EXTENDING THE SCOPE OF CUBE ROOT ASYMPTOTICS

MYUNG HWAN SEO AND TAISUKE OTSU

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 size typically due to a bandwidth sequence. For dependent empirical processes that characterize criterions inducing cube root phenomena, maximal inequalities are established to derive the convergence rates and limit laws of the M-estimators. The limit theory is applied not only to extend existing examples, such as the 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, conditional maximum score estimator for a panel data discrete choice model, and Hough transform estimator with a drifting tuning parameter.

1. Introduction

There is a class of estimation problems where point estimators converge at the cube root rate to some 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 literature suggests these cube root phenomena commonly arise when the criterion functions for point estimation are not continuous in parameters.

A seminal work by Kim and Pollard (1990) explained elegantly these cube root phenomena in a unified framework by means of empirical process theory; they 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 (Prakasa 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 contexts in statistics, such as the Hough transform estimator in image analysis (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 are posed suggesting emergence of the cube root asymptotics but being outside the scope of Kim and Pollard's (1990) framework. Most problems appeared in the course of generalizations of the existing examples listed above. As a prototype, let us consider construction of a minimum volume predictive region, studied by Polonik and Yao (2000), in a simplified manner. A statistician who

The authors would like to thank Javier Hidalgo, Benedikt Pötscher, Peter Robinson, and seminar participants at Cambridge, CREATES in Aarhus University, IHS in Vienna, LSE, and University of Surrey and Vienna for helpful comments.

observes a bivariate dependent process $\{y_t, x_t\}$ wishes to predict y from x by some interval. In this simple case, Polonik and Yao's (2000) minimum volume predictive region of y at $x = c$ with level α may be written as the interval $[\hat{\theta} - \hat{r}, \hat{\theta} + \hat{r}]$, where

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

and $\hat{P}[a, b] = \sum_{t=1}^{n} \mathbb{I}\{a \leq y_t \leq b\} K\left(\frac{x_t - c}{h_n}\right) / \sum_{t=1}^{n} K\left(\frac{x_t - c}{h_n}\right)$ is a nonparametric estimator of the conditional probability $P\{a \leq y_t \leq b | x_t = c\}$. K is a kernel function and h_n is a bandwidth varying with the sample size n. This predictive region is a natural generalization of the shorth to the conditional distribution of dependent observables. Polonik and Yao (2000, Remark 3b) conjectured that this region would converge at the $(nh_n)^{-1/3}$ rate under certain norm. 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 sample size due to the bandwidth. To allow dependent observations, the empirical process theory of Kim and Pollard (1990) for independent observations needs to be adapted. In particular, maximal inequalities to derive a convergence rate and to establish weak convergence of the criterion process need to be modified. Also, to allow drifting criterions, the class of criterion functions in consideration for M-estimation needs to be reformulated.

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 analysis includes monotone density estimation as a special case. Goldenshluger and Zeevi (2004, p. 1916) emphasized importance of generalization of the Hough transform estimator to the case of drifting tuning constants and left it for future research. Honoré and Kyriazidou (2000) proposed the conditional maximum score estimator for a panel data discrete choice model. Although they showed the consistency, the convergence rate and limiting distribution are unknown. Also, 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).

To address these open questions, we extend the scope of cube root asymptotics for M-estimators in two directions: allow weakly dependent observations and criterion functions drifting with the sample size typically due to a bandwidth sequence. In particular, we consider an absolutely regular dependent process characterized by β-mixing coefficients and study M-estimation for a class of criterion functions, named the cube root class, which induces the cube root asymptotics. In this setup, we establish maximal inequalities to derive the cube root rate and weak convergence of the normalized process of the criterion so that a continuous mapping theorem for maximizing values of the criterions delivers limit laws of the M-estimators. Furthermore we extend the cube root class to deal with criterions drifting with the sample size, named the drifting cube root class. The limit theory for the cube root class is adapted to the drifting class. We establish the $(nh_n)^{1/3}$-rate of convergence of the M-estimator, where h_n means a bandwidth sequence, and derive a non-normal limiting distribution. The limit theory is also extended to the M-estimation problems where the criterion functions contain estimated nuisance parameters. Our framework is general enough not only to address the open questions listed above but also to extend existing results to more general setups, such as split point estimation in dynamic decision trees.

The paper is organized as follows. Section 2 develops the cube root asymptotic theory for a class of M-estimators with dependent data. It also considers the case where the criterion contains estimated nuisance parameters. Section 3 extends the asymptotic theory to drifting criterion functions. In Section 4, we illustrate our cube root asymptotic theory by some examples; the maximum score estimator (Section 4.1), nonparametric monotone density estimation (Section 4.2), least median of squares (Section 4.3), conditional maximum score estimator for panel data (Section 4.4), minimum volume predictive region (Section 4.5), and Hough transform estimator (Section 4.6).

2. Cube root asymptotics with dependent observations

This section extends Kim and Pollard's (1990) main theorem on the cube root asymptotics of the M-estimator to allow for dependent data. This section focuses on the case where the criterion function does not vary with the sample size. 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 class 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 induce cube root phenomena (or "sharp edge effects" in the sense of Kim and Pollard, 1990) and is general enough to cover the examples discussed in the introduction. Let $Pf = \int f dP$ for a function f, $|\cdot|$ be the Euclidean norm of a vector, and $\|\cdot\|_2$ be the $L_2(P)$-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*

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

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

$$P \sup_{\theta \in \Theta : |\theta - \theta_0| < \varepsilon} |f_\theta - f_{\theta_0}|^2 \leq C'' \varepsilon,$$

  for all $\varepsilon > 0$ small enough.

Condition (i) contains standard identification conditions for M-estimation (cf. Kim and Pollard, 1990, Conditions (ii) and (iv) of their main theorem). Boundedness of the class $\{f_\theta : \theta \in \Theta\}$ is a major requirement. Kim and Pollard (1990) does not impose boundedness even though all of their examples consider bounded criterions. In our analysis, boundedness is required to establish a maximal inequality for the cube root convergence rate (Lemma M below). In particular, boundedness is used to guarantee the relation $\|f_\theta - f_{\theta_0}\|_2 \sim \|f_\theta - f_{\theta_0}\|_{2,\beta}$, where $\|\cdot\|_{2,\beta}$ is so-called the $L_{2,\beta}(P)$-norm using β-mixing coefficients defined below. It should be noted that $\|f_\theta - f_{\theta_0}\|_2 = \|f_\theta - f_{\theta_0}\|_{2,\beta}$ for independent observations. We provide a detailed discussion on boundedness after Lemma M. Condition (ii) is required not only for the maximal inequality to derive the convergence rate but also for finite dimensional convergence to derive the limiting distribution of the M-estimator. In particular, this condition is used to relate the $L_2(P)$-norm to the Euclidean norm over Θ. This condition is implicit in Kim and Pollard (1990, Condition (v)) and the equivalence $\|f_\theta - f_{\theta_0}\|_2 = \|f_\theta - f_{\theta_0}\|_{2,\beta}$ under independent observations. Condition (iii), which corresponds to Kim and Pollard (1990, Condition (vi)), is an envelope condition for the class $\{f_\theta - f_{\theta_0} : |\theta - \theta_0| \leq \varepsilon\}$. Similar to the case of independent observations, this condition plays a key role for the cube root asymptotics. It should be noted that for the familiar squared root asymptotics, the upper bound in Condition (iii) is of order ε^2.

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 empirical criterion $\mathbb{P}_n f_\theta$ to Pf_θ over Θ. Thus, in this section we assume consistency of $\hat{\theta}$. See illustrations in Section 4 for details to verify consistency.

The next step is to derive the convergence rate of $\hat{\theta}$. A key ingredient for this step is to obtain the modulus of continuity of the centered empirical process $\{\mathbb{G}_n(f_\theta - f_{\theta_0}) : \theta \in \Theta\}$ by certain maximum inequality, 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 task is to establish a maximal inequality for the cube root class with dependent observations.

To proceed, we now characterize the dependence structure of data. Among several notions of dependence, this paper focuses on an absolutely regular process. See Doukhan, Massart and Rio (1995) for a detail on empirical process theory of absolutely regular processes. 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. Throughout the paper, we impose the following assumption on 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.[1] For example, finite-order ARMA processes typically satisfy this assumption. 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 for further discussions. Under this assumption, the maximal inequality for the empirical process $\mathbb{G}_n(f_\theta - f_{\theta_0})$ of the cube root class 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 $x \mapsto P\{|g(z_t)| > x\}$.[2] 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 $\beta(\cdot)$. The $L_{2,\beta}(P)$-norm is defined as

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

We use the following sets defined by different norms:

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

For any $g \in \mathcal{G}_\delta^1$, g is bounded (Condition (i)) and so is Q_g. Thus 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 of positive constants $\{a_j\}$. Now take any $C' > 0$, and then pick any n (so that $n^{-1/2} \leq C'$) and $\delta \in [n^{-1/2}, C']$. Throughout the proof, positive constants C_j $(j = 1, 2, \ldots)$ are independent of n and δ.

---

[1]Indeed, the polynomial decay rates of β_m are often associated with strong dependence and long memory type behavior in sample statistics. See, e.g., Chen, Hansen and Carrasco (2010) and references therein. Therefore, asymptotic analysis for the M-estimator will become very different.

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

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

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

where the first inequality is due to Doukhan, Massart and Rio (1995, Lemma 1), 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

$$\|f_\theta - f_{\theta_0}\|_2 \leq \|f_\theta - f_{\theta_0}\|_{2,\beta} \leq M\|f_\theta - f_{\theta_0}\|_2. \qquad (2)$$

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

$$\mathcal{G}_\delta^1 \subset \mathcal{G}_{C_1\delta^{1/2}}^2 \subset \mathcal{G}_{MC_1\delta^{1/2}}^\beta, \qquad \mathcal{G}_\delta^\beta \subset \mathcal{G}_\delta^2 \subset \mathcal{G}_{\delta C_2}^1, \qquad (3)$$

where the relation $\mathcal{G}_\delta^1 \subset \mathcal{G}_{C_1\delta^{1/2}}^2$ follows from Condition (ii) of the cube root class and the relation $\mathcal{G}_\delta^2 \subset \mathcal{G}_{\delta C_2}^1$ follows from Condition (iii).

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 radius $\nu > 0$ and norm $\|\cdot\|$. By (2) and the second relation in (3),

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

for some positive constant C_3, where the second inequality follows from the argument to derive Andrews (1993, eq. (4.7)) based on Condition (iii) of the cube root class (called the L_2-continuity assumption in Andrews, 1993). Therefore, for some positive constant C_4, it holds

$$\varphi_n(\delta) = \int_0^\delta \sqrt{\log N_{[]}(\nu, \mathcal{G}_\delta^\beta, \|\cdot\|_{2,\beta})}d\nu \leq C_4\delta. \qquad (4)$$

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

$$P\sup_{g\in\mathcal{G}_\delta^\beta} |\mathbb{G}_n g| \leq C_5[1 + \delta^{-1}q_{G_\delta}(\min\{1, v_n(\delta)\})]\varphi_n(\delta), \qquad (5)$$

where $q_{G_\delta}(v) = \sup_{u\leq v} Q_G(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ 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

$$\frac{v_n(\delta)^2}{\int_0^{v_n(\delta)} \beta^{-1}(\tilde{u})d\tilde{u}} = \frac{\varphi_n(\delta)^2}{n\delta^2}.$$

Since $\varphi_n(\delta) \leq C_4\delta$ from (4), it holds $v_n(\delta) \leq C_5 n^{-1}$ for some positive constant C_5. Now take some n_0 such that $v_{n_0}(\delta) \leq 1$, and then pick again any $n \geq n_0$ and $\delta \in [n^{-1/2}, C']$. We have

$$q_G(\min\{1, v_n(\delta)\}) \leq C_6\sqrt{v_n(\delta)} Q_G(v_n(\delta)) \leq C_7 n^{-1/2}, \tag{6}$$

for some positive constants C_6 and C_7. Therefore, combining (4)-(6), the conclusion follows by

$$P \sup_{g\in\mathcal{G}_\delta^1} |\mathbb{G}_n g| \leq P \sup_{g\in\mathcal{G}^\beta_{M\delta^{1/2}}} |\mathbb{G}_n g| \leq C_8\delta^{1/2}, \tag{7}$$

where the first inequality follows from the first relation in (3). □

We now discuss the boundedness requirement on f_θ in Condition (i) of the cube root class and exponential decay requirement on the mixing coefficient β_m in Assumption D. Boundedness is used to obtain the second inequality in (1), which guarantees the norm relation in (2). Without boundedness, the $L_{2,\beta}(P)$-norm is bounded from above only by the $L_{2+\eta}(P)$-norm with any $\eta > 0$ (Doukhan, Massart and Rio, 1995, pp. 403-404). Therefore, the resulting maximal inequality will be

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

provided Condition (iii) of the cube root class is replaced with

$$P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} |f_\theta - f_{\theta_0}|^{2+\eta} \leq C''\varepsilon,$$

for some positive constant C'' and all ε small enough. By applying a similar argument below, we can show $\hat{\theta} - \theta_0 = O_p(n^{-\frac{1}{4}-\frac{1}{6(2+\eta)}})$ although this rate may not be sharp.

The exponential decay of the mixing coefficient β_m is also used in (1). Lemma M can be shown under a slightly weaker condition $\sup_{0<x\leq 1} x^{-1}\int_0^x \beta^{-1}(u)du < \infty$ than $\beta_m = O(\rho^m)$ in Assumption D. However, this weaker condition already excludes polynomial decay of β_m. Note that for any β_m with a polynomial decay rate, it holds $\sup_{g\in\mathcal{G}_\delta^1} \|g\|_{2,\beta} = \infty$. For this point, it is intuitive to consider the case where g is a binary function (0 or 1). In this case, we have $\|g\|_{2,\beta} = \|g\|_2 \sqrt{x^{-1}\int_0^x \beta^{-1}(u)du}$ for $x = P\{g(z_t) = 1\}$. Therefore, we have $\sup_{g\in\mathcal{G}_\delta^1} \|g\|_{2,\beta} = \infty$ unless $\sup_{0<x\leq 1} x^{-1}\int_0^x \beta^{-1}(u)du < \infty$. See a remark on Lemma C below for an additional discussion.

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*

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

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

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

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

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

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

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

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. We adapt this approach to dependent observations satisfying Assumption D.

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. Recall that $Q_g(u)$ means the inverse function of the tail probability function $x \mapsto P\{|g(z_t)| > x\}$.

Lemma C. *Suppose* $Pg_n = 0$ *and*

$$\sup_n \int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 du < \infty. \tag{8}$$

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

Proof. First of all, any β-mixing process is α-mixing with $\alpha_m \leq \beta_m/2$. It is sufficient to check Conditions (a) and (b) of Rio (1997, Corollary 1). Condition (a) is verified by Rio (1997, Proposition 1), which guarantees $\mathrm{Var}(\mathbb{G}_n g_n) \leq \int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 du$ for all n. Since $\mathrm{Var}(\mathbb{G}_n g_n)$ 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 (8) implies Condition (b). □

The finite dimensional convergence of Z_n follows from Lemma C by setting g_n as any finite dimensional projection of the process $\{n^{1/6}\{(f_{\theta_0+sn^{-1/3}} - f_{\theta_0}) - P(f_{\theta_0+sn^{-1/3}} - f_{\theta_0})\} : s\}$. The requirement (8) can be considered as a Lindeberg-type condition to guarantee Rio's (1997, Corollary 1) Lindeberg condition in our setup. The condition (8) excludes polynomial decay of β_m. Therefore, exponential decay of β_m is required not only for the maximal inequality in Lemma M but also for the finite dimensional convergence in Lemma C. Also, Doukhan, Massart and Rio (1994, Theorem 5) provided some result, where any polynomial mixing rate will destroy the asymptotic normality of $\mathbb{G}_n g_n$. 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. This happens because the cube root class $\{f_\theta\}$ typically involves the indicator function. 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.

To establish tightness of the normalized process Z_n, we show the following maximal inequality.

Lemma M'. *Consider a sequence of classes of functions $\mathcal{G}_n = \{g_{n,s} : |s| \leq K\}$ for some $K > 0$ with envelope functions G_n. Suppose there is a universal positive constant C such that*

$$P \sup_{s:|s-s'|<\varepsilon} |g_{n,s} - g_{n,s'}|^2 \leq C\varepsilon, \tag{9}$$

for all n large enough, $|s'| \leq K$, and $\varepsilon > 0$ small enough. Also assume that there exist $0 \leq \kappa < 1/2$ and $C' > 0$ such that $G_n \leq C' n^\kappa$ and $\|G_n\|_2 \leq C'$ for all n large enough. Then for any $\sigma > 0$, there exist $\delta > 0$ and a positive integer N_δ such that

$$P \sup_{|s-s'|<\delta} |\mathbb{G}_n g_{n,s} - \mathbb{G}_n g_{n,s'}| \leq \sigma,$$

for all $n \geq N_\delta$.

Proof. Pick any $K > 0$ and $\sigma > 0$. Let $g_{n,s,s'} = g_{n,s} - g_{n,s'}$, $\mathcal{G}^1_{n,\delta} = \{g_{n,s,s'} : |s - s'| < \delta\}$, $\mathcal{G}^\beta_{n,\delta} = \{g_{n,s,s'} : \|g_{n,s,s'}\|_{2,\beta} < \delta\}$, and $\mathcal{G}^2_{n,\delta} = \{g_{n,s,s'} : \|g_{n,s,s'}\|_2 < \delta\}$. Since $g_{n,s}$ satisfies the condition (9), there exists a positive constant C_1 such that $\mathcal{G}^1_{n,\delta} \subset \mathcal{G}^2_{n,C_1\delta^{1/2}}$ for all n large enough and all $\delta > 0$ small enough. Also, by the same argument to derive (2), there exists a positive constant C_2 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 large enough, $|s| \leq K$, 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 large enough and all $\delta > 0$ small enough. The constant C_2 depends only on the mixing sequence $\{\beta_m\}$. Also note that the bracketing numbers satisfy

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

where $\mathcal{G}^K_n = \{g_{n,s,s'} : |s| \leq K, |s'| \leq K\}$ and the second inequality follows from (9). 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 large enough and all $\eta > 0$ small enough. Based on this entropy condition, we apply the maximal inequality of Doukhan, Massart and Rio (1995, Theorem 3), i.e., there exists a positive constant C_3 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 large enough and all $\eta > 0$ small enough, 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

$$\frac{v_n(\eta)^2}{\int_0^{v_n(\eta)} \beta^{-1}(\tilde{u}) d\tilde{u}} = \frac{\varphi_n^2(\eta)}{n\eta^2}.$$

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

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

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

Tightness of the process Z_n follows by Lemma M' with $g_{n,s} = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})$. Note that the condition (9) is satisfied by Condition (iii) of the cube root class. 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 maximal inequality.[3]

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

Theorem 1. *Suppose $\{z_t\}$ satisfies Assumption D. 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 (8) holds with $(g_{n,s} - P g_{n,s})$ for each s, where $g_{n,s} = n^{1/6}(f_{\theta_0 + sn^{-1/3}} - f_{\theta_0})$. Then*

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

[3]In particular, the process Z_n itself does not satisfy Condition (ii) of the cube root class.

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

This theorem can be considered as an extension of the main theorem of Kim and Pollard (1990) to an absolutely regular dependent process. The Lindeberg-type condition (8) 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 some positive constant C and all n, where the first inequality follows from the facts that $\beta^{-1}(\cdot)$ is monotonically decreasing and f_θ is bounded, and the second inequality follows by Assumption D and $P\{g_{n,s} = 0\} \geq 1 - cn^{-1/3}$.

Once we show that the M-estimator has a proper limiting distribution, Politis, Romano and Wolf (1999, Theorem 3.3.1) justify the use of subsampling to construct confidence intervals and make inference. Our mixing condition in Assumption D satisfies the requirement of their theorem and thus subsampling inference based on b consecutive observations with $b/n \to \infty$ is asymptotically valid. See Politis, Romano and Wolf (1999, Section 3.6) for a discussion on data-dependent choices of b.[4]

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 a situation, 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 $\{f_{\theta,\nu_0} : \theta \in \Theta\}$ be a cube root class. Suppose there are some negative definite matrix V_1 and some finite matrix V_2 such that*

$$P(f_{\theta,\nu} - f_{\theta_0,\nu_0}) = \frac{1}{2}(\theta - \theta_0)' V_1 (\theta - \theta_0) + (\theta - \theta_0)' V_2 (\nu - \nu_0) + o(|\theta - \theta_0|^2 + |\nu - \nu_0|^2), \quad (10)$$

for all θ and ν in neighborhoods of θ_0 and ν_0, respectively. Furthermore, suppose $\{f_{\theta,\nu} : \theta \in \Theta, \nu \in \Lambda\}$ satisfies Condition (iii) of the cube root class. Then $\hat{\theta} = \tilde{\theta} + o_p(n^{-1/3})$. Additionally, if (8) holds with $(g_{n,s} - Pg_{n,s})$ for each s, where $g_{n,s} = n^{1/6}(f_{\theta_0+sn^{-1/3},\nu_0} - f_{\theta_0,\nu_0})$, 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 $H(s_1, s_2) = \lim_{n\to\infty} \sum_{t=-n}^{n} \mathrm{Cov}(g_{n,s_1}(z_0), g_{n,s_2}(z_t)) < \infty$.

Note that we only need to verify that the subclass $\{f_{\theta,\nu_0} : \theta \in \Theta\}$ belongs to the cube root class. The additional conditions in (10) and an expansion in Condition (iii) for the whole class $\{f_{\theta,\nu} : \theta \in \Theta, \nu \in \Lambda\}$ are imposed to guarantee the asymptotic orthogonality between $\hat{\theta}$ and $\hat{\nu}$. The fact that $\hat{\nu} - \nu_0 = o_p(n^{-1/3})$ often requires these additional conditions to be satisfied.

---

[4] Another candidate to conduct inference based on the M-estimator is the bootstrap. However, even for independent observations, it is known that the naive nonparametric bootstrap is typically invalid under the cube root asymptotics (Abrevaya and Huang, 2005, and Sen, Banerjee and Woodroofe, 2010).

Proof. To ease notation, let $\theta_0 = \nu_0 = 0$. First, we show that $\hat{\theta} = O_p(n^{-1/3})$. Since $\{f_{\theta,\nu}\}$ satisfies Condition (iii) of the cube root class, we can apply Lemma M' with $g_{n,s} = n^{1/6}(f_{\theta,cn^{-1/3}} - f_{\theta,0})$ for $s = (\theta', c')'$, which implies

$$\sup_{|\theta|\leq\epsilon,|c|\leq\epsilon} n^{1/6}\mathbb{G}_n(f_{\theta,cn^{-1/3}} - f_{\theta,0}) = O_p(1), \tag{11}$$

for all $\epsilon > 0$. Also from (10) and $\hat{\nu} = o_p(n^{-1/3})$, we have

$$P(f_{\theta,\hat{\nu}} - f_{\theta,0}) - P(f_{0,\hat{\nu}} - f_{0,0}) \leq \theta' V_2 \hat{\nu} + \epsilon|\theta|^2 + O_p(n^{-2/3}), \tag{12}$$

for all θ in a neighborhood of θ_0 and all $\epsilon > 0$. Combining (11), (12), and Lemma 1,

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

for all θ in a neighborhood of θ_0 and all $\epsilon > 0$, where the last inequality follows from (10). From $\mathbb{P}_n(f_{\hat{\theta},\hat{\nu}} - f_{0,\hat{\nu}}) \geq o_p(n^{-2/3})$, negative definiteness of V_1, and $\hat{\nu} = o_p(n^{-1/3})$, we can find $c > 0$ such that

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

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

Next, we show that $\hat{\theta} - \tilde{\theta} = o_p(n^{-1/3})$. By reparametrization,

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

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 (10) implies $P(f_{sn^{-1/3},\hat{\nu}} - f_{0,\hat{\nu}}) - P(f_{sn^{-1/3},0} - f_{0,0}) = o_p(n^{-2/3})$ uniformly in s. Given $\hat{\theta} - \tilde{\theta} = o_p(n^{-1/3})$, an application of Theorem 1 to the cube root class $\{f_{\theta,\nu_0} : \theta \in \Theta\}$ implies the limiting distribution of $\hat{\theta}$. □

3. Cube root asymptotics with drifting criterions

We next investigate the case where the criterion function depends on the sample size typically due to a bandwidth sequence. We maintain Assumption D for the dependence structure of $\{z_t\}$. The cube root class is modified as follows.

Definition (Drifting cube root class). *A class of functions $\{f_{n,\theta} : \theta \in \Theta\}$ containing a sequence $\{h_n\}$ for $n = 1, 2, \ldots$ is called the drifting cube root class if*

- **(i):** *$\{h_n f_{n,\theta} : \theta \in \Theta\}$ is a class of bounded functions for all n. As $n \to \infty$, it holds $h_n \to 0$ and $nh_n \to \infty$. Also, $\lim_{n\to\infty} Pf_{n,\theta}$ is uniquely maximized at θ_0, and $Pf_{n,\theta}$ is twice* *continuously differentiable at θ_0 for all n large enough and admits the expansion*

$$P(f_{n,\theta} - f_{n,\theta_0}) = \frac{1}{2}(\theta - \theta_0)'V(\theta - \theta_0) + o(|\theta - \theta_0|^2) + o((nh_n)^{-2/3}), \tag{13}$$

for a negative definite matrix V.

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

$$|\theta_1 - \theta_2| \leq C h_n^{1/2} \|f_{n,\theta_1} - f_{n,\theta_2}\|_2 \,,$$

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

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

$$P \sup_{\theta \in \Theta : |\theta - \theta_0| < \varepsilon} h_n |f_{n,\theta} - f_{n,\theta_0}|^2 \leq C''\varepsilon,$$

for all n large enough and all $\varepsilon > 0$ small enough.

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. Verifications of Conditions (i)-(iii) require more restrictions on h_n. For example, negligibility of the bias term for nonparametric estimation is implicit in (13). 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 (see, Sections 4.4-4.6 for examples). In this case, boundedness of $\{h_n f_{n,\theta} : \theta \in \Theta\}$ in Condition (i) means boundedness of $K\left(\frac{x-c}{h_n}\right) m(y, x, \theta)$. The expansion in (13) can be understood as a restriction for $P(f_{n,\theta} - f_{n,\theta_0}) = \int\int K(a)\, m(y, c + h_n a, \theta) p_{yx}(y, c + h_n a) da dy$ by a change of variables, where p_{yx} is the joint density of (y, x). The reasons for multiplications 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 drifting cube root class. The M-estimator 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 establish the modulus of continuity of the empirical process $\{\mathbb{G}_n h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) : \theta \in \Theta\}$ for the drifting cube root class defined above. We show the following maximal inequality.

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

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

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 ($j = 1, 2, \ldots$) are independent of n and δ. By changing the notation to indicate the drifting classes of functions, a similar argument to derive (5) implies the bound

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

where $\mathcal{G}^{\beta}_{n,\delta} = \{f_{n,\theta} - f_{n,\theta_0} : \left\| h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) \right\|_{2,\beta} < \delta \text{ for } \theta \in \Theta\}$ with an envelope function $G_{n,\delta}$, and $\varphi_n(\delta) = \int_0^\delta \sqrt{\log N_{[]}(\nu, \mathcal{G}^{\beta}_{n,\delta}, \|\cdot\|_{2,\beta})} d\nu$. By 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

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

for all n large enough. □

Compared to Lemma M, the modulus of continuity of the empirical process $\{\mathbb{G}_n(f_{n,\theta} - f_{n,\theta_0}) : \theta \in \Theta\}$ changes from $\delta^{1/2}$ to $h_n^{-1/2}\delta^{1/2}$ (because of the change of the envelope in Condition (iii) of the drifting cube root class). Consequently, the convergence rate of $\hat{\theta}$ will change from $n^{1/3}$ to $(nh_n)^{1/3}$. In order to derive the convergence rate of $\hat{\theta}$, we modify Lemma 1 as follows. Since the proof is similar, it is omitted.

Lemma 2. *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 $(nh_n)^{-1/3} \leq |\theta - \theta_0| \leq C$.

It should be noted that Lemma Mn is used to derive this lemma. Based on Lemma 2, the convergence rate of $\hat{\theta}$ is obtained as follows. Suppose $|\hat{\theta} - \theta_0| \geq (nh_n)^{-1/3}$. Then we can take $c > 0$ such that

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

for each $\varepsilon > 0$, where the second inequality follows from Lemma 2 and the third inequality follows from Condition (i) of the drifting cube root class. Taking ε small enough to satisfy $c - \varepsilon > 0$ yields the conclusion that $\hat{\theta} - \theta_0 = O_p((nh_n)^{-1/3})$.

In order to derive the limiting distribution, we need to establish tightness of the centered process

$$Z_n(s) = n^{1/6} h_n^{2/3} \mathbb{G}_n(f_{n,\theta_0 + s(nh_n)^{-1/3}} - f_{n,\theta_0}),$$

for $|s| \leq K$ with any $K > 0$. The finite dimensional convergence and tightness of Z_n follows from Lemmas M' and C in the previous section by setting $g_{n,s} = n^{1/6} h_n^{2/3}(f_{n,\theta_0 + s(nh_n)^{-1/3}} - f_{n,\theta_0})$ and $g_n = g_{n,s} - P g_{n,s}$, respectively. Therefore, the argmax theorem (Kim and Pollard, 1990, Theorem 2.7) implies the limiting distribution of the M-estimator $\hat{\theta}$ for the drifting cube root class. The main theorem of this section is summarized as follows.

Theorem 3. *Suppose $\{z_t\}$ satisfies Assumption D. 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 $\theta_0 \in$ intΘ, and (8) holds with $(g_{n,s} - Pg_{n,s})$ for each s, where $g_{n,s} = n^{1/6}h_n^{2/3}(f_{n,\theta_0+s(nh_n)^{-1/3}} - f_{n,\theta_0})$. 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 $H(s_1, s_2) = \lim_{n\to\infty} \sum_{t=-n}^{n} \text{Cov}(g_{n,s_1}(z_0), g_{n,s_2}(z_t)) < \infty$.

This theorem extends Kim and Pollard's (1990) main theorem to the case where the criterion function varies with the sample size typically due to a bandwidth sequence. Since $h_n \to 0$, 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 $\hat{\nu} - \nu_0 = o_p((nh_n)^{-1/3})$, 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 $\{f_{n,\theta,\nu_0} : \theta \in \Theta\}$ be a drifting cube root class. Suppose there are some negative definite matrix V_1 and some finite matrix V_2 such that*

$$P(f_{n,\theta,\nu} - f_{n,\theta_0,\nu_0}) = \frac{1}{2}(\theta-\theta_0)'V_1(\theta-\theta_0) + (\theta-\theta_0)'V_2(\nu-\nu_0) + o(|\theta-\theta_0|^2 + |\nu-\nu_0|^2) + o((nh_n)^{-2/3}), \tag{14}$$

for all θ and ν in neighborhoods of θ_0 and ν_0, respectively. Furthermore, $\{f_{n,\theta,\nu} : \theta \in \Theta, \nu \in \Lambda\}$ satisfies Condition (iii) of the drifting cube root class. Then $\hat{\theta} = \tilde{\theta} + o_p((nh_n)^{-1/3})$. Additionally, if (8) holds with $(g_{n,s} - Pg_{n,s})$ for each s, where $g_{n,s} = n^{1/6}h_n^{2/3}(f_{n,\theta_0+s(nh_n)^{-1/3},\nu_0} - f_{n,\theta_0,\nu_0})$, 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 $s'V_1s/2$ and covariance kernel $H(s_1, s_2) = \lim_{n\to\infty} \sum_{t=-n}^{n} \text{Cov}(g_{n,s_1}(z_0), g_{n,s_2}(z_t)) < \infty$.

Since the proof is similar to that of Theorem 2, it is omitted. In the next section, we illustrate the above theoretical results by several examples.

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. 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 $P\{x_t'\theta_0 = 0, \dot{\kappa}(x_t)'\theta_0 p(x_t) > 0\} > 0$.

Except for Assumption D, which allows dependent observations, all assumptions are similar to the ones in Kim and Pollard (1990, Section 6.4). First, note that the criterion function is written as

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

We can see that $\hat{\theta} = \arg\max_{\theta \in S} \mathbb{P}_n f_\theta$ and $\theta_0 = \arg\max_{\theta \in S} P f_\theta$. Existence and uniqueness of θ_0 are guaranteed by (b) (see, Manski, 1985). Also the uniform law of large numbers for an absolutely regular process by Nobel and Dembo (1993, Theorem 1) implies $\sup_{\theta \in S} |\mathbb{P}_n f_\theta - P f_\theta| \xrightarrow{p} 0$. Therefore, $\hat{\theta}$ is consistent for θ_0.

We next compute the expected value and covariance kernel of the limit process (i.e., V and H in Theorem 1). Due to strict stationarity (in Assumption D), we can apply the same argument to Kim and Pollard (1990, pp. 214-215) to derive the second derivative

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

where σ is the surface measure on the boundary of the set $\{x : x'\theta_0 \geq 0\}$. The matrix V is negative definite under the last condition of (b). Now 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 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

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

The limit of $n^{1/3}\text{Var}(g_{n,t})$ is given in Kim and Pollard (1990, p. 215). For the covariance $\text{Cov}(g_{n,t}, g_{n,t+m})$, note that $g_{n,t}$ can take only three values, -1, 0, or 1. By the definition of β_m, 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 $j, k = -1, 0, 1$, i.e., $\{g_{n,t}\}$ is a β-mixing array whose mixing coefficients are bounded by $n^{-2/3}\beta_m$. In turn, this implies that $\{g_{n,t}\}$ is an α-mixing array whose mixing coefficients are bounded by $2n^{-2/3}\beta_m$. Thus, by applying the α-mixing inequality, the covariance is bounded as

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

for some $C > 0$ and $p > 2$. 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. Condition (i) is already verified. By Jensen's inequality,

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

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 $|\theta_1 - \theta_2|$, the last condition in (a) implies Condition (ii) of the cube root class. For Condition (iii), pick any $\varepsilon > 0$ and observe that

$$P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} |f_\theta - f_{\theta_0}|^2 = P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} \mathbb{I}\{x_t'\theta \geq 0 > x_t'\theta_0 \text{ or } x_t'\theta_0 \geq 0 > x_t'\theta\}.$$

Again, the right hand side is the probability for a pair of wedge shaped regions with an angle of order ε. Thus the last condition in (a) also guarantees Condition (iii) of the cube root class. Since $\{f_\theta : \theta \in S\}$ belongs to the cube root class, Theorem 1 implies 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 monotone density estimation.** Preliminary results (Lemmas M, M', C, and 1) to show Theorem 1 may be applied to establish weak convergence of certain 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 $\hat{\gamma}(c)$ of the density $\gamma(c)$ at a fixed $c > 0$ is given by the left derivative of the concave majorant of the empirical distribution function $\hat{\Gamma}$. It is known that $n^{1/3}(\hat{\gamma}(c)-\gamma(c))$ can be written as the left derivative of the concave majorant of the process $W_n(s) = n^{2/3}\{\hat{\Gamma}(c+sn^{-1/3})-\hat{\Gamma}(c)-\gamma(c)sn^{-1/3}\}$ (Prakasa Rao, 1969). Let $f_\theta(z) = \mathbb{I}\{z \leq c+\theta\}$ and Γ be the distribution function of γ. Decompose

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

A Taylor expansion implies convergence of the second term to $\frac{1}{2}\dot{\gamma}(c)s^2 < 0$. For the first term $Z_n(s) = n^{1/6}\mathbb{G}_n(f_{sn^{-1/3}} - f_0)$, we can apply Lemmas C and M' to establish the weak convergence. Lemma C (setting g_n as any finite dimensional projection of the process $\{n^{1/6}(f_{sn^{-1/3}} - f_0) : s\}$) implies finite dimensional convergence of Z_n to projections of a centered Gaussian process with the covariance kernel

$$H(s_1, s_2) = \lim_{n\to\infty} n^{1/3} \sum_{t=-n}^{n} \{\Gamma_{0t}(c + s_1n^{-1/3}, c + s_2n^{-1/3}) - \Gamma(c + s_1n^{-1/3})\Gamma(c + s_2n^{-1/3})\},$$

where Γ_{0t} is the joint distribution function of (z_0, z_t). For tightness of Z_n, we apply Lemma M' by setting $g_{n,s} = n^{1/6}(f_{sn^{-1/3}} - f_0)$. The envelope condition is clearly satisfied. The condition in (9) is verified as

$$\begin{aligned}
& P \sup_{s:|s-s'|<\varepsilon} |g_{n,s} - g_{n,s'}|^2 \\
= \; & n^{1/3} P \sup_{s:|s-s'|<\varepsilon} |\mathbb{I}\{z \leq c + sn^{-1/3}\} - \mathbb{I}\{z \leq c + s'n^{-1/3}\}| \\
\leq \; & n^{1/3} \max\{\Gamma(c + sn^{-1/3}) - \Gamma(c + (s - \varepsilon)n^{-1/3}), \Gamma(c + (s + \varepsilon)n^{-1/3}) - \Gamma(c + sn^{-1/3})\} \\
\leq \; & \gamma(0)\varepsilon.
\end{aligned}$$

Therefore, by applying Lemmas C and M', W_n weakly converges to Z, a Gaussian process with expected value $\frac{1}{2}\dot{\gamma}(c)s^2$ and covariance kernel H.

The remaining part follows by the same argument to Kim and Pollard (1990, pp. 216-218) (by replacing their Lemma 4.1 with our Lemma 1). Then we can conclude that $n^{1/3}(\hat{\gamma}(c)-\gamma(c))$ 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, consider the least median of squares estimator for the regression model $y_t = x_t'\beta_0 + u_t$, that is

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

We impose the following assumptions. Except for Assumption D, which allows dependent observations, all assumptions are similar to the ones in Kim and Pollard (1990, Section 6.3).

- **(a):** $\{x_t, u_t\}$ satisfies Assumption D. x_t and u_t are independent. $P|x_t|^2 < \infty$, Px_tx_t' is positive definite, and the distribution of x_t puts zero mass on each hyperplane.
- **(b):** The density γ of u_t is bounded, differentiable, and symmetric around zero, and decreases away from zero. $|u_t|$ has the unique median ν_0 and $\dot{\gamma}(\nu_0) < 0$.

It is known 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 $\sup_{\theta,\nu} |\mathbb{P}_n f_{\theta,\nu} - Pf_{\theta,\nu}| = O_p(n^{-1/2})$. Thus, the same argument to Kim and Pollard (1990, pp. 207-208) yields the convergence rate $\hat{\nu} - 1 = O_p(n^{-1/2})$.

By expansions, the condition in (10) is verified as

$$\begin{aligned} P(f_{\theta,\nu} - f_{0,1}) &= P|\{\Gamma(x'\theta+\nu) - \Gamma(\nu)\} - \{\Gamma(x'\theta-\nu) - \Gamma(-\nu)\}| \\ &\quad + P|\{\Gamma(\nu) - \Gamma(1)\} - \{\Gamma(-\nu) - \Gamma(-1)\}| \\ &= \dot{\gamma}(1)\theta' Pxx'\theta + o(|\theta|^2 + |\nu-1|^2). \end{aligned} \tag{15}$$

To check Condition (iii) of the cube root class for $\{f_{\theta,\nu} : \theta \in \mathbb{R}^d, \nu \in \mathbb{R}\}$, pick any $\varepsilon > 0$ and decompose

$$P \sup_{(\theta,\nu):|(\theta,\nu)-(0,1)|<\varepsilon} |f_{\theta,\nu} - f_{0,1}|^2 \le P \sup_{(\theta,\nu):|(\theta,\nu)-(0,1)|<\varepsilon} |f_{\theta,\nu} - f_{\theta,1}|^2 + P \sup_{\theta:|\theta|<\varepsilon} |f_{\theta,1} - f_{0,1}|^2.$$

By similar arguments to (15), these terms are of order $|\nu - 1|^2$ and $|\theta|^2$, respectively, which are bounded by $C\varepsilon$ with some $C > 0$.

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 Condition (i) is satisfied. Since Condition (iii) is already shown, it remains to verify Condition (ii). Some expansions (using symmetry of $\gamma(\cdot)$) yield

$$\begin{aligned} \|f_{\theta_1,1} - f_{\theta_2,1}\|_2^2 &= P|\Gamma(x'\theta_1+1) - \Gamma(x'\theta_2+1) + \Gamma(x'\theta_1-1) - \Gamma(x'\theta_2-1)| \\ &\ge (\theta_2-\theta_1)'P\dot{\gamma}(-1)xx'(\theta_2-\theta_1) + o(|\theta_2-\theta_1|^2), \end{aligned}$$

i.e., Condition (ii) is satisfied under (b). Therefore, $\{f_{\theta,1} : \theta \in \mathbb{R}^d\}$ belongs to the cube root class.

We finally compute the covariance kernel H. 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 $L(s_1, s_2) = \lim_{n\to\infty} n^{4/3}\text{Var}(\mathbb{P}_n g_{n,t})$ and $g_{n,t} = \mathbb{I}\{|x_t' s_1 n^{-1/3} - u_t| \leq 1\} - \mathbb{I}\{|x_t' s_2 n^{-1/3} - u_t| \leq 1\}$. By a similar argument to the maximum score example in Section 4.1, we can show that H is the same as the one for the independent case derived in Kim and Pollard (1990, p. 213). Therefore, by Theorem 2, we conclude that $n^{1/3}(\hat{\beta} - \beta_0)$ converges in distribution to the argmax of $Z(s)$, a Gaussian process with expected value $\dot{\gamma}(1)s'Pxx's$ and the covariance kernel H.

4.4. **Panel data discrete choice model.** As an illustration of Theorem 3, we consider a dynamic panel data model with a binary dependent variable

$$P\{y_{i0} = 1|x_i, \alpha_i\} = F_0(x_i, \alpha_i),$$
$$P\{y_{it} = 1|x_i, \alpha_i, y_{i0}, \ldots, y_{it-1}\} = F(x_{it}'\beta_0 + \gamma_0 y_{it-1} + \alpha_i),$$

for $i = 1, \ldots, n$ and $t = 1, 2, 3$, where y_{it} is binary, x_{it} is a k-vector, and both F_0 and F are unknown functions. Honoré and Kyriazidou (2000) proposed the conditional maximum score estimator

$$(\hat{\beta}, \hat{\gamma}) = \arg\max_{\beta,\gamma} \sum_{i=1}^{n} K\left(\frac{x_{i2} - x_{i3}}{b_n}\right)(y_{i2} - y_{i1})\text{sgn}\{(x_{i2} - x_{i1})'\beta + (y_{i3} - y_{i0})\gamma\},$$

where K is a kernel function and b_n is a bandwidth. Honoré and Kyriazidou (2000) obtained consistency of this estimator but the convergence rate and limiting distribution are unknown. Theorem 3 answers these open questions. Let $z = (z_1', z_2, z_3')'$ with $z_1 = x_2 - x_3$, $z_2 = y_2 - y_1$, and $z_3 = ((x_2 - x_1)', y_3 - y_0)$. Also define $x_{21} = x_2 - x_1$ and $x_{23} = x_2 - x_3$. Based on Honoré and Kyriazidou (2000, Theorem 4), we impose the following assumptions.

- **(a):** $\{z_i\}_{i=1}^n$ is an iid sample. z_1 has a bounded density which is continuously differentiable at zero. The conditional density of $z_1|z_2 \neq 0, z_3$ is positive in a neighborhood of zero, and $P\{z_2 \neq 0|z_3\} > 0$ for almost every z_3. Support of x_{21} conditional on x_{23} in a neighborhood of zero is not contained in any proper linear subspace of $\mathbb{R}^k$. There exists at least one $j \in \{1, \ldots, k\}$ such that $\beta_0^{(j)} \neq 0$ and $x_{21}^{(j)}|x_{21}^{j-}, x_{23}$, where $x_{21}^{j-} = (x_{21}^{(1)}, \ldots, x_{21}^{(j-1)}, x_{21}^{(j+1)}, \ldots, x_{21}^{(k)})$, has everywhere positive conditional density for almost every x_{21}^{j-} and almost every x_{23} in a neighborhood of zero. $E[z_2|z_3, z_1 = 0]$ is differentiable in z_3. $E[z_2\text{sgn}((\beta_0', \gamma_0)'z_3)|z_1]$ is continuously differentiable at $z_1 = 0$. F is strictly increasing.
- **(b):** K is a bounded symmetric density function with $\int s^2 K(s)ds < \infty$. As $n \to \infty$, it holds $nb_n^k/\ln n \to \infty$ and $nb_n^{k+3} \to 0$.

Note that the estimator $\hat{\theta} = (\hat{\beta}', \hat{\gamma})'$ can be written as $\hat{\theta} = \arg\max_\theta \mathbb{P}_n f_{n,\theta}$, where

$$\begin{aligned} f_{n,\theta}(z) &= b_n^{-k} K(b_n^{-1} z_1) z_2\{\text{sgn}(z_3'\theta) - \text{sgn}(z_3'\theta_0)\} \\ &= e_n(z)(\mathbb{I}\{z_3'\theta \geq 0\} - \mathbb{I}\{z_3'\theta_0 \geq 0\}), \end{aligned}$$

and $e_n(z) = 2b_n^{-k}K(b_n^{-1}z_1)z_2$.

We verify that $\{f_{n,\theta}\}$ belongs to the drifting cube root class with $h_n = b_n^k$. We first check Condition (ii) of the drifting cube root class. By the definition of $z_2 = y_2 - y_1$ (which can take -1, 0, or 1) and change of variables $a = b_n^{-1}z_1$,

$$E[e_n(z)^2|z_3] = 4b_n^{-k}\int K(a)^2 p_1(b_n a|z_2 \neq 0, z_3)daP\{z_2 \neq 0|z_3\},$$

almost surely for all n, where p_1 is the conditional density of z_1 given $z_2 \neq 0$ and z_3. Thus under (a), $b_n^k E[e_n(z)^2|z_3] > c$ almost surely, for some $c > 0$. Pick any θ_1 and θ_2. Note that

$$\begin{aligned} h_n^{1/2}\|f_{n,\theta_1} - f_{n,\theta_2}\|_2 &= \left(P\left\{h_n E[e_n(z)^2|z_3]|\mathbb{I}\{z_3'\theta_1 \geq 0\} - \mathbb{I}\{z_3'\theta_2 \geq 0\}|\right\}\right)^{1/2} \\ &\geq c^{1/2}P|\mathbb{I}\{z_3'\theta_1 \geq 0\} - \mathbb{I}\{z_3'\theta_2 \geq 0\}| \\ &\geq c_1|\theta_1 - \theta_2|, \end{aligned}$$

for some $c_1 > 0$, where the last inequality follows from the same argument to the maximum score example in Section 4.1 using Condition (a). Similarly, Condition (iii) of the drifting cube root class is verified as

$$h_n P \sup_{|\theta-\theta_0|<\varepsilon} |f_{n,\theta} - f_{n,\theta_0}|^2 \leq C_1 P \sup_{|\theta-\theta_0|<\varepsilon} |\mathbb{I}\{z_3'\theta \geq 0\} - \mathbb{I}\{z_3'\theta_0 \geq 0\}| \leq C_2\varepsilon,$$

for some positive constants C_1 and C_2 and all n. We now verify Condition (i). Since $h_n f_{n,\theta}$ is clearly bounded, it is enough to verify (13). A change of variables $a = b_n^{-1}z_1$ and Condition (b) imply

$$\begin{aligned} Pf_{n,\theta} &= \int K(a)E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = a]p_1(b_n a)da \\ &= p_1(0)E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = 0] \\ &\quad + b_n^2\int a^2K(a)\left.\frac{\partial E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = t]p_1(t)}{\partial t}\right|_{t=t_a} da, \end{aligned}$$

where t_a is a point on the line joining a and 0, and the second equality follows from the dominated convergence and mean value theorems. Since $b_n^2 = o((nb_n^k)^{-2/3})$ by (b), the second term is negligible. Thus, for the condition in (13), it is enough to derive a second order expansion of $E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = 0]$. Let $\mathcal{Z}_\theta = \{z_3 : \mathbb{I}\{z_3'\theta \geq 0\} \neq \mathbb{I}\{z_3'\theta_0 \geq 0\}\}$. Honoré and Kyriazidou (2000, p. 872) showed that

$$-E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = 0] = 2\int_{\mathcal{Z}_\theta} |E[z_2|z_1 = 0, z_3]|dF_{z_3|z_1=0} > 0,$$

for all $\theta \neq \theta_0$ on the unit sphere and that $\mathrm{sgn}(E[z_2|z_3, z_1 = 0]) = \mathrm{sgn}(z_3'\theta_0)$. Therefore, by applying the same argument as Kim and Pollard (1990, pp. 214-215), we obtain $\frac{\partial}{\partial\theta}E[z_2\mathrm{sgn}(z_3'\theta)|z_1 = 0]\big|_{\theta=\theta_0} = 0$ and

$$-\frac{\partial^2 E[z_2\{\mathrm{sgn}(z_3'\theta) - \mathrm{sgn}(z_3'\theta_0)\}|z_1 = 0]}{\partial\theta\partial\theta'} = \int \mathbb{I}\{z_3'\theta_0 = 0\}\dot{\kappa}(z_3)'\theta_0 z_3 z_3' p_3(z_3|z_1 = 0)d\mu_{\theta_0},$$

where $\dot{\kappa}(z_3) = \frac{\partial}{\partial z_3}E[z_2|z_3, z_1 = 0]$, p_3 is the conditional density of z_3 given $z_1 = 0$, and μ_{θ_0} is the surface measure on the boundary of $\{z_3 : z_3'\theta_0 \geq 0\}$. Combining these results, the condition in

(13) is satisfied with

$$V = -2p_1(0) \int \mathbb{I}\{z_3'\theta_0 = 0\}\dot{\kappa}(z_3)'\theta_0 z_3 z_3' p_3(z_3|z_1 = 0) d\mu_{\theta_0}.$$

Finally the covariance kernel H is obtained in the same manner to Kim and Pollard (1990). That is, decompose z_3 into $r'\theta_0 + \bar{z}_3$ with $\bar{z}_3$ orthogonal to θ_0. Then it holds $H(s_1, s_2) = L(s_1) + L(s_2) - L(s_1 - s_2)$, where

$$L(s) = 4p_1(0) \int |\bar{z}_3' s| p_3(0, \bar{z}_3|z_1 = 0) d\bar{z}_3.$$

4.5. **Minimum volume predictive region.** As an illustration of Theorem 4, 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 $\hat{I} = [\hat{\theta} - \hat{\nu}, \hat{\theta} + \hat{\nu}]$, where

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

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

- **(a):** $\{y_t, x_t\}$ satisfies Assumption D. $I_0 = [\theta_0 - \nu_0, \theta_0 + \nu_0]$ is the unique shortest interval such that $P\{y_t \in I_0 | x_t = c\} \geq \alpha$. The conditional density $\gamma_{y|x=c}$ of y_t given $x_t = c$ is bounded and strictly positive at $\theta_0 \pm \nu_0$, and its derivative satisfies $\dot{\gamma}_{y|x=c}(\theta_0 - \nu_0) - \dot{\gamma}_{y|x=c}(\theta_0 + \nu_0) > 0$.
- **(b):** K is bounded and symmetric, and satisfies $\lim_{a \to \infty} |a| K(a) = 0$. As $n \to \infty$, $nh_n \to \infty$ and $nh_n^4 \to 0$.

For notational convenience, assume $\theta_0 = 0$ and $\nu_0 = 1$. We first derive the convergence rate for $\hat{\nu}$. Note that $\hat{\nu} = \inf\{\nu : \sup_\theta \hat{g}([\theta - \nu, \theta + \nu]) \geq \alpha \hat{\gamma}(c)\}$, 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 $\hat{\gamma}(c) = \frac{1}{nh_n} \sum_{t=1}^n K\left(\frac{x_t - c}{h_n}\right)$. By applying Lemma M' and a central limit theorem, we can obtain uniform convergence rate

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

Thus the same argument to Kim and Pollard (1990, pp. 207-208) yields $\hat{\nu} - 1 = O_p((nh_n)^{-1/2} + h_n^2)$. Let $\hat{\theta} = \arg\min_\theta \hat{g}([\theta - \hat{\nu}, \theta + \hat{\nu}])$. Consistency follows from uniqueness of (θ_0, ν_0) in (a) and the uniform convergence

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

which is obtained by applying Nobel and Dembo (1993, Theorem 1).

Now let $z = (y, x)'$ and

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

Note that $\hat{\theta} = \arg\max_\theta \mathbb{P}_n f_{n,\theta,\hat{\nu}}$. We apply Theorem 4 to obtain the convergence rate of $\hat{\theta}$. For the condition in (14), observe that

$$\begin{aligned} & P(f_{n,\theta,\nu} - f_{n,0,1}) = P(f_{n,\theta,\nu} - f_{n,0,\nu}) + P(f_{n,0,\nu} - f_{n,0,1}) \\ = & -\frac{1}{2}\{-\dot{\gamma}_{y|x}(1|c) + \dot{\gamma}_{y|x}(-1|c)\}\gamma_x(c)\theta^2 + \{\dot{\gamma}_{y|x}(1|c) + \dot{\gamma}_{y|x}(-1|c)\}\gamma_x(c)\theta\nu + o(\theta^2 + |\nu - 1|^2) + O(h_n^2). \end{aligned}$$

The condition (14) holds with $V_1 = \{\dot{\gamma}_{y|x}(1|c) - \dot{\gamma}_{y|x}(-1|c)\}\gamma_x(c)$ and $V_2 = \{\dot{\gamma}_{y|x}(1|c) + \dot{\gamma}_{y|x}(-1|c)\}\gamma_x(c)$. Condition (iii) of the drifting cube root class for $\{f_{n,\theta,\nu} : \theta \in \mathbb{R}, \nu \in \mathbb{R}\}$ is verified in the same manner as in Section 4.3. It remains to verify Condition (ii) of the drifting cube root class for $\{f_{n,\theta,1} : \theta \in \mathbb{R}\}$. Pick any θ_1 and θ_2. Some expansions yield

$$\begin{aligned} & h_n \|f_{n,\theta_1,1} - f_{n,\theta_2,1}\|_2^2 \\ = & \int K(a)^2 \left| \begin{array}{c} \Gamma_{y|x}(\theta_2 + 1|x = c + ah_n) - \Gamma_{y|x}(\theta_1 + 1|x = c + ah_n) \\ +\Gamma_{y|x}(\theta_2 - 1|x = c + ah_n) - \Gamma_{y|x}(\theta_1 - 1|x = c + ah_n) \end{array} \right| \gamma_x(c + ah_n) da \\ \geq & \int K(a)^2 \{\gamma_{y|x}(\dot{\theta} + 1|x = c + ah_n) + \gamma_{y|x}(\ddot{\theta} - 1|x = c + ah_n)\}\gamma_x(c + ah_n) da |\theta_1 - \theta_2|, \end{aligned}$$

where $\Gamma_{y|x}$ is the conditional distribution function of y given x, and $\dot{\theta}$ and $\ddot{\theta}$ are points between θ_1 and θ_2. By (a), Condition (ii) is satisfied. Therefore, we can conclude that $\hat{\nu} - \nu_0 = O_p((nh_n)^{-1/2} + h_n^2)$ and $\hat{\theta} - \theta_0 = O_p((nh_n)^{-1/3} + h_n)$. This result confirms positively the conjecture of Polonik and Yao (2000, Remark 3b) on the exact convergence rate of $\hat{I}$.

4.6. **Hough transform estimator.** As a final illustration, we consider an example where the criterion function does not belong to the (drifting) cube root class and discuss how our main theorems can be modified. In particular, consider the Hough transform estimator for the regression model $y_t = x_t'\beta_0 + u_t$ with the drifting tuning constant h_n,

$$\hat{\beta} = \arg\max_\beta \sum_{t=1}^n \mathbb{I}\{|y_t - x_t'\beta| \leq h_n|x_t|\},$$

where $x_t = (1, \tilde{x}_t)'$ for a scalar $\tilde{x}_t$. Goldenshluger and Zeevi (2004) studied the case when h_n does not vary with n and derived the cube root asymptotics for $\hat{\beta}$ and left the analysis for the case of $h_n \to 0$ as an open question. Here we answer this question. We impose the following assumptions.

- **(a):** $\{x_t, u_t\}$ satisfies Assumption D. x_t and u_t are independent. $P|x_t|^3 < \infty$, Px_tx_t' is positive definite, and the distribution of x_t puts zero mass on each hyperplane. The density γ of u_t is bounded, continuously differentiable in a neighborhood of zero, symmetric around zero, and strictly unimodal at zero.
- **(b):** As $n \to \infty$, $h_n \to 0$ and $nh_n^5 \to \infty$.

Let $z = (x, u)$. Note that $\hat{\theta} = \hat{\beta} - \beta_0$ is written as $\hat{\theta} = \arg\max_\theta \mathbb{P}_n f_{n,\theta}$, where

$$f_{n,\theta}(z) = h_n^{-1}\mathbb{I}\{|u - x'\theta| \leq h_n|x|\}.$$

The consistency of $\hat{\theta}$ follows from the uniform convergence $\sup_\theta |\mathbb{P}_n f_{n,\theta} - P f_{n,\theta}| \xrightarrow{p} 0$ by applying Nobel and Dembo (1993, Theorem 1).

In order to apply Theorem 3, we need to verify that $\{f_{n,\theta}\}$ belongs to the drifting cube root class. Obviously $h_n f_{n,\theta}$ is bounded for all n. Since $\lim_{n\to\infty} P f_{n,\theta} = 2P\gamma(x'\theta)|x|$ and γ is uniquely maximized at zero (by Condition (a)), $\lim_{n\to\infty} P f_{n,\theta}$ is uniquely maximized at $\theta = 0$. Since γ is continuously differentiable in a neighborhood of zero, $P f_{n,\theta}$ is twice continuously differentiable at $\theta = 0$ for all n large enough. Let Γ be the distribution function of γ. An expansion yields

$$\begin{aligned} P(f_{n,\theta} - f_{n,0}) &= h_n^{-1}P\{\Gamma(x'\theta + h_n|x|) - \Gamma(h_n|x|)\} - h_n^{-1}P\{\Gamma(x'\theta - h_n|x|) - \Gamma(-h_n|x|)\} \\ &= \ddot{\gamma}(0)\theta' P(|x|xx')\theta\{1 + O(h_n)\} + o(|\theta|^2), \end{aligned}$$

i.e., the condition in (13) holds with $V = \ddot{\gamma}(0)P(|x|xx')$. Note that $\ddot{\gamma}(0) < 0$ by Condition (a). Therefore, Condition (i) of the drifting cube root class is satisfied.

For Condition (ii), pick any θ_1 and θ_2 and note that

$$\begin{aligned} h_n \|f_{n,\theta_1} - f_{n,\theta_2}\|_2^2 \;\; = \;\; & 2P\{\gamma(x'\theta_1) + \gamma(x'\theta_2)\}|x| \\ & -2h_n^{-1}P\{x'\theta_1 - h_n|x| < u < x'\theta_2 + h_n|x|, \; -2h_n|x| < x'(\theta_2 - \theta_1) < 0\} \\ & -2h_n^{-1}P\{x'\theta_2 - h_n|x| < u < x'\theta_1 + h_n|x|, \; -2h_n|x| < x'(\theta_1 - \theta_2) < 0\}. \end{aligned}$$

Since the second and third terms converge to zero (by a change of variable), Condition (ii) of the drifting cube root class holds true.

However, we can see that Condition (iii) of the drifting cube root class is not satisfied in this case. Although Theorem 3 is not directly applicable, Condition (iii) can be modified as follows.

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

$$P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} h_n^2|f_{n,\theta} - f_{n,\theta_0}|^2 \leq C''\varepsilon,$$

for all n large enough and all $\varepsilon > 0$ small enough.

Compared to Condition (iii) of the drifting cube root class, Condition (iii)' assumes a larger envelope for the class $\{|f_{n,\theta} - f_{n,\theta_0}|^2 : |\theta - \theta_0| < \varepsilon\}$. Thus, Lemma Mn in Section 3 is modified as follows.

Lemma Mn'. *Suppose that Assumption D holds and $\{f_{n,\theta}\}$ satisfies Condition (ii) of the drifting cube root class and Condition (iii)' above. Then there exist positive constants C and C' such that*

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

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

Proof. The proof is similar to that of Lemma Mn except that for some positive constant C''', we have

$$\mathcal{G}_\delta^1 \subset \mathcal{G}^2_{C''h_n^{-1/2}\delta^{1/2}} \subset \mathcal{G}^\beta_{C'''h_n^{-1/2}\delta^{1/2}},$$

which reflects the component "h_n^2" in Condition (iii)' instead of "h_n" in Condition (iii) of the drifting cube root class. As a consequence of this change, the upper bound in the maximal inequality becomes $Ch_n^{-1/2}\delta^{1/2}$ instead of $C\delta^{1/2}$. All the other parts remain the same. $\square$

We now check Condition (iii)'. Observe that

$$
\begin{aligned}
P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} h_n^2|f_{n,\theta}-f_{n,0}|^2 \ \leq\ & P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} \mathbb{I}\{|u|\leq h_n|x|,\ |u-x'\theta|>h_n|x|\} \\
& +P \sup_{\theta\in\Theta:|\theta-\theta_0|<\varepsilon} \mathbb{I}\{|u-x'\theta|\leq h_n|x|,\ |u|>h_n|x|\}.
\end{aligned}
$$

Since the same argument applies to the second term, we focus on the first term (say, T). If $\varepsilon\leq 2h_n$, then an expansion around $\varepsilon=0$ implies

$$T\leq P\{(h_n-\varepsilon)|x|\leq u\leq h_n|x|\}=P\gamma(h_n|x|)|x|\varepsilon+o(\varepsilon).$$

Also, if $\varepsilon>2h_n$, then an expansion around $h_n=0$ implies

$$T\leq P\{-h_n|x|\leq u\leq h_n|x|\}\leq P\gamma(0)|x|\varepsilon+o(h_n).$$

Therefore, Condition (iii)' is satisfied.

Finally, the covariance kernel is obtained by a similar way as Section 4.1. Let $r_n=(nh_n^2)^{1/3}$ be the convergence rate in this example. The covariance kernel is written by $H(s_1,s_2)=\frac{1}{2}\{L(s_1,0)+L(0,s_2)-L(s_1,s_2)\}$, where $L(s_1,s_2)=\lim_{n\to\infty}\mathrm{Var}(r_n^2\mathbb{P}_n g_{n,t})$ with $g_{n,t}=f_{n,s_1/r_n}-f_{n,s_2/r_n}$. An expansion implies $n^{-1}\mathrm{Var}(r_n^2 g_{n,t})\to 2\gamma(0)P|x'(s_1-s_2)|$. We can also see that the covariance term $n^{-1}\sum_{m=1}^{\infty}\mathrm{Cov}(r_n^2 g_{n,t}, r_n^2 g_{n,t+m})$ is negligible. Therefore, by a similar argument to Theorem 3, the limiting distribution of the Hough transform estimator with drifting h_n is obtained as

$$(nh_n^2)^{1/3}(\hat{\beta}-\beta_0)\xrightarrow{d}\arg\max_s Z(s),$$

where $Z(s)$ is a Gaussian process with continuous sample paths, expected value $\ddot{\gamma}(0)s'P(|x|xx')s/2$, and covariance kernel $H(s_1,s_2)=2\gamma(0)P|x'(s_1-s_2)|$.

References

[1] Abrevaya, J. and J. Huang (2005) On the bootstrap of the maximum score estimator, *Econometrica*, 73, 1175-1204.

[2] 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.

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

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

[5] 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.

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

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

[8] Chen, X., Hansen, L. P. and M. Carrasco (2010) Nonlinearity and temporal dependence, *Journal of Econometrics*, 155, 155-169.

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

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

[11] 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.

[12] 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.

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

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

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

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

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

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

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

[20] Politis, D. N., Romano, J. P. and M. Wolf (1999) *Subsampling*, New York: Springer-Verlag.

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

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

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

[24] Sen, B., Banerjee, M. and M. Woodroofe (2010) Inconsistency of bootstrap: the Grenander estimator, *Annals of Statistics*, 38, 1953-1977.

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

[26] 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