kernel function is Lipschitz continuous, which is essential to his proof, while the uniform kernel treated in the present paper is not continuous at the end points. The estimation of the error density in nonparametric regression causes much attention in recent years. Several authors who address this issue include Ahmad (1992), Cheng (2002, 2004, 2005), Efromovich (2005, 2007a,b) and Liang and Niu (2009). Cheng (2005) and Liang and Niu (2009) show asymptotic normality of their kernel estimators; both of them uses the uniform kernel when deriving the asymptotic distributions.

The rest of the paper is organized as follows. In Sect. 2, we prove asymptotic normality of Powell's kernel estimator $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ for i.i.d. data. In Sect. 3, we use the asymptotic distribution to evaluate the AMSE and derive the optimal h that minimizes the AMSE. In Sect. 4, we derive the corresponding results to censored quantile regression. In Sect. 5, we establish asymptotic normality of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ under a weak dependence condition. In Sect. 6, we leave some concluding remarks.

We introduce some notations used in the present paper. Let $I(A)$ denote the indicator of an event A. The symbols "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote "convergence in probability" and "convergence in distribution", respectively. We use the stochastic orders $o_p(\cdot)$ and $O_p(\cdot)$ in the usual sense. For a real number a, $[a]$ denotes the greatest integer not exceeding a. For a $d \times d$ matrix $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_d]$, $\mathrm{vec}(\mathbf{A}) = (\mathbf{a}_1', \ldots, \mathbf{a}_d')'$.

2 Asymptotic normality of Powell's kernel estimator

In this section, we study the first order asymptotic property of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ under the i.i.d. condition. Throughout this and the next sections, we fix τ and suppress the dependence on τ for notational convenience. For example, we simply write $\boldsymbol{\beta}_0$ for $\boldsymbol{\beta}_0(\tau)$. Then, the model (1) may be written as

$$Y = \mathbf{X}'\boldsymbol{\beta}_0 + U, \quad Q_U(\tau|\mathbf{X}) = 0, \tag{2}$$

where $Q_U(\tau|\mathbf{X}) = \inf\{u : \mathrm{P}(U \le u|\mathbf{X}) \ge \tau\}$. It should be noted that the distribution of U generally depends on τ and $\mathbf{X}$. For example, let us consider a linear location scale model

$$Y = \mathbf{X}'\boldsymbol{\theta}_0 + (\mathbf{X}'\boldsymbol{\gamma}_0)\epsilon, \tag{3}$$

where $\mathbf{X}'\boldsymbol{\gamma}_0 > 0$ and ϵ is independent of $\mathbf{X}$. In this model, U corresponds to $\mathbf{X}'\boldsymbol{\gamma}\{\epsilon - F^{-1}(\tau)\}$, where F is the distribution function of ϵ. Typically, the model (3) allows for the heteroscedasticity of U.

We now return to the general model (2). Let $f_0(u|\mathbf{x})$ denote the conditional density of U given $\mathbf{X} = \mathbf{x}$. Then the matrix $\mathbf{J}$ is expressed as $\mathrm{E}[f_0(0|\mathbf{X})\mathbf{X}\mathbf{X}']$. In order to justify our asymptotic theory, we impose the following regularity conditions:

Assumption 1 $\{(U_i, \mathbf{X}_i), i = 1, 2, \ldots\}$ is an i.i.d. sequence whose marginal distribution is the same as $(U, \mathbf{X})$.

Assumption 2 The conditional density $f_0(u|\mathbf{x})$ of U given $\mathbf{X} = \mathbf{x}$ is twice continuously differentiable with respect to u for each $\mathbf{x}$. Furthermore, there exist measurable functions $G_j(\mathbf{x})(j = 0, 1, 2)$ such that $|f_0^{(j)}(u|\mathbf{x})| \le G_j(\mathbf{x})$ $(j = 0, 1, 2)$ for every realization $(u, \mathbf{x})$ of $(U, \mathbf{X})$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^4 + \|\mathbf{X}\|^5)G_0(\mathbf{X})] < \infty$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^3)G_1(\mathbf{X})] < \infty$ and $\mathrm{E}[\|\mathbf{X}\|^2 G_2(\mathbf{X})] < \infty$, where $f_0^{(j)}(u|\mathbf{x}) = \partial^j f_0(u|\mathbf{x})/\partial u^j$ for $j = 0, 1, 2$.

Assumption 3 As $n \to \infty$, $h \to 0$ and $(n^{1/2}h)/\log n \to \infty$.

We state some remarks on the conditions. We substantially assume the existence of the fifth order moment of $\mathbf{X}$, which is slightly stronger than the one assumed in proving consistency of $\hat{\mathbf{J}}_{\mathrm{P}}$. For example, Angrist et al. (2006) assume the fourth order moment of $\mathbf{X}$ to prove (uniform) consistency of $\hat{\mathbf{J}}_{\mathrm{P}}$. The first part of Assumption 2 is standard in the (conditional) density estimation literature (for example, see Fan and Yao 2005, Chapter 5). Unlike the fully nonparametric conditional density estimation, the effect of localization on the $\mathbf{X}$-space does not work in the present situation. Thus, the latter part of Assumption 2 is needed to ensure the dominated convergence. Assumption 3 allows for bandwidth rules such as the rule used in `R` implementation of the kernel estimation in `quantreg` package (Koenker 2009), the Bofinger (1975) and the Hall-Sheather (1988) rules. The Bofinger and the Hall-Sheather rules are originally for the scalar sparsity estimation but also used for the kernel estimation by some authors. Powell (1991) and other authors show consistency of $\hat{\mathbf{J}}_{\mathrm{P}}$ under the condition that $h \to 0$ and $nh^2 \to \infty$.

For any fixed matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$, define

$$\begin{aligned} s_n(\boldsymbol{\beta}) &= \frac{1}{nh}\sum_{i=1}^{n} Z_i K\left(\frac{Y_i - \mathbf{X}_i'\boldsymbol{\beta}}{h}\right) \\ &= \frac{1}{nh}\sum_{i=1}^{n} Z_i K\left(\frac{U_i - \mathbf{X}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{h}\right), \end{aligned} \tag{4}$$

where $Z_i = \mathrm{tr}(\mathbf{A}\mathbf{X}_i\mathbf{X}_i')$. We first show asymptotic normality of $s_n(\hat{\boldsymbol{\beta}})$. Then, we use the Cramér-Wold device to derive the asymptotic distribution of $\hat{\mathbf{J}}_{\mathrm{P}}$. The proof of asymptotic normality of $s_n(\hat{\boldsymbol{\beta}})$ consists of series of lemmas. Lemma 1 uses the empirical process technique to establish the uniform convergence in probability. See, for example, Chapter 2 of van der Vaart and Wellner (1996) for related materials.

Lemma 1 *Suppose that Assumptions 1–3 hold. Then, for any fixed $l > 0$, we have $s_n(\boldsymbol{\beta}) - \mathrm{E}[s_n(\boldsymbol{\beta})] = s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)] + o_p((nh)^{-1/2})$ uniformly in $\|\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le l$.*

Proof We have to show $s_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t}) - \mathrm{E}[s_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t})] = s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)] + o_p((nh)^{-1/2})$ uniformly in $\|\mathbf{t}\| \le l$. Observe that

$$
\begin{aligned}
& h\{s_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t}) - s_n(\boldsymbol{\beta}_0)\} \\
&= \frac{1}{n}\sum_{i=1}^{n} Z_i \left\{ K\left(\frac{U_i - n^{-1/2}\mathbf{X}_i'\mathbf{t}}{h}\right) - K\left(\frac{U_i}{h}\right) \right\} \\
&= \frac{1}{2}\left\{ \frac{1}{n}\sum_{i=1}^{n} Z_i I(h < U_i \le h + n^{-1/2}\mathbf{X}_i'\mathbf{t}) \right. \\
&\qquad - \frac{1}{n}\sum_{i=1}^{n} Z_i I(-h \le U_i < -h + n^{-1/2}\mathbf{X}_i'\mathbf{t}) \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n} Z_i I(-h + n^{-1/2}\mathbf{X}_i'\mathbf{t} \le U_i < -h) \\
&\qquad \left. - \frac{1}{n}\sum_{i=1}^{n} Z_i I(h + n^{-1/2}\mathbf{X}_i'\mathbf{t} < U_i \le h) \right\} \\
&=: \frac{1}{2}\{W_{1n}(\mathbf{t}) - W_{2n}(\mathbf{t}) + W_{3n}(\mathbf{t}) - W_{4n}(\mathbf{t})\}. \qquad (5)
\end{aligned}
$$

It suffices to show that $n^{1/2}h^{-1/2}\{W_{jn}(\mathbf{t}) - \mathrm{E}[W_{jn}(\mathbf{t})]\} \xrightarrow{p} 0$ uniformly in $\|\mathbf{t}\| \le l$ for $j = 1, 2, 3, 4$. We only prove the $j = 1$ case since the proofs for the other cases are completely analogous.

Fix any $\epsilon > 0$. Define $U_i^*(\mathbf{t}) = Z_i I(h < U_i \le h + n^{-1/2}\mathbf{X}_i'\mathbf{t})$. Let $\sigma_1, \ldots, \sigma_n$ be independent and uniformly distributed over $\{-1, 1\}$ and independent of $(U_1, \mathbf{X}_1), \ldots, (U_n, \mathbf{X}_n)$. Using the symmetrization technique (van der Vaart and Wellner 1996, Lemma 2.3.7), we have

$$
\begin{aligned}
& \eta_n \mathrm{P}\left(\sup_{\|\mathbf{t}\| \le l} |W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]| > n^{-1/2}h^{1/2}\epsilon \right) \\
&\le 2\mathrm{P}\left(\sup_{\|\mathbf{t}\| \le l} \left| \frac{1}{n}\sum_{i=1}^{n} \sigma_i U_i^*(\mathbf{t}) \right| > \frac{n^{-1/2}h^{1/2}\epsilon}{4} \right),
\end{aligned}
$$

where $\eta_n = 1 - (4/(\epsilon^2 h)) \sup_{\|\mathbf{t}\| \le l} \mathrm{E}[\{U_1^*(\mathbf{t})\}^2]$. Let $F_0(u|\mathbf{x})$ denote the conditional distribution function of U given $\mathbf{X} = \mathbf{x}$. Then, we have

$$\begin{aligned}
\sup_{\|\mathbf{t}\| \le l} \mathrm{E}[\{U_i^*(\mathbf{t})\}^2] &\le \mathrm{E}[|Z|^2 I(h < U \le h + n^{-1/2} l \|\mathbf{X}\|)] \\
&= \mathrm{E}[|Z|^2 \{F_0(h + n^{-1/2} l \|\mathbf{X}\| |\mathbf{X}) - F_0(h|\mathbf{X})\}] \\
&\le l n^{-1/2} \mathrm{E}[|Z|^2 G_0(\mathbf{X}) \|\mathbf{X}\|],
\end{aligned}$$

where we have used $F_0(h + n^{-1/2} l \|\mathbf{X}\| |\mathbf{X}) - F_0(h|\mathbf{X}) \le l n^{-1/2} G_0(\mathbf{X}) \|\mathbf{X}\|$. Since $nh^2 \to \infty$, $\eta_n = 1 - o(1)$ as $n \to \infty$ and consequently $\eta_n \ge 1/2$ for large n. Thus, for large n,

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\|\mathbf{t}\| \le l} |W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]| > n^{-1/2} h^{1/2} \epsilon \right) \\
&\quad \le 4\mathrm{P}\left(\sup_{\|\mathbf{t}\| \le l} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i U_i^*(\mathbf{t}) \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \right).
\end{aligned}$$

Let $D_n = \{(U_i, \mathbf{X}_i), i = 1, \ldots, n\}$. Given D_n, at most finite elements are contained in the functional set $\{\boldsymbol{\sigma}^{(n)} \mapsto n^{-1} \sum_{i=1}^n \sigma_i U_i^*(\mathbf{t}) : \|\mathbf{t}\| \le l\}$, where $\boldsymbol{\sigma}^{(n)} = (\sigma_1, \ldots, \sigma_n)$, since every element of the functional set is of the form $\boldsymbol{\sigma}^{(n)} \mapsto n^{-1} \sum_{i \in \{\text{subset of } \{1,\ldots,n\}\}} \sigma_i Z_i$. Let k_n denote the cardinality of this set. Then, there exist k_n points $\mathbf{t}_j \in \{\mathbf{t} : \|\mathbf{t}\| \le l\}$, $j = 1, \ldots, k_n$ such that

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\|\mathbf{t}\| \le l} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i U_i^*(\mathbf{t}) \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \middle| D_n \right) \\
&\quad \le \sum_{j=1}^{k_n} \mathrm{P}\left(\left| \frac{1}{n} \sum_{i=1}^n \sigma_i U_i^*(\mathbf{t}_j) \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \middle| D_n \right).
\end{aligned}$$

It is noted that k_n and $\mathbf{t}_j (j = 1, \ldots, k_n)$ depend on D_n. Observe that for any $\|\mathbf{t}\| \le l$,

$$\begin{aligned}
&-|Z_i| I(h < U_i \le h + n^{-1/2} l \|\mathbf{X}_i\|) \le \sigma_i U_i^*(\mathbf{t}) \\
&\quad \le |Z_i| I(h < U_i \le h + n^{-1/2} l \|\mathbf{X}_i\|).
\end{aligned}$$

By Hoeffding's inequality (van der Vaart and Wellner 1996, Lemma 2.2.7),

$$\sup_{\|\mathbf{t}\| \le l} \mathrm{P}\left(\left| \frac{1}{n} \sum_{i=1}^n \sigma_i U_i^*(\mathbf{t}) \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \middle| D_n \right) \le 2 \exp\left(-\frac{\epsilon^2 h}{32 v_n} \right),$$

where $v_n = n^{-1}\sum_{i=1}^n |Z_i|^2 I(h < U_i \le h + n^{-1/2} l\|\mathbf{X}_i\|)$. Hence,

$$\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i U_i^*(\mathbf{t})\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{4}\,\middle|\, D_n\right) \le 2k_n \exp\left(-\frac{\epsilon^2 h}{32 v_n}\right).$$

We now bound k_n. It is not difficult to see that k_n is bounded by the cardinality of the set $\{A \cap \{(U_1, \mathbf{X}_1), \ldots, (U_n, \mathbf{X}_n)\} : A \in \mathcal{A}\}$, where $\mathcal{A} = \{\{(u, \mathbf{x}) : u > h, u \le h + \mathbf{x}'\mathbf{t}\} : h \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^d\}$. Application of Lemma 2.6.15 in van der Vaart and Wellner (1996) shows that the VC dimension $V_{\mathcal{A}}$ of $\mathcal{A}$ is finite, namely $0 < V_{\mathcal{A}} < \infty$. Then, Sauer's lemma (van der Vaart and Wellner 1996, Corollary 2.6.3) implies that k_n is bounded by $cn^{V_{\mathcal{A}}}$ for some constant c not depending on D_n. Therefore, we have

$$\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i U_i^*(\mathbf{t})\right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{4}\,\middle|\, D_n\right) \le 2cn^{V_{\mathcal{A}}-1} \exp\left(-\frac{\epsilon^2 h_n}{32 v_n}\right). \quad (6)$$

Define

$$A_n = \left\{ v_n > \frac{\epsilon^2 h_n}{32 V_{\mathcal{A}} \log n}\right\}.$$

Using (6) and the obvious inequality, we have

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i U_i^*(\mathbf{t})\right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{4}\right) \\
&\le \mathrm{P}(A_n) + 2cn^{V_{\mathcal{A}}-1}\mathrm{E}\left[\exp\left(-\frac{\epsilon^2 h_n}{32 v_n}\right) I(A_n^c)\right] \\
&\le \mathrm{P}(A_n) + 2cn^{-1}.
\end{aligned}$$

To show that $\mathrm{P}(A_n) \to 0$, it suffices to show that $(\log n)h_n^{-1} v_n \xrightarrow{p} 0$. By Markov's inequality, for any $\delta > 0$,

$$\begin{aligned}
\mathrm{P}\left(v_n > \frac{h_n \delta}{\log n}\right) &\le \delta^{-1}(\log n) h_n^{-1}\mathrm{E}[|Z|^2 I(h_n < U \le h_n + n^{-1/2} l\|\mathbf{X}\|)] \\
&\le l\delta^{-1} n^{-1/2}(\log n) h_n^{-1}\mathrm{E}[|Z|^2 G_0(\mathbf{X})\|\mathbf{X}\|] \to 0.
\end{aligned}$$

Therefore, we complete the proof. □

Remark 1 The proof of Lemma 4 in Chai and Li (1993) states that the cardinality of the functional set $\{\boldsymbol{\sigma}^{(n)} \mapsto n^{-1}\sum_{i=1}^n \sigma_i I(a_n < e_i < a_n + h_i) : 0 < h_i \le b_n\}$ is bounded by $(n+1)$, where $\{e_i\}$ is arbitrarily fixed, a_n is the bandwidth such that $a_n \to 0$ and $b_n = Cn^{-1/2}$. However, this statement is incorrect. For example, if $a_n < e_i < a_n + b_n$ for $i = 1, \ldots, n$, the cardinality of the functional set is 2^n.

Lemma 2 *Suppose that Assumptions 1–3 hold. Then, for any fixed $l > 0$, we have* $\mathrm{E}[s_n(\boldsymbol{\beta})] = \mathrm{E}[s_n(\boldsymbol{\beta}_0)] + O(n^{-1/2})$ *uniformly in* $\|\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le l$.

Proof We have to show $\mathrm{E}[s_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t})] = \mathrm{E}[s_n(\boldsymbol{\beta}_0)] + O(n^{-1/2})$ uniformly in $\|\mathbf{t}\| \le l$. Observe that for any $\|\mathbf{t}\| \le l$,

$$\begin{aligned}
&|\mathrm{E}[s_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t})] - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]| \\
&\quad = \left|\mathrm{E}\left[Z\int K(u)\{f_0(uh + n^{-1/2}\mathbf{X}'\mathbf{t}|\mathbf{X}) - f_0(uh|\mathbf{X})\}du\right]\right| \\
&\quad \le l n^{-1/2}\mathrm{E}\left[|Z|G_1(\mathbf{X})\|\mathbf{X}\|\right].
\end{aligned}$$

This yields the desired result. □

Lemma 3 *Under Assumptions 1–3, we have* $(nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 f_0(0|\mathbf{X})]/2)$.

Proof This result can be proved by checking the conditions of the Lindeberg–Feller central limit theorem. Since the argument is standard, we omit the detail. □

Suppose that $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$, namely $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(n^{-1/2})$. Then, by Lemmas 1 and 2,

$$\begin{aligned}
&(nh)^{1/2}\{s_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} \\
&\quad = (nh)^{1/2}\{s_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[s_n(\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\} + (nh)^{1/2}\{\mathrm{E}[s_n(\boldsymbol{\beta})]|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} \\
&\quad = (nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} + o_p(1).
\end{aligned}$$

Using the Taylor expansion, we see that

$$\mathrm{E}[s_n(\boldsymbol{\beta}_0)] = \mathrm{E}[Zf_0(0|\mathbf{X})] + \frac{h^2}{6}\mathrm{E}[Zf_0^{(2)}(0|\mathbf{X})] + o(h^2).$$

We now describe the asymptotic distribution of the matrix estimator $\hat{\mathbf{J}}_{\mathrm{P}}$. Let $\mathbf{S} = \mathbf{X}\mathbf{X}'$. Since $\mathrm{tr}(\mathbf{A}\mathbf{S}) = \mathrm{vec}(\mathbf{A}')'\,\mathrm{vec}(\mathbf{S})$, the asymptotic covariance matrix of $s_n(\hat{\boldsymbol{\beta}})$ is written as $2^{-1}\,\mathrm{vec}(\mathbf{A}')'\mathrm{E}[f_0(0|\mathbf{X})\,\mathrm{vec}(\mathbf{S})\,\mathrm{vec}(\mathbf{S})']\,\mathrm{vec}(\mathbf{A}')$. Therefore, the Cramér-Wold device leads to the next theorem:

Theorem 1 *Suppose that Assumptions 1–3 hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

$$(nh)^{1/2}\left\{\hat{\mathbf{J}}_P - \mathbf{J} - \frac{h^2}{6}\mathrm{E}[f_0^{(2)}(0|\mathbf{X})\mathbf{X}\mathbf{X}'] + o(h^2)\right\}$$

is asymptotically normally distributed with zero mean matrix. The asymptotic covariance of the (j,k)-th and the (l,m)-th elements is given by

$$\frac{1}{2}\mathrm{E}[f_0(0|\mathbf{X})X_jX_kX_lX_m],$$

where $j, k, l, m = 1, \ldots, d$.

We end this section with a remark. While we put the conditional quantile restriction on U, the proof of Theorem 2 does not use the restriction. Therefore, Theorem 2 is valid for any $\hat{\boldsymbol{\beta}}$ such that $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(n^{-1/2})$ for some $\boldsymbol{\beta}_0$. For example, when the model (1) is misspecified, $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$ that uniquely solves $\mathrm{E}[\{\tau - I(Y \le \mathbf{X}'\boldsymbol{\beta}_0)\}\mathbf{X}] = \mathbf{0}$, where the existence and the uniqueness of such $\boldsymbol{\beta}_0$ is assumed. See Angrist et al. (2006) for a proof of this result. Thus, Theorem 2 is valid for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathrm{KB}}$ even when the model is misspecified.

3 Application: bandwidth selection

Since $\hat{\mathbf{J}}_{\mathrm{P}}$ contains the estimated parameter in the sum, the direct calculation of the bias and the variance of $\hat{\mathbf{J}}_{\mathrm{P}}$ is infeasible. However, Theorem 2 enables us to approximate the mean squared error (MSE) of $\hat{\mathbf{J}}_{\mathrm{P}}$. From Theorem 2, we can see that the MSE is approximated as

$$\begin{aligned}\mathrm{MSE}(h) &:= \mathrm{E}[\mathrm{tr}\{(\hat{\mathbf{J}}_{\mathrm{P}} - \mathbf{J})^2\}] \\ &\simeq \frac{h^4}{36}\sum_{j,k=1}^{d}\left(\mathrm{E}[f_0^{(2)}(0|\mathbf{X})X_jX_k]\right)^2 + \frac{1}{2nh}\sum_{j,k=1}^{d}\mathrm{E}[f_0(0|\mathbf{X})X_j^2X_k^2] \\ &=: \mathrm{AMSE}(h).\end{aligned}$$

The optimal h that minimizes $\mathrm{AMSE}(h)$ is given by

$$h_{\mathrm{opt}} = n^{-1/5}\left\{\frac{4.5\sum_{j,k=1}^{d}\mathrm{E}[f_0(0|\mathbf{X})X_j^2X_k^2]}{\sum_{j,k=1}^{d}\left(\mathrm{E}[f_0^{(2)}(0|\mathbf{X})X_jX_k]\right)^2}\right\}^{1/5}$$

where we assume that the denominator is not zero. It should be noted that h_{opt} depends on τ, namely $h_{\mathrm{opt}} = h_{\mathrm{opt}}(\tau)$, since the distribution of U generally depends on τ. We further note that h_{opt} depends on the distribution of $\mathbf{X}$, which is the difference from the scalar (unconditional) density estimation. In the simple case where $f_0(u|\mathbf{x})$ is independent of $\mathbf{x}$, namely $f_0(u|\mathbf{x}) = f_0(u)$, h_{opt} depends on the (unconditional) error density and the second and the fourth order moments of $\mathbf{X}$.

It is well known that convergence in distribution does not necessarily imply moment convergence. In order to make the argument rigorous, we introduce the truncated MSE

$$\mathrm{MSE}_T(h) := \mathrm{E}\left[\min\left[\mathrm{tr}\left\{n^{4/5}\left(\hat{\mathbf{J}}_{\mathrm{P}} - \mathbf{J}\right)^2\right\}, T\right]\right]$$

and take the limit $n \to \infty$ and $T \to \infty$. Andrews (1991) uses the same device to evaluate covariance matrix estimators that contain estimated parameters in the different context. Then, the optimality of h_{opt} is stated as follows.

Proposition 1 *Suppose that Assumptions 1–3 hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then, $\lim_{T\to\infty} \lim_{n\to\infty}\{MSE_T(h) - MSE_T(h_{opt})\} \geq 0$, where the inequality is strict unless $h = h_{opt} + o(n^{-1/5})$.*

Proof The proposition follows from the fact that for a bounded sequence of random variables, convergence in distribution implies moment convergence of any order. □

As in the usual density estimation, h_{opt} involves unknown quantities and is not directly usable. In the density estimation literature, there are several methods, namely rule of thumb, cross validation and plug-in methods, to cope with this difficulty. For a comprehensive treatment on practical aspects of density estimation, see Sheather (2004) and references therein. For example, the optimal bandwidth h_{opt} for a Gaussian location model

$$Y = \mathbf{X}'\boldsymbol{\theta}_0 + \epsilon, \quad \epsilon|\mathbf{X} \sim N(0, 1),$$

is given by

$$h_{\text{opt}} = n^{-1/5} \left\{ \frac{4.5 \sum_{j,k=1}^{d} \mathrm{E}[X_j^2 X_k^2]}{\alpha(\tau) \sum_{j,k=1}^{d} \left(\mathrm{E}[X_j X_k]\right)^2} \right\}^{1/5},$$

where $\alpha(\tau) = \{1 - \Phi^{-1}(\tau)\}^2 \phi(\Phi^{-1}(\tau))$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution function and the density function of the standard normal distribution. Thus, a rule of thumb bandwidth for the Gaussian location model is given by

$$\hat{h}_{\text{ROT}} = n^{-1/5} \left\{ \frac{4.5 \sum_{j,k=1}^{d} \left(n^{-1} \sum_{i=1}^{n} X_{ij}^2 X_{ik}^2\right)}{\alpha(\tau) \sum_{j,k=1}^{d} \left(n^{-1} \sum_{i=1}^{n} X_{ij} X_{ik}\right)^2} \right\}^{1/5}.$$

4 Censored quantile regression

Powell (1984, 1986) originally suggests to use the kernel method to estimate the asymptotic covariance matrix for censored quantile regression. In Powell's censored quantile regression model, the latent variable Y^* is left censored by the observable, possibly random censoring point C. We observe $Y = \max\{Y^*, C\}$, $\mathbf{X}$ and C. Suppose that Y^* is independent of C conditionally on $\mathbf{X}$ and satisfies the τ-th conditional linear quantile restriction

$$Q_{Y^*}(\tau|\mathbf{X}, C) = \mathbf{X}'\boldsymbol{\beta}_0(\tau),$$

which yields the τth conditional censored quantile regression model

$$Q_Y(\tau|\mathbf{X}, C) = \max\{C, \mathbf{X}'\boldsymbol{\beta}_0(\tau)\}.$$

Without loss of generality, we may set $C = 0$. Thus, we consider the model

$$Q_Y(\tau|\mathbf{X}) = \max\{0, \mathbf{X}'\boldsymbol{\beta}_0(\tau)\}. \tag{7}$$

As in Sect. 2, we fix τ and suppress the dependence on τ for notational convenience. Then, the model (7) may be written as

$$Y = \max\{0, \mathbf{X}'\boldsymbol{\beta}_0 + U\}, \quad Q_U(\tau|\mathbf{X}) = 0,$$

where $U = Y^* - \mathbf{X}'\boldsymbol{\beta}_0$ in this case. Powell (1984, 1986) proposes the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{CQ}}$ for $\boldsymbol{\beta}_0$ which minimizes $\sum_{i=1}^n \rho_\tau(Y_i - \max\{0, \mathbf{X}_i'\boldsymbol{\beta}\})$. Assuming that the parameter space of $\boldsymbol{\beta}_0$ is a compact subset of $\mathbb{R}^d$, Powell (1984, 1986) shows that, under suitable regularity conditions, $\hat{\boldsymbol{\beta}}_{\mathrm{CQ}}$ satisfies consistency and asymptotic normality. Let $f_0(u|\mathbf{x})$ denote the conditional density of U given $\mathbf{X} = \mathbf{x}$. Then the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{CQ}} - \boldsymbol{\beta}_0)$ is given by $\mathbf{J}_{\mathrm{CQ}}^{-1}\boldsymbol{\Sigma}_{\mathrm{CQ}}\mathbf{J}_{\mathrm{CQ}}^{-1}$, where $\boldsymbol{\Sigma}_{\mathrm{CQ}} = \tau(1-\tau)\mathrm{E}[I(\mathbf{X}'\boldsymbol{\beta}_0 > 0)\mathbf{X}\mathbf{X}']$ and $\mathbf{J}_{\mathrm{CQ}} = \mathrm{E}[I(\mathbf{X}'\boldsymbol{\beta}_0 > 0) f_0(0|\mathbf{X})\mathbf{X}\mathbf{X}']$. The difference from the ordinary quantile regression is the appearance of the additional indicator term inside the expectations. Powell (1984, 1986) proposes the kernel estimator for $\mathbf{J}_{\mathrm{CQ}}$

$$\hat{\mathbf{J}}_{\mathrm{PCQ}} = \frac{1}{nh}\sum_{i=1}^n I(\mathbf{X}_i'\hat{\boldsymbol{\beta}} > 0) K_+\left(\frac{Y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}}{h}\right)\mathbf{X}_i\mathbf{X}_i',$$

where $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$ and $K_+(\cdot)$ is the one-sided uniform kernel $K_+(u) = I(0 \le u \le 1)$. Powell (1984) shows consistency of $\hat{\mathbf{J}}_{\mathrm{PCQ}}$ under the condition that $h \to 0$ and $nh^2 \to \infty$ but does not investigate the asymptotic distribution.

We now show asymptotic normality of $\hat{\mathbf{J}}_{\mathrm{PCQ}}$. For censored quantile regression, we impose the following regularity conditions:

Assumption 4 $\{(U_i, \mathbf{X}_i), i = 1, 2, \ldots\}$ is an i.i.d. sequence whose marginal distribution is the same as $(U, \mathbf{X})$.

Assumption 5 $\mathrm{E}[\|\mathbf{X}\|^5] < \infty$. In addition,

$$\mathrm{E}[\|\mathbf{X}\|^2 I(|\mathbf{X}'\boldsymbol{\beta}_0| \le z\|\mathbf{X}\|)] = O(z), \quad z \to 0. \tag{8}$$

Assumption 6 The conditional density $f_0(u|\mathbf{x})$ of U given $\mathbf{X} = \mathbf{x}$ is continuously differentiable with respect to u for each $\mathbf{x}$. Furthermore, there exist some constant $A_0 > 0$ and measurable function $G_1(\mathbf{x})$ such that $f_0(u|\mathbf{x}) \le A_0$, $|f_0^{(1)}(u|\mathbf{x})| \le G_1(\mathbf{x})$ for every realization $(u, \mathbf{x})$ of $(U, \mathbf{X})$ and $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^3)G_1(\mathbf{X})] < \infty$ where $f_0^{(1)}(u|\mathbf{x}) = \partial f_0(u|\mathbf{x})/\partial u$.

Assumption 7 As $n \to \infty$, $h \to 0$ and $(n^{1/2}h)/\log n \to \infty$.

The additional condition (8) corresponds to Assumption R.2. in Powell (1984). For an intuitive interpretation of this condition, see Powell (1984, pp.310). The other conditions are almost identical to Assumptions 1–3.

For any fixed $\mathbf{A} \in \mathbb{R}^{d\times d}$, define

$$
\begin{aligned}
s_n(\boldsymbol{\beta}) &= \frac{1}{nh}\sum_{i=1}^{n} Z_i I(\mathbf{X}_i'\boldsymbol{\beta} > 0) K_+\left(\frac{Y_i - \mathbf{X}_i'\boldsymbol{\beta}}{h}\right) \\
&= \frac{1}{nh}\sum_{i=1}^{n} Z_i I(\mathbf{X}_i'\boldsymbol{\beta} > 0) K_+\left(\frac{U_i - \mathbf{X}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{h}\right),
\end{aligned}
$$

where $Z_i = \mathrm{tr}(\mathbf{A}\mathbf{X}_i\mathbf{X}_i')$. The second equality is due to the fact that for $\mathbf{X}_i'\boldsymbol{\beta} > 0$, $I(\mathbf{X}_i'\boldsymbol{\beta} \le Y_i \le \mathbf{X}_i'\boldsymbol{\beta} + h) = I(\mathbf{X}_i'\boldsymbol{\beta} \le Y_i^* \le \mathbf{X}_i'\boldsymbol{\beta} + h)$ by the definition of Y. The next lemma is essential to our purpose.

Lemma 4 *Under Assumptions 4–7, the conclusions of Lemmas 1 and 2 are also valid in the present situation.*

Proof Decompose the difference $s_n(\boldsymbol{\beta}) - s_n(\boldsymbol{\beta}_0)$ as

$$
\begin{aligned}
s_n(\boldsymbol{\beta}) - s_n(\boldsymbol{\beta}_0) &= \frac{1}{nh}\sum_{i=1}^{n} Z_i I(\mathbf{X}_i'\boldsymbol{\beta} > 0)\left\{K_+\left(\frac{U_i - \mathbf{X}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{h}\right) - K_+\left(\frac{U_i}{h}\right)\right\} \\
&\quad + \frac{1}{nh}\sum_{i=1}^{n} Z_i K_+\left(\frac{U_i}{h}\right)\{I(\mathbf{X}_i'\boldsymbol{\beta} > 0) - I(\mathbf{X}_i'\boldsymbol{\beta}_0 > 0)\} \\
&=: \bar{s}_{1n}(\boldsymbol{\beta}) + \bar{s}_{2n}(\boldsymbol{\beta}).
\end{aligned}
$$

It is not difficult to see from the proofs of Lemmas 1 and 2 that for any fixed $l > 0$, $(nh)^{1/2}\{\bar{s}_{1n}(\boldsymbol{\beta}) - \mathrm{E}[\bar{s}_{1n}(\boldsymbol{\beta})]\} = o_p(1)$ and $\mathrm{E}[\bar{s}_{1n}(\boldsymbol{\beta})] = O(n^{-1/2})$ uniformly in $\|\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \le l$. In addition, observe that

$$
I(u+v>0) - I(u>0) = \begin{cases} I(-u < v) & \text{if } u \le 0, \\ -I(u \le -v) & \text{if } u > 0. \end{cases}
$$

Thus, for any $\|\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \le l$,

$$
\begin{aligned}
|\bar{s}_{2n}(\boldsymbol{\beta})| &\le \frac{1}{nh}\sum_{i=1}^{n} |Z_i| K_+\left(\frac{U_i}{h}\right) I(|\mathbf{X}_i'\boldsymbol{\beta}_0| \le |\mathbf{X}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|) \\
&\le \frac{1}{nh}\sum_{i=1}^{n} |Z_i| K_+\left(\frac{U_i}{h}\right) I(|\mathbf{X}_i'\boldsymbol{\beta}_0| \le l n^{-1/2}\|\mathbf{X}_i\|).
\end{aligned} \tag{9}
$$

The expectation on the right-hand side of (9) is bounded by $A_0\mathrm{E}[|Z|I(|\mathbf{X}'\boldsymbol{\beta}_0| \leq ln^{-1/2}\|\mathbf{X}\|)]$, which is of order $O(n^{-1/2})$ by (8). Therefore, the proof is completed. □

By Lemma 4, for any $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\beta}}$, we have $(nh)^{1/2}\{s_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} = (nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} + o_p(1)$. The Lindeberg–Feller central limit theorem shows that $(nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 I(\mathbf{X}'\boldsymbol{\beta}_0 > 0) f_0(0|\mathbf{X})])$. Therefore, arguing as in Sect. 2, we get the following theorem:

Theorem 2 *Suppose that Assumptions 4–7 hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

$$(nh)^{1/2}\left\{\hat{\mathbf{J}}_{PCQ} - \mathbf{J}_{CQ} - \frac{h}{2}\mathrm{E}[f_0^{(1)}(0|\mathbf{X})I(\mathbf{X}'\boldsymbol{\beta}_0 > 0)\mathbf{X}\mathbf{X}'] + o(h)\right\}$$

is asymptotically normally distributed with zero mean matrix. The asymptotic covariance of the (j, k)th and the (l, m)th elements is given by

$$\mathrm{E}[f_0(0|\mathbf{X})I(\mathbf{X}'\boldsymbol{\beta}_0 > 0)X_j X_k X_l X_m],$$

where $j, k, l, m = 1, \ldots, d$.

Based on Theorem 2, we can evaluate the AMSE of $\hat{\mathbf{J}}_{\mathrm{PCQ}}$ and derive the optimal h that minimizes the AMSE. The optimal h for censored quantile regression is given by

$$h_{\mathrm{opt}} = n^{-1/3}\left\{\frac{2\sum_{j,k=1}^{d}\mathrm{E}[f_0(0|\mathbf{X})I(\mathbf{X}'\boldsymbol{\beta}_0 > 0)X_j^2 X_k^2]}{\sum_{j,k=1}^{d}\left(\mathrm{E}[f_0^{(1)}(0|\mathbf{X})I(\mathbf{X}'\boldsymbol{\beta}_0 > 0)X_j X_k]\right)^2}\right\}^{1/3}.$$

Note that the order of the optimal bandwidth is $n^{-1/3}$ because the one-sided uniform kernel is not symmetric about the origin. The estimation of the constant term of h_{opt} is analogous to Sect. 3. For instance, a rule of thumb bandwidth for the Gaussian location model described in Sect. 3 is given by

$$\hat{h}_{\mathrm{ROT}} = n^{-1/3}\left\{\frac{2\sum_{j,k=1}^{d}\left(n^{-1}\sum_{i=1}^{n} I\left(\mathbf{X}_i'\hat{\boldsymbol{\beta}} > 0\right) X_{ij}^2 X_{ik}^2\right)}{\gamma(\tau)\sum_{j,k=1}^{d}\left(n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i'\hat{\boldsymbol{\beta}} > 0)X_{ij}X_{ik}\right)^2}\right\}^{1/3},$$

where $\gamma(\tau) = \{\Phi^{-1}(\tau)\}^2\phi(\Phi^{-1}(\tau))$.

5 Extension to weakly dependent data

So far this paper has considered i.i.d. data. We now make note of sufficient conditions for asymptotic normality of Powell's kernel estimator for weakly dependent data. For simplicity, we only deal with the uncensored quantile regression model (2). Let

$\{(U_i, \mathbf{X}_i), i = 1, 2, \ldots\}$ be a strictly stationary sequence whose marginal distribution is the same as $(U, \mathbf{X})$. Under a sufficient weak dependence condition (and additional regularity conditions), it can be shown that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{KB}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}^{-1}\boldsymbol{\Omega}\mathbf{J}^{-1})$, where $\boldsymbol{\Omega}$ is the asymptotic covariance matrix of $n^{-1/2}\sum_{i=1}^{n}\{\tau - I(U_i \leq 0)\}\mathbf{X}_i$ (of course, if $\{(U_i, \mathbf{X}_i)\}$ is i.i.d., $\boldsymbol{\Omega} = \boldsymbol{\Sigma}$). See, for example, Phillips (1991, pp.459). In this case, the estimation of $\boldsymbol{\Omega}$ is not straightforward. It should be noted that Theorem 1 in Andrews (1991) does not apply to the estimation of $\boldsymbol{\Omega}$ since the smoothness of the moment function is violated in the present situation. However, we concentrate on the estimation of $\mathbf{J}$ in this paper and will discuss the estimation of $\boldsymbol{\Omega}$ in another place.

Here we state some regularity conditions to ensure asymptotic normality of $\hat{\mathbf{J}}_{\mathrm{P}}$.

Assumption 8 $\{(U_i, \mathbf{X}_i), i=1, 2, \ldots\}$ is a strict stationary sequence whose marginal distribution is the same as $(U, \mathbf{X})$.

Assumption 9 The sequence $\{(U_i, \mathbf{X}_i), i = 1, 2, \ldots\}$ is β-mixing; that is

$$\beta(j) := \sup_{i\geq 1} \mathrm{E}\left[\sup_{A\in\mathcal{F}_{i+j}^{\infty}} |\mathrm{P}(A|\mathcal{F}_1^i) - \mathrm{P}(A)|\right] \to 0, \text{ as } j \to \infty,$$

where $\mathcal{F}_i^j$ is the σ-field generated by $\{(U_k, \mathbf{X}_k), k = i, \ldots, j\}(j \geq i)$. In addition,

$$\sum_{j=1}^{\infty} j^{\lambda}\{\beta(j)\}^{1-2/\delta} < \infty, \tag{10}$$

for some $\delta > 2$ and $\lambda > 1 - 2/\delta$.

Assumption 10 $\mathrm{E}[\|\mathbf{X}\|^{\max\{6,2\delta\}}] < \infty$, where δ is given in Assumption 9.

Assumption 11 The conditional density $f_0(u|\mathbf{x})$ of U given $\mathbf{X} = \mathbf{x}$ is twice continuously differentiable with respect to u for each $\mathbf{x}$. Furthermore, there exist $A_0 > 0$ and $G_j(\mathbf{x})(j = 1, 2)$ such that $f_0(u|\mathbf{x}) \leq A_0$, $|f_0^{(j)}(u|\mathbf{x})| \leq G_j(\mathbf{x})(j = 1, 2)$ for every realization $(u, \mathbf{x})$ of $(U, \mathbf{X})$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^3)G_1(\mathbf{X})] < \infty$ and $\mathrm{E}[\|\mathbf{X}\|^2 G_2(\mathbf{X})] < \infty$, where $f_0^{(j)}(u|\mathbf{x}) = \partial^j f_0(u|\mathbf{x})/\partial u^j$ for $j = 1, 2$.

Assumption 12 Let $f_0(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j}; j)$ denote the conditional density of (U_1, U_{1+j}) given $(\mathbf{X}_1, \mathbf{X}_{1+j}) = (\mathbf{x}_1, \mathbf{x}_{1+j})(j \geq 1)$. Then, there exists a constant $A_1 > 0$ independent of j such that $f_0(u_1, u_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j}; j) \leq A_1$ for every realization $(u_1, u_{1+j}, \mathbf{x}_1, \mathbf{x}_{1+j})$ of $(U_1, U_{1+j}, \mathbf{X}_1, \mathbf{X}_{1+j})$.

Assumption 13 As $n \to \infty, h \to 0$ and $(n^{1/2}h)/\log n \to \infty$. In addition, there exists a sequence of positive integers $s = s_n$ satisfying $s \to \infty$ and $s = o((nh)^{1/2})$ as $n \to \infty$ such that

$$(n/h)^{1/2}\beta(s) \to 0 \quad \text{as } n \to \infty. \tag{11}$$

The β-mixing condition is required for establishing the uniform convergence result corresponding to Lemma 1 because our approach uses the blocking technique as in Yu (1994) and Arcones and Yu (1994). The blocking technique enables us to employ the symmetrization technique and an exponential inequality available in the i.i.d. case. In order to validate the blocking technique, we use Lemma 4.1 in Yu (1994), which requires the β-mixing condition. A set of conditions such as (10), $\mathrm{E}[\|\mathbf{X}\|^{2\delta}] < \infty$, the boundedness of the conditional densities (included in Assumptions 11–12) and the latter part of Assumption 13 is typically assumed in the density estimation and the nonparametric regression literature. See Condition 1 of Theorem 6.3 in Fan and Yao (2005) (we note that Theorem 6.3 of Fan and Yao (2005) assumes the α-mixing condition, which is weaker than the β-mixing condition). These conditions are sufficient for asymptotic normality of $(nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\}$, where $s_n(\boldsymbol{\beta})$ is given by (4). A sufficient condition on the mixing coefficient $\beta(j)$ to satisfy the conditions (10) and (11) is provided in Fan and Yao (2005, pp.387).

Below we follow the notations used in Sect. 2. The next lemma is essential to our purpose.

Lemma 5 *Under Assumptions 8–13, the conclusion of Lemma 1 is valid in the present situation.*

Proof Working with the same notations as in the proof of Lemma 1, we show that $n^{1/2}h^{-1/2}\{W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]\} \xrightarrow{p} 0$, uniformly in $\|\mathbf{t}\| \le l$.

Before proceeding to the proof, as in Yu (1994) and Arcones and Yu (1994), we introduce a sequence of independent blocks. Divide the n-sequence $\{1, \dots, n\}$ into blocks of length $a_n = [n^{(1-2/\delta)/(1-2/\delta+\lambda)}]$ one after the other:

$$\begin{aligned} H_k &= \{i : 2(k-1)a_n + 1 \le i \le (2k-1)a_n\}, \\ T_k &= \{i : (2k-1)a_n + 1 \le i \le 2ka_n\}, \end{aligned}$$

for $k = 1, \dots, \mu_n$, where $\mu_n = [n/(2a_n)]$. Let $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in \cup_{k=1}^{\mu_n} H_k\}$ be a set of random vectors such that blocks $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in H_i\}$ $(k = 1, \dots, \mu_n)$ are independent and have the same distribution as $\{(U_i, \mathbf{X}_i), i \in H_1\}$. Replacing $(U_i, \mathbf{X}_i)$ with $(\tilde{U}_i, \tilde{\mathbf{X}}_i)$, define $\tilde{Z}_i$ and $\tilde{U}_i^*(\mathbf{t})$ as Z_i and $U_i^*(\mathbf{t})$, respectively, for $i \in \cup_{k=1}^{\mu_n} H_k$.

Fix any $\epsilon > 0$. Observe that

$$\begin{aligned} &\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} |W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]| > n^{-1/2}h^{1/2}\epsilon\right) \\ &\le \mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left|\frac{1}{n}\sum_{i=1}^{2a_n\mu_n}\{U_i^*(\mathbf{t}) - \mathrm{E}[U_i^*(\mathbf{t})]\}\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{2}\right) \\ &\quad + \mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left|\frac{1}{n}\sum_{i=2a_n\mu_n+1}^{n}\{U_i^*(\mathbf{t}) - \mathrm{E}[U_i^*(\mathbf{t})]\}\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{2}\right). \end{aligned} \tag{12}$$

A simple calculation shows that the second term on the right-hand side of (12) converges to zero (use the fact that $\sup_{\|\mathbf{t}\|\le l} |U_i^*(\mathbf{t})| \le |Z_i| I(h < U_i \le h + n^{-1/2} l \|\mathbf{X}_i\|)$). On the other hand, using the same argument as in Lemma 4.2 of Yu (1994), the first term on the right-hand side of (12) is bounded by

$$2\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left| \frac{1}{n} \sum_{k=1}^{\mu_n} \{\tilde{V}_k(\mathbf{t}) - \mathrm{E}[\tilde{V}_k(\mathbf{t})]\} \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \right) + 2\mu_n \beta(a_n),$$

where $\tilde{V}_k(\mathbf{t}) = \sum_{i \in H_k} \tilde{U}_i^*(\mathbf{t})$. Because of the condition (10), $\mu_n \beta(a_n) = o(1)$. Therefore, it suffices to show that

$$\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left| \frac{1}{n} \sum_{k=1}^{\mu_n} \{\tilde{V}_k(\mathbf{t}) - \mathrm{E}[\tilde{V}_k(\mathbf{t})]\} \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \right) \to 0.$$

Let $\sigma_1, \ldots, \sigma_{\mu_n}$ be independent and uniformly distributed over $\{-1, 1\}$ and independent of $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in H_k\}(k = 1, \ldots, \mu_n)$. Since $\{\mathbf{t} \mapsto \tilde{V}_k(\mathbf{t}), k = 1, \ldots, \mu_n\}$ is a sequence of i.i.d. stochastic processes, the symmetrization technique (van der Vaart and Wellner 1996, Lemma 2.3.7) yields that

$$\begin{aligned} &\xi_n \mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left| \frac{1}{n} \sum_{k=1}^{\mu_n} \{\tilde{V}_k(\mathbf{t}) - \mathrm{E}[\tilde{V}_k(\mathbf{t})]\} \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{4} \right) \\ &\le 2\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} \left| \frac{1}{n} \sum_{k=1}^{\mu_n} \sigma_i \tilde{V}_k(\mathbf{t}) \right| > \frac{n^{-1/2} h^{1/2} \epsilon}{16} \right), \end{aligned}$$

where $\xi_n = 1 - (16\mu_n/(\epsilon^2 nh)) \sup_{\|\mathbf{t}\|\le l} \mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2]$. We show that $\sup_{\|\mathbf{t}\|\le l} \mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2] = O(a_n n^{-1/2})$. By stationarity,

$$\mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2] = a_n \mathrm{E}[\{U_1^*(\mathbf{t})\}^2] + 2a_n \sum_{j=1}^{a_n - 1} (1 - j/a_n) \mathrm{E}[U_1^*(\mathbf{t}) U_{1+j}^*(\mathbf{t})].$$

Observe that $\sup_{\|\mathbf{t}\|\le l} \mathrm{E}[\{U_1^*(\mathbf{t})\}^2] = O(n^{-1/2})$. By conditioning on $(\mathbf{X}_1, \mathbf{X}_{1+j})$, we have

$$\begin{aligned} &|\mathrm{E}[U_1^*(\mathbf{t}) U_{1+j}^*(\mathbf{t})]| \\ &\le \mathrm{E}\left[|Z_1 Z_{1+j}| \int_h^{h+n^{-1/2} l \|\mathbf{X}_1\|} \int_h^{h+n^{-1/2} l \|\mathbf{X}_{1+j}\|} f_0(u_1, u_{1+j} | \mathbf{X}_1, \mathbf{X}_{1+j}; j) du_1 du_{1+j} \right] \\ &\le \text{const.} \times n^{-1} \mathrm{E}[|Z_1 Z_{1+j}| \cdot \|\mathbf{X}_1\| \|\mathbf{X}_{1+j}\|] \\ &= O(n^{-1}), \end{aligned}$$

uniformly in $\|\mathbf{t}\| \leq l$ and $j \geq 1$. This yields that

$$\sup_{\|\mathbf{t}\|\leq l}\left|\sum_{j=1}^{a_n-1}\mathrm{E}[U_1^*(\mathbf{t})U_{1+j}^*(\mathbf{t})]\right| = O(a_n n^{-1}) = o(n^{-1/2}).$$

Thus, we have shown that $\sup_{\|\mathbf{t}\|\leq l}\mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2] = O(a_n n^{-1/2})$, which implies that $\xi_n = 1 - O(n^{-1/2}h^{-1}) = 1 - o(1)$ and consequently $\xi_n \geq 1/2$ for large n. Therefore, for large n,

$$\mathrm{P}\left(\sup_{\|\mathbf{t}\|\leq l}\left|\frac{1}{n}\sum_{k=1}^{\mu_n}\left\{\tilde{V}_k(\mathbf{t}) - \mathrm{E}[\tilde{V}_k(\mathbf{t})]\right\}\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{4}\right)$$
$$\leq 4\mathrm{P}\left(\sup_{\|\mathbf{t}\|\leq l}\left|\frac{1}{n}\sum_{k=1}^{\mu_n}\sigma_i\tilde{V}_k(\mathbf{t})\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{16}\right).$$

The rest of the proof is similar to the latter part of the proof of Lemma 1. Arguing as in the proof of Lemma 1, it is shown that the cardinality of the functional set $\{\boldsymbol{\sigma}^{(\mu_n)} \mapsto n^{-1}\sum_{k=1}^{\mu_n}\sigma_i\tilde{V}_k(\mathbf{t}) : \|\mathbf{t}\| \leq l\}$, where $\boldsymbol{\sigma}^{(\mu_n)} = (\sigma_1, \ldots, \sigma_{\mu_n})$, is bounded by some polynomial of n uniformly over every realization of $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in \cup_{k=1}^{\mu_n}H_k\}$. In addition, Hoeffding's inequality implies that

$$\sup_{\|\mathbf{t}\|\leq l}\mathrm{P}_\sigma\left(\left|\frac{1}{n}\sum_{k=1}^{\mu_n}\sigma_i\tilde{V}_k(\mathbf{t})\right| > \frac{n^{-1/2}h^{1/2}\epsilon}{16}\right) \leq 2\exp\left(-\frac{h\epsilon^2}{512w_n}\right),$$

where P_σ denotes the probability with respect to $\boldsymbol{\sigma}^{(\mu_n)}$ only and $w_n = n^{-1}\sum_{k=1}^{\mu_n}$ $\{\sum_{i\in H_k}|\tilde{Z}_i|I(h < \tilde{U}_i \leq h + n^{-1/2}l\|\tilde{\mathbf{X}}_i\|)\}^2$. Thus, it suffices to show that

$$(\log n)h^{-1}w_n \xrightarrow{p} 0. \tag{13}$$

From the evaluation of $\mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2]$ above, it is shown that

$$\mathrm{E}\left[\left\{\sum_{i\in H_1}|\tilde{Z}_i|I(h < \tilde{U}_i \leq h + n^{-1/2}l\|\tilde{\mathbf{X}}_i\|)\right\}^2\right] = O(a_n n^{-1/2}),$$

which leads to $\mathrm{E}[w_n] = O(\mu_n a_n n^{-3/2}) = O(n^{-1/2})$. Since $n^{1/2}h/\log n \to \infty$, (13) follows from Markov's inequality. Therefore, we complete the proof. □

The proof of Lemma 2 does not use the independence assumption and hence the conclusion of Lemma 2 applies to the present situation. Thus, for any $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\beta}}$, we have $(nh)^{1/2}\{s_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} = (nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} + o_p(1)$. In addition, mimicking the proof of Theorem 6.3 in Fan and Yao (2005), it can be shown that under Assumptions 8–13, $(nh)^{1/2}\{s_n(\boldsymbol{\beta}_0) - \mathrm{E}[s_n(\boldsymbol{\beta}_0)]\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 f(0|\mathbf{X})]/2)$. Therefore, arguing as in Sect. 2, we get the following theorem:

Theorem 3 *Suppose that Assumptions 8–13 hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then, the conclusion of Theorem 1 holds under the present situation.*

The conclusion of Theorem 4 is the same as that of Theorem 2 which assumes the i.i.d. condition. Therefore, the optimal bandwidth that minimizes the AMSE under the weak dependence condition is the same as that under the i.i.d. condition.

6 Concluding remarks

In this paper, we have shown asymptotic normality of Powell's kernel estimator for the asymptotic covariance matrix of the quantile regression estimator for both i.i.d and weakly dependent data. The asymptotic distribution of the kernel estimator enables us to calculate the approximate mean squared error. It should be noted that since the kernel estimator contains the estimated parameter in the sum, the direct calculation of the mean squared error is infeasible. We have derived the optimal bandwidth that minimizes the AMSE. In addition, we have derived the corresponding results to censored quantile regression.

As pointed out by a referee, despite the fact that Lemma 1 exploits the specific property of the uniform kernel, the results of this paper can be extended to, for instance, general kernels of bounded variation, with some obvious modifications. In that case, we can write K as a difference of two non-decreasing functions and apply the essentially same argument as Lemma 1 to such K.

Acknowledgments The author would like to thank Professor Naoto Kunitomo and Professor Tatsuya Kubokawa for their encouragement and suggestions. He also would like to thank the Editor, the Associate Editor and anonymous referees for their careful reading of the manuscript.

References

Ahmad, I. A. (1992). Residual density estimation in nonparametric regression. *Statistics and Probability Letters, 14*, 133–139.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica, 59*, 817–858.

Angrist, J., Chernozhukov, V., Fernandéz-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica, 74*, 539–563.

Arcones, M. A., Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal Theoretical Probability, 7*, 47–71.

Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian Journal of Statistics, 17*, 1–7.

Chai, G. X., Li, Z. Y. (1993). Asymptotic theory for estimation of error distributions of linear model. *Science in China Series A, 36*, 408–419.

Chai, G. X., Li, Z. Y., Tian, H. (1991). Consistent nonparametric estimation of error distributions in linear model. *Acta Mathematicae Applicatae Sinica, 7*, 245–256.

Cheng, F. (2002). Consistency of error density and distribution function estimators in nonparametric regression. *Statistics and Probability Letters, 59*, 257–270.

Cheng, F. (2004). Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression. *Journal of Statistical Planning and Inference, 119*, 95–107.

Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *Journal of Statistical Planning and Inference, 128*, 327–349.

Efromovich, S. (2005). Estimation of the density of regression errors. *Annals of Statistics, 33*, 2194–2227.

Efromovich, S. (2007a). Adaptive estimation of error density in nonparametric regression in small sample. *Journal of Statistical Planning and Inference, 137*, 363–378.

Efromovich, S. (2007b). Optimal nonparametric estimation of the density of regression errors with finite support. *Annals of the Institute of Statistical Mathematics, 59*, 617–654.

Fan, J., Yao, Q. (2005). *Nonlinear time series: Nonparametric and parametric methods*. New York: Springer.

Hall, P., Horowitz, J. L. (1990). Bandwidth selection in semiparametric estimation of censored linear models. *Econometric Theory, 6*, 123–150.

Hall, P., Sheather, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society Series B, 50*, 381–391.

Hendricks, W., Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association, 87*, 58–68.

Koenker, R. (2005). *Quantile regression*. Oxford: Oxford University Press.

Koenker, R. (2009). Quantile regression in R: A vignette. Available from http://cran.r-project.org/web/packages/quantreg/.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica, 46*, 33–50.

Koul, H. L. (1992). *Weighted empiricals and linear models*. In Lecture Notes-Monograph Series (Vol. 12). Hayward: Institute of Mathematical Statistics.

Li, Z. Y. (1995). A study of nonparametric estimation of error distribution in linear model based on L_1-norm. *Hiroshima Mathematical Journal, 25*, 171–205.

Liang, H.-Y., Niu, S.-L. (2009). Asymptotic properties of error density estimator in regression model under α-mixing assumptions. *Communications in Statistics: Theory and Methods, 38*, 761–783.

Liebscher, E. (1999). Estimating the density of the residuals in autoregressive models. *Statistical Inference for Stochastic Processes, 2*, 105–117.

Müller, U. U., Schick, A., Wefelmeyer, W. (2005). Weighted residual-based density estimators for nonlinear autoregressive models. *Statistica Sinica, 15*, 177–195.

Phillips, P. C. B. (1991). A shortcut to LAD estimator asymptotics. *Econometric Theory, 7*, 450–463.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics, 25*, 303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics, 32*, 143–155.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J. Powell, G. Tauchen (Eds.), *Nonparametric and semiparametric models in econometrics*. Cambridge: Cambridge University Press.

Robinson, P. M. (1987). Time series residuals with application to probability density estimation. *Journal of Time Series Analysis, 8*, 329–344.

Schick, A., Wefelmeyer, W. (2007). Uniformly root-n consistent density estimators for weakly dependent invertible linear processes. *Annals of Statistics, 35*, 815–843.

Sheather, S. J. (2004). Density estimation. *Statistical Science, 19*, 588–597.

Siddiqui, M. (1960). Distribution of quantiles from a bivariate population. *Journal of Research of the National Bureau Standards, 64*, 145–150.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. New York: Springer.

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Annals of Probability, 22*, 94–116.