# Asymptotic normality of Powell's kernel estimator

Kengo Kato *

This version: July 13, 2009†

**Abstract**

In this paper, 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.

*Key words*: asymptotic normality; bandwidth selection; density estimation; quantile regression.
*AMS subject classifications*: 62G07, 62J05.

## 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 $(Y_i, \mathbf{X}_i)$ $(i = 1, 2, \dots n)$ 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 $\tau$-th ($\tau \in (0, 1)$) conditional linear quantile regression model is defined as

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

where $Q_Y(\tau|\mathbf{X}) = \inf\{y : \mathrm{P}(Y \leq y|\mathbf{X}) \geq \tau\}$ is the $\tau$-th 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

$$\sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i'\boldsymbol{\beta}), \tag{2}$$

where $\rho_\tau(u) = \{\tau - I(u \leq 0)\}u$ is called the check function. It is well known that, under suitable regularity conditions, $\hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau)$ satisfies consistency and asymptotic normality; see

*Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. Email: kkato@hiroshima-u.ac.jp

†The first version appeared on May 11, 2009. The former Section 4 is removed.

Chapter 4 of Koenker (2005). Letting $f(y|\mathbf{x})$ denote the conditional density of $Y$ given $\mathbf{X}=\mathbf{x}$, 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 $\boldsymbol{\Sigma}(\tau)=\tau(1-\tau)\mathrm{E}[\mathbf{X}\mathbf{X}']$. The estimation of the matrix $\boldsymbol{\Sigma}(\tau)$ is straightforward. However, since the matrix $\mathbf{J}(\tau)$ involves the conditional density, its estimation is not a trivial task. 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 (Siddiqui, 1960). On the other hand, Powell (1991) proposes the kernel estimator

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

where $\hat{\boldsymbol{\beta}}(\tau)$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0(\tau)$, $h_n>0$ is a bandwidth and $K(\cdot)$ is the uniform kernel

$$K(u)=\begin{cases}\frac{1}{2} & \text{if } |u|\le 1,\\ 0 & \text{otherwise.}\end{cases}\tag{3}$$

In usual, we take $\hat{\boldsymbol{\beta}}(\tau)=\hat{\boldsymbol{\beta}}_{\mathrm{KB}}(\tau)$. He shows that $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ is consistent under some regularity conditions. Especially, he imposes the condition on the bandwidth $h_n$ that $h_n\to 0$ and $nh_n^2\to\infty$. The recent study by Angrist et al. (2006) shows that $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ is uniformly consistent over a closed interval of $\tau$ even when the model is misspecified. However, to the author's knowledge, there is no literature that rigorously studies the asymptotic distribution of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$.

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 $h_n$ is such that $h_n\to 0$ and $(n^{1/2}h_n)/\log(n)\to\infty$. 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 $h_n$ 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 a complicated task. This paper is the first result that derives the optimal bandwidth for $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ under a certain criterion. 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. Except for the correctness of the proof, the differences of the present paper from them 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 Section 3. (ii) Chai and Li impose the stringent condition that the covariate vectors are bounded over all observations. Actually, the boundedness of the covariate vectors is essential to their proofs. 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.

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 of a nonlinear autoregressive model. He assumes that the kernel function is Lipschitz continuous (see equation (3.5) of his paper), which is essential to his proof, while the uniform kernel treated in the present paper is not. 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 use the uniform kernel when deriving the asymptotic distributions.

The rest of the paper is organized as follows. In Section 2, we prove asymptotic normality of Powell's kernel estimator $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ for i.i.d. data. In Section 3, we use the asymptotic distribution to evaluate the AMSE and derive the optimal $h$ that minimizes the AMSE. In Section 4, we establish asymptotic normality of $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ under a weak dependence condition. In Section 5, 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 $\hat{\mathbf{J}}_{\mathrm{P}}(\tau)$ for i.i.d. data. Throughout this paper, we fix $\tau$ and suppress the dependence on $\tau$ for notational convenience. For example, we simply write $\boldsymbol{\beta}_0$ for $\boldsymbol{\beta}_0(\tau)$. Then, the model (1) may be written as

$$Y = \mathbf{X}'\boldsymbol{\beta}_0 + U, \ Q_U(\tau|\mathbf{X}) = 0, \tag{4}$$

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

$$Y = \mathbf{X}'\boldsymbol{\theta}_0 + (\mathbf{X}'\boldsymbol{\gamma}_0)\epsilon, \tag{5}$$

where $\mathbf{X}'\boldsymbol{\gamma}_0 > 0$ and $\epsilon$ is independent of $\mathbf{X}$. In this model, $U$ corresponds to $\mathbf{X}'\boldsymbol{\gamma}_0\{\epsilon - F^{-1}(\tau)\}$, where $F$ is the distribution function of $\epsilon$. Typically, the model (5) allows for the heteroscedasticity of $U$.

We now return to the general model (4). Letting $f_0(u|\mathbf{x})$ denote the conditional density of $U$ given $\mathbf{X} = \mathbf{x}$, 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:

**(A1)** $\{(U_i, \mathbf{X}_i), i = 1, 2, \dots\}$ is an i.i.d. sequence whose marginal distribution is same as $(U, \mathbf{X})$.

**(A2)** 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 $|f_0^{(j)}(u|\mathbf{x})| \leq G_j(\mathbf{x})$ $(j = 0, 1, 2)$ for every realization $(u, \mathbf{x})$ of $(U, \mathbf{X})$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^4 + \|\mathbf{X}\|^5)G_0(\mathbf{X})] < \infty$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^3 + \|\mathbf{X}\|^4)G_1(\mathbf{X})] < \infty$ and $\mathrm{E}[\|\mathbf{X}\|^2 G_2(\mathbf{X})] < \infty$,

**(A3)** As $n \to \infty$, $h_n \to 0$ and $(n^{1/2}h_n)/\log(n) \to \infty$.

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 condition (A2) 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 (A2) is needed to ensure the dominated convergence. Condition (A3) 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 and Sheather (1988) rules, although the latter two bandwidth rules are originally for the scalar sparsity estimation. Powell (1991) and other authors show consistency of $\hat{\mathbf{J}}_{\mathrm{P}}$ under the condition that $h_n \to 0$ and $nh_n^2 \to \infty$.

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

$$\begin{aligned} T_n(\boldsymbol{\beta}) &= \frac{1}{nh_n}\sum_{i=1}^n Z_i K\left(\frac{Y_i - \mathbf{X}_i'\boldsymbol{\beta}}{h_n}\right) \\ &= \frac{1}{nh_n}\sum_{i=1}^n Z_i K\left(\frac{U_i - \mathbf{X}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{h_n}\right), \end{aligned} \tag{6}$$

where $Z_i = \mathrm{tr}(\mathbf{A}\mathbf{X}_i\mathbf{X}_i')$. We first show asymptotic normality of $T_n(\hat{\boldsymbol{\beta}})$. Then, we use the Cramér-Wold device to derive the asymptotic distribution of $\hat{\mathbf{J}}_{\mathrm{P}}$. The proof of asymptotic normality of $T_n(\hat{\boldsymbol{\beta}})$ 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 conditions (A1)-(A3) hold. Then, for any fixed $l > 0$, we have*

$$T_n(\boldsymbol{\beta}) - \mathrm{E}[T_n(\boldsymbol{\beta})] = T_n(\boldsymbol{\beta}_0) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)] + o_p((nh_n)^{-1/2})$$

*uniformly in $\|\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le l$.*

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

$$\begin{aligned}
&h\{T_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t}) - T_n(\boldsymbol{\beta}_0)\}\\
&= \frac{1}{n}\sum_{i=1}^n Z_i \left\{ K\left(\frac{U_i - n^{-1/2}\mathbf{X}_i'\mathbf{t}}{h_n}\right) - K\left(\frac{U_i}{h_n}\right)\right\}\\
&= \frac{1}{2}\left\{ \frac{1}{n}\sum_{i=1}^n Z_i I(h_n < U_i \le h_n + n^{-1/2}\mathbf{X}_i'\mathbf{t}) \right.\\
&\quad - \frac{1}{n}\sum_{i=1}^n Z_i I(-h_n \le U_i < -h_n + n^{-1/2}\mathbf{X}_i'\mathbf{t})\\
&\quad + \frac{1}{n}\sum_{i=1}^n Z_i I(-h_n + n^{-1/2}\mathbf{X}_i'\mathbf{t} \le U_i < -h_n)\\
&\quad \left. - \frac{1}{n}\sum_{i=1}^n Z_i I(h_n + n^{-1/2}\mathbf{X}_i'\mathbf{t} < U_i \le h_n) \right\}\\
&=: \frac{1}{2}\{W_{1n}(\mathbf{t}) - W_{2n}(\mathbf{t}) + W_{3n}(\mathbf{t}) - W_{4n}(\mathbf{t})\}. 
\end{aligned} \tag{7}$$

It suffices to show that $n^{1/2}h_n^{-1/2}\{W_{jn}(\mathbf{t}) - \mathrm{E}[W_{jn}(\mathbf{t})]\} \xrightarrow{p} 0$ 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 $U_i^*(\mathbf{t}) = Z_i I(h_n < U_i \le h_n + n^{-1/2}\mathbf{X}_i'\mathbf{t})$. 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

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

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

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

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

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

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

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

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

$$\begin{aligned}&-|Z_i| I(h_n < U_i \le h_n + n^{-1/2} l\|\mathbf{X}_i\|) \le \sigma_i U_i^*(\mathbf{t}) \\ &\qquad\qquad \le |Z_i| I(h_n < U_i \le h_n + n^{-1/2} l\|\mathbf{X}_i\|). \quad (8)\end{aligned}$$

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

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

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

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

We now bound $k_n$. It is not difficult to see that $k_n$ is bounded by the cardinality of the set $\big\{A \cap \{(U_1, \mathbf{X}_1), \ldots, (U_n, \mathbf{X}_n)\} : A \in \mathcal{A}\big\}$, where $\mathcal{A} = \big\{\{(u, \mathbf{x}) : u > h, u \le h + \mathbf{x}'\mathbf{t}\} : h > 0, \mathbf{t} \in \mathbb{R}^d\big\}$. Application of Lemma 2.6.15 in van der Vaart and Wellner (1996) shows that the Vapnik-Červonenkis (VC) index $V_{\mathcal{A}}$ of $\mathcal{A}$ is finite, namely $0 < V_{\mathcal{A}} < \infty$; see van der Vaart and Wellner (1996), pp. 135 for the definition of the VC index. Then, Sauer's lemma (van der Vaart and Wellner, 1996, Corollary 2.6.3) implies that $k_n$ is bounded by $cn^{V_{\mathcal{A}}-1}$ for some constant $c$ not depending on $D_n$. Therefore, we have

$$\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i U_i^*(\mathbf{t})\right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{4}\,\middle|\, D_n\right) \le 2cn^{V_{\mathcal{A}}-1}\exp\left(-\frac{\epsilon^2 h_n}{32 v_n}\right). \tag{9}$$

Define

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

Using (9) 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 $\delta > 0$,

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

where we have used $n^{1/2}h_n/\log(n) \to \infty$. 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 $2^n$.

*Remark 2.* It is not possible to directly apply Theorem II 37 in Pollard (1984) to obtain the uniform convergence result of Lemma 1 since $Z_i$ is not bounded random variable. Instead of relying on Lemma II 33 in Pollard (1984), we use the explicit bound (8) when using Hoeffding's inequality in the proof of Lemma 1.

**Lemma 2.** *Suppose that conditions (A1)-(A3) hold. Then, for any fixed $l > 0$, we have*

$$\mathrm{E}[T_n(\boldsymbol{\beta})] = \mathrm{E}[T_n(\boldsymbol{\beta}_0)] + O(n^{-1/2})$$

*uniformly in $\|\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le l$.*

*Proof.* We have to show $\mathrm{E}[T_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t})] = \mathrm{E}[T_n(\boldsymbol{\beta}_0)] + O(n^{-1/2})$ uniformly in $\|\mathbf{t}\| \le l$. Using the relation

$$
\begin{aligned}
&\mathrm{E}[T_n(\boldsymbol{\beta})] - \mathrm{E}[Zf_0(0|\mathbf{X})] \\
&= \mathrm{E}\left[ Z \int K(u - h_n^{-1}\mathbf{X}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0))\{f_0(uh|\mathbf{X}) - f_0(0|\mathbf{X})\}du \right] \\
&= h\mathrm{E}\left[ Z \int uK(u - h_n^{-1}\mathbf{X}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0))g_{h_n}(u|\mathbf{X})du \right],
\end{aligned}
$$

where $g_{h_n}(u|\mathbf{x}) = (uh_n)^{-1}\{f_0(uh_n|\mathbf{x}) - f_0(0|\mathbf{x})\}$ for $u \ne 0$ and $g_{h_n}(0|\mathbf{x}) = 0$, the absolute value of the difference $\mathrm{E}[T_n(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{t})] - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]$ is evaluated as

$$
\begin{aligned}
&\left| h\mathrm{E}\left[ Z \int u\{K(u - n^{-1/2}h_n^{-1}\mathbf{X}'\mathbf{t}) - K(u)\}g_{h_n}(u|\mathbf{X})du \right] \right| \\
&\le h\mathrm{E}\left[ |Z| \cdot G_1(\mathbf{X}) \int |u\{K(u - n^{-1/2}h_n^{-1}\mathbf{X}'\mathbf{t}) - K(u)\}|du \right], \qquad (10)
\end{aligned}
$$

where we have used $|g_{h_n}(u|\mathbf{x})| \le \sup_u |f_0^{(1)}(u|\mathbf{x})| \le G_1(\mathbf{x})$. Using the identity

$$
\begin{aligned}
I(|u - v| \le 1) - I(|u| \le 1) = &\{I(1 < u \le 1 + v) - I(-1 \le u < -1 + v)\}I(v > 0) \\
&+ \{I(-1 + v \le u < -1) - I(1 + v < u \le 1)\}I(v < 0),
\end{aligned}
$$

we have

$$
\begin{aligned}
&\int |u\{I(|u - v| \le 1) - I(|u| \le 1)\}|du \\
&= \left\{ \int_1^{1+v} udu + \int_{-1}^{-1+v} |u|du \right\} I(v > 0) \\
&\quad + \left\{ \int_{-1+v}^{-1} |u|du + \int_{1+v}^{1} |u|du \right\} I(v < 0) \\
&\le 2(1 + |v|)|v|.
\end{aligned}
$$

Since $nh_n^2 \to \infty$, $n^{-1/2}h_n^{-1} \le 1$ for large $n$. Therefore, the right hand side of (10) is bounded by

$$
ln^{-1/2}\mathrm{E}\left[|Z|G_1(\mathbf{X})(1 + l\|\mathbf{X}\|)\|\mathbf{X}\|\right]
$$

for any $\|\mathbf{t}\| \le l$. This yields the desired result. □

**Lemma 3.** *Under conditions (A1)-(A3), we have*

$$
(nh_n)^{1/2}\{T_n(\boldsymbol{\beta}_0) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 f_0(0|\mathbf{X})]/2).
$$

*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,

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

Using the Taylor expansion, we see that

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

Therefore, by Lemma 3, we get the following theorem:

**Theorem 1.** *Suppose that conditions (A1)-(A3) hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

$$(nh_n)^{1/2}\left\{T_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[Zf_0(0|\mathbf{X})] - \frac{h_n^2}{6}\mathrm{E}[Zf_0^{(2)}(0|\mathbf{X})] + o(h_n^2)\right\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 f_0(0|\mathbf{X})]/2).$$

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 $T_n(\hat{\boldsymbol{\beta}})$ 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 2.** *Suppose that conditions (A1)-(A3) hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

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

*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

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

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

$$h_n^{\mathrm{opt}} = n^{-1/5} \left\{ \frac{4.5 \sum_{j,k=1}^{d} \mathrm{E}[f_0(0|\mathbf{X})X_j^2 X_k^2]}{\sum_{j,k=1}^{d} \left(\mathrm{E}[f_0^{(2)}(0|\mathbf{X})X_j X_k]\right)^2} \right\}^{1/5}, \tag{11}$$

where we assume that the denominator is not zero. It should be noted that $h_n^{\mathrm{opt}}$ depends on $\tau$, namely $h_n^{\mathrm{opt}} = h_n^{\mathrm{opt}}(\tau)$, since the distribution of $U$ generally depends on $\tau$. We further note that $h_n^{\mathrm{opt}}$ 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 $f_0(u|\mathbf{x}) = f_0(u)$, $h_{\mathrm{opt}}$ 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

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

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

**Proposition 1.** *Suppose that conditions (A1)-(A3) hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

$$\lim_{T\to\infty} \lim_{n\to\infty} \{MSE_T(h_n) - MSE_T(h_n^{opt})\} \geq 0,$$

*where the inequality is strict unless $h_n = h_n^{opt} + o(n^{-1/5})$.*

*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 well as the usual density estimation, $h_n^{\mathrm{opt}}$ 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 $h_n^{\text{opt}}$ for a Gaussian location model

$$Y = \mathbf{X}'\boldsymbol{\theta}_0 + \epsilon, \ \ \epsilon|\mathbf{X} \sim N(0,1), \tag{12}$$

is given by

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

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

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

# 4 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. Let $\{(U_i, \mathbf{X}_i), i = 1, 2, \dots\}$ be a strictly stationary sequence whose marginal distribution is same as $(U, \mathbf{X})$. Under a sufficient weak dependence condition (and additional regularity conditions), it can be shown that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{KB}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}^{-1}\boldsymbol{\Omega}\mathbf{J}^{-1}),$$

where $\boldsymbol{\Omega}$ is the asymptotic covariance matrix of $n^{-1/2}\sum_{i=1}^{n}\{\tau - I(U_i \le 0)\}\mathbf{X}_i$. Of course, $\boldsymbol{\Omega}$ reduces to $\boldsymbol{\Sigma}$ when $\{(U_i, \mathbf{X}_i)\}$ is i.i.d. 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 $\hat{\mathbf{J}}_{\text{P}}$.

**(B1)** $\{(U_i, \mathbf{X}_i), i = 1, 2, \dots\}$ is a strict stationary sequence whose marginal distribution is same as $(U, \mathbf{X})$.

**(B2)** The sequence $\{(U_i, \mathbf{X}_i), i = 1, 2, \dots\}$ is $\beta$-mixing; that is

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

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

$$\sum_{j=1}^{\infty} j^{\lambda} \{\beta(j)\}^{1-2/\delta} < \infty, \tag{13}$$

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

**(B3)** $\mathrm{E}[\|\mathbf{X}\|^{\max\{6,2\delta\}}] < \infty$, where $\delta$ is given in condition (B2).

**(B4)** 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 constant $A_0 > 0$ and measurable functions $G_j(\mathbf{x})$ $(j = 1, 2)$ 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 $(U, \mathbf{X})$, $\mathrm{E}[(\|\mathbf{X}\|^2 + \|\mathbf{X}\|^3 + \|\mathbf{X}\|^4)G_1(\mathbf{X})] < \infty$ 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 $j = 1, 2$.

**(B5)** 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 $(U_1, U_{1+j}, \mathbf{X}_1, \mathbf{X}_{1+j})$.

**(B6)** As $n \to \infty$, $h_n \to 0$ and $(n^{1/2}h_n)/\log(n) \to \infty$. In addition, there exists a sequence of positive integers $s_n$ satisfying $s_n \to \infty$ and $s_n = o((nh_n)^{1/2})$ as $n \to \infty$ such that

$$(n/h_n)^{1/2}\beta(s_n) \to 0 \text{ as } n \to \infty. \tag{14}$$

The $\beta$-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 $\beta$-mixing condition. A set of conditions such as (13), $\mathrm{E}[\|\mathbf{X}\|^{2\delta}] < \infty$, the boundedness of the conditional densities (included in conditions (B4)-(B5)) and the latter part of condition (B6) is often assumed in density estimation and nonparametric regression. 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 corresponding $\alpha$-mixing condition, which is weaker than the current $\beta$-mixing condition). These conditions are sufficient for asymptotic normality of $(nh_n)^{1/2}\{T_n(\boldsymbol{\beta}_0) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]\}$, where $T_n(\boldsymbol{\beta})$ is given by (6). A sufficient condition on the mixing coefficient $\beta(j)$ to satisfy the conditions (13) and (14) is provided in Fan and Yao (2005, pp.387).

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

**Lemma 4.** *Under conditions (B1)-(B6), 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

$$n^{1/2}h_n^{-1/2}\{W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]\} \xrightarrow{p} 0,$$

uniformly in $\|\mathbf{t}\| \leq l$.

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

$$H_k = \{i : 2(k-1)a_n + 1 \leq i \leq (2k-1)a_n\},$$
$$T_k = \{i : (2k-1)a_n + 1 \leq i \leq 2ka_n\},$$

for $k = 1, \ldots, \mu_n$, where $\mu_n = [n/(2a_n)]$. Let $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in \bigcup_{k=1}^{\mu_n} H_k\}$ be a set of random vectors such that blocks $\{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in H_k\}$ $(k = 1, \ldots, \mu_n)$ 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 $i \in \bigcup_{k=1}^{\mu_n} H_k$.

Fix any $\epsilon > 0$. Observe that

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l} |W_{1n}(\mathbf{t}) - \mathrm{E}[W_{1n}(\mathbf{t})]| > n^{-1/2}h_n^{1/2}\epsilon\right) \\
&\le \mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=1}^{2a_n\mu_n}\{U_i^*(\mathbf{t}) - \mathrm{E}[U_i^*(\mathbf{t})]\}\right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{2}\right) \\
&\quad + \mathrm{P}\left(\sup_{\|\mathbf{t}\|\le l}\left|\frac{1}{n}\sum_{i=2a_n\mu_n+1}^{n}\{U_i^*(\mathbf{t}) - \mathrm{E}[U_i^*(\mathbf{t})]\}\right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{2}\right). \qquad (15)
\end{aligned}
$$

A simple calculation shows that the second term of the right hand side of (15) converges to zero; use the fact that $\sup_{\|\mathbf{t}\|\le l} |U_i^*(\mathbf{t})| \le |Z_i| I(h_n < U_i \le h_n + n^{-1/2} l \|\mathbf{X}_i\|)$. On the other hand, using the same argument as in Lemma 4.2 of Yu (1994), we may bound the first term of the right hand side of (15) by

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

where

$$
\tilde{V}_k(\mathbf{t}) = \sum_{i\in H_k} \tilde{U}_i^*(\mathbf{t}).
$$

Because of the condition (13), $\mu_n\beta(a_n) = o(1)$. Therefore, it suffices to show that

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

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

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

where $\xi_n = 1 - (16\mu_n/(\epsilon^2 n h_n)) \sup_{\|\mathbf{t}\|\le l} \mathrm{E}[\{\tilde{V}_1(\mathbf{t})\}^2]$. 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

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

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

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

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

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

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 $\{\boldsymbol{\sigma}^{(\mu_n)} \mapsto n^{-1}\sum_{k=1}^{\mu_n} \sigma_k \tilde{V}_k(\mathbf{t}) : \|\mathbf{t}\| \le l\}$, where $\boldsymbol{\sigma}^{(\mu_n)} = (\sigma_1, \ldots, \sigma_{\mu_n})$, is bounded by some polynomial of $n$ uniformly over every realization of $\tilde{D}_n := \{(\tilde{U}_i, \tilde{\mathbf{X}}_i), i \in \bigcup_{k=1}^{\mu_n} H_k\}$. In addition, Hoeffding's inequality implies that

$$
\sup_{\|\mathbf{t}\|\le l} \mathrm{P}\left( \left| \frac{1}{n} \sum_{k=1}^{\mu_n} \sigma_k \tilde{V}_k(\mathbf{t}) \right| > \frac{n^{-1/2}h_n^{1/2}\epsilon}{16} \;\middle|\; \tilde{D}_n \right) \le 2\exp\left( -\frac{h_n \epsilon^2}{512 w_n} \right),
$$

where $w_n = n^{-1}\sum_{k=1}^{\mu_n} \{\sum_{i\in H_k} |\tilde{Z}_i| I(h_n < \tilde{U}_i \le h_n + n^{-1/2}l\|\tilde{\mathbf{X}}_i\|)\}^2$. Thus, it suffices to show that

$$
\log(n) h_n^{-1} w_n \xrightarrow{p} 0. \tag{16}
$$

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

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

which leads to

$$
\mathrm{E}[w_n] = O(\mu_n a_n n^{-3/2}) = O(n^{-1/2}).
$$

Since $n^{1/2}h_n/\log(n) \to \infty$, (16) 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

$$(nh_n)^{1/2}\{T_n(\hat{\boldsymbol{\beta}}) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]\} = (nh_n)^{1/2}\{T_n(\boldsymbol{\beta}_0) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]\} + o_p(1).$$

In addition, mimicking the proof of Theorem 6.3 in Fan and Yao (2005), it can be shown that under conditions (B1)-(B6),

$$(nh_n)^{1/2}\{T_n(\boldsymbol{\beta}_0) - \mathrm{E}[T_n(\boldsymbol{\beta}_0)]\} \xrightarrow{d} N(0, \mathrm{E}[Z^2 f(0|\mathbf{X})]/2).$$

Therefore, arguing as in Section 2, we get the following theorem:

**Theorem 3.** *Suppose that conditions (B1)-(B6) hold and $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent for $\boldsymbol{\beta}_0$. Then,*

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

*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*

$$\frac{1}{2}\mathrm{E}[f_0(0|\mathbf{X})X_j X_k X_l X_m],$$

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

The result of Theorem 4 is same as that of Theorem 2 which deals with the i.i.d. case. Hence, the optimal bandwidth that minimizes the AMSE under the current weak dependence condition is the same as that for the i.i.d. case.

# 5 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 this paper, we have treated the uniform kernel, which is suited to apply the VC theory to obtain the uniform convergence results (see the proof of Lemma 1; see also Angrist et al., 2006, Appendix). As Powell (1984) states, however, more elaborating kernel functions could be devised. Although the kernel selection is less important than the bandwidth selection in density estimation (cf. Fan and Yao, 2005, Section 5.2), the extension of the present paper's results to a general kernel remains in the future research.

# Acknowledgments

The author would like to thank Professor Naoto Kunitomo and Professor Tatsuya Kubokawa for their encouragement and suggestions.

## 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. and Fernandéz-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **74** 539-563.

Arcones, M.A. and 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. and 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. and 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. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

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

Hendricks, W. and 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. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33-50.

Koul, H.L. (1992). *Weighted Empiricals and Linear Models.* Lecture Notes-Monograph Series, Vol. 12. Hayward, CA: 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. and Niu, S.-L. (2009). Asymptotic properties of error density estimator in regression model under $\alpha$-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. and 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.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer-Verlag.

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

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J. Powell and G. Tauchen (Ed.) *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.

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

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

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. and Wellner, J.A. (1996). *Weak convergence and empirical processes: With applications to statistics.* New York: Springer-Verlag.

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