ASYMPTOTICS FOR MAXIMUM SCORE METHOD UNDER GENERAL CONDITIONS

MYUNG HWAN SEO AND TAISUKE OTSU

Abstract. Since Manski's (1975) seminal work, the maximum score method for discrete choice models has been applied to various econometric problems. Kim and Pollard (1990) established the cube root asymptotics for the maximum score estimator. Since then, however, econometricians posed several open questions and conjectures in the course of generalizing the maximum score approach, such as (a) asymptotic distribution of the conditional maximum score estimator for a panel data dynamic discrete choice model (Honoré and Kyriazidou, 2000), (b) convergence rate of the modified maximum score estimator for an identified set of parameters of a binary choice model with an interval regressor (Manski and Tamer, 2002), and (c) asymptotic distribution of the conventional maximum score estimator under dependent observations. To address these questions, this article extends the cube root asymptotics into four directions to allow (i) criterions drifting with the sample size typically due to a bandwidth sequence, (ii) partially identified parameters of interest, (iii) weakly dependent observations, and/or (iv) nuisance parameters with possibly increasing dimension. 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. This limit theory is applied not only to address the open questions listed above but also to develop a new econometric method, the random coefficient maximum score. Furthermore, our limit theory is applied to address other open questions in econometrics and statistics, such as (d) convergence rate of the minimum volume predictive region (Polonik and Yao, 2000), (e) asymptotic distribution of the least median of squares estimator under dependent observations, (f) asymptotic distribution of the nonparametric monotone density estimator under dependent observations, and (g) asymptotic distribution of the mode regression and related estimators containing bandwidths.

1. Introduction

In a seminal paper, Manski (1975) introduced the maximum score method to estimate parameters in discrete choice models without imposing parametric assumptions on the error terms. Indeed the maximum score estimator is the first semiparametric estimator for limited dependent variable models and has been drawing considerable attention in various contexts of econometrics. One distinguishing feature of the maximum score estimator is that it obeys the cube root asymptotics instead of the conventional squared root (Kim and Pollard, 1990). This feature has also inspired several research

The authors would like to thank Aureo de Paula, Marine Carrasco, Javier Hidalgo, Dennis Kristensen, Benedikt Pötscher, Peter Robinson, Kyungchul Song, and seminar participants at Cambridge, CIREQ Econometrics Conference in Montreal, CREATES in Aarhus University, IHS in Vienna, ISNPS Conference in Cádiz, LSE, Microeconometrics Conference in Hakone, UCL, and University of Surrey and Vienna for helpful comments. Financial support from the ERC Consolidator Grant (SNP 615882) is gratefully acknowledged (Otsu).

agenda in econometrics, such as the smoothed maximum score method (Horowitz, 1992) and failure of bootstrap (Abrevaya and Huang, 2005).

A key of the maximum score method is to explore median or quantile restrictions in disturbances of latent variable models to construct a population criterion that identifies structural parameters of interest. Since Manski (1975), this idea has been generalized to various contexts to cope with different data environments, and these generalizations succeed in producing consistent estimators for the parameters of interest. However, their theoretical properties beyond consistency are often unknown perhaps because the limit theory of Kim and Pollard (1990) is not applicable. There are at least three important examples. First, Honoré and Kyriazidou (2000) considered estimation of a dynamic panel data discrete choice model with unknown error distribution and developed the conditional maximum score estimator. Although they showed consistency of the estimator, its convergence rate and limiting distribution are unknown. The theory of Kim and Pollard (1990) cannot be applied directly due to a bandwidth sequence in the criterion to deal with the unknown error distribution. Second, Manski and Tamer (2002) studied the problem of partial identification for regression models with interval data. For a binary choice model with an interval regressor, Manski and Tamer (2002) proposed the modified maximum score estimator for the identified set of structural parameters. Although they showed consistency of the set estimator, its convergence rate is unknown. The theory of Kim and Pollard (1990) cannot be applied directly due to partial identification of the parameter of interest. In addition, Manski and Tamer's (2002) set estimator contains estimated nuisance parameters for a nonparametric component, which are not allowed in Kim and Pollard (1990). Third, the original maximum score estimator by Manski (1975) calls for extensions to more general data environments. For example, de Jong and Woutersen (2011) studied the asymptotic property of the smoothed maximum score estimator (Horowitz, 1992) under dependent observations. Moon (2004) derived consistency of the maximum score estimator under nonstationary observations. To best of our knowledge, however, the convergence rate and limiting distribution of the original maximum score estimator under dependent observations is still an open question. The limit theory of Kim and Pollard (1990) is not applicable because it is confined to independent observations.

It should be emphasized that addressing these open questions has considerable impacts on the econometrics literature. In the first example, we investigate whether there is a nonparametric counterpart of the cube root asymptotics (i.e., $(nh_n)^{1/3}$-rate with a bandwidth h_n). To best of our knowledge, there is no paper that establishes such a non-standard rate and limiting distribution. Development of such limit theory would be useful for inference and bandwidth selection. In the second example, we complement the set estimation theory of Chernozhukov, Hong and Tamer (2007) by investigating the maximum score-type criterions that involve discontinuities. It is crucial for set estimation to establish its convergence rate since it clarifies the basic requirement on the cutoff value of level set estimation. In the third example, we can study the issue of how data dependence will change the limiting distribution from the independent case. In particular, (lack of) emergence of the long-run covariance in the cube root context has not been explored yet.

To address at least these issues, we extend the scope of cube root asymptotics for M-estimators into four directions to allow (i) criterions drifting with the sample size typically due to a bandwidth sequence, (ii) partially identified parameters, (iii) weakly dependent observations, and/or (iv) nuisance parameters with possibly increasing dimension. We first establish point estimation theory. In particular, we consider an absolutely regular dependent process characterized by β-mixing coefficients and study asymptotic properties of the M-estimator for a class of criterion functions, named the *cube root class*, which induces the cube root asymptotics and allows the criterions to depend on the sample size. In this setup, we establish maximal inequalities to derive the (possibly nonparametric) 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. We establish the $(nh_n)^{1/3}$-rate of convergence of the M-estimator, where h_n usually means a bandwidth sequence, and derive a non-normal limiting distribution. The limit theory is also extended to the M-estimation problems where the criterions contain estimated nuisance parameters. We emphasize that the limit theory with the nonparametric cube root $(nh_n)^{1/3}$-rate is new in the econometrics and statistics literature.

Also, based on the point estimation theory, we extend the cube root class to accommodate partially identified models (named set identified cube root class), such as Manski and Tamer (2002). As in Manski and Tamer (2002) and Chernozhukov, Hong and Tamer (2007), we consider the set estimator defined by a level set of a criterion function. By modifying the maximal inequalities to deal with the set identified class, we characterize the convergence rate of the set estimator under the Hausdorff distance. In our setup, the set estimator converges at the $(nh_n)^{1/4}/\sqrt{\log(nh_n)}$-rate. In contrast to Chernozhukov, Hong and Tamer (2007), our theory allows nuisance parameters with increasing dimension (which is required to cover the model of Manski and Tamer, 2002). We emphasize that even if we do not have a nonparametric nuisance component, it is not trivial to verify the high level assumptions of Chernozhukov, Hong and Tamer (2007), particularly their Condition C.2 (computation of polynomial minorant), in the cube root class. The verification involves the maximal inequality developed in this paper.

Our limit theory is general enough to cover the open questions listed above. In particular, we derive the limiting distribution of Honoré and Kyriazidou's (2000) estimator for the dynamic panel discrete choice model, convergence rate of Manski and Tamer's (2002) set estimator for the binary choice model with an interval regressor, and limiting distribution of Manski's (1975) estimator under dependent observations. Furthermore, our theory can be applied to investigate a new model. As an example, we extend the maximum score estimator for a binary choice model to allow random coefficients (or unknown functions of observables) and derive its asymptotic distribution. In addition, our limit theory is applied to address other open questions in econometrics and statistics. In particular, we establish the convergence rate of the minimum volume predictive region by

Polonik and Yao (2000). Polonik and Yao (2000) established consistency of their predictive region and conjectured its convergence rate. We confirm their conjecture. Also, we derive the asymptotic distribution of the least median of squares estimator (Rousseeuw, 1984) and nonparametric monotone density estimator (Prakasa Rao, 1969) under dependent observations. Finally, we investigate the mode regression and related estimator (Lee, 1989) and establish the limiting distribution of the Hough transform estimator (Goldenshluger and Zeevi, 2004) under drifting tuning constants. All these extensions are important and novel in the literature.[1]

The paper is organized as follows. Section 2 considers point estimation and develops general cube root asymptotic theory. We also consider the case where the criterion contains estimated nuisance parameters. Section 3 extends the obtained theory to set identified models. In Section 4, we apply the generalized cube root asymptotic theory to address the open questions for the maximum score method, such as the conditional maximum score estimator by Honoré and Kyriazidou (2000) (Section 4.1), modified maximum score set estimator by Manski and Tamer (2002) (Section 4.2), conventional maximum score under dependent observations (Section 4.3), and maximum score estimator with random coefficients (Section 4.4). Section 5 presents further illustrations for minimum volume predictive region (Section 5.1), least median of squares (Section 5.2), monotone density estimator (Section 5.3), and mode and related estimators (Section 5.4). All proofs are contained in the Appendix.

2. General cube root asymptotics

This section extends Kim and Pollard's (1990) main theorem on the cube root asymptotics of the M-estimator to allow for drifting criterions typically due to a bandwidth sequence and dependent data. Consider the M-estimator $\hat{\theta}$ that maximizes the random criterion

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

where $\{f_{n,\theta} : \theta \in \Theta\}$ is a sequence of classes 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), which 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 $\{f_{n,\theta} : \theta \in \Theta\}$ is called the cube root class if Conditions (i)-(iii) below are satisfied with a sequence $\{h_n\}$ of positive numbers such that $nh_n \to \infty$.*

[1]For the least median of squares, Zinde-Walsh (2002) employed a generalized function approach to characterize the limiting behavior of the least median of squares estimator under dependent observations by using a smoothed estimator. The asymptotic analysis based on empirical process theory as in Kim and Pollard (1990) under dependent observations is still open.

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

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

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*

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

for all n *large enough,* $\varepsilon > 0$ *small enough, and* θ' *in a neighborhood of* θ_0*.*

This definition covers not only the case of $h_n = 1$ where the criterion function $f_{n,\theta}$ does not vary with n (called the non-drifting case), but also the case of $h_n \to 0$ where it does due to the sequence h_n (called the drifting case). For the drifting case, the sequence $\{h_n\}$ is usually a bandwidth sequence to deal with some nonparametric component. Kim and Pollard (1990) focused on the non-drifting case with independent observations. Compared to Kim and Pollard (1990), our conditions consist of directly verifiable moment conditions without demanding knowledge on the empirical process theory such as "uniform manageability". This is due to the insight that the envelope condition (iii) is sufficient to control the size of entropy.

Condition (i) contains boundedness and point identification conditions. Boundedness of the class $\{h_n f_{n,\theta} : \theta \in \Theta\}$ is a major requirement. 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 norm equivalence relationship between the $L_2(P)$-norm $\|f_{n,\theta} - f_{n,\theta_0}\|_2$ and so-called $L_{2,\beta}(P)$-norm $\|f_{n,\theta} - f_{n,\theta_0}\|_{2,\beta}$ using β-mixing coefficients for the observations $\{z_t\}$ (defined in (14) in the Appendix). It should be noted that $\|f_{n,\theta} - f_{n,\theta_0}\|_2 = \|f_{n,\theta} - f_{n,\theta_0}\|_{2,\beta}$ for independent observations.[2] See a remark on Theorem 1 for further discussion. The identification condition for θ_0 is standard and similar to Kim and Pollard (1990, Conditions (ii) and (iv)) of their main theorem. This condition should be checked for each application. In the next section we relaxes the point identification assumption.

When the criterion $f_{n,\theta}$ involves a kernel estimate for a nonparametric component, h_n is considered as a bandwidth parameter. In this case, the criterion typically 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 $z = (y, x)$ with some function m and kernel K (see Sections 4.1, 4.4, and 5.1

[2] Because of this norm equivalence, Kim and Pollard (1990) (focused on the non-drifting case with independent observations) did not need to impose boundedness.

for examples). In this case, boundedness of $\{h_n f_{n,\theta} : \theta \in \Theta\}$ means that of $K\left(\frac{x-c}{h_n}\right) m(y,x,\theta)$. The expansion in (1) is understood as a restriction for the integral $P(f_{n,\theta} - f_{n,\theta_0}) = \int\int K(a)\, m(y, c + h_n a, \theta) p_{yx}(y, c + h_n a) da dy$ 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 Condition (iii) are understood in the same manner.

Condition (ii) is required not only for the maximal inequality to derive the convergence rate of the M-estimator but also for finite dimensional convergence to derive the limiting distribution. In particular, this condition is used to relate the $L_2(P)$-norm $\|f_{n,\theta} - f_{n,\theta_0}\|_2$ of the contrast of criterions to the Euclidean norm $|\theta - \theta_0|$ over Θ. This condition is implicit in Kim and Pollard (1990, Condition (v)) on non-degeneracy of the covariance kernel combined with the norm equivalence $\|f_{n,\theta} - f_{n,\theta_0}\|_2 = \|f_{n,\theta} - f_{n,\theta_0}\|_{2,\beta}$ under independent observations. Condition (ii) is often verified in the course of verifying the expansion (1) in Condition (i).

Condition (iii) is an envelope condition for the class $\{f_{n,\theta} - f_{n,\theta'} : |\theta - \theta'| \leq \varepsilon\}$ of contrasts. 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 instead of ε. This condition merges three conditions in Kim and Pollard (1990): their envelope conditions ((vi) and (vii)) and uniform manageability of the class $\{f_{n,\theta} - f_{n,\theta_0} : |\theta - \theta_0| \leq \varepsilon\}$. It is often the case that verifying the envelope condition for arbitrary θ' in a neighborhood of θ_0 is not more demanding than for θ_0.[3]

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

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

The first step is to establish weak consistency of the M-estimator, i.e., $\hat{\theta}$ converges in probability to the unique maximizer θ_0 of $\lim_{n\to\infty} P f_{n,\theta}$. The technical argument to derive the weak consistency is rather standard and usually shown by establishing uniform convergence of the objective function $\sup_{\theta\in\Theta} |\mathbb{P}_n f_{n,\theta} - P f_{n,\theta}| \xrightarrow{p} 0$. Thus, in this section we assume the consistency of $\hat{\theta}$. See illustrations in Sections 4 and 5 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 $\{\mathbb{G}_n h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) : \theta \in \Theta\}$ by certain maximum inequality, where $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$ for a function f. For non-drifting criterion functions under independent observations, several maximal inequalities are available in the literature (see, e.g., Kim and Pollard, 1990, p. 199). For drifting criterion functions under possibly dependent observations, to best of our knowledge, there is no maximal inequality which

[3]Condition (iii) may be relaxed slightly as follows. The upper bound in Condition (iii) is $C''\varepsilon$ for $\theta' = \theta_0$ and is $C''\varepsilon^{1/p}$ with some $p > 0$ for $\theta' \neq \theta_0$.

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.[4] 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 $0 < \rho < 1$.*

Obviously this assumption covers the case of independent observations studied in Kim and Pollard (1990). Assumption D says the mixing coefficient β_m should decay at an exponential rate.[5] For example, various Markov, GARCH, and stochastic volatility models satisfy this assumption (Carrasco and Chen, 2002). 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.

Under Assumption D, we obtain the following maximal inequality for the empirical process $\mathbb{G}_n h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0})$ of the cube root class.

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

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

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

This maximal inequality is applied to obtain the following lemma.

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

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

for all $\theta \in \Theta$ satisfying $(nh_n)^{-1/3} \leq |\theta - \theta_0| \leq C$.

Based on Lemma 1, the convergence rate of $\hat{\theta}$ is obtained as follows. Suppose $|\hat{\theta} - \theta_0| \geq (nh_n)^{-1/3}$. Then we can take a positive constant c such that

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

[4] See Doukhan, Massart and Rio (1995) for a detail on empirical process theory of absolutely regular processes.

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

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

Given the convergence rate of $\hat{\theta}$, the final step is to derive the limiting distribution. To this end, we 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 empirical 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 $K > 0$. Weak convergence of the process Z_n may be characterized by its finite dimensional convergence and tightness (or stochastic equicontinuity). If $f_{n,\theta}$ does not vary with n and $\{z_t\}$ 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 our cube root class under possibly 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. 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 $\beta(\cdot)$. Also let $Q_g(u)$ be the inverse function of the tail probability function $x \mapsto P\{|g(z_t)| > x\}$.

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

$$\sup_n \int_0^1 \beta^{-1}(u) Q_{g_n}(u)^2 du < \infty. \tag{3}$$

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

The finite dimensional convergence of Z_n follows from Lemma C by setting g_n as any finite dimensional projection of the process $\{g_{n,s} - Pg_{n,s} : s\}$ with $g_{n,s} = n^{1/6} h_n^{2/3}(f_{n,\theta_0+s(nh_n)^{-1/3}} - f_{n,\theta_0})$. The requirement in (3) is an adaptation of the Lindeberg-type condition of Rio's (1997, Corollary 1) central limit theorem to our setup. The condition (3) 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) showed that any polynomial mixing rate will destroy the asymptotic normality of $\mathbb{G}_n g_n$. It should be noted that for the rescaled object $g_{n,s} = n^{1/6} h_n^{2/3}(f_{n,\theta_0+s(nh_n)^{-1/3}} - f_{n,\theta_0})$ based on the cube root class $\{f_{n,\theta}\}$, the $(2+\delta)$-th moments $P|g_{n,s}|^{2+\delta}$ for $\delta > 0$ typically diverge. This happens because the cube root class $\{f_{n,\theta}\}$ 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 conditions higher than two. To verify the Lindeberg-type condition (3), the following lemma is often useful.

Lemma 2. *Suppose there is a positive constant c such that $P\{|g_{n,s}| \geq c\} \leq c(nh_n^{-2})^{-1/3}$ for all n large enough and s. Then (3) holds true.*

Note that in the cube root class, $g_{n,s}$ is typically a difference between two indicators multiplied by $n^{1/6}h_n^{2/3}$ and possibly a kernel weight. Therefore, $g_{n,s}$ is zero or close to zero with high probability so that the condition $P\{|g_{n,s}| \geq c\} \leq c(nh_n^{-2})^{-1/3}$ can be satisfied.

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 positive constant C such that*

$$P \sup_{s:|s-s'|<\varepsilon} |g_{n,s} - g_{n,s'}|^2 \leq C\varepsilon, \tag{4}$$

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

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

for all $n \geq N_\delta$.

Tightness of the process Z_n follows from Lemma M' by setting $g_{n,s} = n^{1/6}h_n^{2/3}(f_{n,\theta_0+s(nh_n)^{-1/3}} - f_{n,\theta_0})$. Note that the condition (4) is satisfied by Condition (iii) of the cube root class.[6] Compared to Lemma M used to derive the convergence rate of $\hat{\theta}$, 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.[7]

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. *Let $\{f_{n,\theta} : \theta \in \Theta\}$ be a cube root class. Suppose that Assumption D holds, $\hat{\theta}$ defined in (2) converges in probability to $\theta_0 \in \text{int}\Theta$, and (3) 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*

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

[6]The upper bound in (4) can be relaxed to $\varepsilon^{1/p}$ for some $1 \leq p < \infty$. However, for the cube root class it is typically satisfied with $p = 1$.

[7]In particular, the process Z_n itself does not satisfy Condition (ii).

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 the cases where the criterion function $f_{n,\theta}$ can vary with the sample size n and/or the observations $\{z_t\}$ can obey a dependent process. To best of our knowledge, the cube root (nonparametric) convergence rate $(nh_n)^{1/3}$ with $h_n \to 0$ is new in econometrics and statistics literature. It is interesting to note that similar to standard nonparametric estimation, nh_n still plays a role as the "effective sample size." Also, similar to Kim and Pollard (1990), the limiting distribution is characterized by the maximizer of the Gaussian process Z. However, the covariance kernel H takes the form of the long-run variance. On the other hand, when the condition of Lemma 2 holds true, the maximal inequality in Lemma M' often implies that the asymptotic temporal covariances are negligible. For example, in Section 4.3, we find that the covariance kernel H of the maximum score estimator under dependent observations is same as the independent case.

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. 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). It is beyond the scope of this paper to investigate bootstrap inference in our setup.

We now discuss the boundedness requirement on $h_n f_{n,\theta}$ in Condition (i). Boundedness is used to show the maximal inequality in Lemma M particularly to guarantee the norm relation $\|\cdot\|_{2,\beta} \le C\|\cdot\|_2$ for some positive constant C (see (17) in the Appendix). Note that it always holds $\|\cdot\|_2 \le \|\cdot\|_{2,\beta}$ (Doukhan, Massart and Rio, 1995, Lemma 1). Thus, boundedness of $h_n f_{n,\theta}$ is used to guarantee the norm equivalence relationship between the $L_2(P)$- and $L_{2,\beta}(P)$-norms. 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, instead of Lemma M, the resulting maximal inequality will be

$$P \sup_{\theta\in\Theta:|\theta-\theta_0|<\delta} |\mathbb{G}_n h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0})| \le C\delta^{1/(2+\eta)},$$

provided Condition (iii) is replaced with

$$P \sup_{\theta\in\Theta:|\theta-\theta'|<\varepsilon} h_n^{1+\eta/2} |f_{n,\theta} - f_{n,\theta'}|^{2+\eta} \le C''\varepsilon,$$

for some positive constant C'' and all n large enough, $\varepsilon > 0$ small enough, and θ' in a neighborhood of θ_0. By applying a similar argument to show Lemma M, we can show $\hat{\theta} - \theta_0 = O_p((nh_n)^{-\frac{1}{4}-\frac{1}{6(2+\eta)}})$ for any $\eta > 0$ without boundedness of $h_n f_{n,\theta}$. However, this convergence rate may not be sharp.

Also note that the exponential decay of the mixing coefficient β_m in Assumption D is used to show the norm equivalence (17) in the Appendix. Instead of Assumption D (which imposes $\beta_m = O(\rho^m)$), Lemma M can be shown under a slightly weaker condition: $\sup_{0<x\leq 1} x^{-1}\int_0^x \beta^{-1}(u)du < \infty$.[8] However, this weaker condition already excludes polynomial decay of β_m.

Let us close this section with an extension of Theorem 1. It is often the case that the criterion function contains some nuisance parameters which can be estimated with rates faster than $(nh_n)^{-1/3}$. For the rest of this section, let $\hat{\theta}$ and $\tilde{\theta}$ satisfy

$$\begin{aligned}\mathbb{P}_n f_{n,\hat{\theta},\hat{\nu}} &\geq \sup_{\theta\in\Theta}\mathbb{P}_n f_{n,\theta,\hat{\nu}} + o_p((nh_n)^{-2/3}),\\ \mathbb{P}_n f_{n,\tilde{\theta},\nu_0} &\geq \sup_{\theta\in\Theta}\mathbb{P}_n f_{n,\theta,\nu_0} + o_p((nh_n)^{-2/3}),\end{aligned}$$

respectively, where ν_0 is a vector of nuisance parameters and $\hat{\nu}$ is its estimator satisfying $\hat{\nu}-\nu_0 = o_p((nh_n)^{-1/3})$. Theorem 1 is extended as follows.

Theorem 2. *Let $\{f_{n,\theta,\nu_0} : \theta\in\Theta\}$ be a cube root class and $\{f_{n,\theta,\nu} : \theta\in\Theta, \nu\in\Lambda\}$ satisfies Condition (iii). Suppose there exists some negative definite matrix V_1 such that*

$$P(f_{n,\theta,\nu} - f_{n,\theta_0,\nu_0}) = \frac{1}{2}(\theta-\theta_0)'V_1(\theta-\theta_0) + o(|\theta-\theta_0|^2) + O(|\nu-\nu_0|^2) + o((nh_n)^{-2/3}), \quad (5)$$

for all θ and ν in neighborhoods of θ_0 and ν_0, respectively. Then $\hat{\theta} = \tilde{\theta} + o_p((nh_n)^{-1/3})$. Additionally, if (3) 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 $s'V_1 s/2$ and covariance kernel $H(s_1,s_2) = \lim_{n\to\infty}\sum_{t=-n}^{n}\mathrm{Cov}(g_{n,s_1}(z_0), g_{n,s_2}(z_t)) < \infty$.

3. Set identification

In this section, we relax the assumption of point identification of θ_0, the maximizer of the limiting population criterion $\lim_{n\to\infty} Pf_{n,\theta}$, and consider the case where the limiting criterion is maximized at any element of a set $\Theta_I \subset \Theta$. The set Θ_I is called the identified set. In order to estimate Θ_I, we consider a collection of approximate maximizers of the sample criterion function $\mathbb{P}_n f_{n,\theta}$, that is

$$\hat{\Theta} = \{\theta\in\Theta : \max_{\theta\in\Theta}\mathbb{P}_n f_{n,\theta} - \mathbb{P}_n f_{n,\theta} \leq \hat{c}(nh_n)^{-1/2}\},$$

i.e., the level set based on the criterion $f_{n,\theta}$ from the maximum with a cutoff value $\hat{c}(nh_n)^{-1/2}$. This section studies the convergence rate of $\hat{\Theta}$ to Θ_I under the Hausdorff distance defined below. We assume that Θ_I is convex. Then the projection $\pi_\theta = \arg\min_{\theta'\in\Theta_I}|\theta'-\theta|$ of $\theta\in\Theta$ on Θ_I is

[8] For example, consider the case where g is a binary function (takes 0 or 1). In this case, we have $\|g\|_{2,\beta} = \|g\|_2\sqrt{x^{-1}\int_0^x \beta^{-1}(u)du}$ with $x = P\{g(z_t) = 1\}$. Therefore, we cannot bound $\|g\|_{2,\beta}$ from the above unless $\sup_{0<x\leq 1} x^{-1}\int_0^x \beta^{-1}(u)du < \infty$.

uniquely defined. To deal with the set identified case, we modify the definition of the cube root class as follows.

Definition (Set identified cube root class). *A class of functions $\{f_{n,\theta} : \theta \in \Theta\}$ is called the set identified cube root class if Conditions (i)-(iii) below are satisfied with a sequence $\{h_n\}$ of positive numbers such that $nh_n \to \infty$.*

(i): *$\{h_n f_{n,\theta} : \theta \in \Theta\}$ is a class of uniformly bounded functions. Also, $\lim_{n\to\infty} Pf_{n,\theta}$ is maximized at any θ in a bounded convex set Θ_I. There exist positive constants c and c' such that*

$$P(f_{n,\pi_\theta} - f_{n,\theta}) \geq c|\theta - \pi_\theta|^2 + o(|\theta - \pi_\theta|^2) + o((nh_n)^{-2/3}), \tag{6}$$

for all n large enough and all $\theta \in \{\theta \in \Theta : 0 < |\theta - \pi_\theta| \leq c'\}$.

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

$$|\theta - \pi_\theta| \leq Ch_n^{1/2} \|f_{n,\theta} - f_{n,\pi_\theta}\|_2 ,$$

for all n large enough and all $\theta \in \{\theta \in \Theta : 0 < |\theta - \pi_\theta| \leq C'\}$.

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

$$P \sup_{\theta\in\Theta:0<|\theta-\pi_\theta|<\varepsilon} h_n|f_{n,\theta} - f_{n,\pi_\theta}|^2 \leq C''\varepsilon,$$

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

For the non-drifting case, we set as $h_n = 1$. Similar comments to the cube root class apply. The main difference is that the conditions are imposed on the contrast $f_{n,\theta} - f_{n,\pi_\theta}$ using the projection π_θ. Condition (i) contains boundedness and expansion conditions. The inequality in (6) can be verified by a one-sided Taylor expansion based on the directional derivative. Conditions (ii) and (iii) play similar roles and are verified by similar arguments to the point identified case. In contrast to the point identified case, Condition (iii) does not require θ' in a neighborhood of θ_0.

We establish the following maximal inequality for the set identified cube root class. Let $r_n = nh_n/\log(nh_n)$.

Lemma MS. *There exist positive constants C and $C' < 1$ such that*

$$P \sup_{\theta\in\Theta:0<|\theta-\pi_\theta|<\delta} |\mathbb{G}_n h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta})| \leq C(\delta \log(1/\delta))^{1/2},$$

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

Compared to Lemma M, the additional log term on the right hand side is due to the fact that the supremum is taken over the δ-tube (or manifold) instead of the δ-ball, which increases the entropy. This maximal inequality is applied to obtain the following analog of Lemma 1.

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

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

for all $\theta \in \{\theta \in \Theta : r_n^{-1/3} \leq |\theta - \pi_\theta| \leq C\}$.

We now establish the convergence rate of the set estimator $\hat{\Theta}$ to Θ_I. Let $\rho(A, B) = \sup_{a \in A} \inf_{b \in B} |a - b|$ and $H(A, B) = \max\{\rho(A, B), \rho(B, A)\}$ be the Hausdorff distance of sets $A, B \subset \mathbb{R}^d$. Based on Lemmas MS and 3, the asymptotic property of the set estimator $\hat{\Theta}$ is obtained as follows.

Theorem 3. *Let $\{f_{n,\theta} : \theta \in \Theta\}$ be a set identified cube root class, and $\{h_n^{1/2} f_{n,\theta} : \theta \in \Theta_I\}$ be a P-Donsker class. Assume $H(\hat{\Theta}, \Theta_I) \xrightarrow{p} 0$ and $\hat{c} = o_p((nh_n)^{1/2})$. Then*

$$\rho(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4} + r_n^{-1/3}).$$

Furthermore, if $\hat{c} \to \infty$, then $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ and

$$H(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4}).$$

Note that ρ is asymmetric in its arguments. The first part of this theorem says $\rho(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4} + r_n^{-1/3})$. On the other hand, in the second part, we show $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ (i.e., $\rho(\Theta_I, \hat{\Theta})$ can converge to zero at an arbitrary rate) as far as $\hat{c} \to \infty$. For example, we may set at $\hat{c} = \log(nh_n)$. These results are combined to imply the convergence rate $H(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4})$ under the Hausdorff distance. The cube root term of order $r_n^{-1/3}$ in the rate of $\rho(\hat{\Theta}, \Theta_I)$ is dominated by the term of order $\hat{c}^{1/2}(nh_n)^{-1/4}$.

We next consider the case where the criterion function contains nuisance parameters. In particular, we allow that the dimension k_n of the nuisance parameters ν can grow as the sample size increases. For instance, the nuisance parameters might be coefficients in sieve estimation. It is important to allow the growing dimension of ν to cover Manski and Tamer's (2002) set estimator, where the criterion contains some nonparametric estimate and its transform by the indicator. The rest of this section considers the set estimator

$$\hat{\Theta} = \{\theta \in \Theta : \max_{\theta \in \Theta} \mathbb{P}_n f_{n,\theta,\hat{\nu}} - \mathbb{P}_n f_{n,\theta,\hat{\nu}} \leq \hat{c}(nh_n)^{-1/2}\},$$

with some preliminary estimator $\hat{\nu}$ and cutoff value $\hat{c}$. To derive the convergence rate of $\hat{\Theta}$, we establish a maximal inequality over the set $\mathcal{G}_n = \{g_{n,\theta,\nu} - Pg_{n,\theta,\nu} : |\theta - \theta_0| \leq K_1, |\nu - \nu_0| \leq K_2 a_n\}$ for some $K_1, K_2 > 0$ with envelope functions G_n, where $k_n \to \infty$, $a_n \to 0$, and $k_n a_n^{2/3} \to 0$. Let $g_{n,s} = g_{n,\theta,\nu}$ with $s = (\theta', \nu')'$. Suppose the preliminary estimator $\hat{\nu}$ satisfies $\hat{\nu} - \nu_0 = O_p(n^{-1/2}(k_n \log k_n)^{1/2})$. Then, we may set as $a_n = n^{-1/2}(k_n \log k_n)^{1/2}$ and $g_{n,s} = a_n^{1/2} h_n^{1/2}(f_{n,\theta,\nu} - f_{n,\theta,\nu_0})$, and the condition $k_n a_n^{2/3} \to 0$ is guaranteed by $k_n^4 \log k_n / n \to 0$. Lemma MS is modified as follows.

Lemma MS'. *Suppose there exists a positive constant C such that*

$$P \sup_{s:|s-s'|<\varepsilon} |g_{n,s} - g_{n,s'}|^2 \leq C\varepsilon, \tag{7}$$

for all n large enough 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

$$(\log k_n)^{-1} P \sup_{|s-s'|<\delta} |\mathbb{G}_n g_{n,s} - \mathbb{G}_n g_{n,s'}| \leq \sigma,$$

for all $n \geq N_\delta$.

Similar comments to Lemma M' apply. The term $(\log k_n)^{-1}$ is the cost due to the growing dimension of nuisance parameters. Based on this lemma, the convergence rate of the set estimator $\hat{\Theta}$ is characterized as follows.

Theorem 4. *Let $\{f_{n,\theta,\nu_0} : \theta \in \Theta\}$ be a set identified cube root class, and $\{h_n^{1/2} f_{n,\theta,\nu_0} : \theta \in \Theta_I\}$ be a P-Donsker class. Suppose there exists a positive constant C'' such that*

$$P \sup_{|\nu-\nu_0|<\varepsilon} \sup_{\theta\in\Theta:|\theta-\pi_\theta|<\varepsilon} h_n |f_{n,\theta,\nu} - f_{n,\pi_\theta,\nu_0}|^2 \leq C''\varepsilon, \tag{8}$$

for all n large enough and all $\varepsilon > 0$ small enough. Assume $\rho(\hat{\Theta}, \Theta_I) \xrightarrow{p} 0$, $\hat{c} = o_p((nh_n)^{1/2})$, $k_n \to \infty$, $|\hat{\nu} - \nu_0| = o_p(a_n)$ for some $\{a_n\}$ such that $h_n/a_n \to \infty$ and $k_n a_n^{2/3} \to 0$. Furthermore, for some $\varepsilon > 0$ and for each $\theta \in \{\theta \in \Theta : |\theta - \pi_\theta| < \varepsilon\}$ and ν in a neighborhood of ν_0, it holds

$$P(f_{n,\theta,\nu} - f_{n,\theta,\nu_0}) - P(f_{n,\pi_\theta,\nu} - f_{n,\pi_\theta,\nu_0}) = o(|\theta - \pi_\theta|^2) + O(|\nu - \nu_0|^2 + r_n^{-2/3}). \tag{9}$$

Then

$$\rho(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4} + r_n^{-1/3} + (nh_n a_n^{-1})^{-1/4}(\log k_n)^{1/2}) + o(a_n).$$

Furthermore, if $\hat{c} \to \infty$, then $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ and

$$H(\hat{\Theta}, \Theta_I) = O_p(\hat{c}^{1/2}(nh_n)^{-1/4} + (nh_n a_n^{-1})^{-1/4}(\log k_n)^{1/2}) + o(a_n).$$

Compared to Theorem 3, we have an extra term in the convergence rate of $H(\hat{\Theta}, \Theta_I)$ due to (nonparametric) estimation of ν_0. For example, if $h_n = 1$ and $\hat{\nu}$ is a vector of coefficients for sieve estimation, then it is often the case that $a_n = o(n^{-1/4})$ (see, e.g., Chen, 2007).[9] In this case, the convergence rate under the Hausdorff distance becomes $H(\hat{\Theta}, \Theta_I) = O_p(n^{-1/4}\hat{c}^{1/2})$ as far as $\hat{c} \to \infty$. For example, we can set as $\hat{c} = \log n$. The envelope condition in (8) allows for step functions containing some nonparametric estimates.

Let us close this section by some remarks on the relationship to Chernozhukov, Hong and Tamer (2007), which established general asymptotic theory for the criterion-based set inference method.

[9] Alternatively ν_0 can be estimated by some high-dimensional method (e.g. Belloni, Chen, Chernozhukov and Hansen, 2012), which also typically guarantees $a_n = o(n^{-1/4})$.

First, our Theorem 3 basically verifies their high level condition (Condition C.2 of Chernozhukov, Hong and Tamer, 2007) for the convergence rate of the general level set estimator in the set identified cube root class. Second, as motivating examples, they focused on the moment condition models (Section 4 of Chernozhukov, Hong and Tamer, 2007) and applied their general set inference theory. Here we focus on the criterion for the M-estimation. Consider the non-drifting case $h_n = 1$. Although the GMM-type criterion function is typically of order $O_p(n^{-1})$ on Θ_I, the M-estimation criterion is of order $O_p(n^{-1/2})$. This difference yields a rather slower convergence rate $H(\hat{\Theta}, \Theta_I) = O_p((\log n)^{1/2} n^{-1/4})$ in Theorem 3 instead of $H(\hat{\Theta}, \Theta_I) = O_p((\log n)^{1/2} n^{-1/2})$ in Chernozhukov, Hong and Tamer (2007, Theorem 4.1). However, in our setup, this is an efficient rate in the sense of the fastest convergence rate that preserves consistency. Third, we allow the criterion to contain the drifting bandwidth h_n and nuisance parameters with increasing dimension. Finally, the maximal inequality in Lemma MS' and the assumption that $\{h_n^{1/2} f_{n,\theta,\nu_0} : \theta \in \Theta_I\}$ is P-Donsker are sufficient to verify Conditions C.4 and C.5 of Chernozhukov, Hong and Tamer (2007). Thus their subsampling confidence set is also valid in our setup.

4. Generalizations of maximum score method

4.1. **Dynamic panel: Asymptotic distribution of Honoré and Kyriazidou's (2000) estimator.** As an illustration of Theorem 1, we consider a dynamic panel data model with a binary dependent variable

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

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. We observe $\{y_{it}, x_{it}\}$ but do not observe α_i. Honoré and Kyriazidou (2000) proposed the conditional maximum score estimator for (β_0, γ_0),

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

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 1 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 cube root class with $h_n = b_n^k$. We first check Condition (ii). By the definition of $z_2 = y_2 - y_1$ (which can take -1, 0, or 1) and change of variables $a = b_n^{-1} z_1$, we obtain

$$E[e_n(z)^2|z_3] = 4h_n^{-1} \int K(a)^2 p_1(b_n a | z_2 \neq 0, z_3) da P\{z_2 \neq 0|z_3\},$$

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), $h_n E[e_n(z)^2|z_3] > c$ 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 4.3 below using (a). Similarly, Condition (iii) is verified as

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

for some positive constants C_1 and C_2 and all ϑ in a neighborhood of θ_0 and n large enough. We now verify Condition (i). Since $h_n f_{n,\theta}$ is clearly bounded, it is enough to verify (1). A change of variables $a = b_n^{-1} z_1$ and (b) imply

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

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 (1), it is enough to derive a second order expansion of $E[z_2 \{\text{sgn}(z_3'\theta) - \text{sgn}(z_3'\theta_0)\}|z_1 = 0]$. 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

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

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

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

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 (1) is satisfied with the negative definite matrix

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

Finally the covariance kernel H is obtained in the same manner as 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

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

4.2. **Partial identification under interval regressor: Convergence rate of Manski and Tamer's (2002) set estimator.** As an illustration of Theorem 4, we consider Manski and Tamer's (2002) modified maximum score estimator for the identified set of parameters in a binary choice model with an interval regressor. The binary outcome is determined by $y=\mathbb{I}\{x'\theta_0+w+u\geq 0\}$, where x is a vector of observable regressors, w is an unobservable regressor, and u is an unobservable error term satisfying $P\{u\leq 0|x,w\}=\alpha$ (we set $\alpha=.5$ to simplify the notation). Instead of w, we observe the interval $[w_l,w_u]$ such that $P\{w_l\leq w\leq w_u\}=1$. Here we normalize that the coefficient of w to determine y equals one. In this setup, the parameter θ_0 is partially identified and its identified set is written as (Manski and Tamer 2002, Proposition 2)

$$\Theta_I=\{\theta\in\Theta:P\{x'\theta+w_u\leq 0<x'\theta_0+w_l\text{ or }x'\theta_0+w_u\leq 0<x'\theta+w_l\}=0\}.$$

Let $\tilde{x}=(x',w_l,w_u)'$ and $q_{\hat{\nu}}(\tilde{x})$ be an estimator for $P\{y=1|\tilde{x}\}$ with the estimated parameters $\hat{\nu}$. Suppose $P\{y=1|\tilde{x}\}=q_{\nu_0}(\tilde{x})$. By exploring the maximum score approach, Manski and Tamer (2002) developed the estimator $\hat{\Theta}=\{\theta\in\Theta:\max_{\theta\in\Theta}S_n(\theta)-S_n(\theta)\leq\epsilon_n\}$ for Θ_I, where

$$S_n(\theta)=\mathbb{P}_n(y-.5)[\mathbb{I}\{q_{\hat{\nu}}(\tilde{x})>.5\}\text{sgn}(x'\theta+w_u)+\mathbb{I}\{q_{\hat{\nu}}(\tilde{x})\leq .5\}\text{sgn}(x'\theta+w_l)].$$

Manski and Tamer (2002) established the consistency of $\hat{\Theta}$ to Θ_I under the Hausdorff distance. To establish the consistency, they assumed the cutoff value ϵ_n is bounded from below by the (almost sure) decay rate of $\sup_{\theta\in\Theta}|S_n(\theta)-S(\theta)|$, where $S(\theta)$ is the limiting object of $S_n(\theta)$. As Manski and Tamer (2002, Footnote 3) argued, characterization of the decay rate is a complex task because

$S_n(\theta)$ is a step function and $\mathbb{I}\{q_{\hat{\nu}}(\tilde{x}) > .5\}$ is a step function transform of the nonparametric estimate of $P\{y = 1|\tilde{x}\}$. Therefore, it has been an open question. Obtaining the lower bound rate of ϵ_n is important because we wish to minimize the volume of the estimator $\hat{\Theta}$ without losing the asymptotic validity. Here we explicitly characterize the decay rate for the lower bound of ϵ_n and establish the convergence rate of this estimator.

A little algebra shows that the estimator is written as

$$\hat{\Theta} = \{\theta \in \Theta : \max_{\theta \in \Theta} \mathbb{P}_n f_{\theta,\hat{\nu}} - \mathbb{P}_n f_{\theta,\hat{\nu}} \leq \hat{c} n^{-1/2}\},$$

where $z = (x', w, w_l, w_u, u)'$, $h(x, w, u) = \mathbb{I}\{x'\theta_0 + w + u \geq 0\} - \mathbb{I}\{x'\theta_0 + w + u < 0\}$, and

$$f_{\theta,\nu}(z) = h(x, w, u)[\mathbb{I}\{x'\theta + w_u \geq 0, q_\nu(\tilde{x}) > .5\} - \mathbb{I}\{x'\theta + w_l < 0, q_\nu(\tilde{x}) \leq .5\}].$$

We impose the following assumptions. Let $\partial\Theta_I$ be the boundary of Θ_I, $\kappa_u(\tilde{x}) = (2q_{\nu_0}(\tilde{x}) - 1)\mathbb{I}\{q_{\nu_0}(\tilde{x}) > .5\}$, and $\kappa_l(\tilde{x}) = (1 - 2q_{\nu_0}(\tilde{x}))\mathbb{I}\{q_{\nu_0}(\tilde{x}) \leq .5\}$.

- **(a):** $\{x_t, w_t, w_{lt}, w_{ut}, u_t\}$ satisfies Assumption D. $x|w_u$ has a bounded and continuous conditional density $p(\cdot|w_u)$ for almost every w_u. There exists an element x_j of x whose conditional density $p(x_j|w_u)$ is bounded away from zero over the support of of w_u. The same condition holds for $x|w_l$. The conditional densities of $w_u|w_l, x$ and $w_l|w_u, x$ are bounded. $q_\nu(\cdot)$ is continuously differentiable at ν_0 a.s. and the derivative is bounded for almost every $\tilde{x}$.
- **(b):** For each $\theta \in \partial\Theta_I$, $\kappa_u(\tilde{x})$ is non-negative for $x'\theta + w_u \geq 0$, $\kappa_l(\tilde{x})$ is non-positive for $x'\theta + w_l \leq 0$, $\kappa_u(\tilde{x})$ and $\kappa_l(\tilde{x})$ are continuously differentiable, and it holds

$$P\{x'\theta + w_u = 0, q_{\nu_0}(\tilde{x}) > .5, (\theta'\partial\kappa_u(\tilde{x})/\partial x)p(x|w_l, w_u) > 0\} > 0, \text{ or}$$
$$P\{x'\theta + w_l = 0, q_{\nu_0}(\tilde{x}) \leq .5, (\theta'\partial\kappa_l(\tilde{x})/\partial x)p(x|w_l, w_u) > 0\} > 0.$$

To apply Theorem 4, we verify that $\{f_{\theta,\nu_0} : \theta \in \Theta\}$ belongs to the set identified cube root class with $h_n = 1$. We first check Condition (i). This class is clearly bounded. From Manski and Tamer (2002, Lemma 1 and Corollary (a)), Pf_{θ,ν_0} is maximized at any $\theta \in \Theta_I$ and Θ_I is a bounded convex set. By applying the argument in Kim and Pollard (1990, pp. 214-215), the second directional derivative at $\theta \in \partial\Theta_I$ with the orthogonal direction outward from Θ_I is

$$-2P\int \mathbb{I}\{x'\theta = -w_u\}\theta'\frac{\partial\kappa_u(\tilde{x})}{\partial x}p(x|w_l, w_u)(x'\theta)^2 d\sigma_u - 2P\int \mathbb{I}\{x'\theta = -w_l\}\theta'\frac{\partial\kappa_u(\tilde{x})}{\partial x}p(x|w_l, w_u)(x'\theta)^2 d\sigma_l,$$

where σ_u and σ_l are the surface measures on the boundaries of the sets $\{x : x'\pi_\theta + w_u \geq 0\}$ and $\{x : x'\pi_\theta + w_l \geq 0\}$, respectively. Since this matrix is negative definite by (b), Condition (i) is verified. We next check Condition (ii). By $h(x, w, u)^2 = 1$, observe that

$$\|f_{\theta,\nu_0} - f_{\pi_\theta,\nu_0}\|_2 \geq \sqrt{2}\min\left\{\begin{array}{c} P\{x'\theta \geq -w_u \geq x'\pi_\theta \text{ or } x'\theta < -w_u < x'\pi_\theta\}\mathbb{I}\{q_{\nu_0}(\tilde{x}) > .5\}, \\ P\{x'\theta \geq -w_l \geq x'\pi_\theta \text{ or } x'\theta < -w_l < x'\pi_\theta\}\mathbb{I}\{q_{\nu_0}(\tilde{x}) \leq .5\} \end{array}\right\}.$$

for any $\theta \in \Theta$. Since the right hand side is the minimum of probabilities for pairs of wedge shaped regions with angles of order $|\theta - \pi_\theta|$, (a) implies Condition (ii). We now check Condition (iii). By

$h(x,w,u)^2 = 1$, the triangle inequality, and $|\mathbb{I}\{q_{\nu_0}(\tilde{x}) > 0.5\}| \leq 1$, we obtain

$$\begin{aligned} & P \sup_{\theta\in\Theta:0<|\theta-\pi_\theta|<\varepsilon} |f_{\theta,\nu_0} - f_{\pi_\theta,\nu_0}|^2 \\ \leq \; & P \sup_{\theta\in\Theta:0<|\theta-\pi_\theta|<\varepsilon} \mathbb{I}\{x'\theta \geq -w_u \geq x'\pi_\theta \text{ or } x'\theta < -w_u < x'\pi_\theta\} \\ & +P \sup_{\theta\in\Theta:0<|\theta-\pi_\theta|<\varepsilon} \mathbb{I}\{x'\theta \geq -w_l \geq x'\pi_\theta \text{ or } x'\theta < -w_l < x'\pi_\theta\}, \end{aligned} \tag{10}$$

for any $\varepsilon > 0$. Again, the right hand side is the sum of the probabilities for pairs of wedge shaped regions with angles of order ε. Thus, (a) also guarantees Condition (iii).

It remains to check (8) and (9). The condition (8) is verified in the same manner as Condition (iii) by modifying the bound in (10). Let $I_\nu(\tilde{x}) = \mathbb{I}\{q_\nu(\tilde{x}) > .5 \geq q_{\nu_0}(\tilde{x}) \text{ or } q_\nu(\tilde{x}) \leq .5 < q_{\nu_0}(\tilde{x})\}$. For (9), note that

$$\begin{aligned} & |P(f_{\theta,\nu} - f_{\theta,\nu_0}) - P(f_{\pi_\theta,\nu} - f_{\theta,\nu_0})| \\ \leq \; & P\mathbb{I}\{x'\theta \geq -w_u \geq x'\pi_\theta \text{ or } x'\theta < -w_u < x'\pi_\theta\}I_\nu(\tilde{x}) \\ & +P\mathbb{I}\{x'\theta \geq -w_l \geq x'\pi_\theta \text{ or } x'\theta < -w_l < x'\pi_\theta\}I_\nu(\tilde{x}), \end{aligned} \tag{11}$$

for each $\theta \in \{\theta \in \Theta : |\theta - \pi_\theta| < \varepsilon\}$ and ν in a neighborhood of ν_0. For the first term of (11), the law of iterated expectation and an expansion of $q_\nu(\tilde{x})$ around ν_0 based on (a) imply

$$\begin{aligned} & P\mathbb{I}\{x'\theta \geq -w_u \geq x'\pi_\theta \text{ or } x'\theta < -w_u < x'\pi_\theta\}I_\nu(\tilde{x}) \\ \leq \; & P\mathbb{I}\{x'\theta \geq -w_u \geq x'\pi_\theta \text{ or } x'\theta < -w_u < x'\pi_\theta\}A(w_u,x)|v - \nu_0|, \end{aligned}$$

for some bounded function A. The second term of (11) is bounded in the same manner. Therefore, $|P(f_{\theta,\nu} - f_{\theta,\nu_0}) - P(f_{\pi_\theta,\nu} - f_{\theta,\nu_0})| = O(|\theta - \pi_\theta||v - \nu_0|)$ and (9) is verified.

Manski and Tamer (2002) proved the consistency of $\hat{\Theta}$ to Θ_I in terms of the Hausdorff distance. We provide a sharper lower bound on the their tuning parameter ϵ_n, which is in our notation $\hat{c}n^{-1/2}$ with $\hat{c} \to \infty$. For example, if we set $\hat{c} = \log n$, then the convergence rate becomes $H(\hat{\Theta}, \Theta_I) = O_p(n^{-1/4}(\log n)^{1/2})$. As we discussed in the end of Section 3, this rather slow convergence rate comes from the stochastic order of the criterion function on Θ_I. We basically verify the high level assumption of Chernozhukov, Hong and Tamer (2007, Condition C.2) in the cube root context. However, we mention that in the above setup, the criterion contains nuisance parameters with increasing dimension and the result in Chernozhukov, Hong and Tamer (2007) is not directly applicable.

4.3. **Dependent observations.** 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$(u|x) = 0$, 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 $P\{x'\theta_0 = 0, \dot{\kappa}(x)'\theta_0 p(x) > 0\} > 0$.

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 $q_{n,t} = f_{\theta_0 + n^{-1/3}s_1}(x_t, u_t) - f_{\theta_0 + n^{-1/3}s_2}(x_t, u_t)$. 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

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

The limit of $n^{1/3}\mathrm{Var}(q_{n,t})$ is given in Kim and Pollard (1990, p. 215). For the covariance $\mathrm{Cov}(q_{n,t}, q_{n,t+m})$, note that $q_{n,t}$ can take only three values, -1, 0, or 1. By the definition of β_m, Assumption D implies

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

for all $n, m \geq 1$ and $j, k = -1, 0, 1$, i.e., $\{q_{n,t}\}$ is a β-mixing array whose mixing coefficients are bounded by $n^{-2/3}\beta_m$. In turn, this implies that $\{q_{n,t}\}$ 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

$$\mathrm{Cov}(q_{n,t}, q_{n,t+m}) \leq C n^{-2/3} \beta_m \left\| q_{n,t} \right\|_p^2,$$

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

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

Combining these results, $n^{1/3} \sum_{m=1}^{\infty} \mathrm{Cov}(q_{n,t}, q_{n,t+m}) \to 0$ 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 with $h_n = 1$. Condition (i) is already verified. By Jensen's inequality,

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

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). For Condition (iii), pick any $\varepsilon > 0$ and observe that

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

for all ϑ in a neighborhood of θ_0. 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). 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.

4.4. **Random coefficient: New econometric model.** As a new econometric model which can be covered by our cube root asymptotic theory, let us consider the regression model with a random coefficient $y_t = x_t'\theta(w_t) + u_t$. We observe $x_t \in \mathbb{R}^d$, $w_t \in \mathbb{R}^k$, and the sign of y_t. We wish to estimate $\theta_0 = \theta(c)$ at some given $c \in \mathbb{R}^k$.[10] In this setup, we can consider a localized version of the maximum score estimator

$$\hat{\theta} = \arg\max_{\theta \in S} \sum_{t=1}^{n} K\left(\frac{w_t - c}{b_n}\right) [\mathbb{I}\{y_t \geq 0, x_t'\theta \geq 0\} + \mathbb{I}\{y_t < 0, x_t'\theta < 0\}],$$

where S is the surface of the unit sphere in $\mathbb{R}^d$. We adapt the assumptions in Section 4.3 for the maximum score estimator to localized counterparts. Let $h(x,u) = \mathbb{I}\{x'\theta_0 + u \geq 0\} - \mathbb{I}\{x'\theta_0 + u < 0\}$.

- **(a):** $\{x_t, w_t, u_t\}$ satisfies Assumption D. The density $p(x,w)$ of (x_t, w_t) is continuous at all x and $w = c$. The conditional distribution $x|w = c$ has compact support and continuously differentiable conditional density. The angular component of $x|w = c$, considered as a

[10] Gautier and Kitamura (2013) studied identification and estimation of the random coefficient binary choice model, where $\theta_t = \theta(w_t)$ is unobservable. Here we study the model where heterogeneity in the slope is caused by the observables w_t.

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$(u|x, w = c) = 0$, the function $\kappa(x, w) = E[h(x_t, u_t)|x_t = x, w_t = w]$ is continuous at all x and $w = c$, $\kappa(x, c)$ is non-negative for $x'\theta_0 \geq 0$ and non-positive for $x'\theta_0 < 0$ and is continuously differentiable in x, and $P\{x'\theta_0 = 0, \left(\frac{\partial\kappa(x,w)}{\partial x}\right)' \theta_0 p(x, w) > 0|w = c\} > 0$.

(c): K is a bounded symmetric density function with $\int s^2 K(s)ds < \infty$. As $n \to \infty$, it holds $nb_n^{k'} \to \infty$ for some $k' > k$.

Note that the criterion function is written as

$$f_{n,\theta}(x, w, u) = \frac{1}{h_n} K\left(\frac{w - c}{h_n^{1/k}}\right) h(x, u)[\mathbb{I}\{x'\theta \geq 0\} - \mathbb{I}\{x'\theta_0 \geq 0\}],$$

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

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

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 $q_{n,t} = f_{n,\theta_0+(nh_n)^{-1/3}s_1}(x_t, w_t, u_t) - f_{n,\theta_0+(nh_n)^{-1/3}s_2}(x_t, w_t, u_t)$. 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

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

The limit of $(nh_n)^{1/3}\mathrm{Var}(q_{n,t})$ is obtained in the same manner as Kim and Pollard (1990, p. 215). For the covariance, the α-mixing inequality implies

$$|\mathrm{Cov}(q_{n,t}, q_{n,t+m})| \leq C\beta_m \|q_{n,t}\|_p^2 = O(\rho^m) O((nh_n)^{-\frac{2}{3p}} h_n^{\frac{2(1-p)}{p}}),$$

for some $C > 0$ and $p > 2$, where the equality follows from the change of variables and Assumption D. Also, by the change of variables $|\mathrm{Cov}(q_{n,t}, q_{n,t+m})| = |P q_{n,t} q_{n,t+m} - (P q_{n,t})^2| = O((nh_n)^{-2/3})$. By using these bounds (note: if $0 < A \leq \min\{B_1, B_2\}$, then $A \leq B_1^{\ell} B_2^{1-\ell}$ for any $\ell \in [0, 1]$), there exists a positive constant C' such that

$$(nh_n)^{1/3}\sum_{m=1}^{\infty}|\text{Cov}(q_{n,t},q_{n,t+m})| \le C'(nh_n)^{-\frac{1}{3}+\frac{2(p-1)\ell}{3}}h_n^{-\frac{2(p-1)\ell}{p}}\sum_{m=1}^{\infty}\rho^{\ell m},$$

for any $\ell \in [0,1]$. Thus, by taking ℓ sufficiently small, we obtain $\lim_{n\to\infty}(nh_n)^{1/3}\sum_{m=1}^{\infty}\text{Cov}(q_{n,t},q_{n,t+m}) = 0$ due to $nb_n^{k'} \to \infty$.

We now verify that $\{f_{n,\theta} : \theta \in S\}$ belongs to the cube root class with $h_n = b_n^k$. Condition (i) is already verified. By the change of variables and Jensen's inequality (also note that $h(x,u)^2 = 1$), there exists a positive constant C such that

$$\begin{aligned} h_n^{1/2}\|f_{n,\theta_1}-f_{n,\theta_2}\|_2 &= \sqrt{\int\int K(s)^2|\mathbb{I}\{x'\theta_1 \ge 0\}-\mathbb{I}\{x'\theta_2 \ge 0\}|p(x,c+sb_n)dxds} \\ &\ge CE\left[|\mathbb{I}\{x'\theta_1 \ge 0\}-\mathbb{I}\{x'\theta_2 \ge 0\}|\,\middle|\, w=c\right] \\ &= CP\{x'\theta_1 \ge 0 > x'\theta_2 \text{ or } x'\theta_2 \ge 0 > x'\theta_1 | w=c\}, \end{aligned}$$

for all $\theta_1, \theta_2 \in S$ and all n large enough. Since the right hand side is the conditional 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). For Condition (iii), pick any $\varepsilon > 0$ and there exists a positive constant C' such that

$$\begin{aligned} & P\sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon} h_n|f_{n,\theta}-f_{n,\vartheta}|^2 \\ = & \int\int K(s)^2 \sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon}|[\mathbb{I}\{x'\theta \ge 0\}-\mathbb{I}\{x'\vartheta \ge 0\}]|^2 p(x,c+sb_n)dxds \\ \le & C'E\left[\sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon}|[\mathbb{I}\{x'\theta \ge 0\}-\mathbb{I}\{x'\vartheta \ge 0\}]|^2 \,\middle|\, w=c\right], \end{aligned}$$

for all ϑ in a neighborhood of θ_0 and n large enough. Again, the right hand side is the conditional probability for a pair of wedge shaped regions with an angle of order ε. Thus the last condition in (a) also guarantees Condition (iii). Since $\{f_{n,\theta} : \theta \in S\}$ belongs to the cube root class, Theorem 1 implies the limiting distribution of $(nh_n)^{1/3}(\hat{\theta} - \theta_0)$ for the random coefficient model.

5. Further examples

5.1. **Minimum volume predictive region.** As an illustration of Theorem 2, 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

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

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 2 to obtain the convergence rate of $\hat{\theta}$. For the condition in (5), 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 (5) holds with $V_1 = \{\dot{\gamma}_{y|x}(1|c) - \dot{\gamma}_{y|x}(-1|c)\}\gamma_x(c)$. Condition (iii) for $\{f_{n,\theta,\nu} : \theta \in \mathbb{R}, \nu \in \mathbb{R}\}$ is verified in the same manner as in Section 5.2 below. It remains to verify Condition (ii) for the class $\{f_{n,\theta,1} : \theta \in \mathbb{R}\}$. 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 $\hat{I}$.

5.2. **Least median of squares.** As another 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

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

and $\hat{\nu} = \inf\{\nu : \sup_\theta \mathbb{P}_n f_{\theta,\nu} \geq \frac{1}{2}\}$. 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 (1) is verified as

$$\begin{aligned} P(f_{\theta,\nu} - f_{0,1}) &= P|\{\Gamma(x'\theta+\nu) - \Gamma(\nu)\} - \{\Gamma(x'\theta-\nu) - \Gamma(-\nu)\}| \\ &\quad +P|\{\Gamma(\nu) - \Gamma(1)\} - \{\Gamma(-\nu) - \Gamma(-1)\}| \\ &= \dot{\gamma}(1)\theta' Pxx'\theta + o(|\theta|^2 + |\nu-1|^2). \end{aligned} \tag{12}$$

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

$$P \sup_{(\theta,\nu):|(\theta,\nu)-(\theta',\nu')|<\varepsilon} |f_{\theta,\nu} - f_{(\theta',\nu')}|^2 \leq P \sup_{(\theta,\nu):|(\theta,\nu)-(\theta',\nu')|<\varepsilon} |f_{\theta,\nu} - f_{\theta,\nu'}|^2 + P \sup_{\theta:|\theta-\theta'|<\varepsilon} |f_{\theta,\nu'} - f_{\theta',\nu'}|^2,$$

for (θ', ν') in a neighborhood of $(0, 1)$. By similar arguments to (12), these terms are of order $|\nu - \nu'|^2$ and $|\theta - \theta'|^2$, 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 with $h_n = 1$. 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

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

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 4.3, 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.

5.3. **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

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

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 (4) is verified as

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

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.

5.4. **Mode and related estimator.** As a final illustration, we consider estimation of mode and related methods. In a seminal paper, Chernoff (1964) studied the asymptotic property of the mode estimator that maximizes $\sum_{t=1}^{n} \mathbb{I}\{|y_t - \beta| \le h\}$ with respect to β for some fixed h. Indeed this is a first example of the cube root asymptotics. By extending this approach, Lee (1989) proposed the mode regression estimator for a regression model with a truncated dependent variable that maximizes $\sum_{t=1}^{n} \mathbb{I}\{|y_t - \max\{x_t'\beta, c+h\}| \le h\}$ with respect to β for some fixed h and known truncation point c. Lee (1989) established the consistency of the mode regression estimator and conjectured the cube root convergence rate. Also, in the statistics literature on computer vision algorithm, Goldenshluger and Zeevi (2004) investigated the so-called Hough transform estimator for the regression model

$$\hat{\beta} = \arg\max_{\beta} \sum_{t=1}^{n} \mathbb{I}\{|y_t - x_t'\beta| \le h|x_t|\}, \tag{13}$$

where $x_t = (1, \tilde{x}_t)'$ for a scalar $\tilde{x}_t$ and h is a fixed tuning constant. Goldenshluger and Zeevi (2004) derived the cube root asymptotics for $\hat{\beta}$. All these papers treat the tuning constant h as fixed and discuss carefully about the practical choice of h. However, for these estimators, h plays a role of the bandwidth and the analysis for the case of $h_n \to 0$ is a substantial open question (see, Goldenshluger and Zeevi, 2004, pp. 1915-1916). Here we focus on the Hough transform estimator in (13) with $h = h_n \to 0$ and study its asymptotic property. The estimators by Chernoff (1964) and Lee (1989) can be analyzed in the same manner. Indeed Condition (iii) of the cube root class is not satisfied for these cases and we show how to modify our framework.

Let us 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 $\sup_\theta |\mathbb{P}_n f_{n,\theta} - Pf_{n,\theta}| \xrightarrow{p} 0$ by applying Nobel and Dembo (1993, Theorem 1).

In order to apply Theorem 1, we verify that $\{f_{n,\theta}\}$ belongs to the cube root class. Obviously $h_n f_{n,\theta}$ is bounded. Since $\lim_{n\to\infty} Pf_{n,\theta} = 2P\gamma(x'\theta)|x|$ and γ is uniquely maximized at zero (by (a)), $\lim_{n\to\infty} Pf_{n,\theta}$ is uniquely maximized at $\theta = 0$. Since γ is continuously differentiable in a neighborhood of zero, $Pf_{n,\theta}$ 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 (1) holds with $V = \ddot{\gamma}(0)P(|x|xx')$. Note that $\ddot{\gamma}(0) < 0$ by (a). Therefore, Condition (i) is satisfied.

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

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

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

However, we can see that Condition (iii) is not satisfied in this case. Although Theorem 1 is not directly applicable, Condition (iii) can be modified as follows.

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

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

for all n large enough, $\varepsilon > 0$ small enough, and θ' in a neighborhood of θ_0.

Compared to Condition (iii), this condition assumes a larger envelope for the class $\{|f_{n,\theta} - f_{n,\theta'}|^2 : |\theta - \theta'| < \varepsilon\}$. Thus, Lemma M in Section 2 is modified as follows.

Lemma M1. *Suppose that Assumption D holds and $\{f_{n,\theta}\}$ satisfies Condition (ii) of the cube root class and Condition (iii)' above. Then there exist positive constants C and C' such that*

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

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

We now check Condition (iii)'. Observe that

$$P \sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon} h_n^2|f_{n,\theta}-f_{n,\vartheta}|^2 \leq P \sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon} \mathbb{I}\{|u-x'\vartheta|\leq h_n|x|,\ |u-x'\theta|>h_n|x|\} + P \sup_{\theta\in\Theta:|\theta-\vartheta|<\varepsilon} \mathbb{I}\{|u-x'\theta|\leq h_n|x|,\ |u-x'\vartheta|>h_n|x|\},$$

for all ϑ in a neighborhood of 0. 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 4.3. 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 1, 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 $H(s_1,s_2) = 2\gamma(0)P|x'(s_1-s_2)|$.

Appendix A. Mathematical Appendix

Notation: Recall that $Q_g(u)$ is the inverse function of the tail probability function $x \mapsto P\{|g(z_t)| > x\}$,[11] and that $\{\beta_m\}$ is the β-mixing coefficients used in Assumption D. 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

$$\|g\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u) Q_g(u)^2 du}. \tag{14}$$

A.1. **Proof of Lemma M.** Pick any $C' > 0$ and then pick any n satisfying $(nh_n)^{-1/2} \leq C'$ and any $\delta \in [(nh_n)^{-1/2}, C']$. Throughout the proof, positive constants C_j $(j = 1, 2, \ldots)$ are independent of n and δ.

First, we introduce some notation. Consider the sets defined by different norms:

$$\begin{aligned}
\mathcal{G}^1_{n,\delta} &= \left\{h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) : |\theta - \theta_0| < \delta \text{ for } \theta \in \Theta\right\}, \\
\mathcal{G}^2_{n,\delta} &= \left\{h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) : \left\|h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0})\right\|_2 < \delta \text{ for } \theta \in \Theta\right\}, \\
\mathcal{G}^\beta_{n,\delta} &= \left\{h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0}) : \left\|h_n^{1/2}(f_{n,\theta} - f_{n,\theta_0})\right\|_{2,\beta} < \delta \text{ for } \theta \in \Theta\right\}.
\end{aligned}$$

For any $g \in \mathcal{G}^1_{n,\delta}$, g is bounded (by 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

$$Q_{\hat{g}}(u) = \sum_{j=1}^m a_j \mathbb{I}\{(j-1)/m \leq u < j/m\}, \tag{15}$$

satisfying $Q_g \leq Q_{\hat{g}}$, for some positive integer m and sequence of positive constants $\{a_j\}$.

Next, based on the above notation, we derive the set inclusion relationships

$$\mathcal{G}^\beta_{n,\delta} \subset \mathcal{G}^2_{n,\delta} \subset \mathcal{G}^1_{n,C_1\delta}, \qquad \mathcal{G}^1_{n,\delta} \subset \mathcal{G}^\beta_{n,C_2\delta^{1/2}}, \tag{16}$$

for some positive constants C_1 and C_2. The relation $\mathcal{G}^\beta_{n,\delta} \subset \mathcal{G}^2_{n,\delta}$ follows from $\|\cdot\|_2 \leq \|\cdot\|_{2,\beta}$ (Doukhan, Massart and Rio, 1995, Lemma 1). The relation $\mathcal{G}^2_{n,\delta} \subset \mathcal{G}^1_{n,C_1\delta}$ follows from Condition (ii). Pick any $g \in \mathcal{G}^1_\delta$. The relation $\mathcal{G}^1_{n,\delta} \subset \mathcal{G}^\beta_{n,C_2\delta^{1/2}}$ follows by

$$\begin{aligned}
\|g\|_{2,\beta}^2 &\leq \sum_{j=1}^m a_j^2 \left\{\int_{(j-1)/m}^{j/m} \beta^{-1}(u) du\right\} \leq \left\{m \int_0^{1/m} \beta^{-1}(u) du\right\} \int_0^1 Q_{\hat{g}}(u)^2 du \\
&\leq \left\{\sup_{0 < a \leq 1} a \int_0^{1/a} \beta^{-1}(u) du\right\} 2\|g\|_2^2 \leq C_2^2 \delta, \qquad (17)
\end{aligned}$$

for some positive constant C_2, where the first inequality follows from $Q_g \leq Q_{\hat{g}}$, the second inequality follows from monotonicity of $\beta^{-1}(u)$ and $\int_0^1 Q_{\hat{g}}(u)^2 du = \frac{1}{m}\sum_{j=1}^m a_j^2$, the third inequality follows by

[11] 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.

$\int_0^1 Q_{\hat{g}}(u)^2 du = \|\hat{g}\|_2^2 \leq 2\|g\|_2^2$, and the last inequality follows from $\sup_{0<a\leq 1} a\int_0^{1/a} \beta^{-1}(u)du < \infty$ (by Assumption D) and Condition (iii).

Third, based on (16), 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 $\|\cdot\|$. Note that

$$N_{[]}(\nu, \mathcal{G}_{n,\delta}^{\beta}, \|\cdot\|_{2,\beta}) \leq N_{[]}(\nu, \mathcal{G}_{n,C_1\delta}^{1}, \|\cdot\|_2) \leq C_3 \left(\frac{\delta}{\nu}\right)^{2d},$$

for some positive constant C_3, where the first inequality follows from $\mathcal{G}_{n,\delta}^{\beta} \subset \mathcal{G}_{n,C_1\delta}^{1}$ (by (16)) and $\|\cdot\|_2 \leq \|\cdot\|_{2,\beta}$ (Doukhan, Massart and Rio, 1995, Lemma 1), and 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, by the indefinite integral formula $\int \log x dx = \text{const.} + x(\log x - 1)$, there exists a positive constant C_4 such that

$$\varphi_n(\delta) = \int_0^{\delta} \sqrt{\log N_{[]}(\nu, \mathcal{G}_{n,\delta}^{\beta}, \|\cdot\|_{2,\beta})} d\nu \leq C_4\delta. \tag{18}$$

Finally, based on the entropy condition (18), we apply the maximal inequality of Doukhan, Massart and Rio (1995, Theorem 3), i.e., there exists a positive constant C_5 such that

$$P \sup_{g\in\mathcal{G}_{n,\delta}^{\beta}} |\mathbb{G}_n g| \leq C_5[1 + \delta^{-1} q_{G_{n,\delta}}(\min\{1, v_n(\delta)\})]\varphi_n(\delta), \tag{19}$$

where $q_{G_{n,\delta}}(v) = \sup_{u\leq v} Q_{G_{n,\delta}}(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ with the envelope function $G_{n,\delta}$ of $\mathcal{G}_{n,\delta}^{\beta}$, 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 (18), 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 $\delta \in [(nh_n)^{-1/2}, C']$. We have

$$q_{G_{n,\delta}}(\min\{1, v_n(\delta)\}) \leq C_6 Q_{G_{n,\delta}}(v_n(\delta))\sqrt{v_n(\delta)} \leq C_7(nh_n)^{-1/2}, \tag{20}$$

for some positive constants C_6 and C_7. Therefore, combining (18)-(20), we obtain

$$P \sup_{g\in\mathcal{G}_{n,C_2\delta^{1/2}}^{\beta}} |\mathbb{G}_n g| \leq C_8\delta^{1/2}, \tag{21}$$

for some positive constant C_8. The conclusion follows from the second relation in (16).

A.2. **Proof of Lemma 1.** Pick any $C > 0$ and $\varepsilon > 0$. Define $A_n = \{\theta \in \Theta : (nh_n)^{-1/3} \leq |\theta - \theta_0| \leq C\}$ and

$$R_n^2 = (nh_n)^{2/3} \sup_{\theta\in A_n} \{|\mathbb{P}_n(f_{n,\theta} - f_{n,\theta_0}) - P(f_{n,\theta} - f_{n,\theta_0})| - \varepsilon|\theta - \theta_0|^2\}.$$

It is enough to show $R_n = O_p(1)$. Letting $A_{n,j} = \{\theta \in \Theta : (j-1)(nh_n)^{-1/3} \leq |\theta - \theta_0| < j(nh_n)^{-1/3}\}$, there exists a positive constant C' such that

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

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

A.3. **Proof of Lemma C.** First of all, any β-mixing process is α-mixing with the mixing coefficient $\alpha_m \leq \beta_m/2$. Thus it is sufficient to check Conditions (a) and (b) of Rio (1997, Corollary 1). Condition (a) is verified under (3) 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 (by (3)) and $\{z_t\}$ is strictly stationary under Assumption D, 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 $n \to \infty$. Pick any $u \in (0,1)$. Since $\beta^{-1}(u)Q_{g_n}(u)^2$ is non-increasing in $u \in (0,1)$, the condition (3) implies $\beta^{-1}(u)Q_{g_n}(u)^2 < C < \infty$ for all n. Therefore, for each $u \in (0,1)$, it holds $n^{-1/2}\beta^{-1}(u)Q_{g_n}(u) \to 0$ as $n \to \infty$. Then the dominated convergence theorem based on (3) implies Condition (b).

A.4. **Proof of Lemma 2.** By Condition (i), it holds $|g_{n,s}| \leq 2C(nh_n^{-2})^{1/6}$ for all n and s, which implies $Q_{g_{n,s} - Pg_{n,s}}(u)^2 \leq 16C^2(nh_n^{-2})^{1/3}$ for all n, s, and $u \in (0,1)$. By the condition of this lemma, it holds $Q_{g_{n,s}}(u) \leq c$ for all n large enough and $u > c(nh_n^{-2})^{-1/3}$. By the triangle inequality and the definition of Q_g,

$$P\{|g_{n,s} - Pg_{n,s}| \geq Q_{g_{n,s}}(u) + |Pg_{n,s}|\} \leq P\{|g_{n,s}| \geq Q_{g_{n,s}}(u)\} = P\{|g_{n,s} - Pg_{n,s}| > Q_{g_{n,s} - Pg_{n,s}}(u)\},$$

which implies $Q_{g_{n,s} - Pg_{n,s}}(u) \leq Q_{g_{n,s}}(u) + |Pg_{n,s}|$. Thus, for all n large enough, s, and $u > c(nh_n^{-2})^{-1/3}$, it holds

$$Q_{g_{n,s} - Pg_{n,s}}(u)^2 \leq c^2 + |Pg_{n,s}|^2 + 2c|Pg_{n,s}|.$$

Combining these bounds, (3) is verified as

$$\begin{aligned}
& \int_0^1 \beta^{-1}(u) Q_{g_{n,s} - Pg_{n,s}}(u)^2 du \\
\leq\ & 16C^2(nh_n^{-2})^{1/3} \int_0^{c(nh_n^{-2})^{-1/3}} \beta^{-1}(u) du + \{c^2 + (Pg_{n,s})^2 + 2c|Pg_{n,s}|\} \int_{c(nh_n^{-2})^{-1/3}}^1 \beta^{-1}(u) du < \infty,
\end{aligned}$$

for all n large enough, where the second inequality follows by (1) and Assumption D (which guarantees $\sup_n (nh_n^{-2})^{1/3} \int_0^{c(nh_n^{-2})^{-1/3}} \beta^{-1}(u)du < \infty$, and $\int_0^1 \beta^{-1}(u)du < \infty$).

A.5. **Proof of Lemma M'.** Pick any $K > 0$ and $\sigma > 0$. Let $g_{n,s,s'} = g_{n,s} - g_{n,s'}$,

$$\begin{aligned} \mathcal{G}_n^K &= \{g_{n,s,s'} : |s| \leq K, |s'| \leq K\}, \\ \mathcal{G}_{n,\delta}^1 &= \{g_{n,s,s'} \in \mathcal{G}_n^K : |s - s'| < \delta\}, \\ \mathcal{G}_{n,\delta}^\beta &= \{g_{n,s,s'} \in \mathcal{G}_n^K : \left\|g_{n,s,s'}\right\|_{2,\beta} < \delta\}. \end{aligned}$$

Since $g_{n,s}$ satisfies the condition (4), there exists a positive constant C_1 such that $\mathcal{G}_{n,\delta}^1 \subset \{g_{n,s,s'} \in \mathcal{G}_n^K : \left\|g_{n,s,s'}\right\|_2 < C_1\delta^{1/2}\}$ for all n large enough and all $\delta > 0$ small enough. Also, by the same argument to derive (17), there exists a positive constant C_2 such that $\left\|g_{n,s,s'}\right\|_{2,\beta} \leq C_2 \left\|g_{n,s,s'}\right\|_2$ for all n large enough, $|s| \leq K$, and $|s'| \leq K$. The constant C_2 depends only on the mixing sequence $\{\beta_m\}$. Combining these results, we obtain the set inclusion relationship

$$\mathcal{G}_{n,\delta}^1 \subset \mathcal{G}_{n,C_1C_2\delta^{1/2}}^\beta, \tag{22}$$

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

Also note that the bracketing numbers satisfy

$$N_{[]}(\nu, \mathcal{G}_{n,\delta}^\beta, \|\cdot\|_{2,\beta}) \leq N_{[]}(\nu, \mathcal{G}_n^K, \|\cdot\|_2) \leq C_3\nu^{-d/2},$$

where the first inequality follows from $\mathcal{G}_{n,\delta}^\beta \subset \mathcal{G}_n^K$ (by the definitions) and $\|\cdot\|_2 \leq \|\cdot\|_{2,\beta}$ (Doukhan, Massart and Rio, 1995, Lemma 1), and the second inequality follows from the argument to derive Andrews (1993, eq. (4.7)) based on (4) (called the L_2-continuity assumption in Andrews, 1993). Thus, there is a function $\varphi(\eta)$ such that $\varphi(\eta) \to 0$ as $\eta \to 0$ and

$$\varphi_n(\eta) = \int_0^\eta \sqrt{\log N_{[]}(\nu, \mathcal{G}_{n,\eta}^\beta, \|\cdot\|_{2,\beta})} d\nu \leq \varphi(\eta),$$

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

Based on the above entropy condition, we can 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

$$P \sup_{g \in \mathcal{G}_{n,\eta}^\beta} |\mathbb{G}_n g| \leq C_4[1 + \eta^{-1} q_{G_n}(\min\{1, v_n(\eta)\})]\varphi(\eta),$$

for all n large enough and all $\eta > 0$ small enough, where $q_{2G_n}(v) = \sup_{u \leq v} Q_{2G_n}(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ with the envelope function $2G_n$ of $\mathcal{G}_{n,\eta}^\beta$ (note: by the definition of $\mathcal{G}_{n,\eta}^\beta$, the envelope $2G_n$ does not depend on 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 $2C_4\varphi(\eta) < \sigma$. Since $\varphi_n(\eta) \leq \varphi(\eta)$, there is a positive constant C_5 such that $v_n(\eta) \leq C_5 \frac{\varphi(\eta)}{n\eta^2}$ 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_6 such that $q_{2G_n}(\min\{1, v_n(\eta)\}) \leq C_6\sqrt{\varphi(\eta)}\eta^{-1}n^{\kappa-1/2}$ with $0 < \kappa < 1/2$ for all n large enough. Therefore, by setting $\eta = C_1C_2\delta^{1/2}$, we obtain

$$P \sup_{g\in\mathcal{G}^\beta_{n,C_1C_2\delta^{1/2}}} |\mathbb{G}_n g| \leq \sigma,$$

for all n large enough. The conclusion follows by (22).

A.6. **Proof of Theorem 1.** As discussed in the main body, Lemma 1 yields the convergence rate of the estimator. This enables us to consider the centered and normalized process $Z_n(s)$, which can be defined on arbitrary compact parameter space. 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 $\hat{\theta}$.

A.7. **Proof of Theorem 2.** To ease notation, let $\theta_0 = \nu_0 = 0$. First, we show that $\hat{\theta} = O_p((nh_n)^{-1/3})$. Since $\{f_{n,\theta,\nu}\}$ satisfies Condition (iii) of the drifting cube root class, we can apply Lemma M' with $g_{n,s} = n^{1/6}h_n^{2/3}(f_{n,\theta,c(nh_n)^{-1/3}} - f_{n,\theta,0})$ for $s = (\theta', c')'$, which implies

$$\sup_{|\theta|\leq\epsilon,|c|\leq\epsilon} n^{1/6}h_n^{2/3}\mathbb{G}_n(f_{n,\theta,c(nh_n)^{-1/3}} - f_{n,\theta,0}) = O_p(1), \tag{23}$$

for all $\epsilon > 0$. Also from (5) and $\hat{\nu} = o_p((nh_n)^{-1/3})$, we have

$$P(f_{n,\theta,\hat{\nu}} - f_{n,\theta,0}) - P(f_{n,0,\hat{\nu}} - f_{n,0,0}) \leq 2\epsilon|\theta|^2 + O_p((nh_n)^{-2/3}), \tag{24}$$

for all θ in a neighborhood of θ_0 and all $\epsilon > 0$. Combining (23), (24), and Lemma 1,

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

for all θ in a neighborhood of θ_0 and all $\epsilon > 0$, where the last inequality follows from (5). From $\mathbb{P}_n(f_{n,\hat{\theta},\hat{\nu}} - f_{n,0,\hat{\nu}}) \geq o_p((nh_n)^{-2/3})$, negative definiteness of V_1, and $\hat{\nu} = o_p((nh_n)^{-1/3})$, we can find $c > 0$ such that

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

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

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

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

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

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

uniformly in s. Also (5) implies $P(f_{n,s(nh_n)^{-1/3},\hat{\nu}} - f_{n,0,\hat{\nu}}) - P(f_{n,s(nh_n)^{-1/3},0} - f_{n,0,0}) = o_p((nh_n)^{-2/3})$ uniformly in s. Given $\hat{\theta} - \tilde{\theta} = o_p((nh_n)^{-1/3})$, an application of Theorem 1 to the cube root class $\{f_{n,\theta,\nu_0} : \theta \in \Theta\}$ implies the limiting distribution of $\hat{\theta}$.

A.8. **Proof of Lemma MS.** First, we introduce some notation. Let

$$\begin{aligned}
\mathcal{G}^{\beta}_{n,\delta} &= \{h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta}) : \left\| h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta}) \right\|_{2,\beta} < \delta \text{ for } \theta \in \Theta\}, \\
\mathcal{G}^{1}_{n,\delta} &= \{h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta}) : |\theta - \pi_\theta| < \delta \text{ for } \theta \in \Theta\}, \\
\mathcal{G}^{2}_{n,\delta} &= \{h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta}) : \left\| h_n^{1/2}(f_{n,\theta} - f_{n,\pi_\theta}) \right\|_{2} < \delta \text{ for } \theta \in \Theta\}.
\end{aligned}$$

For any $g \in \mathcal{G}^1_{n,\delta}$, 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 $\{a_j\}$. Let $r_n = nh_n/\log(nh_n)$. Pick any $C' > 0$ and then pick any n satisfying $r_n^{-1/2} \leq C'$ and any $\delta \in [(r_n^{-1/2}, C']$. Throughout the proof, 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 $g \in \mathcal{G}^1_\delta$, it holds

$$\begin{aligned}
\|g\|_2^2 &\leq \int_0^1 \beta^{-1}(u) Q_g(u)^2 du \leq \frac{1}{m}\sum_{j=1}^{m} a_j^2 \left\{ m \int_{(j-1)/m}^{j/m} \beta^{-1}(u)du \right\} \\
&\leq \left\{ m \int_0^{1/m} \beta^{-1}(u)du \right\} \int_0^1 Q_{\hat{g}}(u)^2 du \\
&\leq M \|g\|_2^2,
\end{aligned} \tag{25}$$

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 $\|\hat{g}\|_2^2 \leq 2\|g\|_2^2$. Therefore,

$$\|f_{n,\theta} - f_{n,\pi_\theta}\|_2 \leq \|f_{n,\theta} - f_{n,\pi_\theta}\|_{2,\beta} \leq M^{1/2}\|f_{n,\theta} - f_{n,\pi_\theta}\|_2, \tag{26}$$

for each $\theta \in \{\theta \in \Theta : |\theta - \pi_\theta| < \delta\}$, where the first inequality follows from Doukhan, Massart and Rio (1995, Lemma 1) and the second inequality follows from (25). Based on this, we can deduce the inclusion relationships: there exist positive constants C_1 and C_2 such that

$$\mathcal{G}^1_{n,\delta} \subset \mathcal{G}^2_{n,C_1\delta^{1/2}} \subset \mathcal{G}^\beta_{n,M^{1/2}C_1\delta^{1/2}}, \qquad \mathcal{G}^\beta_{n,\delta} \subset \mathcal{G}^2_{n,\delta} \subset \mathcal{G}^1_{n,C_2\delta}, \tag{27}$$

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

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 (26) and the second relation in (27),

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

for some positive constant C_3. Note that the upper bound here is different from the point identified case. Therefore, for some positive constant C_4, it holds

$$\varphi_n(\delta) = \int_0^\delta \sqrt{\log N_{[]}(\nu, \mathcal{G}^\beta_{n,\delta}, \|\cdot\|_{2,\beta})} d\nu \leq C_4 \delta \log \delta^{-1}. \tag{28}$$

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

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

where $q_{G_{n,\delta}}(v) = \sup_{u \leq v} Q_{G_{n,\delta}}(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ with the envelope function $G_{n,\delta}$ of $\mathcal{G}^\beta_{n,\delta}$ (note: $\mathcal{G}^\beta_{n,\delta}$ 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 $\varphi_n(\delta) \leq C_4\delta \log \delta^{-1}$ from (28), it holds $v_n(\delta) \leq C_5 n^{-1}(\log \delta^{-1})^2 \leq C_5 n^{-1}\{\log(nh_n)^{1/2}\}^2$ 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 $\delta \in [r_n^{-1/2}, C']$. We have

$$q_{G_{n,\delta}}(\min\{1, v_n(\delta)\}) \leq C_6\sqrt{v_n(\delta)} Q_{G_{n,\delta}}(v_n(\delta)) \leq C_7 n^{-1/2} \log(nh_n)^{1/2},$$

for some positive constants C_6 and C_7. Therefore, combining (28)-(20), the conclusion follows by

$$P \sup_{g \in \mathcal{G}^1_{n,\delta}} |\mathbb{G}_n g| \leq P \sup_{g \in \mathcal{G}^\beta_{n,M^{1/2}C_1\delta^{1/2}}} |\mathbb{G}_n g| \leq C_8(\delta \log \delta^{-1})^{1/2},$$

where the first inequality follows from the first relation in (27).

A.9. **Proof of Lemma 3.** Pick any $C > 0$ and $\varepsilon > 0$. Then define $A_n = \{\theta \in \Theta \setminus \Theta_I : r_n^{-1/3} \leq |\theta - \pi_\theta| \leq C\}$ and

$$R_n^2 = r_n^{2/3} \sup_{\theta \in A_n} \{|\mathbb{P}_n(f_{n,\theta} - f_{n,\pi_\theta}) - P(f_{n,\theta} - f_{n,\pi_\theta})| - \varepsilon|\theta - \pi_\theta|^2\}.$$

It is enough to show $R_n = O_p(1)$. Letting $A_{n,j} = \{\theta \in \Theta : (j-1)r_n^{-1/3} \leq |\theta - \pi_\theta| < jr_n^{-1/3}\}$, there exists a positive constant C' such that

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

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

A.10. **Proof of Theorem 3.** Pick any $\vartheta \in \hat{\Theta}$. By the definition of $\hat{\Theta}$,

$$\mathbb{P}_n(f_{n,\vartheta} - f_{n,\pi_\vartheta}) \geq \max_{\theta \in \Theta} \mathbb{P}_n f_{n,\theta} - (nh_n)^{-1/2}\hat{c} - \mathbb{P}_n f_{n,\pi_\vartheta} \geq -(nh_n)^{-1/2}\hat{c}.$$

Now, suppose $H(\vartheta, \Theta_I) = |\vartheta - \pi_\vartheta| > r_n^{-1/3}$. By Lemma 3 and Condition (i) of the set identified cube root class,

$$\begin{aligned}
\mathbb{P}_n(f_{n,\vartheta} - f_{n,\pi_\vartheta}) &\leq P(f_{n,\vartheta} - f_{n,\pi_\vartheta}) + \varepsilon|\vartheta - \pi_\vartheta|^2 + r_n^{-2/3}R_n^2 \\
&\leq (-c + \varepsilon)|\vartheta - \pi_\vartheta|^2 + o(|\vartheta - \pi_\vartheta|^2) + O_p(r_n^{-2/3}),
\end{aligned}$$

for any $\varepsilon > 0$. Note that c, ε, and R_n do not depend on ϑ. By taking ε small enough, the convergence rate of $\rho(\hat{\Theta}, \Theta_I)$ is obtained as

$$\rho(\hat{\Theta}, \Theta_I) = \sup_{\vartheta \in \hat{\Theta}} |\vartheta - \pi_\vartheta| \leq O_p(\hat{c}^{1/2}(nh_n)^{-1/4} + r_n^{-1/3}).$$

Furthermore, for the maximizer $\hat{\theta}$ of $\mathbb{P}_n f_{n,\theta}$, then it holds $\mathbb{P}_n(f_{n,\hat{\theta}} - f_{n,\pi_{\hat{\theta}}}) \geq 0$ and this implies $\hat{\theta} - \pi_{\hat{\theta}} = O_p(r_n^{-1/3})$.

For the convergence rate of $\rho(\Theta_I, \hat{\Theta})$, we show $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ for $\hat{c} \to \infty$, which implies that $\rho(\Theta_I, \hat{\Theta})$ can converge at arbitrarily fast rate. To see this, note that

$$\begin{aligned}
& (nh_n)^{1/2} \max_{\theta' \in \Theta_I} |(\max_{\theta \in \Theta} \mathbb{P}_n f_{n,\theta} - \mathbb{P}_n f_{n,\theta'})| \\
\leq \;& |\mathbb{G}_n(f_{n,\hat{\theta}} - f_{n,\pi_{\hat{\theta}}})| + (nh_n)^{1/2}|P(f_{n,\hat{\theta}} - f_{n,\pi_{\hat{\theta}}})| + 2(nh_n)^{1/2}|\max_{\theta' \in \Theta_I}(\mathbb{P}_n f_{n,\theta'} - Pf_{n,\theta'})| \\
= \;& 2h_n^{1/2}|\max_{\theta' \in \Theta_I} \mathbb{G}_n f_{n,\theta'}| + o_p(1), \qquad (29)
\end{aligned}$$

where the inequality follows from the triangle inequality and the equality follows from Lemmas MS and 3, Condition (i) of the set identified cube root class, and the rate $\hat{\theta} - \pi_{\hat{\theta}} = O_p(r_n^{-1/3})$ obtained above. Therefore, since $\{h_n^{1/2} f_{n,\theta}, \theta \in \Theta_I\}$ is P-Donsker (Condition (i)), it follows $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ if $\hat{c} \to \infty$.

A.11. **Proof of Lemma MS'.** Pick any $\sigma > 0$. Define $g_{n,s,s'} = g_{n,s} - g_{n,s'}$,

$$\begin{aligned}
\mathcal{G}_n^1 &= \{g_{n,s,s'} : g_{n,s} \in \mathcal{G}_n, g_{n,s'} \in \mathcal{G}_n\}, \\
\mathcal{G}_{n,\delta}^1 &= \{g_{n,s,s'} \in \mathcal{G}_n^1 : |s - s'| < \delta\}, \\
\mathcal{G}_{n,\delta}^\beta &= \{g_{n,s,s'} \in \mathcal{G}_n^1 : \|g_{n,s,s'}\|_{2,\beta} < \delta\}.
\end{aligned}$$

Since $g_{n,s}$ satisfies (7), there exists a positive constant C_1 such that $\mathcal{G}_{n,\delta}^1 \subset \{g_{n,s,s'} \in \mathcal{G}_n^1 : \|g_{n,s,s'}\|_2 < C_1 \delta^{1/2}\}$ for all n large enough and all $\delta > 0$ small enough. Also, by the same argument to derive (17), there exists a positive constant C_2 such that $\|g_{n,s,s'}\|_{2,\beta} \leq C_2 \|g_{n,s,s'}\|_2$ for all n large enough, $|s| \leq K$ and $|s'| \leq K$. The constant C_2 depends only on the mixing sequence $\{\beta_m\}$. Combining these results, we obtain the set inclusion relationship

$$\mathcal{G}_{n,\delta}^1 \subset \mathcal{G}_{n,C_1 C_2 \delta^{1/2}}^\beta, \tag{30}$$

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

$$N_{[]}(\nu, \mathcal{G}_{n,\delta}^\beta, \|\cdot\|_{2,\beta}) \leq N_{[]}(\nu, \mathcal{G}_n^1, \|\cdot\|_2) \leq N_{[]}(\nu, \mathcal{G}_n, \|\cdot\|_2)^2,$$

where the first inequality follows from $\mathcal{G}_{n,\delta}^\beta \subset \mathcal{G}_n^1$ (by the definitions) and $\|\cdot\|_2 \leq \|\cdot\|_{2,\beta}$ (Doukhan, Massart and Rio, 1995, Lemma 1). Following the argument to derive Andrews (1993, eq. (4.7)) based on (7), we get a bound for $N_{[]}(\nu, \mathcal{G}_n, \|\cdot\|_2)$ by the covering number of the parameter space, say, $N_\Theta(\nu, [-K_1, K_1]^d \times [-K_2 a_n, K_2 a_n]^{k_n})$. By direct calculation, this covering number is bounded[12] by $(2K_1)^d \left(\frac{\sqrt{d+k_n}}{2\nu}\right)^{d+k_n} (2K_2 a_n)^{k_n}$ if $\nu < K_2 a_n \sqrt{d+k_n}$ and by $(2K_1)^d \left(\frac{\sqrt{d+k_n}}{2\nu}\right)^d$ otherwise. Thus,

$$\begin{aligned}
\varphi_n(\eta) &= \int_0^\eta \sqrt{\log N_{[]}(\nu, \mathcal{G}_{n,\eta}^\beta, \|\cdot\|_{2,\beta})} d\nu \\
&\leq \int_0^{K_2 a_n \sqrt{d+k_n}} \log (2K_1)^d \left(\frac{\sqrt{d+k_n}}{2\nu}\right)^{d+k_n} (2K_2 a_n)^{k_n} d\nu + \int_{K_2 a_n \sqrt{d+k_n}}^\eta d \log (2K_1)^d \left(\frac{\sqrt{d+k_n}}{2\nu}\right)^d d\nu,
\end{aligned}$$

for any η and large n as $K_2 a_n \sqrt{d+k_n} \to 0$. A straightforward algebra yields that the first term after the inequality goes to zero as $k_n a_n^{2/3} \to 0$, while the second term is bounded by $(c_1 \log k_n)\eta + c_2 \eta(\log \eta + 1)$ for some $c_1, c_2 > 0$. In other words, there is a function $\varphi(\eta)$ such that $\varphi(\eta) \to 0$ as $\eta \to 0$ and

$$\varphi_n(\eta) = \int_0^\eta \sqrt{\log N_{[]}(\nu, \mathcal{G}_{n,\eta}^\beta, \|\cdot\|_{2,\beta})} d\nu \leq (\log k_n) \varphi(\eta),$$

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

[12]The circumradius of the unit s-dimensional hypercube is $\sqrt{s}/2$. Or $\sqrt{\sum_{i=1}^s a_i^2}/2$ for the hypercube of side lengths $(a_1, \ldots, a_s)$.

Based on the above entropy condition, we can 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

$$P \sup_{g\in\mathcal{G}_{n,\eta}^{\beta}} |\mathbb{G}_n g| \leq C_4[1+\eta^{-1}q_{G_n}(\min\{1,v_n(\eta)\})]\varphi(\eta)(\log k_n),$$

for all n large enough and all $\eta>0$ small enough, where $q_{2G_n}(v)=\sup_{u\leq v} Q_{2G_n}(u)\sqrt{\int_0^u \beta^{-1}(\tilde{u})d\tilde{u}}$ with the envelope function $2G_n$ of $\mathcal{G}_{n,\eta}^{\beta}$ (note: by the definition of $\mathcal{G}_{n,\eta}^{\beta}$, the envelope $2G_n$ does not depend on 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 $2C_4\varphi(\eta)<\sigma$. Since $\varphi_n(\eta)\leq\varphi(\eta)$, there is a positive constant C_5 such that $v_n(\eta)\leq C_5\frac{\varphi(\eta)^2(\log k_n)^2}{n\eta^2}\leq 1$ for all n large enough and for any given $\eta>0$. Since $G_n\leq C'n^{\kappa}$ by the definition of $\mathcal{G}_{n,\eta}^{\beta}$, there exists a positive constant C_6 such that $q_{2G_n}(\min\{1,v_n(\eta)\})\leq C_6\varphi(\eta)\eta^{-1}n^{\kappa-1/2}(\log k_n)$ with $0<\kappa<1/2$ for all n large enough. Therefore, by setting $\eta=C_1C_2\delta^{1/2}$, we obtain

$$P \sup_{g\in\mathcal{G}_{n,C_1C_2\delta^{1/2}}^{\beta}} |\mathbb{G}_n g|\leq \sigma\log k_n,$$

for all n large enough. The conclusion follows by (30).

A.12. **Proof of Theorem 4.** To ease notation, let $\nu_0=0$. By (8), we can apply Lemma MS' with $g_{n,s}=\sqrt{h_n/a_n}(f_{n,\theta,v}-f_{n,\theta,0})$ for $s=(\theta',v')'$, which implies

$$\sup_{|\theta-\pi_\theta|\leq\epsilon,|v|\leq a_nK}\sqrt{h_n/a_n}\mathbb{G}_n(f_{n,\theta,v}-f_{n,\theta,0})=O_p(\log k_n), \tag{31}$$

for all small $\varepsilon>0$ and $K<\infty$. Also from (9), we have

$$P(f_{n,\theta,\hat{\nu}}-f_{n,\theta,0})-P(f_{n,\pi_\theta,\hat{\nu}}-f_{n,\pi_\theta,0})=o(|\theta-\pi_\theta|^2)+O(|\hat{\nu}|^2)+O_p(r_n^{-2/3}), \tag{32}$$

for all θ in a neighborhood of Θ_I and all $\epsilon>0$. Combining (31), (32), and Condition (i) of the set identified cube root class, and Lemma 3,

$$\begin{aligned}\mathbb{P}_n(f_{n,\theta,\hat{\nu}}-f_{n,\pi_\theta,\hat{\nu}}) &= n^{-1/2}\{\mathbb{G}_n(f_{n,\theta,\hat{\nu}}-f_{n,\theta,0})-\mathbb{G}_n(f_{n,\pi_\theta,\hat{\nu}}-f_{n,\pi_\theta,0})+\mathbb{G}_n(f_{n,\theta,0}-f_{n,\pi_\theta,0})\}\\ &\quad +P(f_{n,\theta,\hat{\nu}}-f_{n,\theta,0})-P(f_{n,\pi_\theta,\hat{\nu}}-f_{n,\pi_\theta,0})+P(f_{n,\theta,0}-f_{n,\pi_\theta,0})\\ &\leq O_p((nh_na_n^{-1})^{-1/2}\log k_n)+\epsilon|\theta-\pi_\theta|^2+O_p(r_n^{-2/3})\\ &\quad -c|\theta-\pi_\theta|^2+\epsilon|\theta-\pi_\theta|^2+O_p(|\hat{\nu}|^2)+O_p(r_n^{-2/3}),\end{aligned} \tag{33}$$

for all θ in a neighborhood of Θ_I and all $\epsilon>0$, where the last inequality follows from (5).

Let $\hat{\theta} = \arg\max_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta,\hat{v}}$. If $|\hat{\theta} - \pi_{\hat{\theta}}| > a_n + r_n^{-1/3}$, then $\mathbb{P}_n(f_{n,\theta,\hat{\nu}} - f_{n,\pi_\theta,\hat{\nu}}) \geq 0$ and thus by (33),

$$|\hat{\theta} - \pi_{\hat{\theta}}| \leq o(a_n) + O_p(r_n^{-1/3}) + O_p((nh_n a_n^{-1})^{-1/4}(\log k_n)^{1/2}). \tag{34}$$

And for any $\theta' \in \hat{\Theta}$, if $|\theta' - \pi_{\theta'}| > a_n + r_n^{-1/3}$, then

$$-(nh_n)^{-1/2}\hat{c} \leq \max_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta,\hat{v}} - \mathbb{P}_n f_{n,\pi_{\theta'},\hat{v}} - c_n^{-1}\hat{c} \leq \mathbb{P}_n f_{n,\theta',\hat{v}} - \mathbb{P}_n f_{n,\pi_{\theta'},\hat{v}},$$

and thus by (33),

$$|\theta' - \pi_{\theta'}| \leq o(a_n) + O_p(r_n^{-1/3}) + O_p((nh_n a_n^{-1})^{-1/4}(\log k_n)^{1/2}) + (nh_n)^{-1/4}\hat{c}^{1/2}.$$

It remains to show that $P\{\Theta_I \subset \hat{\Theta}\} \to 1$ for $\hat{c} \to \infty$. Proceeding as in (29), we get

$$\begin{aligned}
& (nh_n)^{1/2} \max_{\theta'\in\Theta_I} |(\max_{\theta\in\Theta} \mathbb{P}_n f_{n,\theta,\hat{v}} - \mathbb{P}_n f_{n,\theta',\hat{v}})| \\
\leq\ & |h_n^{1/2}\mathbb{G}_n(f_{n,\hat{\theta},\hat{v}} - f_{n,\pi_{\hat{\theta}},\hat{v}})| + (nh_n)^{1/2}|P(f_{n,\hat{\theta},\hat{v}} - f_{n,\pi_{\hat{\theta}},\hat{v}})| + 2(nh_n)^{1/2}|\max_{\theta'\in\Theta_I}(\mathbb{P}_n f_{n,\theta',\hat{v}} - P f_{n,\theta',\hat{v}})| \\
=\ & 2|\max_{\theta'\in\Theta_I} h_n^{1/2}\mathbb{G}_n f_{n,\theta',\hat{v}}| + o_p(1),
\end{aligned}$$

where the first term after the inequality being $o_p(1)$ is due to (31) and Lemma 3 and the second term is to (32) and Condition (i) of the set identified cube root class together with the rate for $\hat{\theta}$ in (34). Finally, due to (31) and the class $\{h_n^{1/2} f_{n,\theta}, \theta \in \Theta_I\}$ being a P-Donsker, we conclude $\Pr\{\Theta_I \subset \hat{\Theta}_I\} \to 1$.

A.13. **Proof of Lemma M1.** The proof is similar to that of Lemma M 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 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.

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] Belloni, A., Chen, D., Chernozhukov, V. and C. Hansen (2012) Sparse models and methods for optimal instruments with application to eminent domain, *Econometrica*, 80, 2369-2429.

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

[9] Carrasco, M. and X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory*, 17-39.

[10] Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models, *Handbook of Econometrics*, vol. 6B, ch. 76, Elsevier.

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

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

[13] Chernozhukov, V., Hong, H. and E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models, *Econometrica*, 75, 1243-1284.

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

[15] 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.

[16] 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.

[17] Gautier, E. and Y. Kitamura (2013) Nonparametric estimation in random coefficients binary choice models, *Econometrica*, 81, 581-607.

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

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

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

[21] Lee, M.-J. (1989) Mode regression, *Journal of Econometrics*, 42, 337-349.

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

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

[24] Moon, H. R. (2004) Maximum score estimation of a nonstationary binary choice model, *Journal of Econometrics*, 122, 385-403.

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

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

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

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

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

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

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

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

[33] 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