EXTENDING THE SCOPE OF CUBE ROOT ASYMPTOTICS

MYUNG HWAN SEO AND TAISUKE OTSU

Preliminary and incomplete. Please do not distribute.

Abstract. This article extends the scope of cube root asymptotics for M-estimators in two directions: allow weakly dependent observations and criterion functions drifting with the sample sizes typically due to bandwidth sequences. For dependent empirical processes that characterize criterions inducing cube root phenomena, maximal inequalities are established so that a modified continuous mapping theorem for maximizing values of the criterions delivers limit laws of the M-estimators. The limit theory is applied not only to extend existing examples, such as maximum score estimator, nonparametric maximum likelihood density estimator under monotonicity, and least median of squares, toward weakly dependent observations, but also to address some open questions, such as asymptotic properties of the minimum volume predictive region, nonparametric Hough transform estimator, and smoothed maximum score estimator for dynamic panel data.

1. Introduction

There is a class of estimation problems where point estimators converge at the cube root rate to non-normal distributions instead of the familiar squared root rate to normals. Since Chernoff's (1964) study on estimation of the mode at least, several papers reported emergence of the cube root phenomena; see Prakasa Rao (1969) and Andrews *et al.* (1972), among others. The cube root convergence commonly arises when the criterion functions for point estimation are not smooth in parameters.

A seminal work by Kim and Pollard (1990) explained these cube root phenomena in a unified framework by means of empirical process theory, and established a limit theory for a general class of M-estimators defined by maximization of random processes that induces the cube root asymptotics. The limit theory of Kim and Pollard (1990) is general enough to encompass existing examples, such as the shorth (Andrews *et al.,* 1972), least median of squares (Rousseeuw, 1984), nonparametric monotone density estimator (Rao, 1969), and maximum score estimator (Manski, 1975), which are all illustrated in Kim and Pollard (1990). Also their theory has been applied to other statistical contexts, such as the Hough transform estimator (Goldenshluger and Zeevi, 2004) and split point estimator in decision trees (Bühlmann and Yu, 2002, and Banerjee and McKeague, 2007).

Since Kim and Pollard (1990), in spite of the generality, several statistical problems suggesting emergence of the cube root asymptotics but being outside the scope of Kim and Pollard's (1990) framework are posed. Most problems appeared in the course of generalizations of the existing

Date: October 2013.
The authors would like to thank Benedikt Pötscher and seminar participants at IHS at Vienna and University of Vienna for helpful comments.

examples discussed above. As a prototype, let us consider construction of conditional minimum volume predictive regions, studied generally by Polonik and Yao (2000), in a simplified manner. A statistician who observes a dependent process $\{y_t, x_t\}$ wishes to predict $y \in \mathbb{R}$ from $x \in \mathbb{R}$ by an interval on the real line $\mathbb{R}$. In this simple case, Polonik and Yao's (2000) minimum volume predictor of y at $x = c$ with level α may be written as the interval $[\hat{\theta} - \hat{r}, \hat{\theta} + \hat{r}]$, where

$$\hat{\theta} = \arg\min_{\theta} \hat{P}[\theta - \hat{r}, \theta + \hat{r}], \qquad \hat{r} = \inf\left\{ r : \sup_{\theta} \hat{P}[\theta - r, \theta + r] \geq \alpha \right\},$$

$\hat{P}[a,b] = \sum_{t=1}^{n} \mathbb{I}\{a \leq y_t \leq b\} K\left(\frac{x_t - c}{h_n}\right) / \sum_{t=1}^{n} K\left(\frac{x_t - c}{h_n}\right)$ is a nonparametric estimator of the conditional probability of $\{a \leq y_t \leq b\}$ given $x_t = c$. K is a kernel function and h_n is a bandwidth varying with the sample size n. This predictor is a natural generalization of the shorth to the conditional distribution of dependent observations. Polonik and Yao (2000, Remark 3b) conjectured that this predictor would converge at the $(nh_n)^{-1/3}$ rate. The framework of Kim and Pollard (1990) cannot be applied directly to address this question by two reasons: the observations are taken from a dependent process and the criterion function drifts with the bandwidth parameter. To allow dependent observations, the empirical process theory of Kim and Pollard (1990) for independent observations, in particular maximal inequalities to establish weak convergence of the criterion process, needs to be modified. To allow bandwidth sequences, the class of criterion functions for M-estimation needs to be reconsidered.

It should be emphasized that the above example is not an exception; several existing works call for development of such generalizations. Anevski and Hössjer (2006) extended the limit theory of nonparametric maximum likelihood under order restrictions toward weakly dependent and long range dependent data. Their extension include monotone density estimation as a special case. Goldenshluger and Zeevi (2004, p. 1916) mentioned possibility and importance of a generalized Hough transform estimator with a radius sequence tending to zero and left it for future research. Honoré and Kyriazidou (2000) proposed a maximum score-type estimator containing a bandwidth for dynamic longitudinal or panel data models with binary dependent variables and showed its consistency. However, the convergence rate and limiting distribution is an open question. Finally extensions of the classical least median of squares and maximum score estimators to dependent observations are still open questions (Zinde-Walsh, 2002, and de Jong and Woutersen, 2011).

This article extends the scope of cube root asymptotics for M-estimators in two directions: allow weakly dependent observations and criterion functions drifting with the sample sizes typically due to bandwidth sequences. For dependent empirical processes that characterize criterions inducing cube root phenomena, maximal inequalities are established so that a modified continuous mapping theorem for maximizing values of the criterions delivers limit laws of the M-estimators. The limit theory is applied not only to extend existing examples, such as maximum score estimator, nonparametric maximum likelihood density estimator under monotonicity, and least median of squares, toward weakly dependent observations, but also to address some open questions, such as asymptotic properties of the minimum volume predictive region, nonparametric Hough transform estimator, and smoothed maximum score estimator under panel data.

2. Cube root asymptotics with dependent observations

This section extends Kim and Pollard's (1990) main theorem on cube root asymptotics of the M-estimator to allow dependent data. This section focuses on the case where the criterion function is independent of the sample size and consider an extension to dependent data. The M-estimator $\hat{\theta}$ maximizes the random criterion

$$\mathbb{P}_n f_\theta = \frac{1}{n}\sum_{t=1}^{n} f_\theta(z_t),$$

where $\{f_\theta : \theta \in \Theta\}$ is a collection of functions indexed by a subset Θ of $\mathbb{R}^d$ and $\{z_t\}$ is a strictly stationary sequence of random variables with marginal P. We characterize a class of criterion functions that induces cube root phenomena (or *sharp edge effects* in the sense of Kim and Pollard, 1990) and is general enough to cover all examples listed above. Let $Pf = \int f dP$ for a function f, $|\cdot|$ be the Euclidean norm of a vector, and $\|\cdot\|_2$ be the L_2-norm of a random variable. The class of criterions of our interest is defined as follows.

Definition (Cube root class). *A class of functions $\{f_\theta : \theta \in \Theta\}$ is called the cube root class if*

- **(i):** *$\{f_\theta : \theta \in \Theta\}$ is a class of bounded functions and Pf_θ is uniquely maximized and twice continuously differentiable at θ_0 with a negative definite second derivative matrix V.*
- **(ii):** *There exist positive constants C and C' such that*

$$|\theta_1 - \theta_2| \leq C \|f_{\theta_1} - f_{\theta_2}\|_2\,, \tag{1}$$

  for all $\theta_1, \theta_2 \in \{\Theta : |\theta - \theta_0| \leq C'\}$.
- **(iii):** *There exists a positive constant C'' such that*

$$P \sup_{\theta\in\Theta:|\theta-\theta'|<\varepsilon} |f_\theta - f_{\theta'}|^2 \leq C''\varepsilon. \tag{2}$$

  for all $\theta' \in \Theta$ and $\varepsilon > 0$.

Condition (i), related to Kim and Pollard (1990, Conditions (ii) and (iv) of the main theorem), is a standard identification condition for M-estimation and local shape restriction to obtain the cube root rate. Condition (ii), which does not appear in Kim and Pollard (1990), is an additional condition to deal with dependent observations. Using the notation in Lemma M below, this condition is used to obtain the entropy relation $N_{[]}(\nu, \mathcal{G}_\delta^\beta, \|\cdot\|_{2,\beta}) \leq N_{[]}(\nu, \mathcal{G}_{C_2\delta}^1, \|\cdot\|_2)$. For independent observations, this condition is not required because the $L_{2,\beta}$-norm $\|\cdot\|_{2,\beta}$ in the sense of Doukhan, Massart and Rio (1995) is equivalent to the L_2-norm $\|\cdot\|_2$. Condition (ii) is often guaranteed by the identification condition in (i). Also this condition can be verified by an expansion of $\|f_{\theta_1} - f_{\theta_2}\|_2$ around $\theta_1, \theta_2 = \theta_0$. See Section 4 for specific illustrations. Condition (iii) is a key condition for the cube root asymptotics. This condition implies an envelope condition of Kim and Pollard (1990, Condition (vi)), $PF_\varepsilon^2 = P\sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} |f_\theta - f_{\theta_0}|^2 \leq C\varepsilon$ for some $C > 0$. For independent observations, upper bounds of maximal inequalities for uniformly manageable classes (Kim and Pollard, 1990, p. 199) are characterized by PF_ε^2, and thus the condition $PF_\varepsilon^2 \leq C\varepsilon$ suffices to control empirical processes for M-estimation. However, for dependent observations, to best of our knowledge, there is no such general maximal inequality and Condition (iii) cannot be replaced with $PF_\varepsilon^2 \leq C\varepsilon$.

Throughout this section, let $\{f_\theta : \theta \in \Theta\}$ be a cube root class. We now study the limit behavior of the M-estimator, which is precisely defined as a random variable $\hat{\theta}$ satisfying

$$\mathbb{P}_n f_{\hat{\theta}} \geq \sup_{\theta \in \Theta} \mathbb{P}_n f_\theta - o_p(n^{-2/3}).$$

The first step is to establish consistency of the M-estimator, i.e., $\hat{\theta}$ converges in probability to the unique maximizer θ_0 of Pf_θ. The technical argument to derive the consistency is rather standard and typically shown by uniform convergence of the criterion $\mathbb{P}_n f_\theta$ to Pf_θ over Θ. In this section we assume consistency of $\hat{\theta}$. See illustrations below for details to verify consistency.

The next step is to derive the convergence rate of $\hat{\theta}$. A key ingredient for this is to establish tightness of the centered empirical process $\{\mathbb{G}_n(f_\theta - f_{\theta_0}) : \theta \in \Theta\}$, where $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$ for a function f. For independent observations, several maximal inequalities are available in the literature (see, e.g., Kim and Pollard, 1990, p. 199). For dependent observations, to best of our knowledge, there is no maximal inequality which can be applied to the cube root class. Our first contribution is to establish a maximal inequality for the cube root class under some dependent observations.

To characterize dependence of observations, this paper considers an absolutely regular process. Let $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_m^\infty$ be σ-fields of $\{\ldots, z_{t-1}, z_0\}$ and $\{z_m, z_{m+1}, \ldots\}$, respectively. Define the β-mixing coefficient as $\beta_m = \frac{1}{2} \sup \sum_{(i,j)\in I\times J} |P\{A_i \cap B_j\} - P\{A_i\}P\{B_j\}|$, where the supremum is taken over all the finite partitions $\{A_i\}_{i\in I}$ and $\{B_j\}_{j\in J}$ respectively $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_m^\infty$ measurable. Also let $\beta(\cdot)$ be a function such that $\beta(t) = \beta_{[t]}$ if $t \geq 1$ and $\beta(t) = 1$ otherwise, and $\beta^{-1}(\cdot)$ be the càdlàg inverse of $\beta(\cdot)$. Throughout the paper, we impose the following assumption to the observations.

Assumption D. *$\{z_t\}$ is a strictly stationary and absolutely regular process with β-mixing coefficients $\{\beta_m\}$ such that $\beta_m = O(\rho^m)$ for some $0 < \rho < 1$.*

This assumption says the mixing coefficient β_m should decay at an exponential rate. For example, finite-order ARMA processes typically satisfy this assumption. For the minimum volume prediction discussed in Introduction, Polonik and Yao (2000) assumed an exponential decay rate. This assumption is required not only to establish the maximal inequality in Lemma M below but also to establish a central limit theorem in Lemma C for finite dimensional convergence. See remarks on Lemmas M and C below for further discussions. Under this assumption, the maximal inequality for the empirical process $\mathbb{G}_n(f_\theta - f_{\theta_0})$ is obtained as follows.

Lemma M. *There exist positive constants C and C' such that*

$$P \sup_{|\theta-\theta_0|<\delta} |\mathbb{G}_n(f_\theta - f_{\theta_0})| \leq C\delta^{1/2},$$

for all n large enough and $\delta \in [n^{-1/2}, C']$.

Proof. For any function g, let $Q_g(u)$ be the inverse function of the tail probability function $x \mapsto \Pr\{|g(z_t)| > x\}$. Then we define the norm

$$\|g\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u) Q_g(u)^2 du}.$$

Let

$$\begin{aligned}
\mathcal{G}_\delta^\beta &= \{f_\theta - f_{\theta_0} : \|f_\theta - f_{\theta_0}\|_{2,\beta} < \delta \text{ for } \theta \in \Theta\}, \\
\mathcal{G}_\delta^1 &= \{f_\theta - f_{\theta_0} : |\theta - \theta_0| < \delta \text{ for } \theta \in \Theta\}, \\
\mathcal{G}_\delta^2 &= \{f_\theta - f_{\theta_0} : \|f_\theta - f_{\theta_0}\|_2 < \delta \text{ for } \theta \in \Theta\}.
\end{aligned}$$

Since g is bounded for any $g \in \mathcal{G}_\delta^1$ and so is Q_g, we can always find a function $\hat{g}$ such that $\|g\|_2^2 \leq \|\hat{g}\|_2^2 \leq 2\|g\|_2^2$ and

$$Q_{\hat{g}}(u) = \sum_{j=1}^m a_j \mathbb{I}\{(j-1)/m \leq u < j/m\},$$

satisfying $|Q_g| \leq Q_{\hat{g}}$, for some positive integer m and sequence $\{a_j\}$. Now take any $C' > 0$, and then pick any n and $\delta \in [n^{-1/2}, C']$. Hereafter positive constants C_j $(j = 1, 2, \ldots)$ are independent of n and δ.

Next, based on the above notation, we derive some set inclusion relationships. Let $M = \frac{1}{2}\sup_{0<x\leq 1} x^{-1}\int_0^x \beta^{-1}(u)du$. For any $g \in \mathcal{G}_\delta^1$, it holds

$$\begin{aligned}
\|g\|_2^2 &\leq \int_0^1 \beta^{-1}(u) Q_g(u)^2 du \leq \frac{1}{m}\sum_{j=1}^m a_j^2 \left\{ m \int_{(j-1)/m}^{j/m} \beta^{-1}(u)du \right\} \\
&\leq \left\{ m \int_0^{1/m} \beta^{-1}(u)du \right\} \int_0^1 Q_{\hat{g}}(u)^2 du \\
&\leq M\|g\|_2^2,
\end{aligned}$$

where the first inequality is due to Doukhan, Massart and Rio (1995, Lemma 1) and the second inequality follows from $|Q_g| \leq Q_{\hat{g}}$, the third inequality follows from monotonicity of $\beta^{-1}(u)$, and the last inequality follows by $\|\hat{g}\|_2^2 \leq 2\|g\|_2^2$. This inequality implies[1]

$$\|f_\theta - f_{\theta_0}\|_2 \leq \|f_\theta - f_{\theta_0}\|_{2,\beta} \leq M\|f_\theta - f_{\theta_0}\|_2. \tag{3}$$

Based on the above inequalities, we can deduce the inclusion relationships: there are positive constants C_1 and C_2 such that

$$\mathcal{G}_\delta^1 \subset \mathcal{G}_{C_1\delta^{1/2}}^2 \subset \mathcal{G}_{MC_1\delta^{1/2}}^\beta, \qquad \mathcal{G}_\delta^\beta \subset \mathcal{G}_\delta^2 \subset \mathcal{G}_{\delta C_2}^1, \tag{4}$$

where the relation $\mathcal{G}_\delta^1 \subset \mathcal{G}_{C_1\delta^{1/2}}^2$ follows from (2) and the relation $\mathcal{G}_\delta^2 \subset \mathcal{G}_{\delta C_2}^1$ follows from (1).

Third, based on the above set inclusion relationships, we derive some relationships for the bracketing numbers. Let $N_{[]}(\nu, \mathcal{G}, \|\cdot\|)$ be the bracketing number for a class of functions $\mathcal{G}$ with

[1] If g is a binary function, then $m^{-1} = \Pr\{g(z_t) = 1\}$ and $\|g\|_{2,\beta} = \|g\|_2 \sqrt{m\int_0^{1/m}\beta^{-1}(u)du}$.

radius $\nu > 0$ and norm $\|\cdot\|$. By (3) and the second relation in (4),

$$N_{[]}(\nu, \mathcal{G}_\delta^\beta, \|\cdot\|_{2,\beta}) \le N_{[]}(\nu, \mathcal{G}_{C_2\delta}^1, \|\cdot\|_2) \le C_3 \left(\frac{\delta}{\nu}\right)^{2d},$$

for some $C_3 > 0$, where the second inequality follows from the argument to derive Andrews (1993, eq. (4.7)) based on the L_2-continuity assumption in (2). Therefore, we have $\varphi_n(\delta) = \int_0^\delta \sqrt{\log N_{[]}(\nu, \mathcal{G}_\delta^\beta, \|\cdot\|_{2,\beta})} d\nu \le C_4\delta$ for some C_4.

Finally, based on the above entropy condition, we apply the maximal inequality of Doukhan, Massart and Rio (1995, Theorem 3), i.e., there exists a positive constant C_5 depending only on the mixing sequence $\{\beta_m\}$ such that

$$P \sup_{g \in \mathcal{G}_\delta^\beta} |\mathbb{G}_n g| \le C_5[1 + \delta^{-1} q_{G_\delta}(\min\{1, v_n(\delta)\})]\varphi_n(\delta),$$

where $q_{G_\delta}(v) = \sup_{u \le v} Q_G(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ with the envelope function G of $\mathcal{G}_\delta^\beta$ (note: $\mathcal{G}_\delta^\beta$ is a class of bounded functions) and $v_n(\delta)$ is the unique solution of $\frac{v_n(\delta)}{v_n(\delta)^{-1}\int_0^{v_n(\delta)} \beta^{-1}(\tilde{u})d\tilde{u}} = \frac{\varphi_n(\delta)^2}{n\delta^2}$. Since $\varphi_n(\delta) \le C_4\delta$, it holds $v_n(\delta) \le C_5 n^{-1}$ for some $C_5 > 0$. Now take some n_0 such that $v_{n_0}(\delta) \le 1$, and then pick any $n \ge n_0$ and $\delta \in [n^{-1/2}, C']$. We have $q_G(\min\{1, v_n(\delta)\}) \le C_6\sqrt{v_n(\delta)}Q_G(v_n(\delta)) \le C_7 n^{-1/2}$. Therefore, the conclusion follows by

$$P \sup_{g \in \mathcal{G}_\delta^1} |\mathbb{G}_n g| \le P \sup_{g \in \mathcal{G}_{M\delta^{1/2}}^\beta} |\mathbb{G}_n g| \le C_8 \delta^{1/2},$$

where the first inequality follows from the first relation in (4). □

Lemma M can be shown under a slightly weaker condition $\sup_{0<x\le 1} x^{-1}\int_0^x \beta^{-1}(u)du < \infty$ than $\beta_m = O(\rho^m)$ in Assumption D. However, this weaker condition already excludes polynomial decay of β_m. To establish the convergence rate (and consistency as well), the following analog of Kim and Pollard (1990, Lemma 4.1) is useful.

Lemma 1. *For each $\varepsilon > 0$, there exist random variables $\{R_n\}$ of order $O_p(1)$ and a positive constant C such that*

$$|\mathbb{P}_n(f_\theta - f_{\theta_0}) - P(f_\theta - f_{\theta_0})| \le \varepsilon|\theta - \theta_0|^2 + n^{-2/3}R_n^2,$$

for any $|\theta - \theta_0| \le C$.

Proof. Define $A_{n,j} = \{\theta : (j-1)n^{-1/3} \le |\theta - \theta_0| < jn^{-1/3}\}$ and

$$R_n^2 = n^{2/3} \inf_{n^{-1/3} \le |\theta - \theta_0| \le C} \{|\mathbb{P}_n(f_\theta - f_{\theta_0}) - P(f_\theta - f_{\theta_0})| - \varepsilon|\theta - \theta_0|^2\}.$$

There exists a positive constant C such that

$$\begin{aligned}
P\{R_n > m\} &= P\left\{|\mathbb{P}_n(f_\theta - f_{\theta_0}) - P(f_\theta - f_{\theta_0})| > \varepsilon|\theta - \theta_0|^2 + n^{-2/3}m^2 \quad \text{for some } \theta\right\} \\
&\leq \sum_{j=1}^{\infty} P\left\{n^{2/3}|\mathbb{P}_n(f_\theta - f_{\theta_0}) - P(f_\theta - f_{\theta_0})| > \varepsilon(j-1)^2 + m^2 \quad \text{for some } \theta \in A_{n,j}\right\} \\
&\leq \sum_{j=1}^{\infty} \frac{C\sqrt{j}}{\varepsilon(j-1)^2 + m^2},
\end{aligned}$$

for all $m > 0$, where the last equality is due to the Markov inequality and Lemma M. Since the above sum is finite for all $m > 0$, the conclusion follows. □

Based on this lemma, the cube root convergence rate of $\hat{\theta}$ is obtained as follows. For $|\hat{\theta} - \theta_0| \geq n^{-1/3}$, we can take $c > 0$ such that

$$\begin{aligned}
0 \leq \mathbb{P}_n(f_{\hat{\theta}} - f_{\theta_0}) &\leq P(f_{\hat{\theta}} - f_{\theta_0}) + \varepsilon|\hat{\theta} - \theta_0|^2 + n^{-2/3}R_n^2 \\
&\leq (-c + \varepsilon)|\hat{\theta} - \theta_0|^2 + O_p(n^{-2/3}),
\end{aligned}$$

for each $\varepsilon > 0$, where the second inequality follows from Lemma 1 and the third inequality follows from the condition (ii) of the cube root class. Taking ε small enough to satisfy $c - \varepsilon > 0$ yields contradiction and thus we obtain $\hat{\theta} - \theta_0 = O_p(n^{-1/3})$.

Given the cube root convergence rate of $\hat{\theta}$, the final step is to derive its limiting distribution. To this end, it is common to apply a continuous mapping theorem of an argmax element (e.g., Kim and Pollard, 1990, Theorem 2.7). A key ingredient for this argument is to establish weak convergence of the centered and normalized process

$$Z_n(s) = n^{1/6}\mathbb{G}_n(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0}),$$

for $|s| \leq K$ with any $K > 0$. Weak convergence of the process Z_n may be characterized by its finite dimensional convergence and tightness (or stochastic equicontinuity). If $\{z_t\}$ is independently and identically distributed as in Kim and Pollard (1990), a classical central limit theorem combined with the Cramér-Wold device implies finite dimensional convergence, and a maximal inequality on a suitably regularized class of functions guarantees tightness of the process of criterion functions. For finite dimensional convergence, we employ the following central limit theorem, which is based on Rio's (1997, Corollary 1) central limit theorem for an α-mixing array. Let $Q_g(u)$ be the inverse function of the tail probability function $x \mapsto P\{|g(z_t)| > x\}$.[2]

Lemma C. *Suppose*

$$\sup_n \int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 du < \infty. \tag{5}$$

Then $V = \lim_{n\to\infty} \mathrm{Var}(\mathbb{G}_n g_n)$ *exists and* $\mathbb{G}_n g_n \xrightarrow{d} N(0, V)$.

Proof. First of all, any β-mixing process is α-mixing with $\alpha_m \leq \beta_m/2$. We now check Conditions (a) and (b) of Rio (1997, Corollary 1). Condition (a) is verified by Rio (1997, Proposition 1),

[2] The function $Q_g(u)$, called the quantile function in Doukhan, Massart and Rio (1995), is different from a familiar function $u \mapsto \inf\{x : u \leq P\{|g(z_t)| \leq x\}\}$.

which guarantees $\text{Var}(\mathbb{G}_n g_n) \leq \int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 du$ for all n. Since $\text{Var}(\mathbb{G}_n g_n)$ is bounded and z_t is strictly stationary in our case, Condition (b) of Rio (1997, Corollary 1) can be written as

$$\int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 \inf_n \{n^{-1/2} \beta^{-1}(u) Q_{g_n}(u), 1\} du \to 0,$$

as $n \to \infty$. Note that for each $u \in (0,1)$, it holds $n^{-1/2} \beta^{-1}(u) Q_{g_n}(u) \to 0$ as $n \to \infty$. Thus, the dominated convergence theorem based on (5) implies Condition (b). □

The finite dimensional convergence of Z_n follows from this lemma by setting $g_n = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})$. The requirement (5) can be considered as a Lindeberg-type condition to guarantee Rio's (1997, Corollary 1) Lindeberg condition in our setup. It should be noted that for the rescaled object $g_n = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})$, the moments $P|g_n|^{2+\delta}$ with $\delta > 0$ typically diverge. Thus we cannot apply central limit theorems for mixing sequences with higher than second moments. The Lindeberg condition is one of the weakest conditions, if any, for the central limit theorem of mixing sequences without moment condition higher than two. The condition (5) excludes polynomial decay of β_m. Also, Doukhan, Massart and Rio (1994, Theorem 5) provided some result, where any polynomial mixing rate will destroy the asymptotic normality of $\mathbb{G}_n g_n$.

Lemma M'. *For any $\epsilon > 0$, there exist $\delta > 0$ and a positive integer n_δ such that*

$$P \sup_{|s-s'|<\delta} |Z_n(s) - Z_n(s')| \leq \epsilon,$$

for all $n \geq n_\delta$.

Compared to Lemma M used to derive the convergence rate of the estimator, Lemma M' is applied only to establish tightness of the process Z_n. Therefore, we do not need an exact decay rate on the right hand side of the above inequality. In particular, the process $Z_n(s)$ itself does not satisfy the condition (1) of the cube root class.

Proof. Pick any $K > 0$ and $\epsilon > 0$. Let $g_{n,s,s'} = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0 + s'n^{-1/3}})$, $\mathcal{G}^1_{n,\delta} = \{g_{n,s,s'} : |s - s'| < \delta\}$, $\mathcal{G}^\beta_{n,\delta} = \{g_{n,s,s'} : \|g_{n,s,s'}\|_{2,\beta} < \delta\}$, and $\mathcal{G}^2_{n,\delta} = \{g_{n,s,s'} : \|g_{n,s,s'}\|_2 < \delta\}$. Since $g_{n,s,s'}$ satisfies the condition (2), there is $C_1 > 0$ such that $\mathcal{G}^1_{n,\delta} \subset \mathcal{G}^2_{n,C_1\delta^{1/2}}$ for all n and $\delta > 0$. Also, by the same argument to derive (3), there exists $C_2 > 0$ such that $\|g_{n,s,s'}\|_2 \leq \|g_{n,s,s'}\|_{2,\beta} \leq C_2 \|g_{n,s,s'}\|_2$ for all n, $|s| \leq K$, and $|s'| \leq K$, which implies

$$\mathcal{G}^1_{n,\delta} \subset \mathcal{G}^2_{n,C_1\delta^{1/2}} \subset \mathcal{G}^\beta_{n,C_1C_2\delta^{1/2}},$$

for all n and $\delta > 0$. Also note that the bracketing numbers satisfy

$$N_{[]}(\nu, \mathcal{G}^\beta_{n,\delta}, \|\cdot\|_{2,\beta}) \leq N_{[]}(\nu, \mathcal{G}_n, C_2 \|\cdot\|_2) \leq C_1 C_2 \nu^{-d/2},$$

where $\mathcal{G}_n = \{g_{n,s,s'} : |s| \leq K, |s'| \leq K\}$ is the original space of $g_{n,s,s'}$ and the second inequality follows from the L_2-continuity assumption in (2). Thus letting $\eta = C_1 C_2 \delta^{1/2}$, there is a function $\varphi(\eta)$ such that $\varphi(\eta) \to 0$ as $\eta \to 0$ and $\varphi_n(\eta) = \int_0^\eta \sqrt{\log N_{[]}(\nu, \mathcal{G}^\beta_{n,\eta}, \|\cdot\|_{2,\beta})} d\nu \leq \varphi(\eta)$ for all n and $\eta > 0$. Based on this entropy condition, we apply the maximal inequality of Doukhan, Massart and Rio (1995, Theorem 3), i.e., there exists $C_3 > 0$ depending only on the mixing sequence $\{\beta_m\}$ such that

$$P \sup_{g \in \mathcal{G}^{\beta}_{n,\eta}} |\mathbb{G}_n g| \leq C_3 [1 + \eta^{-1} q_{G_n}(\min\{1, v_n(\eta)\})] \varphi(\eta),$$

for all n and $\eta > 0$, where $q_{G_n}(\cdot)$ with the envelope G_n of $\mathcal{G}^{\beta}_{n,\eta}$ is defined in the same way as the proof of Lemma M (note: by the definition of $\mathcal{G}^{\beta}_{n,\eta}$, we can take the envelope G_n independently from η), and $v_n(\eta)$ is the unique solution of $\frac{v_n(\eta)}{v_n(\eta)^{-1} \int_0^{v_n(\eta)} \beta^{-1}(\tilde{u}) d\tilde{u}} = \frac{\varphi_n^2(\eta)}{n\eta^2}$.

Now pick any $\eta > 0$ small enough so that $2C_3 \varphi(\eta) < \epsilon$. Since $\varphi_n(\eta) \leq \varphi(\eta)$, there is $C_4 > 0$ such that $v_n(\eta) \leq C_4 \frac{\varphi(\eta)}{n\eta^2}$ for all n and $\eta > 0$. Since $G_n = O(n^{1/6})$ by the definition of $\mathcal{G}^{\beta}_{n,\eta}$, there exists $C_5 > 0$ such that $q_{G_n}(\min\{1, v_n(\eta)\}) \leq C_5 \sqrt{\varphi(\eta)} \eta^{-1} n^{-1/3}$ for all n large enough. Therefore, the conclusion follows by

$$P \sup_{g \in \mathcal{G}^{1}_{n,\eta}} |\mathbb{G}_n g| \leq P \sup_{g \in \mathcal{G}^{\beta}_{n, C_1 \eta^{1/2}}} |\mathbb{G}_n g| \leq \epsilon,$$

for all n large enough, where the first inequality follows from $\mathcal{G}^{1}_{n,\delta} \subset \mathcal{G}^{\beta}_{n, C_1 C_2 \delta^{1/2}}$. □

Based on finite dimensional convergence and Lemma M', we establish weak convergence of Z_n. Then a continuous mapping theorem of an argmax element (Kim and Pollard, 1990, Theorem 2.7) implies the limiting distribution of the M-estimator $\hat{\theta}$. The main result of this section is summarized as follows.

Theorem 1. *Let $\{f_\theta : \theta \in \Theta\}$ be a cube root class and $\hat{\theta}$ satisfy $\mathbb{P}_n f_{\hat{\theta}} \geq \sup_{\theta \in \Theta} \mathbb{P}_n f_\theta - o_p(n^{-2/3})$. Assume $\hat{\theta}$ converges in probability to $\theta_0 \in \text{int}\Theta$, and $\sup_n \int_0^1 \beta^{-1}(u) Q_{g_{n,s}}(u)^2 du < \infty$ for each s, where $g_{n,s} = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})$. Then*

$$n^{1/3}(\hat{\theta} - \theta_0) \xrightarrow{d} \arg\max_s Z(s),$$

where $Z(s)$ is a Gaussian process with continuous sample paths, expected value $-s'Vs/2$, and covariance kernel $H(s_1, s_2) = \lim_{n \to \infty} \sum_{t=-n}^{n} P g_{n,s_1}(z_0) g_{n,s_2}(z_t)$.

This theorem can be considered as an extension of the main theorem of Kim and Pollard (1990) to the absolutely regular dependent process. The Lindeberg-type condition $\sup_n \int_0^1 \beta^{-1}(u) Q_{g_{n,s}}(u)^2 du < \infty$ needs to be verified for each application. It is often the case that $P\{g_{n,s} = 0\} \geq 1 - cn^{-1/3}$ for some $c > 0$ and all n large enough. In this case, this condition can be verified by

$$\int_0^1 \beta^{-1}(u) Q_{g_{n,s}}(u)^2 du \leq Cn^{1/3} \int_0^{cn^{-1/3}} \beta^{-1}(u) du + n^{1/3} \{P(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})\}^2 \int_0^1 \beta^{-1}(u) du < \infty,$$

for all n, where the second inequality follows from the fact that $\beta^{-1}(\cdot)$ is monotonically decreasing and the last inequality follows by Assumption D.

It is often the case that the criterion function contains some nuisance parameters which can be estimated by faster rates than $O_p(n^{-1/3})$. For such situations, Theorem 1 is extended as follows. For the rest of this section, let $\hat{\theta}$ and $\tilde{\theta}$ satisfy $\mathbb{P}_n f_{\hat{\theta}, \hat{\nu}} \geq \sup_{\theta \in \Theta} \mathbb{P}_n f_{\theta, \hat{\nu}} + o_p(n^{-2/3})$, where $\hat{\nu} - \nu_0 = o_p(n^{-1/3})$, and $\mathbb{P}_n f_{\tilde{\theta}, \nu_0} \geq \sup_{\theta \in \Theta} \mathbb{P}_n f_{\theta, \nu_0} + o_p(n^{-2/3})$, respectively.

Theorem 2. *Let $\{f_{\theta,\nu} : \theta \in \Theta, \nu \in \Lambda\}$ be a cube root class, where the condition (i) is replaced with*

(i)': *The class is bounded and for some negative definite matrix V_1 and some finite matrix V_2,*

$$P(f_{\theta,\nu} - f_{\theta_0,\nu_0}) = \frac{1}{2}(\theta-\theta_0)'V_1(\theta-\theta_0) + (\theta-\theta_0)'V_2(\nu-\nu_0) + o(|\theta-\theta_0|^2 + |\nu-\nu_0|^2).$$

Then $\hat{\theta} = \tilde{\theta} + o_p(n^{-1/3})$. Furthermore, if $\sup_n \int_0^1 \beta^{-1}(u)Q_{g_{n,s}}(u)^2 du < \infty$ for each s with $g_{n,s} = n^{1/6}(f_{\theta_0+sn^{-1/3},\nu_0} - f_{\theta_0,\nu_0})$, then $n^{1/3}(\hat{\theta}-\theta_0) \xrightarrow{d} \arg\max_s Z(s)$, where $Z(s)$ is a Gaussian process with continuous sample paths, expected value $s'V_1 s/2$ and covariance kernel H.

Proof. To ease notation, let $\theta_0 = \nu_0 = 0$. This proof shows $|\hat{\theta}| = O_p(n^{-1/3})$. Then Theorem 1 and Lemma 2 below imply the conclusion. By Lemma 1 and the condition (i)', for each $\epsilon > 0$ there is $C > 0$ such that

$$\begin{aligned}
\mathbb{P}_n(f_{\theta,\nu} - f_{0,0}) &\leq P(f_{\theta,\nu} - f_{0,0}) + 2\epsilon(|\theta|^2 + |\nu|^2) + O_p(n^{-2/3}) \\
&\leq \frac{1}{2}\theta'V_1\theta + \theta'V_2\nu + 2\epsilon(|\theta|^2 + |\nu|^2) + O_p(n^{-2/3}),
\end{aligned}$$

for all $|(\theta,\nu)| \leq C$. From $\mathbb{P}_n(f_{\hat{\theta},\hat{\nu}} - f_{0,0}) \geq 0$, negative definiteness of V_1, and $\hat{\nu} = o_p(n^{-1/3})$, we can find $c > 0$ such that

$$0 \leq -c|\hat{\theta}|^2 + |\hat{\theta}|o_p(n^{-1/3}) + O_p(n^{-2/3}),$$

which implies $|\hat{\theta}| = O_p(n^{-1/3})$. □

Lemma 2. *Suppose $\hat{\theta} - \theta_0 = O_p(n^{-1/3})$, $\hat{\nu} - \nu_0 = o_p(n^{-1/3})$, $\{f_{\theta,\nu}\}$ is a cube root class with the condition (i), and $Pf_{\theta,\nu}$ is twice continuously differentiable at (θ_0,ν_0). Then $\hat{\theta} - \tilde{\theta} = o_p(n^{-1/3})$.*

Proof. To ease notation, let $\theta_0 = \nu_0 = 0$. By reparametrization,

$$n^{1/3}\arg\max_\theta \mathbb{P}_n(f_{\theta,\hat{\nu}} - f_{,0\hat{\nu}}) = \arg\max_s [n^{2/3}(\mathbb{P}_n - P)(f_{sn^{-1/3},\hat{\nu}} - f_{0,\hat{\nu}}) + n^{2/3}P(f_{sn^{-1/3},\hat{\nu}} - f_{0,\hat{\nu}})] + o_p(1).$$

By Lemma M (replace θ with (θ,ν)) and $\hat{\nu} = o_p(n^{-1/3})$,

$$n^{2/3}(\mathbb{P}_n - P)(f_{sn^{-1/3},\hat{\nu}} - f_{0,0}) - n^{2/3}(\mathbb{P}_n - P)(f_{sn^{-1/3},0} - f_{0,0}) = o_p(1).$$

uniformly in s. Also by an expansion around $s = 0$,

$$\begin{aligned}
P(f_{sn^{-1/3},\hat{\nu}} - f_{0,\hat{\nu}}) &= n^{-1/3}\frac{\partial Pf_{0,\hat{\nu}}}{\partial s}s + n^{-2/3}s'\frac{\partial^2 Pf_{0,0}}{\partial s\partial s'}s + o_p(n^{-2/3}) \\
&= n^{-2/3}s'\frac{\partial^2 Pf_{0,0}}{\partial s\partial \nu'}\hat{\nu} + n^{-2/3}s'\frac{\partial^2 Pf_{0,0}}{\partial s\partial s'}s + o_p(n^{-2/3}),
\end{aligned}$$

where the second inequality follows from $\left.\frac{\partial Pf_{\theta,0}}{\partial\theta}\right|_{\theta=0} = 0$ and $\hat{\nu} = o_p(n^{-1/3})$. Comparing this with a Taylor expansion of $P(f_{sn^{-1/3},0} - f_{0,0})$ yields that $P(f_{sn^{-1/3},\hat{\nu}} - f_{0,\hat{\nu}}) - P(f_{sn^{-1/3},0} - f_{0,0}) = o_p(n^{-2/3})$ uniformly in s. Thus, the proof is complete. □

3. Cube root asymptotics with drifting criterions

We next investigate the case where the criterion function depends on n and typically contains a bandwidth parameter to deal with some nonparametric component. The cube root class is modified as follows.

Definition (Drifting cube root class). *A class of functions* $\{f_{n,\theta} : \theta \in \Theta\}$ *belongs to the drifting cube root class if*

(i): $\{f_{n,\theta} : \theta \in \Theta\}$ *is a class of bounded functions for all* n, $\sup_{z,\theta} |f_{n,\theta}(z)| = O(h_n^{-1})$ *for a sequence* h_n *satisfying* $h_n \to 0$ *and* $n^{1/2}h_n \to \infty$, *and* $\lim_{n\to\infty} Pf_{n,\theta}$ *is uniquely maximized at* θ_0 *and* $Pf_{n,\theta}$ *is twice continuously differentiable at* θ_0 *and admits the expansion*

$$P(f_{n,\theta} - f_{n,\theta_0}) = \frac{1}{2}(\theta - \theta_0)'V(\theta - \theta_0) + o(|\theta - \theta_0|^2) + o((nh_n)^{-2/3}), \tag{6}$$

for a negative definite matrix V.

(ii): *There exist positive constants* C *and* C' *such that*

$$|\theta_1 - \theta_2| \leq Ch_n^{1/2} \|f_{n,\theta_1} - f_{n,\theta_2}\|_2 \quad \text{uniformly over } n,$$

for all $\theta_1, \theta_2 \in \{\Theta : |\theta - \theta_0| \leq C'\}$.

(iii): *There exists a positive constant* C'' *such that*

$$P \sup_{\theta\in\Theta:|\theta-\theta'|<\varepsilon} h_n|f_{n,\theta} - f_{n,\theta'}|^2 \leq C''\varepsilon \quad \text{uniformly over } n,$$

for all $\theta' \in \Theta$ *and* $\varepsilon > 0$.

Similar comments to the ones for the cube root class apply. When the criterion $f_{n,\theta}$ involves some kernel estimate for a nonparametric component, h_n is considered as a bandwidth parameter. Typically the criterion takes the form of $f_{n,\theta}(z) = \frac{1}{h_n}K\left(\frac{x-c}{h_n}\right)m(y, x, \theta)$ for $z = (y, x)$ and some function m (e.g., the minimum volume prediction in Section 4.4 and smoothed maximum score estimator for panel data by Honoré and Kyriazidou, 2000). In this case, the requirement $\sup_{z,\theta} |f_{n,\theta}(z)| = O(h_n^{-1})$ in Condition (i) means boundedness of $K\left(\frac{x-c}{h_n}\right)m(y, x, \theta)$. The condition in (6) can be understood as a restriction for $P(f_{n,\theta} - f_{n,\theta_0}) = \int K(\tilde{x})\, m(y, c+h_n\tilde{x}, \theta) f_{yx}(y, c+h_n\tilde{x})d\tilde{x}dy$ by change of variables, where f_{yx} is the joint density. The reasons for multiplication of $h_n^{1/2}$ in Condition (ii) and h_n in (iii) are understood in the same manner.

Throughout this section, let $\{f_{n,\theta} : \theta \in \Theta\}$ be a cube root class. In this modified class of functions, we study the limit behavior of the M-estimator, which is precisely defined as a random variable $\hat{\theta}$ satisfying

$$\mathbb{P}_n f_{n,\hat{\theta}} \geq \sup_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta} - o_p((nh_n)^{-2/3}).$$

Similar to the previous section, we assume consistency of $\hat{\theta}$ to θ_0 and focus on the convergence rate and limiting distribution. To derive the convergence rate of $\hat{\theta}$, we need to establish tightness of the centered empirical process $\{\mathbb{G}_n(f_{n,\theta} - f_{n,\theta_0}) : \theta \in \Theta\}$ for the drifting cube root class defined above. We show the following maximal inequality.

Lemma Mn. *There exists positive constant* C *and* C' *such that*

$$P \sup_{|\theta-\theta_0|<\delta} |\mathbb{G}_n(f_{n,\theta} - f_{n,\theta_0})| \leq C\delta^{1/2},$$

for all n *large enough and* $\delta \in [(nh_n)^{-1/2}, C']$.

Proof. The proof is similar to that of Lemma M. Pick any $C' > 0$ and then pick any n and $\delta \in [(nh_n)^{-1/2}, C']$. Hereafter positive constants C_j $(j = 1, 2, \dots)$ are independent of n and δ. By changing the notation to indicate the drifting classes of functions, we can reach the following bound

$$P \sup_{g_n \in \mathcal{G}^{\beta}_{n,\delta}} |\mathbb{G}_n g_n| \leq C_1[1 + \delta^{-1} q_{G_n}(\min\{1, v_n(\delta)\})]\varphi_n(\delta),$$

where $\mathcal{G}^{\beta}_{n,\delta} = \{f_{n,\theta} - f_{n,\theta'} : \|f_{n,\theta} - f_{n,\theta'}\|_{2,\beta} < \delta \text{ for } \theta, \theta' \in \Theta\}$ with an envelope function G_n, and $\varphi_n(\delta) = \int_0^{\delta} \sqrt{\log N_{[]}(\nu, \mathcal{G}^{\beta}_{n,\delta}, \|\cdot\|_{2,\beta})} d\nu$. By the condition (iii) of the drifting cube root class, we can conclude $\varphi_n(\delta) \leq C_2 \delta$, which in turn implies $v_n(\delta) \leq C_3 n^{-1}$ as in the proof of Lemma M. Therefore, the conclusion follows by

$$\delta^{-1} q_{G_n}(\min\{1, v_n(\delta)\}) \leq C_4 \delta^{-1} h_n^{-1/2} n^{-1/2},$$

for all n large enough. □

Lemma 1 for the cube root class can be modified as follows. Since the proof is similar, it is omitted.

Lemma 3. *For each $\varepsilon > 0$, there exist random variables $\{R_n\}$ of order $O_p(1)$ and a positive constant C such that*

$$|\mathbb{P}_n(f_{n,\theta} - f_{n,\theta_0}) - P(f_{n,\theta} - f_{n,\theta_0})| \leq \varepsilon|\theta - \theta_0|^2 + (nh_n)^{-2/3} R_n^2,$$

for any $(nh_n)^{-1/3} \leq |\theta - \theta_0| \leq C$.

Based on this lemma, the cube root convergence rate of $\hat{\theta}$ is obtained as follows. For $|\hat{\theta} - \theta_0| \geq (nh_n)^{-1/3}$, we can take $c > 0$ such that

$$\begin{aligned} 0 &\leq \mathbb{P}_n(f_{n,\hat{\theta}} - f_{n,\theta_0}) \leq P(f_{n,\hat{\theta}} - f_{n,\theta_0}) + \varepsilon|\hat{\theta} - \theta_0|^2 + (nh_n)^{-2/3} R_n^2 \\ &\leq (-c + \varepsilon)|\hat{\theta} - \theta_0| + o(|\hat{\theta} - \theta_0|) + O_p((nh_n)^{-2/3}), \end{aligned}$$

for each $\varepsilon > 0$, where the second inequality follows from Lemma 3 and the third inequality follows from the condition (i) of the cube root class. Taking ε small enough to satisfy $c - \varepsilon > 0$ yields contradiction and thus we obtain $\hat{\theta} - \theta_0 = O_p((nh_n)^{-1/3})$.

In order to derive the limiting distribution, we need to establish tightness of the centered process $Z_n(s) = (nh_n)^{1/6} \mathbb{G}_n(f_{n,\theta_0 + s(nh_n)^{-1/3}} - f_{n,\theta_0})$ for $|s| \leq K$ with any $K > 0$. The finite dimensional convergence of Z_n follows from Lemma C in the previous section by setting $g_n = (nh_n)^{1/6}(f_{n,\theta_0 + s(nh_n)^{-1/3}} - f_{n,\theta_0})$. For the maximal inequality, note that the process Z_n itself does not satisfy the condition (ii) in the drifting cube root class. On the other hand, we do not need a sharp characterization of the upper bound for the maximal deviations of the process and the following maximal inequality suffices.

Lemma Mn'. *For any $\epsilon > 0$, there exist $\delta > 0$ and a positive integer n_δ such that*

$$P \sup_{|s - s'| < \delta} |Z_n(s) - Z_n(s')| \leq \epsilon,$$

for all $n \geq n_\delta$.

Based on this inequality, we can apply the argmax theorem (Kim and Pollard, 1990, Theorem 2.7) to derive the limiting distribution of the M-estimator $\hat{\theta}$ for the criterion function drifting with n. The main theorem of this section is summarized as follows.

Theorem 3. *Let $\{f_{n,\theta} : \theta \in \Theta\}$ be a drifting cube root class and $\hat{\theta}$ satisfy $\mathbb{P}_n f_{n,\hat{\theta}} \geq \sup_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta} - o_p((nh_n)^{-2/3})$. Assume $\hat{\theta}$ converges in probability to $\theta_0 \in \mathrm{int}\Theta$, and $\sup_n \int_0^1 \beta^{-1}(u)Q_{g_{n,s}}(u)^2 du < \infty$ for each s, where $g_{n,s} = (nh_n)^{1/6}(f_{n,\theta_0+s(nh_n)^{-1/3}} - f_{n,\theta_0})$. Then*

$$(nh_n)^{1/3}(\hat{\theta} - \theta_0) \xrightarrow{d} \arg\max_s Z(s),$$

where $Z(s)$ is a Gaussian process with continuous sample paths, expected value $-s'Vs/2$, and covariance kernel $H(s_1, s_2) = \lim_{n\to\infty} \sum_{t=-n}^{n} Pg_{n,s_1}(z_0)g_{n,s_2}(z_t)$.

This theorem extends the main theorem of Kim and Pollard to the case where the criterion function contains a bandwidth parameter. In this case, the convergence rate $O_p((nh_n)^{-1/3})$ is slower than the conventional $O_p(n^{-1/3})$ rate. This theorem can be extended to the case where the criterion function contains estimated nuisance parameters that converge faster than the $O_p((nh_n)^{-1/3})$ rate. Let $\hat{\theta}$ and $\tilde{\theta}$ satisfy $\mathbb{P}_n f_{n,\hat{\theta},\hat{\nu}} \geq \sup_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta,\hat{\nu}} + o_p((nh_n)^{-2/3})$, where $\hat{\nu} - \nu_0 = o_p((nh_n)^{-2/3})$, and $\mathbb{P}_n f_{n,\tilde{\theta},\nu_0} \geq \sup_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta,\nu_0} + o_p((nh_n)^{-2/3})$, respectively.

Theorem 4. *Let $\{f_{n,\theta,\nu} : \theta \in \Theta, \nu \in \Lambda\}$ be a drifting cube root class, where the condition (i) is replaced with*

(i)': *The class is bounded and for some negative definite matrix V_1 and some finite matrix V_2,*

$$P(f_{n,\theta,\nu} - f_{n,\theta_0,\nu_0}) = \frac{1}{2}(\theta - \theta_0)'V_1(\theta - \theta_0) + (\theta - \theta_0)'V_2(\nu - \nu_0) + o(|\theta - \theta_0|^2 + |\nu - \nu_0|^2).$$

Then $\hat{\theta} = \tilde{\theta} + o_p((nh_n)^{-1/3})$. Furthermore, if $\sup_n \int_0^1 \beta^{-1}(u)Q_{g_{n,s}}(u)^2 du < \infty$ for each s with $g_{n,s} = (nh_n)^{1/6}(f_{n,\theta_0+s(nh_n)^{-1/3},\nu_0} - f_{n,\theta_0,\nu_0})$, then $n^{1/3}(\hat{\theta} - \theta_0) \xrightarrow{d} \arg\max_s Z(s)$, where $Z(s)$ is a Gaussian process with continuous sample paths, expected value $s'V_1 s/2$ and covariance kernel H.

4. Applications

4.1. **Maximum score estimator.** As an application of Theorem 1, consider the maximum score estimator for the regression model $y_t = x_t'\theta_0 + u_t$, that is

$$\hat{\theta} = \arg\max_{\theta\in S} \sum_{t=1}^{n} [\mathbb{I}\{y_t \geq 0, x_t'\theta \geq 0\} + \mathbb{I}\{y_t < 0, x_t'\theta < 0\}],$$

where S is the surface of the unit sphere in $\mathbb{R}^d$. Since $\hat{\theta}$ is determined only up to scalar multiples, we standardize it to be unit length. We impose the following assumptions. Let $h(x,u) = \mathbb{I}\{x'\theta_0 + u \geq 0\} - \mathbb{I}\{x'\theta_0 + u < 0\}$.

- **(a):** $\{x_t, u_t\}$ satisfies Assumption D. x_t has compact support and a continuously differentiable density $p(\cdot)$, the angular component of x_t, considered as a random variable on S, has a bounded and continuous density, and the density for the orthogonal angle to θ_0 is bounded away from zero.

(b): Assume that $|\theta_0| = 1$, median$(u_t|x_t) = 0$, the function $\kappa(x) = E[h(x_t, u_t)|x_t = x]$ is non-negative for $x'\theta_0 \geq 0$ and non-positive for $x'\theta_0 < 0$ and is continuously differentiable, and $P\{x_t'\theta_0 = 0, \dot{\kappa}(x_t)'\theta_0 p(x_t) > 0\} > 0$.

We can write as $\hat{\theta} = \arg\max_{\theta\in S} \mathbb{P}_n f_\theta$ and $\theta_0 = \arg\max_{\theta\in S} Pf_\theta$, where

$$f_\theta(x, u) = h(x, u)[\mathbb{I}\{x'\theta \geq 0\} - \mathbb{I}\{x'\theta_0 \geq 0\}].$$

Existence and uniqueness of θ_0 are guaranteed by (b) (Manski, 1985). Also the uniform law of large numbers for an absolutely regular process by Nobel and Dembo (1993, Theorem 1) implies $\sup_{\theta\in S} |\mathbb{P}_n f_\theta - Pf_\theta| \xrightarrow{p} 0$. Therefore, $\hat{\theta}$ is consistent for θ_0.

We next compute V and H in Theorem 1. Under strict stationarity, we can apply the same argument to Kim and Pollard (1990, pp. 214-215) to derive the second derivative

$$V = \left.\frac{\partial^2 Pf_\theta}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0} = -\int \mathbb{I}\{x'\theta_0 = 0\}\dot{\kappa}(x)'\theta_0 p(x)xx'd\sigma,$$

where σ is the surface measure on the boundary of $\{x : x'\theta_0 \geq 0\}$. This matrix is negative definite under the last condition of (b). Pick any s_1 and s_2, and define $g_{n,t} = f_{\theta_0+n^{-1/3}s_1}(x_t, u_t) - f_{\theta_0+n^{-1/3}s_2}(x_t, u_t)$. The covariance kernel characterized by $H(s_1, s_2) = \frac{1}{2}\{L(s_1, 0) + L(0, s_2) - L(s_1, s_2)\}$, where

$$L(s_1, s_2) = \lim_{n\to\infty} n^{4/3}\text{Var}(\mathbb{P}_n g_{n,t}) = \lim_{n\to\infty} n^{1/3}\{\text{Var}(g_{n,t}) + \sum_{m=1}^{\infty} \text{Cov}(g_{n,t}, g_{n,t+m})\}.$$

The limit $n^{1/3}\text{Var}(g_{n,t})$ is given in Kim and Pollard (1990, p. 215). Assumption D implies

$$|P\{g_{n,t} = j, g_{n,t+m} = k\} - P\{g_{n,t} = j\}P\{g_{n,t+m} = k\}| \leq n^{-2/3}\beta_m,$$

for all $n, m \geq 1$ and $j, k = -1, 0, 1$. Thus, $\{g_{n,t}\}$ is an α-mixing array whose mixing coefficients are bounded by $2n^{-2/3}\beta_m$. By applying the α-mixing inequality,

$$\text{Cov}(g_{n,t}, g_{n,t+m}) \leq Cn^{-2/3}\beta_m \|g_{n,t}\|_p^2,$$

for some $p > 2$ and $C > 0$. Note that

$$\|g_{n,t}\|_p^2 \leq [P|\mathbb{I}\{x_t'(\theta_0 + s_1 n^{-1/3}) > 0\} - \mathbb{I}\{x_t'(\theta_0 + s_2 n^{-1/3}) > 0\}|]^{2/p} = O(n^{-2/(3p)}).$$

Combining these results, $n^{1/3}\sum_{m=1}^{\infty} \text{Cov}(g_{n,t}, g_{n,t+m}) \to 0$ as $n \to \infty$. Therefore, the covariance kernel H is same as the independent case in Kim and Pollard (1990, p. 215).

We now verify that $\{f_\theta : \theta \in S\}$ belongs to the cube root class. The first requirement is already verified. By Jensen's inequality,

$$\|f_{\theta_1} - f_{\theta_2}\|_2 = \sqrt{P|\mathbb{I}\{x_t'\theta_1 \geq 0\} - \mathbb{I}\{x_t'\theta_2 \geq 0\}|} \geq P\{x_t'\theta_1 \geq 0 > x_t'\theta_2\} + P\{x_t'\theta_2 \geq 0 > x_t'\theta_1\},$$

for any $\theta_1, \theta_2 \in S$. Since the right hand side is the probability for a pair of wedge shaped regions with an angle of order $|\theta_1 - \theta_2|$ and the density for the orthogonal angle to θ_0 is bounded away from zero by (b), the second requirement is satisfied. For the third requirement, pick any $\varepsilon > 0$

and $\bar{\theta} \in \Theta$, and note that

$$P \sup_{\theta \in \Theta : |\theta - \bar{\theta}| < \varepsilon} |f_\theta - f_{\bar{\theta}}|^2 = P \sup_{\theta \in \Theta : |\theta - \bar{\theta}| < \varepsilon} \mathbb{I}\{x_t'\theta \geq 0 > x_t'\bar{\theta} \text{ or } x_t'\bar{\theta} \geq 0 > x_t'\theta\}.$$

Again, the right hand side is the probability for a pair of wedge shaped regions with an angle of order ε. Therefore, $\{f_\theta : \theta \in S\}$ is in the cube root class, and we can conclude that even if the data obey a dependence process specified in Assumption D, the maximum score estimator possesses the same limiting distribution as the independent sampling case.

4.2. **Nonparametric estimation under order restrictions.** Auxiliary results to show Theorem 1 may be applied to establish weak convergence of processes. As an example, consider estimation of a decreasing marginal density function of z_t with support $[0, \infty)$. We impose Assumption D for $\{z_t\}$. The nonparametric maximum likelihood estimator $\hat{f}(c)$ of the density $f(c)$ at a fixed $c > 0$ is given by the left derivative of the concave majorant of the empirical distribution function $\hat{F}$. It is known that $n^{1/3}(\hat{f}(c) - f(c))$ can be written as the left derivative of the concave majorant of the process $Z_n(s) = n^{2/3}\{\hat{F}(c + sn^{-1/3}) - \hat{F}(c) - f(c)sn^{-1/3}\}$. Let $f_\theta(z) = \mathbb{I}\{z \leq c + \theta\}$ and F be the distribution function of f. Decompose

$$Z_n(s) = n^{1/6}\mathbb{G}_n(f_{sn^{-1/3}} - f_0) + n^{2/3}\{F(c + sn^{-1/3}) - F(c) - f(c)sn^{-1/3}\}.$$

A Taylor expansion implies that convergence of the second term to $\frac{1}{2}\dot{f}(c)s^2$. We establish weak convergence of $W_n(s) = n^{1/6}\mathbb{G}_n(f_{sn^{-1/3}} - f_0)$. Lemma C (setting $w_{tn} = n^{1/6}\{(f_{sn^{-1/3}}(z_t) - f_0(z_t)) - P(f_{sn^{-1/3}} - f_0)\}$) implies finite dimensional convergence of $W_n(s)$ to projections of a centered Gaussian with the covariance kernel

$$H(s_1, s_2) = \lim_{n \to \infty} n^{1/3} \sum_{t=-n}^{n} \{F_{0t}(c + s_1 n^{-1/3}, c + s_2 n^{-1/3}) - F(c + s_1 n^{-1/3})F(c + s_2 n^{-1/3})\},$$

where F_{0t} is the joint distribution function of (z_0, z_t). For tightness of $W_n(s)$, it is enough to show that $\{f_\theta : \theta \in \mathbb{R}\}$ fulfills the conditions (ii) and (iii) of the the cube root class so that the maximal inequality in Lemma M can be applied.

Note that

$$\|f_{\theta_1} - f_{\theta_2}\|_2^2 = |F(c + \theta_1) - F(c + \theta_2)| = f(c + \tilde{\theta})|\theta_1 - \theta_2|,$$

for some $\tilde{\theta}$ between θ_1 and θ_2. Thus the condition (ii) is satisfied by choosing C' small enough. Also, for any $\bar{\theta} \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned} P \sup_{\theta : |\theta - \bar{\theta}| < \varepsilon} |f_\theta - f_{\bar{\theta}}|^2 &= P \sup_{\theta : |\theta - \bar{\theta}| < \varepsilon} |\mathbb{I}\{x_t \leq c + \theta\} - \mathbb{I}\{x_t \leq c + \bar{\theta}\}| \\ &\leq \max\{F(c + \bar{\theta}) - F(c + \bar{\theta}) - \varepsilon, F(c + \bar{\theta}) + \varepsilon - F(c + \bar{\theta})\} \\ &\leq f(0)\varepsilon, \end{aligned}$$

i.e., the condition (iii) is verified. Therefore, $Z_n(s)$ weakly converges to $Z(s)$, a Gaussian process with expected value $\frac{1}{2}\dot{f}(c)s^2$ and covariance kernel H.

For the remaining part, we can apply the same argument to Kim and Pollard (1990, pp. 216-218) (by replacing their Lemma 4.1 with 1), and conclude that $n^{1/3}(\hat{f}(c) - f(c))$ converges in distribution to the derivative of the concave majorant of Z evaluated at 0.

4.3. **Least median of squares.** As an application of Theorem 2 (i), consider the least median of squares estimator for the regression model $y_t = x_t'\beta_0 + u_t$, that is

$$\hat{\beta} = \arg\min_{\theta} \operatorname{median}\{(y_1 - x_1'\beta)^2, \ldots, (y_n - x_n'\beta)^2\}.$$

We impose the following assumptions.

(a): $\{x_t, u_t\}$ satisfies Assumption D. x_t and u_t are independent. $P|x_t|^2 < \infty$, $Px_t x_t'$ is positive definite, and the distribution of x_t puts zero mass on each hyperplane.

(b): The density $\gamma(\cdot)$ of u_t is bounded, differentiable, and symmetric around zero, and decreases away from zero. $\gamma(0)$ is bounded away from zero. $|u_t|$ has the unique median ν_0 and $\dot{\gamma}(\nu_0) < 0$.

Note that $\hat{\theta} = \hat{\beta} - \beta_0$ is written as $\hat{\theta} = \arg\max_\theta \mathbb{P}_n f_{\theta,\hat{\nu}}$, where

$$f_{\theta,\nu}(x, u) = \mathbb{I}\{x'\theta - \nu \le u \le x'\theta + \nu\},$$

and $\hat{\nu} = \inf\{\nu : \sup_\theta \mathbb{P}_n f_{\theta,\nu} \ge \frac{1}{2}\}$. Let $\nu_0 = 1$ to simplify the notation. Since $\{f_{\theta,\nu} : \theta \in \mathbb{R}^d, \nu \in \mathbb{R}\}$ is a VC subgraph class, Arcones and Yu (1994, Theorem 1) implies the uniform convergence $\sup_{\theta,\nu} |\mathbb{P}_n f_{\theta,\nu} - P f_{\theta,\nu}| = O_p(n^{-1/2})$. Thus, the same argument to Kim and Pollard (1990, p. 212) yields the convergence rates $\hat{\nu} - 1 = O_p(n^{-1/2})$ and $\hat{\theta}_n = O_p(n^{-1/3})$ (by applying Lemma 1).

We next compute V and H in Corollary 1 (i). Observe that

$$V = \left.\frac{\partial^2 P\{x_t'\theta - 1 \le u_t \le x_t'\theta + 1\}}{\partial\theta\partial\theta'}\right|_{\theta=0} = 2\dot{\gamma}(1) P x_t x_t',$$

which is negative definite by assumptions. Pick any s_1 and s_2. The covariance kernel is written as $H(s_1, s_2) = \frac{1}{2}\{L(s_1, 0) + L(0, s_2) - L(s_1, s_2)\}$, where $L(s_1, s_2) = \lim_{n\to\infty} n^{4/3}\mathrm{Var}(\mathbb{P}_n g_{n,t})$ and $g_{n,t} = \mathbb{I}\{|x_t's_1 n^{-1/3} - u_t| \le 1\} - \mathbb{I}\{|x_t's_2 n^{-1/3} - u_t| \le 1\}$. By a similar argument to the maximum score example in Section 4.1, we can show that H is same as the one for the independent case derived in Kim and Pollard (1990).

We now verify that $\{f_{\theta,1} : \theta \in \mathbb{R}^d\}$ belongs to the cube root class. By (b), $Pf_{\theta,1}$ is uniquely maximized at $\theta_0 = 0$. So the first requirement is satisfied. Let $p(\cdot)$ and $\Gamma(\cdot)$ be the marginal density of x_t and the distribution function of u_t, respectively. Some expansions (using symmetry of $\gamma(\cdot)$) yield

$$\begin{aligned}
\|f_{\theta_1,1} - f_{\theta_2,1}\|_2^2 &= P\{x_t'\theta_1 - 1 \le u_t \le x_t'\theta_1 + 1\} + P\{x_t'\theta_2 - 1 \le u_t \le x_t'\theta_2 + 1\} \\
&\quad -2P\{\{x_t'\theta_1 - 1 \le u_t \le x_t'\theta_1 + 1\} \cap \{x_t'\theta_2 - 1 \le u_t \le x_t'\theta_2 + 1\}\} \\
&= P|\Gamma(x_t'\theta_1 + 1) - \Gamma(x_t'\theta_2 + 1) + \Gamma(x_t'\theta_1 - 1) - \Gamma(x_t'\theta_2 - 1)| \\
&\ge \frac{1}{2}(\theta_2 - \theta_1)'[P\dot{\gamma}(x_t'\bar{\theta}_2 - 1)x_t x_t'](\theta_2 - \theta_1).
\end{aligned}$$

Thus, the second condition can be verified under (b). For the third requirement, pick any $\varepsilon > 0$ and $\bar{\theta} \in \Theta$, and note that

$$\begin{aligned} P \sup_{\theta:|\theta-\bar{\theta}|<\varepsilon} |f_{\theta,1} - f_{\bar{\theta},1}|^2 &= P \sup_{\theta:|\theta-\bar{\theta}|<\varepsilon} |\mathbb{I}\{x_t'\theta - 1 \leq u_t \leq x_t'\theta + 1\} - \mathbb{I}\{x_t'\bar{\theta} - 1 \leq u_t \leq x_t'\bar{\theta} + 1\}| \\ &\leq P \sup_{\theta:|\theta-\bar{\theta}|<\varepsilon} \{|\Gamma(x_t'\theta + 1) - \Gamma(x_t'\bar{\theta} + 1)| + |\Gamma(x_t'\theta - 1) - \Gamma(x_t'\bar{\theta} - 1)|\}. \end{aligned}$$

By expansions around $\theta = \bar{\theta}$ and boundedness of $\gamma(\cdot)$, the third requirement is also verified. Therefore, $\{f_{\theta,1} : \theta \in \Theta\}$ is in the cube root class, and we can conclude that $n^{1/3}(\hat{\beta}-\beta_0)$ converges in distribution to the argmax of $Z(s)$, a Gaussian process with expected value $-\dot{\gamma}(1)s'Px_tx_t's$ and the covariance kernel H derived above.

4.4. **Minimum volume predictive region.** As an illustration of Corollary 4 (ii), consider a minimum volume predictor for a strictly stationary process proposed by Polonik and Yao (2000). Suppose we are interested in predicting $y \in \mathbb{R}$ from $x \in \mathbb{R}$ based on the observations $\{y_t, x_t\}$. The minimum volume predictor of y at $x = c$ in the class $\mathcal{I}$ of intervals of $\mathbb{R}$ at level $\alpha \in [0, 1]$ is defined as

$$\hat{I} = \arg\min_{S\in\mathcal{I}} \mu(S) \quad \text{s.t. } \hat{P}(S) \geq \alpha,$$

where μ is the Lebesgue measure and $\hat{P}(S) = \sum_{t=1}^n \mathbb{I}\{y_t \in S\}K\left(\frac{x_t-c}{h_n}\right) / \sum_{t=1}^n K\left(\frac{x_t-c}{h_n}\right)$ is the kernel estimator of the conditional probability $P\{y_t \in S|x_t = c\}$. Since $\hat{I}$ is an interval, it can be written as $\hat{I} = [\hat{\theta} - \hat{r}, \hat{\theta} + \hat{r}]$, where

$$\hat{\theta} = \arg\min_{\theta} \hat{P}([\theta - \hat{r}, \theta + \hat{r}]), \qquad \hat{r} = \inf\{r : \sup_{\theta} \hat{P}([\theta - r, \theta + r]) \geq \alpha\}.$$

To study the asymptotic property of $\hat{I}$, we impose the following assumptions.

- **(a):** $\{y_t, x_t\}$ satisfies Assumption D. $I_0 = [\theta_0 - r_0, \theta_0 + r_0]$ is the unique shortest interval such that $P\{y_t \in I_0|x_t = c\} \geq \alpha$, and the conditional density $f_{y|x=c}$ of y_t given $x_t = c$ is bounded, strictly positive at $\theta_0 \pm r_0$, and satisfies $f'_{y|x=c}(\theta_0 - r_0) - f'_{y|x=c}(\theta_0 + r_0) > 0$.
- **(b):** K is bounded and symmetric, and satisfies $\lim_{x\to\infty} |x|K(x) = 0$. As $n \to \infty$, $nh_n \to \infty$ and $nh_n^4 \to 0$.

For notational convenience, assume $\theta_0 = 0$ and $r_0 = 1$. We first derive the convergence rate for $\hat{r}$. Note that $\hat{r} = \inf\{r : \sup_\theta \hat{g}([\theta - r, \theta + r]) \geq \alpha\hat{f}(c)\}$, where $\hat{g}(S) = \frac{1}{nh_n}\sum_{t=1}^n \mathbb{I}\{y_t \in S\}K\left(\frac{x_t-c}{h_n}\right)$ and $\hat{f}(c) = \frac{1}{nh_n}\sum_{t=1}^n K\left(\frac{x_t-c}{h_n}\right)$. By applying Nobel and Dembo (1993, Theorem 1), we can obtain uniform convergence rate

$$\max\left\{|\hat{f}(c) - f(c)|, \sup_{\theta,r} |\hat{g}([\theta - r, \theta + r]) - P\{y_t \in [\theta - r, \theta + r]|x_t = c\}f(c)|\right\} = O_p((nh_n)^{-1/2}+h_n^2).$$

Thus the same argument to Kim and Pollard (1990, pp. 207-208) yields $\hat{r} - 1 = O_p((nh_n)^{-1/2} + h_n^2)$.

We now consider the convergence rate for $\hat{\theta}$, which is written as $\hat{\theta} = \arg\min_\theta \hat{g}([\theta - \hat{r}, \theta + \hat{r}])$. Consistency follows from uniqueness of (θ_0, r_0) in (a) and the uniform convergence

$$\sup_\theta |\hat{g}([\theta - \hat{r}, \theta + \hat{r}]) - P\{y_t \in [\theta - 1, \theta + 1] | x_t = c\} f(c)| \xrightarrow{p} 0,$$

which follows by applying Nobel and Dembo (1993, Theorem 1). Let us reparametrize as $\nu = r - 1$ and define the normalized objective function as $\mathbb{P}_n f_{n,\theta,\nu}$, where $z = (y, x)'$ and

$$f_{n,\theta,\nu}(z) = \frac{1}{h} K\left(\frac{x-c}{h}\right) [\mathbb{I}\{y \in [\theta - 1 - \nu, \theta + 1 + \nu]\} - \mathbb{I}\{y \in [-1 - \nu, 1 + \nu]\}].$$

Note that $\hat{\theta} = \arg\max_\theta \mathbb{P}_n f_{n,\theta,\hat{r}-1}$.

We now apply Theorem 3 to $\{f_{n,\theta,\nu}\}$. By boundedness of K, it holds $\sup_{z,\theta,\nu} |f_{n,\theta,\nu}(z)| = O(h_n^{-1})$. The condition (i') is verified as follows

$$\begin{aligned}
& P(f_{n,\theta,\nu} - f_{n,\theta_0,\nu_0}) \\
= & \int K(a) \left[\begin{array}{c} \{F_{y|x}(\theta + 1 + \nu | c + h_n a) - F_{y|x}(\theta - 1 - \nu | c + h_n a)\} \\ -\{F_{y|x}(1 + \nu | c + h_n a) - F_{y|x}(-1 - \nu | c + h_n a)\} \end{array} \right] da \\
= & \{F_{y|x}(\theta + 1 + \nu | c) - F_{y|x}(\theta - 1 - \nu | c)\} - \{F_{y|x}(1 + \nu | c) - F_{y|x}(-1 - \nu | c)\} + O(h_n^2) \\
= & -\frac{1}{2}\{-f'_{y|x}(1|c) + f'_{y|x}(-1|c)\}\theta^2 + \{f'_{y|x}(1|c) + f'_{y|x}(-1|c)\}\theta\nu + o(\theta^2 + \nu^2) + O(h_n^2),
\end{aligned}$$

where the second and third equalities follow from expansions around $a = 0$ and $(\theta, \nu) = (0, 0)$, respectively.

For the condition (ii), note that

$$\begin{aligned}
& h_n \|f_{n,\theta_1,\nu_1} - f_{n,\theta_2,\nu_2}\|_2^2 \\
= & \int K(a)^2 |F_{y|x}(\theta_2 - 1 - \nu_2 | x = c + ah) - F_{y|x}(\theta_1 - 1 - \nu_1 | x = c + ah)| f_x(c + ah) da \\
& + \int K(a)^2 |F_{y|x}(\theta_2 + 1 + \nu_2 | x = c + ah) - F_{y|x}(\theta_1 + 1 + \nu_1 | x = c + ah)| f_x(c + ah) da \\
= & T_1 + T_2.
\end{aligned}$$

Suppose $\theta_2 - \nu_2 > \theta_1 - \nu_1$. For T_1,

$$\begin{aligned}
T_1 = & \int K(a)^2 da [f_{y|x}(-1|c)\{(\theta_2 - \theta_1) - (\nu_2 - \nu_1)\} + \frac{1}{2} f'_{y|x}(-1|c)\{(\theta_2 - \theta_1) - (\nu_2 - \nu_1)\}^2] f_X(c) \\
& + o(\{(\theta_2 - \theta_1) - (\nu_2 - \nu_1)\}^2) + O(h) \\
\geq & \frac{1}{2} \int K(a)^2 da f'_{y|x}(-1|c)\{(\theta_2 - \theta_1) - (\nu_2 - \nu_1)\}^2.
\end{aligned}$$

For T_2,

$$\begin{aligned}
T_2 = & \int K(a)^2 da \{F_{y|x}(\theta_2 + 1 + \nu_2 | c) - F_{y|x}(\theta_1 + 1 + \nu_1 | c)\} f_X(c) + O(h) \\
= & \int K(a)^2 da f_{y|x}(1|c) |(\theta_2 - \theta_1) - (\nu_2 - \nu_1)| f_X(c) + o(\{(\theta_2 - \theta_1) - (\nu_2 - \nu_1)\}^2).
\end{aligned}$$

Therefore, the condition (ii) holds. The condition (iii) follows from the law of iterated expectation and Taylor expansions. Therefore, we obtain $\hat{r} - r_0 = O_p((nh_n)^{-1/2} + h_n^2)$ and $\hat{\theta} - \theta_0 = O_p((nh_n)^{-1/3} + h_n)$.

References

[1] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and J. W. Tukey (1972) Robust Estimates of Location, Princeton University Press, Princeton.

[2] Andrews, D. W. K. (1993) An introduction to econometric applications of empirical process theory for dependent random variables, *Econometric Reviews*, 12, 183-216.

[3] Anevski, D. and O. Hössjer (2006) A general asymptotic scheme for inference under order restrictions, *Annals of Statistics*, 34, 1874-1930.

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

[5] Banerjee, M. and I. W. McKeague (2007) Confidence sets for split points in decision trees, *Annals of Statistics*, 35, 543-574.

[6] Bühlmann, P. and B. Yu (2002) Analyzing bagging, *Annals of Statistics*, 30, 927-961.

[7] Chernoff, H. (1964) Estimation of the mode, *Annals of the Institute of Statistical Mathematics*, 16, 31-41.

[8] de Jong, R. M. and T. Woutersen (2011) Dynamic time series binary choice, *Econometric Theory*, 27, 673-702.

[9] Doukhan, P., Massart, P. and E. Rio (1994) The functional central limit theorems for strongly mixing processes, *Annales de l'Institut Henri Poincaré, Probability and Statistics*, 30, 63-82.

[10] Doukhan, P., Massart, P. and E. Rio (1995) Invariance principles for absolutely regular empirical processes, *Annales de l'Institut Henri Poincaré, Probability and Statistics*, 31, 393-427.

[11] Goldenshluger, A. and A. Zeevi (2004) The Hough transform estimator, *Annals of Statistics*, 32, 1908-1932.

[12] Honoré, B. E. and Kyriazidou, E. (2000) Panel data discrete choice models with lagged dependent variables, *Econometrica*, 68, 839-874.

[13] Kim, J. and D. Pollard (1990) Cube root asymptotics, *Annals of Statistics*, 18 , 191-219.

[14] Manski, C. F. (1975) Maximum score estimation of the stochastic utility model of choice, *Journal of Econometrics*, 3, 205-228.

[15] Manski, C. F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, *Journal of Econometrics*, 27, 313-333.

[16] Nobel, A. and A. Dembo (1993) A note on uniform laws of averages for dependent processes, *Statistics and Probability Letters*, 17, 169-172.

[17] Prakasa Rao, B. L. S. (1969) Estimation of a unimodal density, *Sankhyā*, A, 31, 23-36.

[18] Polonik, W. and Q. Yao (2000) Conditional minimum volume predictive regions for stochastic processes, *Journal of the American Statistical Association*, 95, 509-519.

[19] Rio, E. (1997) About the Lindeberg method for strongly mixing sequences, *ESAIM: Probability and Statistics*, 1, 35-61.

[20] Rousseeuw, P. J. (1984) Least median of squares regression, *Journal of the American Statistical Association*, 79, 871-880.

[21] van der Vaart, A. W. and J. A. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer, New York.

[22] Zinde-Walsh, V. (2002) Asymptotic theory for some high breakdown point estimators, *Econometric Theory*, 18, 1172-1196.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `m.seo@lse.ac.uk`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `t.otsu@lse.ac.uk`