Journal of Multivariate Analysis 100 (2009) 1816–1829

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotics for argmin processes: Convexity arguments

Kengo Kato

Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

ARTICLE INFO

Article history:
Received 3 April 2008
Available online 25 February 2009

AMS 2000 subject classifications:
60B12
62E20

Keywords:
Argmin process
Convexity argument
Parametrized objective function
Regression quantile process
Representation theorem
Threshold regression

ABSTRACT

The convexity arguments developed by Pollard [D. Pollard, Asymptotics for least absolute deviation regression estimators, Econometric Theory 7 (1991) 186–199], Hjort and Pollard [N.L. Hjort, D. Pollard, Asymptotics for minimizers of convex processes, 1993 (unpublished manuscript)], and Geyer [C.J. Geyer, On the asymptotics of convex stochastic optimization, 1996 (unpublished manuscript)] are now basic tools for investigating the asymptotic behavior of M-estimators with non-differentiable convex objective functions. This paper extends the scope of convexity arguments to the case where estimators are obtained as stochastic processes. Our convexity arguments provide a simple proof for the asymptotic distribution of regression quantile processes. In addition to quantile regression, we apply our technique to LAD (least absolute deviation) inference for threshold regression.

1. Introduction

In this paper, we extend the scope of so-called "convexity arguments" to the case where estimators are obtained as stochastic processes. Suppose we have random functions $f_n(x, \tau)$ and $f_\infty(x, \tau)$ defined on $\mathbb{R}^d \times T$ that are convex in x. Here, $\tau \in T$ is a parameter and $T \subset \mathbb{R}^q$ is a compact set. We call these functions parametrized convex objective functions. Suppose $f_n(\cdot, \tau)$ and $f_\infty(\cdot, \tau)$ take minimum values at $x_n(\tau)$ and $x_\infty(\tau)$ for each τ, respectively. The problem to be addressed is, if f_n converges to f_∞ in some sense, under what conditions does $x_n(\cdot)$ converge weakly to $x_\infty(\cdot)$ as a process?

A canonical example appears in quantile regression [1], where the coefficient estimator is indexed by a quantile and called the regression quantile process. Another example appears in threshold regression with an unknown threshold parameter [2,3]. Consider testing the null hypothesis of no threshold under which the threshold parameter is not identified. In such a situation, we typically construct test statistics which may depend on the threshold parameter and reject the null hypothesis if the supremum of the test statistics is larger than a pre-specified value [4,5]. When we construct the Wald-type test statistics, we need to derive the asymptotic null distribution of the coefficient estimator as a stochastic process indexed by the threshold parameter, in order to calculate approximate critical values of the supremum of the test statistics.

The asymptotics of convex optimization has been studied by several authors including Pollard [6], Hjort and Pollard [7] and Geyer [8], whose convexity arguments appear attractive due to their simplicity. Let us explain a version of the convexity arguments briefly. Let $g_n(x)$ and $g_\infty(x)$ be random convex functions taking minimum values at x_n and x_∞, respectively. If all finite dimensional distributions of g_n converge weakly to those of g_∞ and x_∞ is the unique minimum point of g_∞ with probability one, then x_n converges weakly to x_∞. It seems now that the convexity arguments are basic tools for investigating asymptotic behavior of M-estimators with non-differentiable convex objective functions. For example, based on Geyer's [8] result, Knight [9] investigated the asymptotic behavior of the least absolute deviation (LAD) estimator under fairly general conditions; Knight and Fu [10] investigated the asymptotic properties of the Lasso [11]. However, to the

E-mail address: ee077006@mail.ecc.u-tokyo.ac.jp.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.02.008

author's best knowledge, existing literature on convexity arguments only deals with the case where objective functions are not parametrized and estimators are random vectors. It is thus a challenge to construct an asymptotic theory for argmin processes of parametrized convex objective functions.

One of the main theorems (Theorem 1) goes as follows. The notations will be explained later. Suppose (i) $f_n(x, \tau)(n \geq 1)$ and $f_\infty(x, \tau)$ are convex in x for each τ and continuous in τ for each x; (ii) $x_\infty(\tau)$ is the unique minimum point of $f_\infty(\cdot, \tau)$ for each $\tau \in T$; (iii) $x_n(\cdot) \in (\ell^\infty(T))^d$ and $x_\infty(\cdot) \in (C(T))^d$. Then, $x_\infty(\cdot)$ is a random element of $(C(T))^d$ and if $(f_n(x_1, \cdot), f_n(x_2, \cdot), \ldots, f_n(x_k, \cdot))$ converges weakly to $(f_\infty(x_1, \cdot), f_\infty(x_2, \cdot), \ldots, f_\infty(x_k, \cdot))$ in $(C(T))^k$ for each $k \geq 1$ where $\{x_1, x_2, \ldots\}$ is a countable dense subset of $\mathbb{R}^d$, $x_n(\cdot)$ converges weakly to $x_\infty(\cdot)$ in $(\ell^\infty(T))^d$.

This theorem makes no assumption on the preliminary asymptotic behavior of $x_n(\cdot)$. Showing the weak convergence of $(f_n(x_1, \cdot), f_n(x_2, \cdot), \ldots, f_n(x_k, \cdot))$ for $k \geq 1$ is the only substantially difficult point to check the conditions of Theorem 1. Moreover, it does not require that the limit process $x_\infty(\cdot)$ is Gaussian. If $f_\infty(x, \tau)$ is not quadratic in x, $x_\infty(\cdot)$ may be non-Gaussian.

In usual, when we show the weak convergence of a sequence of stochastic processes, we have to show the convergence of finite dimensional distributions and the asymptotic tightness of the sequence. However, when we do not know explicit forms of these stochastic processes, it is difficult to implement this procedure, especially to show the asymptotic tightness. By the usual convexity argument stated above, the convergence of finite dimensional distributions of $\{x_n(\cdot)\}$ is deduced from the convergence of finite dimensional distributions of $\{(f_n(x_1, \cdot), f_n(x_2, \cdot), \ldots, f_n(x_k, \cdot))\}$ for $k \geq 1$. So roughly speaking, the essential implication of Theorem 1 is "the asymptotic tightness of $\{(f_n(x_1, \cdot), f_n(x_2, \cdot), \ldots, f_n(x_k, \cdot))\}$ for $k \geq 1$ implies the asymptotic tightness of $\{x_n(\cdot)\}$". In addition to the above result, Theorem 2 shows that if $f_n(x, \tau)$ is asymptotically quadratic in x, we can derive an asymptotic representation of $x_n(\cdot)$. Theorem 2 is valid even when $f_n(x, \tau)$ is discontinuous in τ. So it is rather useful in some cases.

The organization of this paper is as follows. In Section 2, we present a general asymptotic theory for argmin processes of parametrized convex objective functions. In Section 3, we apply our techniques to some examples. Section 3.2 deals with the case where the limit process is non-Gaussian.

Here we explain some notations used in the present paper. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be the underlying probability space, P^* be the outer probability and E^* be the outer expectation. For details of outer probability and outer expectation, consult Pollard [12] or van der Vaart and Wellner [13]. Let $\rightsquigarrow$ denote "weak convergence" and $\xrightarrow{p}$ denote "convergence in probability" with respect to the outer probability. For any compact set $T \subset \mathbb{R}^q$, $C(T)$ denotes the space of real-valued continuous functions on T endowed with the uniform topology; $\ell^\infty(T)$ is the space of real-valued bounded functions on T endowed with the uniform topology; $C(\mathbb{R}^d \times T)$ is the space of real-valued continuous functions on $\mathbb{R}^d \times T$ endowed with the topology of locally uniform convergence. For any $a < b$, let $D[a, b]$ denote the space of càdlàg functions endowed with the Skorohod topology [14]. The spaces $C(T)$, $C(\mathbb{R}^d \times T)$ and $D[a, b]$ endowed with the above topologies are Polish. For any topological space S, S^k denotes the k-fold product space endowed with the product topology. Let $\mathbb{S}^{d-1}$ denote the set of d-dimensional unit vectors: $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$.

2. Asymptotics for argmin processes

2.1. Continuous mapping theorem for argmin processes

Let $f_n(x, \tau, \omega) : \mathbb{R}^d \times T \times \Omega \to \mathbb{R}$ $(n \geq 1)$ and $f_\infty(x, \tau, \omega) : \mathbb{R}^d \times T \times \Omega \to \mathbb{R}$ be random functions, i.e., $f_n(x, \tau, \cdot)$ and $f_\infty(x, \tau, \cdot)$ are random variables for each $(x, \tau) \in \mathbb{R}^d \times T$. For each (τ, ω), we define $x_n(\tau, \omega)$ and $x_\infty(\tau, \omega)$ by

$$x_n(\tau, \omega) \in \arg\min_{x \in \mathbb{R}^d} f_n(x, \tau, \omega), \qquad x_\infty(\tau, \omega) \in \arg\min_{x \in \mathbb{R}^d} f_\infty(x, \tau, \omega).$$

For simplicity, we assume that each argmin set is nonempty. We do not assume the measurability of the map $\omega \mapsto x_n(\tau, \omega)$ for each τ. Usually, we omit the argument ω.

We present the first main theorem, which may be considered as a suitably modified form of the continuous mapping theorem. The proof of this theorem uses the notion of a "perfect map". Take an arbitrary probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$ different from $(\Omega, \mathcal{F}, \mathrm{P})$. A measurable map $\phi : \tilde{\Omega} \to \Omega$ is called perfect if

$$\mathrm{E}^*[H] = \tilde{\mathrm{E}}^*[H \circ \phi]$$

for every bounded function H on Ω.

Theorem 1. *Suppose* (i) $f_n(x, \tau)$ $(n \geq 1)$ *and* $f_\infty(x, \tau)$ *are convex in* x *for each* τ *and continuous in* τ *for each* x; (ii) $x_\infty(\tau)$ *is the unique minimum point of* $f_\infty(\cdot, \tau)$ *for each* $\tau \in T$; (iii) $x_n(\cdot) \in (\ell^\infty(T))^d$ $(n \geq 1)$ *and* $x_\infty(\cdot) \in (C(T))^d$. *Then,* $x_\infty(\cdot)$ *is a random element of* $(C(T))^d$ *and if*

$$(f_n(x_1, \cdot), f_n(x_2, \cdot), \ldots, f_n(x_k, \cdot)) \rightsquigarrow (f_\infty(x_1, \cdot), f_\infty(x_2, \cdot), \ldots, f_\infty(x_k, \cdot)) \quad \text{in } (C(T))^k \tag{1}$$

for each $k \geq 1$ *where* $\{x_1, x_2, \ldots\}$ *is a countable dense subset of* $\mathbb{R}^d$, *we have*

$$x_n(\cdot) \rightsquigarrow x_\infty(\cdot) \quad \text{in } (\ell^\infty(T))^d. \tag{2}$$

Before proving the theorem, we add some remarks.

Remark 1. If a function $g(x,\tau):\mathbb{R}^d\times T\to\mathbb{R}$ is convex in x and continuous in τ, then $g(x,\tau)$ is jointly continuous in (x,τ) ([15], Theorem 10.7). Thus, the condition (i) implies that f_n and f_∞ are random elements of $C(\mathbb{R}^d\times T)$.

Remark 2. Since $x_\infty(\tau)$ is the unique minimum point of $f_\infty(\cdot,\tau)$ for each τ, Corollary 1 of Niemiro [16] shows that the map $\omega\mapsto x_\infty(\tau,\omega)$ is measurable for each τ. Combined with (ii), it is shown that $x_\infty(\cdot)$ is a random element of $(C(T))^d$.

Remark 3. Let $(C(T))^\infty$ be the countable product of $C(T)$ endowed with the product topology. The space $(C(T))^\infty$ is Polish ([14], Appendix M6). Let $\pi_k:(C(T))^\infty\to(C(T))^k$ be the natural projection. Then, by Theorem 2.4 of Billingsley [14], it is shown that the collection of every set of the form $\pi_k^{-1}(A)$ for Borel measurable $A\subset(C(T))^k$ and $k\geq 1$ is a convergence-determining class in $(C(T))^\infty$; see also Problem 3.7 in the 1st edition of Billingsley [14]. Therefore, the weak convergence (1) for every $k\geq 1$ is equivalent to

$$(f_n(x_1,\cdot),f_n(x_2,\cdot),\ldots)\rightsquigarrow(f_\infty(x_1,\cdot),f_\infty(x_2,\cdot),\ldots)\quad\text{in }(C(T))^\infty. \tag{3}$$

We are now in position to prove Theorem 1.

Proof of Theorem 1. Applying Dudley's [17] form of representation theorem to (3), there exist another probability space $(\tilde\Omega,\tilde{\mathcal{F}},\tilde{\mathrm{P}})$, perfect maps $\phi_n:\tilde\Omega\to\Omega$ and $\phi_\infty:\tilde\Omega\to\Omega$ such that

$$\sup_{\tau\in T}|\tilde f_n(x_i,\tau)-\tilde f_\infty(x_i,\tau)|\to 0 \tag{4}$$

almost surely for each i, where $\tilde f_n$ and $\tilde f_\infty$ are defined as $\tilde f_n(\cdot,\cdot,\tilde\omega)=f_n(\cdot,\cdot,\phi_n(\tilde\omega))$ and $\tilde f_\infty(\cdot,\cdot,\tilde\omega)=f_\infty(\cdot,\cdot,\phi_\infty(\tilde\omega))$ for each $\tilde\omega\in\tilde\Omega$. By the definition of a perfect map, $\tilde f_n$ and $\tilde f$ are random elements of $C(\mathbb{R}^d\times T)$ whose distributions are same as those of f_n and f_∞, respectively. Moreover, Lemma 3 in Appendix A.1 strengthens the almost sure convergence (4) for each i to the almost convergence of $\tilde f_n$ to $\tilde f_\infty$ in $C(\mathbb{R}^d\times T)$.

Define $\tilde x_n(\cdot,\tilde\omega)=x_n(\cdot,\phi_n(\tilde\omega))$ and $\tilde x_\infty(\cdot,\tilde\omega)=x_\infty(\cdot,\phi_\infty(\tilde\omega))$. It is straightforward to see that

$$\tilde x_n(\tau)\in\arg\min_{x\in\mathbb{R}^d}\tilde f_n(x,\tau),\qquad \tilde x_\infty(\tau)\in\arg\min_{x\in\mathbb{R}^d}\tilde f_\infty(x,\tau)$$

for each τ. To show the weak convergence (2), it suffices to show

$$\lim_{n\to\infty}\tilde{\mathrm{P}}^*\left(\sup_{\tau\in T}\|\tilde x_n(\tau)-\tilde x_\infty(\tau)\|>\delta\right)=0$$

for every $\delta>0$.

Fix any $\delta>0$. Define

$$\tilde\eta=\inf_{\tau\in T}\inf_{u\in\mathbb{S}^{d-1}}\{\tilde f_\infty(\tilde x_\infty(\tau)+\delta u,\tau)-\tilde f_\infty(\tilde x_\infty(\tau),\tau)\}.$$

Since $(x,\tau)\mapsto\tilde f_\infty(x,\tau)$ is continuous, so is the map $\mathbb{R}^d\times T\ni(u,\tau)\mapsto\tilde f_\infty(\tilde x_\infty(\tau)+\delta u,\tau)-\tilde f_\infty(\tilde x_\infty(\tau),\tau)$. Because of the compactness of the set $\mathbb{S}^{d-1}\times T$ and (ii), $\tilde\eta$ is a positive random variable on $(\tilde\Omega,\tilde{\mathcal{F}},\tilde{\mathrm{P}})$.

Since $\tilde f_n(x,\tau)$ is convex in x, for every $u\in\mathbb{S}^{d-1}$ and every $l>\delta$,

$$\left(1-\frac{\delta}{l}\right)\tilde f_n(\tilde x_\infty(\tau),\tau)+\frac{\delta}{l}\tilde f_n(\tilde x_\infty(\tau)+lu,\tau)\geq\tilde f_n(\tilde x_\infty(\tau)+\delta u,\tau).$$

Let $\tilde\Delta_n(x,\tau)=\tilde f_n(x,\tau)-\tilde f_\infty(x,\tau)$. Then

$$\begin{aligned}\frac{\delta}{l}\left\{\tilde f_n(\tilde x_\infty(\tau)+lu,\tau)-\tilde f_n(\tilde x_\infty(\tau),\tau)\right\}&\geq\tilde f_n(\tilde x_\infty(\tau)+\delta u,\tau)-\tilde f_n(\tilde x_\infty(\tau),\tau)\\&=\left\{\tilde f_\infty(\tilde x_\infty(\tau)+\delta u,\tau)-\tilde f_\infty(\tilde x_\infty(\tau),\tau)\right\}+\left\{\tilde\Delta_n(\tilde x_\infty(\tau)+\delta u,\tau)-\tilde\Delta_n(\tilde x_\infty(\tau),\tau)\right\}.\end{aligned} \tag{5}$$

Therefore, for every $u\in\mathbb{S}^{d-1}$, $l>\delta$ and $\tau\in T$,

$$\frac{\delta}{l}\left\{\tilde f_n(\tilde x_\infty(\tau)+lu,\tau)-\tilde f_n(\tilde x_\infty(\tau),\tau)\right\}\geq\tilde\eta-2\tilde\Delta_n, \tag{6}$$

where $\tilde\Delta_n=\sup_{\tau\in T}\sup_{x:\|x-\tilde x_\infty(\tau)\|\leq\delta}|\tilde\Delta_n(x,\tau)|$. If $\tilde x_n(\tau)$ lies outside the set $\{x:\|x-\tilde x_\infty(\tau)\|\leq\delta\}$ for some τ, the right-hand side (henceforth, rhs) of (6) must be non-positive since $\tilde x_n(\tau)$ is a minimum point of $\tilde f_n(\cdot,\tau)$. This implies

$$\tilde{\mathrm{P}}^*\left(\sup_{\tau\in T}\|\tilde x_n(\tau)-\tilde x_\infty(\tau)\|>\delta\right)\leq\tilde{\mathrm{P}}(\tilde\Delta_n\geq\tilde\eta/2).$$

Since $\tilde{\eta}$ is a positive random variable, it suffices to show $\tilde{\Delta}_n \xrightarrow{p} 0$.

For any $\epsilon > 0$, take $M > 0$ such that

$$\tilde{\mathrm{P}}\left(\sup_{\tau \in T} \|\tilde{x}_\infty(\tau)\| > M\right) \le \epsilon.$$

The existence of such M is guaranteed since $\tilde{x}_\infty(\cdot)$ is a random element of $(C(T))^d$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$. Define $K = \{x : \|x - y\| \le \delta, \|y\| \le M\} = \{x : \|x\| \le \delta + M\}$. Clearly, K is a compact set in $\mathbb{R}^d$. Then for any $\xi > 0$,

$$\begin{aligned}\tilde{\mathrm{P}}(\tilde{\Delta}_n > \xi) &= \tilde{\mathrm{P}}(\tilde{\Delta}_n > \xi, \sup_{\tau \in T} \|\tilde{x}_\infty(\tau)\| \le M) + \tilde{\mathrm{P}}(\tilde{\Delta}_n > \xi, \sup_{\tau \in T} \|\tilde{x}_\infty(\tau)\| > M) \\ &\le \tilde{\mathrm{P}}\left(\sup_{\tau \in T}\sup_{x \in K} |\tilde{f}_n(x, \tau) - \tilde{f}_\infty(x, \tau)| > \xi\right) + \epsilon. \end{aligned} \tag{7}$$

Since the first term of the rhs of (7) converges to 0, we obtain

$$\limsup_{n \to \infty} \tilde{\mathrm{P}}(\tilde{\Delta}_n > \xi) \le \epsilon.$$

Because $\epsilon > 0$ is arbitrary, the proof ends. □

The key probabilistic tool in the above proof is the representation theorem. Applications of the representation theorem to the asymptotics of M-estimators are found in Kim and Pollard [18], Davis et al. [19], Geyer [8]. For an exposition of Dudley's form of representation theorem, see also Pollard [20] or van der Vaart and Wellner [13].

It should be noted that there is a notable difference between Theorem 1 of the present paper and the argmax theorem in [13]. Theorem 3.2.2 of van der Vaart and Wellner [13] allows the case that the estimator is a stochastic process; however, this theorem states the weak convergence of the stochastic process that maximizes the objective function. Under our formulation, $x_n(\cdot)$ does not minimize any objective function as a stochastic process; $x_n(\tau)$ does minimize $f_n(x, \tau)$ with respect to x for each τ.

In Theorem 1, we assume the existence of a process $x_n(\cdot)$ such that $x_n(\tau)$ is a minimum point of $f_n(\cdot, \tau)$ for each τ and $\tau \mapsto x_n(\tau)$ is bounded. In examples below (see Sections 3.1–3.3), it is possible to show explicitly the existence of such a process. In general, this condition can be checked in the course of proving consistency: Assume f_∞ is non-stochastic and the conditions (i)–(iii) of Theorem 1 except for the condition on $x_n(\cdot)$ are satisfied. If $f_n(x, \cdot) \xrightarrow{p} f_\infty(x, \cdot)$ in $C(T)$ for each x, then it can be shown that there exists a sequence of bounded stochastic processes $x_n(\cdot)$ uniformly converging in probability to $x_\infty(\cdot)$ such that with probability approaching one, $x_n(\tau)$ is a minimum point of $f_n(\cdot, \tau)$ for each τ. This result can be deduced from the proof of Theorem 1. Then, Theorem 1 is typically applied to the local objective function

$$g_n(x, \tau) = r_n\{f_n(x_\infty(\tau) + a_n^{-1}x, \tau) - f_n(x_\infty(\tau), \tau)\}$$

to obtain the asymptotic distribution of the normalized process $a_n(x_n(\cdot) - x_\infty(\cdot))$, where a_n is the convergence rate of the process and r_n is determined according to a_n.

Finally, we remark that the conditions of Theorem 1 are high-level and more primitive conditions could be derived in concrete examples.

2.2. *Asymptotic representation of argmin processes*

In many applications, $f_n(x, \tau)$ is asymptotically quadratic in x. In this situation, we can derive an asymptotic representation of $x_n(\cdot)$ under suitable regularity conditions. We follow the notations used in the previous section. The next lemma is a slight generalization of the famous "CONVEXITY LEMMA" in Pollard [6].

Lemma 1. *Suppose $f_n(x, \tau)$ and $f_\infty(x, \tau)$ are convex in x for each τ and bounded in τ for each x. Furthermore, we assume that $f_\infty(x, \tau)$ is a non-stochastic function. If*

$$\sup_{\tau \in T} |f_n(x, \tau) - f_\infty(x, \tau)| \xrightarrow{p} 0 \tag{8}$$

for each x, then

$$\sup_{\tau \in T}\sup_{x \in K} |f_n(x, \tau) - f_\infty(x, \tau)| \xrightarrow{p} 0 \tag{9}$$

for every compact set K in $\mathbb{R}^d$.

Proof. From the proof of Lemma 3 in Appendix A.1, there exists a constant $\alpha > 0$ such that

$$|f_\infty(y, \tau) - f_\infty(x, \tau)| \le \alpha\|y - x\|, \quad \forall x, y \in K, \forall \tau \in T.$$

Using this property, a slight modification of the proof of CONVEXITY LEMMA in Pollard [6] yields the desired result. □

Remark 4. Under a suitable measurability assumption, the assertion of Lemma 1 is true for a stochastic limit function. This can be shown by combining the diagonal argument and Lemma 3 in Appendix A.1.

The second main theorem goes as follows.

Theorem 2. *Suppose $f_n(x,\tau)$ $(n \geq 1)$ are convex in x for each τ and bounded in τ for each x. Let $g_n(x,\tau) = -x'W_n(\tau) + \frac{1}{2}x'Q(\tau)x$, where $\{W_n(\cdot)\}$ is a sequence of bounded stochastic processes and $Q(\tau)$ is a $d \times d$ non-stochastic symmetric positive definite matrix for each τ. Furthermore, we assume that the maximum eigenvalue of $Q(\tau)$ is bounded from above and the minimum eigenvalue of $Q(\tau)$ is bounded away from 0 over $\tau \in T$. If*

$$\sup_{\tau \in T} |f_n(x,\tau) - g_n(x,\tau)| \xrightarrow{p} 0 \tag{10}$$

for each x and if for every $\eta > 0$, there exists a constant $M > 0$ such that

$$\limsup_{n\to\infty} \mathrm{P}^*(\sup_{\tau \in T} \|W_n(\tau)\| > M) \leq \eta, \tag{11}$$

then

$$x_n(\tau) = \{Q(\tau)\}^{-1} W_n(\tau) + r_n(\tau),$$

where $\sup_{\tau\in T} \|r_n(\tau)\| = o_p(1)$.

Proof. Let $y_n(\tau) = \{Q(\tau)\}^{-1}W_n(\tau)$, which is the unique minimum point of $g_n(\cdot,\tau)$ for each τ. Then a simple calculation shows that

$$\begin{aligned} g_n(x,\tau) - g_n(y_n(\tau),\tau) &= \frac{1}{2}(x - y_n(\tau))'Q(\tau)(x - y_n(\tau)) \\ &\geq c\|x - y_n(\tau)\|^2 \end{aligned} \tag{12}$$

for some constant $c > 0$.

Let $\delta > 0$ be an arbitrary positive constant. Taking $\tilde{f}_n(x,\tau) = f_n(x,\tau), \tilde{f}_\infty(x,\tau) = g_n(x,\tau)$ and $\tilde{x}_\infty(\tau) = y_n(\tau)$ in (5) and applying (12) to the first term of the rhs of (5), we have for every $u \in \mathbb{S}^{d-1}$, $l > \delta$ and $\tau \in T$,

$$\frac{\delta}{l}\{f_n(y_n(\tau) + lu,\tau) - f_n(y_n(\tau),\tau)\} \geq c\delta^2 - 2\Delta_n,$$

where $\Delta_n = \sup_{\tau\in T}\sup_{x:\|x-y_n(\tau)\|\leq\delta} |f_n(x,\tau) - g_n(x,\tau)|$. Therefore letting $r_n(\tau) = x_n(\tau) - y_n(\tau)$, we have

$$\mathrm{P}^*(\sup_{\tau\in T}\|r_n(\tau)\| > \delta) \leq \mathrm{P}^*(\Delta_n \geq (c\delta^2)/2).$$

So it suffices to show $\Delta_n \xrightarrow{p} 0$.

Let $\eta > 0$ be an arbitrary positive constant. Take $M > 0$ such that

$$\limsup_{n\to\infty} \mathrm{P}^*(\sup_{\tau\in T}\|y_n(\tau)\| > M) \leq \eta.$$

Define $K = \{x : \|x - y\| \leq \delta, \|y\| \leq M\} = \{x : \|x\| \leq \delta + M\}$. Then for every $\epsilon > 0$,

$$\mathrm{P}^*(\Delta_n > \epsilon) \leq \mathrm{P}^*(\sup_{\tau\in T}\sup_{x\in K} |f_n(x,\tau) - g_n(x,\tau)| > \epsilon) + \mathrm{P}^*(\sup_{\tau\in T}\|y_n(\tau)\| > M).$$

Since, by Lemma 1,

$$\sup_{\tau\in T}\sup_{x\in K} |f_n(x,\tau) - g_n(x,\tau)| = \sup_{\tau\in T}\sup_{x\in K} \left|\{f_n(x,\tau) + x'W_n(\tau)\} - \frac{1}{2}x'Q(\tau)x\right| \xrightarrow{p} 0,$$

we conclude that

$$\limsup_{n\to\infty} \mathrm{P}^*(\Delta_n > \epsilon) \leq \eta.$$

Since $\eta > 0$ is arbitrary, the proof ends. □

If $W_n(\cdot) \rightsquigarrow W(\cdot)$ in $(\ell^\infty(T))^d$ for some tight random element $W(\cdot)$ of $(\ell^\infty(T))^d$, then (11) is satisfied by the continuous mapping theorem. In this case, we have $\hat{x}_n(\cdot) \rightsquigarrow \{Q(\cdot)\}^{-1}W(\cdot)$ in $(\ell^\infty(T))^d$, where $\hat{x}_n(\cdot,\omega) = x_n(\cdot,\omega)$ if $x_n(\cdot,\omega) \in (\ell^\infty(T))^d$ and 0 otherwise.

3. Examples

3.1. Quantile regression

We first consider the asymptotic distribution of regression quantile processes. Quantile regression was originally proposed in Koenker and Bassett [1] and has been used in many areas. For a comprehensive treatment of quantile regression, see Koenker [21].

We here consider the following linear location-scale model as in Gutenbrunner and Jurečková [22]:

$$y_{in} = x_{in}'\beta + (x_{in}'\gamma)\epsilon_i, \quad i = 1, 2, \ldots, n,$$

where x_{in} are non-stochastic covariates with $x_{in1} = 1$, $\beta \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^p$ are unknown coefficient vectors and ϵ_i are i.i.d. random variables with a common distribution function F. Let $X_n = [x_{1n} \cdots x_{nn}]'$ be the design matrix. The term $x_{in}'\gamma$ (> 0) corresponds to the scale function of y_{in}. A regression quantile process $\{\hat{\beta}(\tau), \tau \in (0, 1)\}$ is defined by

$$\hat{\beta}(\tau) \in \arg\min_{b\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{in} - x_{in}'b),$$

for each τ, where $\rho_\tau(r) = \tau(r)_+ + (1-\tau)(-r)_+$ and $(t)_+ = \max\{0, t\}$. As indicated in Gutenbrunner and Jurečková [22], it is possible to select $\hat{\beta}(\cdot)$ such that the path $\tau \mapsto \hat{\beta}(\tau)$ is càdlàg.

Let $\beta(\tau) = \beta + F^{-1}(\tau)\gamma$. Since $\mathrm{P}(y_{in} \le x_{in}'\beta(\tau)) = \mathrm{P}(\epsilon_i \le F^{-1}(\tau)) = \tau$, $\beta(\tau)$ is actually the "true" value of the regression τ-quantile. Observe that

$$\begin{aligned} y_{in} - x_{in}'b &= x_{in}'\beta + (x_{in}'\gamma)\epsilon_i - x_{in}'b \\ &= x_{in}'(\beta(\tau) - F^{-1}(\tau)\gamma) + (x_{in}'\gamma)\epsilon_i - x_{in}'b \\ &= x_{in}'(\beta(\tau) - b) + (x_{in}'\gamma)(\epsilon_i - F^{-1}(\tau)). \end{aligned}$$

Then, the local objective function may be defined as

$$Z_n(u, \tau) = \sum_{i=1}^{n} \left\{ \rho_\tau\left((x_{in}'\gamma)(\epsilon_i - F^{-1}(\tau)) - n^{-1/2}x_{in}'u\right) - \rho_\tau\left((x_{in}'\gamma)(\epsilon_i - F^{-1}(\tau))\right) \right\}, \tag{13}$$

for $u \in \mathbb{R}^p$ and $\tau \in (0, 1)$. The normalized estimator $n^{1/2}(\hat{\beta}(\tau) - \beta(\tau))$ satisfies

$$n^{1/2}(\hat{\beta}(\tau) - \beta(\tau)) \in \arg\min_{u\in\mathbb{R}^p} Z_n(u, \tau)$$

for each τ. It is not difficult to show that $u \mapsto Z_n(u, \tau)$ is convex for each τ and $\tau \mapsto Z_n(u, \tau)$ is continuous for each u under the condition (i) stated below. Hence, we can apply both Theorems 1 and 2 to this example.

We impose the following conditions, which seem to be standard in quantile regression. Put $\sigma_{in} = x_{in}'\gamma$ and $\Sigma_n = \mathrm{diag}\{\sigma_{1n}, \ldots, \sigma_{nn}\}$.

Assumption 1. (i) F has continuous Lebesgue density f, which is positive on $\{t : 0 < F(t) < 1\}$.
(ii) $\max_{1\le i\le n} \|x_{in}\| = o(n^{1/2})$.
(iii) There exists a symmetric positive definite matrix Q such that $n^{-1}X_n'X_n \to Q$.
(iv) There exists a symmetric positive definite matrix D such that $n^{-1}X_n'\Sigma_n^{-1}X_n \to D$.
(v) There exist positive constants $\sigma_L \le \sigma_U$ such that $\sigma_L \le \sigma_{in} \le \sigma_U$ for all $1 \le i \le n$ and $n \ge 1$.

The next theorem is concerned with the asymptotic behavior of $Z_n(u, \tau)$. As a corollary to the theorem, we obtain the asymptotic distribution of the regression quantile process. Gutenbrunner and Jurečková [22] originally established the asymptotic distribution of the regression quantile process under the same setting. The proof below is actually another proof of their result. They showed in advance that $\hat{\beta}(\tau) - \beta(\tau) = O_p(n^{-1/2})$ uniformly in $\tau \in [\alpha, 1-\alpha]$ for any $\alpha \in (0, 1/2)$. Then, they used the computational property of regression quantile processes to derive the uniform asymptotic representation of $n^{1/2}(\hat{\beta}(\cdot) - \beta(\cdot))$. The asymptotic theory of regression quantile processes has been an important subject and studied by several authors including Koenker and Portnoy [23], Portnoy [24], Gutenbrunner et al. [25], Koul and Saleh [26] and Koltchinskii [27]. More recently, based on the empirical process theory, Angrist et al. [28] developed a novel proof to this subject for a quantile regression model with stochastic covariates. They first showed the uniform consistency of regression quantile processes. Then, they used the computational property of regression quantile processes and the fact that the functional class $\{g(y, x) = (\tau - I(y \le x'\beta))x_j, \beta \in B, \tau \in [\alpha, 1-\alpha], 1 \le j \le p\}$ for any compact $B \subset \mathbb{R}^p$ is P-Donsker to derive the asymptotic distribution of regression quantile processes.

The contribution of this paper is to bridge the gap between the proof for the vector case and the proof for the process case. The proof below shows that the asymptotic distribution of the regression quantile process can be obtained by merely showing the asymptotic tightness of the objective function as a stochastic process with index τ in addition to Knight's [9] proof for the vector case. This proof does not use, for example, the uniform consistency of $\hat{\beta}(\tau)$. For reference, we make the proof as self-contained as possible.

Theorem 3. *Under Assumption 1, for any* $\alpha \in (0, 1/2)$,

$$Z_n(u,\tau) = -n^{-1/2}\sum_{i=1}^n x_{in}' u\{\tau - I(\epsilon_i \le F^{-1}(\tau))\} + f(F^{-1}(\tau))\frac{u'Du}{2} + \Delta_n(u,\tau),$$

where $\sup_{\tau\in[\alpha,1-\alpha]} |\Delta_n(u,\tau)| = o_p(1)$ *for each u and*

$$n^{-1/2}\sum_{i=1}^n x_{in}\{\tau - I(\epsilon_i \le F^{-1}(\cdot))\} \rightsquigarrow Q^{1/2}W^*(\cdot) \quad in\ (\ell^\infty[\alpha, 1-\alpha])^p. \tag{14}$$

Here W^* *is a vector of p independent Brownian bridges in* $C[0, 1]$.

Proof. Using Knight's [9] identity

$$\rho_\tau(x-y) - \rho_\tau(x) = -y\{\tau - I(x\le 0)\} + y\int_0^1 \{I(x \le ys) - I(x\le 0)\}\mathrm{d}s, \tag{15}$$

we decompose $Z_n(u,\tau)$ as

$$Z_n(u,\tau) = Z_n^{(1)}(u,\tau) + Z_n^{(2)}(u,\tau),$$

where

$$Z_n^{(1)}(u,\tau) = -n^{-1/2}\sum_{i=1}^n x_{in}'u\{\tau - I(\epsilon_i\le F^{-1}(\tau))\},$$

$$Z_n^{(2)}(u,\tau) = n^{-1}\sum_{i=1}^n x_{in}'u\int_0^1 n^{1/2}\{I(\epsilon_i \le F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'us) - I(\epsilon_i \le F^{-1}(\tau))\}\mathrm{d}s.$$

We further decompose $Z_n^{(2)}(u,\tau)$ as

$$\begin{aligned} Z_n^{(2)}(u,\tau) &= \mathrm{E}[Z_n^{(2)}(u,\tau)] + \{Z_n^{(2)}(u,\tau) - \mathrm{E}[Z_n^{(2)}(u,\tau)]\}, \\ &=: Z_n^{(21)}(u,\tau) + Z_n^{(22)}(u,\tau). \end{aligned}$$

First, we show the weak convergence (14). Koul [29] establishes the weak convergence of weighted empirical distribution functions under more general conditions. In this case, (14) can be shown as follows: Use the Lindeberg–Feller central limit theorem to show the finite dimensional convergence. Then, check the condition of Theorem 13.5 in Billingsley [14] to show the asymptotic tightness in $(D[\alpha, 1-\alpha])^p$. In fact, it can be shown that $\mathrm{E}[\{Z_n(u,\tau) - Z_n(u,\tau_1)\}^2\{Z_n(u,\tau_2) - Z_n(u,\tau)\}^2] \le 3\{n^{-1}\sum_{i=1}^n (x_{in}'u)^2\}^2(\tau_2-\tau_1)^2$ for any $\tau_1 \le \tau \le \tau_2$ and $u \in \mathbb{R}^p$ (see Billingsley [14], p. 150). This proves the asymptotic tightness in $(D[\alpha, 1-\alpha])^p$. Since the limit process is continuous, we obtain the weak convergence in $(\ell^\infty[\alpha, 1-\alpha])^p$.

Next, consider the asymptotic behavior of $Z_n^{(21)}(u,\tau)$. Let $G(t,\tau) = n^{1/2}\{F(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'t) - \tau\}$. Since

$$\frac{\mathrm{d}}{\mathrm{d}s}G(ts,\tau) = \sigma_{in}^{-1}x_{in}'tf(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'ts)$$

and $G(0,\tau) = 0$, we have

$$G(t,\tau) = \sigma_{in}^{-1}x_{in}'t\int_0^1 f(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'ts)\mathrm{d}s,$$

which implies

$$|G(t,\tau) - \sigma_{in}^{-1}x_{in}'tf(F^{-1}(\tau))| \le |\sigma_{in}^{-1}x_{in}'t|\int_0^1 |f(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'ts) - f(F^{-1}(\tau))|\mathrm{d}s.$$

We note $|n^{-1/2}\sigma_{in}^{-1}x_{in}'t| \le n^{-1/2}\sigma_L^{-1}\|t\|\max_{1\le i\le n}\|x_{in}\|$ and $n^{-1/2}\max_{1\le i\le n}\|x_{in}\| = o(1)$. Since f is uniformly continuous on each bounded interval,

$$\max_{1\le i\le n}\sup_{\tau\in[\alpha,1-\alpha]}\sup_{t:\|t\|\le M}\int_0^1 |f(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'ts) - f(F^{-1}(\tau))|\mathrm{d}s \to 0,$$

for each $M > 0$. Therefore we have

$$\sup_{\tau\in[\alpha,1-\alpha]}\left|Z_n^{(21)}(u,\tau) - f(F^{-1}(\tau))\frac{u'Du}{2}\right| \to 0,$$

for each u.

Finally, we show $\sup_{\tau\in[\alpha,1-\alpha]}|Z_n^{(22)}(u,\tau)| = o_p(1)$ for each u. Suppose for a moment that u is arbitrarily fixed. Define

$$W_n(t,\tau) = n^{-1/2}\sum_{i=1}^n x_{in}'u\{I(\epsilon_i \le F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'t) - F(F^{-1}(\tau) + n^{-1/2}\sigma_{in}^{-1}x_{in}'t)\}$$

for $t \in \mathbb{R}^p$ and $\tau \in (0, 1)$. Then,

$$Z_n^{(22)}(u,\tau) = \int_0^1 \{W_n(us,\tau) - W_n(0,\tau)\}\mathrm{d}s.$$

Therefore it is enough to show that

$$\sup_{\tau\in[\alpha,1-\alpha]}\ \sup_{t:\|t\|\le M}|W_n(t,\tau) - W_n(0,\tau)| \xrightarrow{p} 0, \tag{16}$$

for each $M > 0$. A proof of (16) is found in Appendix of Koul [30] (see also [31], Theorem 1.2). For reference, we derive its direct proof in Appendix A.2. □

Combining Theorems 2 and 3, we obtain the next corollary.

Corollary 1. *Under Assumption 1, for any $\alpha \in (0, 1/2)$,*

$$n^{1/2}(\hat\beta(\tau) - \beta(\tau)) = \{f(F^{-1}(\tau))\}^{-1}D^{-1}n^{-1/2}\sum_{i=1}^n x_{in}\{\tau - I(\epsilon_i \le F^{-1}(\tau))\} + r_n(\tau),$$

where $\sup_{\tau\in[\alpha,1-\alpha]}\|r_n(\tau)\| = o_p(1)$. Especially,

$$n^{1/2}(\hat\beta(\cdot) - \beta(\cdot)) \rightsquigarrow \{f(F^{-1}(\cdot))\}^{-1}D^{-1}Q^{1/2}W^*(\cdot)\quad \text{in } (\ell^\infty[\alpha, 1-\alpha])^p,$$

where W^* is a vector of p independent Brownian bridges in $C[0, 1]$.

It is worthwhile to see how the conditions of this example relate to the conditions of Theorem 2. The conditions (i)–(iii) of Assumption 1 guarantee the weak convergence (14), which satisfies the condition (11). The condition (i) and the positive definiteness of D are used to satisfy the condition on $Q(\tau)$ in Theorem 2, where in this case $Q(\tau)$ corresponds to $f(F^{-1}(\tau))D$. To guarantee the condition (10), all the conditions of Assumption 1 are used.

3.2. Quantile regression with ℓ_1 penalization

Since the Lasso was proposed by Tibshirani [11], the ℓ_1 penalization has attracted much attention as it enables us to implement simultaneous estimation and variable selection. Asymptotic properties of the Lasso are studied in Knight and Fu [10] based on Geyer's [8] convexity argument. In this subsection we investigate asymptotic properties of ℓ_1 penalized regression quantile processes.

For simplicity, consider the following linear model:

$$y_i = x_i'\beta + \epsilon_i,\quad i = 1, 2, \dots, n,$$

where x_i are non-stochastic covariates, $\beta \in \mathbb{R}^p$ is an unknown coefficient vector and ϵ_i are i.i.d. random variables with a common distribution function F. Without loss of generality, we assume that the covariates are centered and the intercept term is not included in the above linear model. For each $\tau \in (0, 1)$, we consider the estimator $(\hat a(\tau), \hat\beta(\tau))$ that minimizes the ℓ_1 penalized objective function:

$$(\hat a(\tau), \hat\beta(\tau)) \in \operatorname*{arg\,min}_{a\in\mathbb{R},b\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i - a - x_i'b) + \frac{\lambda_n}{n}\sum_{j=1}^p |b_j|.$$

As well as regression quantile processes, it is possible to select $(\hat a(\tau), \hat\beta(\tau))$ such that it has càdlàg paths. Suppose $\lambda_n/n^{1/2} \to \lambda_0 \in [0,\infty)$. Under the conditions (i) and (iii) of Assumption 1 with $x_{in} = x_i$, combining Theorems 1 and 3 of the present paper and Theorem 2 of Knight and Fu [10], it can be shown that $n^{1/2}(a(\cdot) - F^{-1}(\cdot))$ converges weakly in $\ell^\infty[\alpha, 1-\alpha]$ to a $\{f(F^{-1}(\cdot))\}^{-1}$ multiple of a Brownian bridge on [0, 1] and $n^{1/2}(\hat\beta(\cdot) - \beta)$ converges weakly in $(\ell^\infty[\alpha, 1-\alpha])^p$ to $U(\cdot)$, where $U(\tau)$ uniquely minimizes

$$Z_\infty(u,\tau) = -u'Q^{1/2}W^*(\tau) + f(F^{-1}(\tau))\frac{u'Qu}{2} + \lambda_0\sum_{j=1}^p\{u_j\,\mathrm{sgn}(\beta_j)I(\beta_j \ne 0) + |u_j|I(\beta_j = 0)\},$$

with respect to $u \in \mathbb{R}^p$ for each $\tau \in [\alpha, 1-\alpha]$ [note: $n^{1/2}(a(\cdot) - F^{-1}(\cdot))$ and $n^{1/2}(\hat\beta(\cdot) - \beta)$ are asymptotically independent]. Here $\alpha \in (0, 1/2)$ is arbitrary, $\mathrm{sgn}(\cdot)$ is the sign function and W^* is a vector of p independent Brownian bridges on [0, 1].

The continuity of the map $\tau \mapsto U(\tau)$ can be shown by combining the fact that $\inf_{\tau\in[\alpha,1-\alpha]} Z_\infty(u,\tau) \to \infty$ as $\|u\| \to \infty$ and Berge's [32] maximum theorem. In the simplest case where $Q = I_p$,

$$U_j(\tau) = \begin{cases} \{f(F^{-1}(\tau))\}^{-1}(W_j^*(\tau) - \lambda_0 \operatorname{sgn}(\beta_j)), & \text{if } \beta_j \neq 0, \\ \{f(F^{-1}(\tau))\}^{-1} \operatorname{sgn}(W_j^*(\tau))(|W_j^*(\tau)| - \lambda_0)_+, & \text{if } \beta_j = 0. \end{cases}$$

It is seen that when $\beta_j = 0$, the sample path of U_j tends to be degenerate at 0.

3.3. LAD inference for threshold regression

In this subsection, we consider the following threshold regression model as in Hansen [2,3]:

$$\begin{aligned} y_i &= x_i'\theta^{(1)} + \epsilon_i, \quad \text{if } q_i \le \gamma_0, \\ &= x_i'\theta^{(2)} + \epsilon_i, \quad \text{if } q_i > \gamma_0, \end{aligned} \tag{17}$$

for $i = 1, 2, \ldots, n$, where x_i are stochastic covariates, $\theta^{(1)}, \theta^{(2)} \in \mathbb{R}^p$ are unknown coefficient vectors and γ_0 is an unknown threshold parameter. We assume the parameter space of γ_0 is a bounded closed interval $[\gamma_L, \gamma_U]$. If we put $\beta^{(1)} = \theta^{(2)}$ and $\beta^{(2)} = \theta^{(1)} - \theta^{(2)}$, the above threshold regression model (17) can be rewritten as

$$\begin{aligned} y_i &= x_i'\beta^{(1)} + I(q_i \le \gamma_0)x_i'\beta^{(2)} + \epsilon_i, \\ &= z_i(\gamma_0)'\beta + \epsilon_i, \end{aligned}$$

where $z_i(\gamma_0) = (x_i', I(q_i \le \gamma_0)x_i')'$ and $\beta = (\beta^{(1)'}, \beta^{(2)'})'$.

We consider the statistical inference based on quantile regression, which is natural when the main purpose is to estimate the conditional quantile instead of the conditional mean. For simplicity, we deal here with LAD estimation. Specifically, we consider testing the null hypothesis $\beta^{(2)} = 0$ (no threshold) and derive the asymptotic null distribution of the sup-Wald test statistic specified below. To do this, we derive the asymptotic null distribution of the process $\hat\beta(\cdot)$ defined by (18). Chan [33] and Hansen [2] considered testing the null of no-threshold but they used the least square estimation. We remark that Caner [34] considered LAD estimation of the threshold parameter but did not study the testing problem considered in this subsection.

Let $\hat\beta(\gamma)$ be the LAD estimator of β for each given $\gamma \in [\gamma_L, \gamma_U]$:

$$\hat\beta(\gamma) \in \operatorname*{arg\,min}_{b\in\mathbb{R}^{2p}} \sum_{i=1}^n |y_i - z_i(\gamma)'b|. \tag{18}$$

First, we verify that $\hat\beta(\cdot)$ can be chosen such that it has bounded sample paths, i.e., $\hat\beta(\cdot) \in (\ell^\infty[\gamma_L, \gamma_U])^{2p}$. Actually, $\hat\beta(\cdot)$ can be chosen such that it has càdlàg sample paths.

Lemma 2. *There exists an optimal solution $\hat\beta(\gamma)$ of (18) for each $\gamma \in [\gamma_L, \gamma_U]$ such that $\hat\beta(\cdot) \in (D[\gamma_L, \gamma_U])^{2p}$.*

Proof. It suffices to show the lemma when q_i $(1 \le i \le n)$ have no tie. Let $q_{(1)} < q_{(2)} < \cdots < q_{(n)}$ be order statistics of q_i $(1 \le i \le n)$ and i_j $(1 \le j \le n)$ be indices such that $q_{i_j} = q_{(j)}$ for $1 \le j \le n$. Define $\hat b(r)$ $(0 \le r \le n)$ by

$$\hat b(r) \in \operatorname*{arg\,min}_{b\in\mathbb{R}^{2p}} \left\{ \sum_{j=1}^r |y_{i_j} - x_{i_j}'(b^{(1)} + b^{(2)})| + \sum_{j=r+1}^n |y_{i_j} - x_{i_j}'b^{(1)}| \right\}.$$

Then,

$$\hat\beta(\gamma) = \begin{cases} \hat b(0), & \text{if } \gamma < q_{(1)}, \\ \hat b(r), & \text{if } q_{(r)} \le \gamma < q_{(r+1)} \ (1 \le r \le n-1), \\ \hat b(n), & \text{if } \gamma \ge q_{(n)} \end{cases}$$

is an optimal solution of (18) for each γ and $\gamma \mapsto \hat\beta(\gamma)$ is càdlàg. □

Consider testing the null hypothesis $H_0 : \beta^{(2)} = 0$ against the alternative $H_1 : \beta^{(2)} \neq 0$. We consider the Wald-type test statistic

$$T_n(\gamma) = n\hat\beta^{(2)}(\gamma)'[\mathrm{Avar}\{\hat\beta^{(2)}(\gamma)\}]^{-1}\hat\beta^{(2)}(\gamma)$$

for each γ where $\mathrm{Avar}\{\hat\beta^{(2)}(\gamma)\}$ is the asymptotic covariance matrix of $n^{1/2}\hat\beta^{(2)}(\gamma)$ under H_0. Then we reject the null hypothesis if $\sup_{\gamma\in[\gamma_L,\gamma_U]} T_n(\gamma) > c_0$ for some constant $c_0 > 0$, as suggested by Davies [4,5].

In order to derive the asymptotic null distribution of the test statistic $\sup_{\gamma\in[\gamma_L,\gamma_U]} T_n(\gamma)$, we derive the asymptotic distribution of $n^{1/2}(\hat{\beta}(\cdot)-\beta)$ under H_0. As in Section 3.1, we consider the asymptotic behavior of the local objective function

$$Z_n(u,\gamma)=\frac{1}{2}\sum_{i=1}^{n}\{|\epsilon_i-n^{-1/2}z_i(\gamma)'u|-|\epsilon_i|\},\quad u\in\mathbb{R}^{2p},\gamma\in[\gamma_L,\gamma_U].$$

Under H_0, the normalized estimator $n^{1/2}(\hat{\beta}(\gamma)-\beta)$ minimizes $Z_n(u,\gamma)$ with respect to u for each γ. Moreover, $u\mapsto Z_n(u,\tau)$ is convex for each τ. Since $Z_n(u,\gamma)$ is not continuous in γ, Theorem 1 does not apply to this case. However, we can use Theorem 2.

We state some regularity conditions. Let $M(\gamma)=\mathrm{E}[I(q_i\le\gamma)x_ix_i']$ and

$$K(\gamma_1,\gamma_2)=\mathrm{E}[z_i(\gamma_1)z_i(\gamma_2)']=\begin{pmatrix}\mathrm{E}[x_ix_i'] & M(\gamma_2)\\ M(\gamma_1) & M(\gamma_1\wedge\gamma_2)\end{pmatrix}.$$

Assumption 2. (i) (x_i,q_i,ϵ_i) are independent and identically distributed. Furthermore, (x_i,q_i) and ϵ_i are independent for each i.
(ii) The common distribution function F of ϵ_i satisfies $F(0)=1/2$. Furthermore, F has positive and continuous Lebesgue density f in a neighborhood of 0.
(iii) $\mathrm{E}[\|x_i\|^4]<\infty$.
(iv) q_i has continuous distribution.
(v) $K(\gamma,\gamma)$ is positive definite for each $\gamma\in[\gamma_L,\gamma_U]$.

One could weaken the condition (i), for instance, allow time series data or allow that (x_i,q_i) and ϵ_i are dependent. However, we do not pursue this since the primal object of this section is the application of the convexity arguments. The condition (ii) is standard in LAD estimation. Note that we do not assume ϵ_i has any moment. The conditions (i) and (iii) imply $\max_{1\le i\le n}\|x_i\|=o_p(n^{1/2})$ and hence $\max_{1\le i\le n}\sup_{\gamma\in[\gamma_L,\gamma_U]}\|z_i(\gamma)\|=o_p(n^{1/2})$. We note that under the conditions (iii) and (iv), the map $\gamma\mapsto M(\gamma)$ is continuous.

Theorem 4. *Under the condition* (i)–(iv) *of Assumption 2,*

$$Z_n(u,\gamma)=-n^{-1/2}\sum_{i=1}^{n}z_i(\gamma)'u\left\{\frac{1}{2}-I(\epsilon_i\le 0)\right\}+f(0)\frac{u'K(\gamma,\gamma)u}{2}+\Delta_n(u,\gamma),$$

where $\sup_{\gamma\in[\gamma_L,\gamma_U]}|\Delta_n(u,\gamma)|=o_p(1)$ *for each u and*

$$n^{-1/2}\sum_{i=1}^{n}z_i(\cdot)\left\{\frac{1}{2}-I(\epsilon_i\le 0)\right\}\rightsquigarrow G(\cdot)\quad \text{in } (\ell^\infty[\gamma_L,\gamma_U])^{2p}. \tag{19}$$

Here $G(\cdot)$ is a zero-mean, continuous Gaussian process with covariance kernel $\mathrm{E}[G(\gamma_1)G(\gamma_2)']=K(\gamma_1,\gamma_2)/4$.

Proof. Most of the argument will follow the lines in the proof of Theorem 3. Using Knight's [9] identity (15) again, we obtain

$$\begin{aligned}Z_n(u,\gamma)&=-n^{-1/2}\sum_{i=1}^{n}z_i(\gamma)'u\left\{\frac{1}{2}-I(\epsilon_i\le 0)\right\}+n^{-1}\sum_{i=1}^{n}z_i(\gamma)'u\int_0^1 n^{1/2}\{I(\epsilon_i\le n^{-1/2}z_i(\gamma)'us)-I(\epsilon_i\le 0)\}\mathrm{d}s\\ &=: Z_n^{(1)}(u,\gamma)+Z_n^{(2)}(u,\gamma)\\ &= Z_n^{(1)}(u,\gamma)+\mathrm{E}[Z_n^{(2)}(u,\gamma)|x_i,q_i,1\le i\le n]+\{Z_n^{(2)}(u,\gamma)-\mathrm{E}[Z_n^{(2)}(u,\gamma)|x_i,q_i,1\le i\le n]\}\\ &=: Z_n^{(1)}(u,\gamma)+Z_n^{(21)}(u,\gamma)+Z_n^{(22)}(u,\gamma).\end{aligned}$$

First, the weak convergence (19) is proved in Appendix of Hansen [2] under the additional condition that q_i has bounded density function (see also Hansen [3], Lemma A4). Under the conditions (i)–(iv) of Assumption 2, it is possible to show (19) without the density of q_i. Since the proof is a simple modification of the proof of Theorem 2 in Koul [29], we omit it.

Next, consider the asymptotic behavior of $Z_n^{(21)}(u,\gamma)$. Since $\max_{1\le i\le n}\sup_{\gamma\in[\gamma_L,\gamma_U]}\|z_i(\gamma)\|=o_p(n^{1/2})$,

$$\sup_{\gamma\in[\gamma_L,\gamma_U]}\left|Z_n^{(21)}(u,\gamma)-\frac{f(0)}{2}u'\left\{n^{-1}\sum_{i=1}^{n}z_i(\gamma)z_i(\gamma)'\right\}u\right|\xrightarrow{p}0.$$

Then, Lemma 1 of Hansen [2] implies that $Z_n^{(21)}(u,\gamma)$ converges in probability to $\{f(0)u'K(\gamma,\gamma)u\}/2$ uniformly in $\gamma\in[\gamma_L,\gamma_U]$ for each u.

The remaining task is to show $\sup_{\gamma\in[\gamma_L,\gamma_U]}|Z_n^{(22)}(u,\gamma)| = o_p(1)$ for each $u \in \mathbb{R}^{2p}$. To show this, it suffices to show that

$$\sup_{\gamma\in[\gamma_L,\gamma_U]}\left| n^{-1/2}\sum_{i=1}^n x_i' v I(q_i\le\gamma)\int_0^1 [\{I(\epsilon_i \le n^{-1/2}x_i'vs) - I(\epsilon_i\le 0)\} - \{F(n^{-1/2}x_i'vs) - F(0)\}]\mathrm{d}s\right| \xrightarrow{p} 0$$

and

$$n^{-1/2}\sum_{i=1}^n x_i' v\int_0^1 [\{I(\epsilon_i \le n^{-1/2}x_i'vs) - I(\epsilon_i\le 0)\} - \{F(n^{-1/2}x_i'vs) - F(0)\}]\mathrm{d}s \xrightarrow{p} 0$$

for each $v \in \mathbb{R}^p$. Suppose for a moment that v is fixed. Define

$$e_{in} = \int_0^1 [\{I(\epsilon_i \le n^{-1/2}x_i'vs) - I(\epsilon_i\le 0)\} - \{F(n^{-1/2}x_i'vs) - F(0)\}]\mathrm{d}s.$$

We have to show $W_n(\gamma) = n^{-1/2}\sum_{i=1}^n x_i' v e_{in} I(q_i\le\gamma) \xrightarrow{p} 0$ uniformly in $\gamma\in[\gamma_L,\gamma_U]$. Since

$$\mathrm{E}[e_{in}^2|x_i, q_i] \le \int_0^1 |F(n^{-1/2}x_i'vs) - F(0)|\mathrm{d}s,$$

we have

$$\mathrm{E}[W_n(\gamma)^2] \le \mathrm{E}\left[(x_i'v)^2\int_0^1 |F(n^{-1/2}x_i'vs) - F(0)|\mathrm{d}s\right].$$

Then, the continuity of F at 0 and the dominated convergence theorem implies that $W_n(\gamma)$ converges in probability to 0 for each γ. Furthermore, a simple calculation shows that for any $\gamma_1\le\gamma\le\gamma_2$,

$$\mathrm{E}[\{W_n(\gamma) - W_n(\gamma_1)\}^2\{W_n(\gamma_2) - W_n(\gamma)\}^2] \le 4[v'\{M(\gamma_2) - M(\gamma_1)\}v]^2,$$

where we have used $\mathrm{E}[e_{in}^2|x_i,q_i]\le 2$. Therefore, by Theorem 13.5 of Billingsley [14], we conclude that $W_n(\gamma)$ converges in probability to 0 uniformly in $\gamma\in[\gamma_L,\gamma_L]$. It is also seen that $n^{-1/2}\sum_{i=1}^n x_i' v e_{in} \xrightarrow{p} 0$. Therefore we complete the proof. □

Combining Theorems 2 and 4, we obtain the next corollary.

Corollary 2. *Suppose Assumption 2 is satisfied. Then under H_0,*

$$n^{1/2}(\hat\beta(\cdot) - \beta) \rightsquigarrow \{f(0)\}^{-1}\{K(\cdot,\cdot)\}^{-1}G(\cdot)\quad \text{in } (\ell^\infty[\gamma_L,\gamma_U])^{2p},$$

where $G(\cdot)$ is a zero-mean, continuous Gaussian process with covariance kernel $K(\gamma_1,\gamma_2)/4$.

Again, let us see the relationship between the conditions of this example and Theorem 2. The conditions (i)–(iv) of Assumption 2 guarantee the weak convergence (19), which satisfies the condition (11) of Theorem 2. The condition (ii) and (v) are used to satisfy the condition on $Q(\tau)$ in Theorem 2, where in this case $Q(\tau)$ corresponds to $f(0)K(\gamma,\gamma)$. To guarantee the condition (10), the conditions (i)–(iv) of Assumption 2 are used.

By Corollary 2, $\mathrm{Avar}\{\hat\beta^{(2)}(\gamma)\}$ is $\frac{1}{4\{f(0)\}^2}R\{K(\gamma,\gamma)\}^{-1}R'$ where $R = [0\ I_p]$. The asymptotic null distribution of the test statistic $\sup_{\gamma\in[\gamma_L,\gamma_U]} T_n(\gamma)$ is

$$\sup_{\gamma\in[\gamma_L,\gamma_U]} S(\gamma)'[R\{K(\gamma,\gamma)\}^{-1}R']^{-1}S(\gamma),$$

where $S(\cdot)$ is a zero-mean, continuous Gaussian process with covariance kernel $\mathrm{E}[S(\gamma_1)S(\gamma_2)'] = R\{K(\gamma_1,\gamma_1)\}^{-1}K(\gamma_1,\gamma_2)$ $\{K(\gamma_2,\gamma_2)\}^{-1}R'$.

In usual, $f(0)$ and $K(\gamma_1,\gamma_2)$ are unknown and so they are replaced by their consistent estimators. It is natural to adopt $\hat K(\gamma_1,\gamma_2) = n^{-1}\sum_{i=1}^n z_i(\gamma_1)z_i(\gamma_2)'$ as an estimator of $K(\gamma_1,\gamma_2)$. For estimation of $f(0)$, consult Section 3.4 of Koenker [21].

Acknowledgments

The author would like to thank Professor Tatsuya Kubokawa, Professor Naoto Kunitomo and Professor Hidehiko Ichimura for their encouragement and helpful suggestions. He also would like to thank the Editor, the Associate Editor and the reviewers for their constructive comments. One of the reviewers suggested some references on the asymptotics of regression quantile processes. This research was supported by Grant-in-Aid for JSPS Fellows, the Ministry of Education, Culture, Sports, Science and Technology.

Appendix

A.1. Technical appendix for Section 2

In this subsection, we derive a technical lemma which is used in the proof of Theorem 1. In Lemma 3 below, all functions are assumed to be non-stochastic. Lemma 3 is a slight generalization of Theorem 10.8 in Rockafellar [15], which is roughly stated as follows: If a sequence of convex functions $\{g_n(x)\}$ converges to some function $g_\infty(x)$ for each x in a dense subset of $\mathbb{R}^d$, then g_n converges to g_∞ uniformly on each compact set in $\mathbb{R}^d$.

Lemma 3. *Suppose $f_n(x, \tau)$ $(n \geq 1)$ and $f_\infty(x, \tau)$ are convex in x for each τ and bounded in τ for each x. If*

$$\sup_{\tau \in T} |f_n(x, \tau) - f_\infty(x, \tau)| \to 0$$

as $n \to \infty$ for each $x \in D$ where D is a dense subset of $\mathbb{R}^d$, then,

$$\sup_{\tau \in T} \sup_{x \in K} |f_n(x, \tau) - f(x, \tau)| \to 0,$$

as $n \to \infty$ for every compact set $K \subset \mathbb{R}^d$.

Proof. Take an arbitrary compact set K in $\mathbb{R}^d$. Since $\{f_n(x, \tau) : \tau \in T, n \geq 1\}$ is bounded for each fixed $x \in D$, applying Theorem 10.6 of Rockafellar [15] to the collection of convex functions $\{f_n(\cdot, \tau) : \tau \in T, n \geq 1\}$ implies that there exists a constant $\alpha_1 > 0$ such that

$$|f_n(y, \tau) - f_n(x, \tau)| \leq \alpha_1 \|y - x\|, \quad \forall x, y \in K, \forall \tau \in T, \forall n \geq 1.$$

Similarly, there exists a constant $\alpha_2 > 0$ such that

$$|f_\infty(y, \tau) - f_\infty(x, \tau)| \leq \alpha_2 \|y - x\|, \quad \forall x, y \in K, \forall \tau \in T.$$

The rest of the proof is a simple modification of the proof of Theorem 10.8 of Rockafellar [15]. However, since this lemma is a key result for the proof of Theorem 1, we provide the whole proof. Define $\alpha_0 = \max\{\alpha_1, \alpha_2\}$. Fix any $\epsilon > 0$. Since D is dense in $\mathbb{R}^d$, there exists a finite set $D_0 \subset D \cap K$ such that each point of K lies within the distance $\epsilon/(3\alpha_0)$ of at least one point of D_0. Since D_0 is finite and $f_n(x, \tau)$ converge to $f_\infty(x, \tau)$ uniformly in τ for each $x \in D_0$, there exists a positive integer n_0 such that

$$|f_n(y, \tau) - f_\infty(y, \tau)| \leq \epsilon/3, \quad \forall \tau \in T, \forall n \geq n_0, \forall y \in D_0.$$

Given any $x \in K$, let y be a point of D_0 such that $\|x - y\| \leq \epsilon/(3\alpha_0)$. Then, for every $n \geq n_0$ and every $\tau \in T$,

$$\begin{aligned} |f_n(x, \tau) - f_\infty(x, \tau)| &\leq |f_n(x, \tau) - f_n(y, \tau)| + |f_n(y, \tau) - f_\infty(y, \tau)| + |f_\infty(y, \tau) - f_\infty(x, \tau)| \\ &\leq \alpha_1 \|x - y\| + (\epsilon/3) + \alpha_2 \|y - x\| \leq \epsilon. \end{aligned}$$

This proves that for every $n \geq n_0$,

$$\sup_{\tau \in T} \sup_{x \in K} |f_n(x, \tau) - f_\infty(x, \tau)| \leq \epsilon.$$

Therefore we complete the proof. □

A.2. Proof of (16)

Because of the compactness of the set $\{t : \|t\| \leq M\}$, it suffices to show (a) for every $\epsilon > 0$, there exists a constant $\delta > 0$ such that for each fixed $t_0 \in \{t : \|t\| \leq M\}$,

$$\lim_{n \to \infty} \mathrm{P}\left(\sup_{\tau \in [\alpha, 1-\alpha]} \sup_{t : \|t - t_0\| \leq \delta} |W_n(t, \tau) - W_n(t_0, \tau)| > \epsilon \right) = 0;$$

(b) for every $\epsilon > 0$ and each fixed $t_0 \in \{t : \|t\| \leq M\}$,

$$\lim_{n \to \infty} \mathrm{P}\left(\sup_{\tau \in [\alpha, 1-\alpha]} |W_n(t_0, \tau) - W_n(0, \tau)| > \epsilon \right) = 0.$$

We first show (a). Let

$$J_n(t, \tau) = n^{-1/2} \sum_{i=1}^{n} x_{in}' u I(\epsilon_i \leq F^{-1}(\tau) + n^{-1/2} \sigma_{in}^{-1} x_{in}' t).$$

Then, observe that

$$W_n(t,\tau) - W_n(t_0,\tau) = \{J_n(t,\tau) - J_n(t_0,\tau)\} - \{\mathrm{E}[J_n(t,\tau)] - \mathrm{E}[J_n(t_0,\tau)]\}. \tag{20}$$

The absolute value of the second term of rhs of (20) is bounded by a constant multiple of δ uniformly in t such that $\|t-t_0\| \le \delta$ and in $\tau \in [\alpha, 1-\alpha]$ for large n, where we have used $\max_{1\le i\le n}\|x_{in}\| = o(n^{1/2})$, the Lipschitz continuity of F on each bounded interval and $n^{-1}\sum_{i=1}^n \|x_{in}\|^2 = O(1)$.

It remains to bound the first term of the rhs of (20). Assume first $x_{in}'u \ge 0$ for $1 \le i \le n$. Then,

$$\begin{aligned} J_n(t,\tau) - J_n(t_0,\tau) &\le n^{-1/2}\sum_{i=1}^n x_{in}'uI(n^{-1/2}\sigma_{in}^{-1}x_{in}'t_0 < \epsilon_i - F^{-1}(\tau) \le n^{-1/2}\sigma_{in}^{-1}(x_{in}'t_0 + \|x_{in}\|\delta)) \\ &=: \bar{J}_n(\tau,\delta) \end{aligned}$$

for any t such that $\|t - t_0\| \le \delta$. Decompose $\bar{J}_n(\tau,\delta)$ as

$$\bar{J}_n(\tau,\delta) = \mathrm{E}[\bar{J}_n(\tau,\delta)] + \{\bar{J}_n(\tau,\delta) - \mathrm{E}[\bar{J}_n(\tau,\delta)]\}. \tag{21}$$

The first term of the rhs of (21) is bounded from above by a constant multiple of δ uniformly in $\tau \in [\alpha, 1-\alpha]$ for large n. Moreover, we show that $\sup_{\tau\in[\alpha,1-\alpha]}|\bar{J}_n(\tau,\delta) - \mathrm{E}[\bar{J}_n(\tau,\delta)]| \xrightarrow{p} 0$ for any $\delta > 0$. First, the pointwise convergence follows from the fact that the variance of $\bar{J}(\tau,\delta)$ converges to zero for each fixed τ. To show the uniform convergence, it suffices to show the asymptotic equicontinuity of $\{\tilde{J}_n(\tau), \tau \in [\alpha, 1-\alpha]\}$ where $\tilde{J}_n(\tau)$ is given by

$$\tilde{J}_n(\tau) = n^{-1/2}\sum_{i=1}^n x_{in}'u\{I(\epsilon_i \le F^{-1}(\tau) + n^{-1/2}\|x_{in}\|h_{in}) - F(F^{-1}(\tau) + n^{-1/2}\|x_{in}\|h_{in})\} \tag{22}$$

and $\{h_{in}\}$ is any bounded triangular sequence of constants. We establish this fact in Lemma 4 below. Hence, $J_n(t,\tau) - J_n(t_0,\tau) \le C\delta + o_p(1)$ for some constant $C > 0$ uniformly in t such that $\|t - t_0\| \le \delta$ and in $\tau \in [\alpha, 1-\alpha]$. The lower bound can be handled similarly. Therefore, we have proved (a) when $x_{in}'u \ge 0$, $1 \le i \le n$. For $\{x_{in}'u\}$ with variable sign, use the decomposition into positive and negative parts. Then, (a) follows because of linearity of $J_n(t,\tau)$ in $\{x_{in}'u\}$.

Next, we show (b). The variance of $W_n(t_0,\tau) - W_n(0,\tau)$ converges to 0. Furthermore, observe that

$$W_n(t_0,\tau) - W_n(0,\tau) = \{J_n(t_0,\tau) - \mathrm{E}[J_n(t_0,\tau)]\} + \{J_n(0,\tau) - \mathrm{E}[J_n(0,\tau)]\}. \tag{23}$$

Because of Lemma 4 below, each term of the rhs of (23) is asymptotically equicontinuous in probability as a stochastic process with index $\tau \in [\alpha, 1-\alpha]$. Therefore, (b) is established. □

Lemma 4. *A sequence of stochastic processes $\{\tilde{J}_n(\tau), \tau \in [\alpha, 1-\alpha]\}$ is asymptotically equicontinuous in probability, where $\tilde{J}_n(\tau)$ is given by (22) and $\{h_{in}\}$ is any bounded triangular sequence of constants.*

Remark 5. We do not assume $x_{in}'u \ge 0$, $1 \le i \le n$ in this lemma.

Proof of Lemma 4. We show that $\tilde{J}_n(\cdot)$ converges weakly to a $(u'Qu)^{1/2}$ multiple of a Brownian bridge on $[0, 1]$ in $D[\alpha, 1-\alpha]$. Since the limit process is continuous, the assertion of the lemma will follow from this result.

The finite dimensional convergence is verified by the Lindeberg–Feller central limit theorem. Furthermore, a direct calculation (see Billingsley [14], p. 150) shows that for any $\tau_1 \le \tau \le \tau_2$,

$$\begin{aligned} &\mathrm{E}[\{\tilde{J}_n(\tau) - \tilde{J}_n(\tau_1)\}^2\{\tilde{J}_n(\tau_2) - \tilde{J}_n(\tau)\}^2] \\ &\quad\le 3\left[n^{-1}\sum_{i=1}^n (x_{in}'u)^2\{F(F^{-1}(\tau_2) + n^{-1/2}\|x_{in}\|h_{in}) - F(F^{-1}(\tau_1) + n^{-1/2}\|x_{in}\|h_{in})\}\right]^2. \end{aligned}$$

Using $\max_{1\le i\le n}\|x_{in}\| = o(n^{1/2})$, the Lipschitz continuity of F on each bounded interval and $n^{-1}\sum_{i=1}^n\|x_{in}\|^2 = O(1)$, we obtain the desired result by Theorem 13.5 of Billingsley [14]. □

References

[1] R. Koenker, G. Bassett, Regression quantiles, Econometrica 46 (1978) 33–50.
[2] B. Hansen, Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica 64 (1996) 413–430.
[3] B. Hansen, Sample splitting and threshold estimation, Econometrica 68 (2000) 575–603.
[4] R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative, Biometrika 64 (1977) 247–254.
[5] R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative, Biometrika 74 (1987) 34–43.
[6] D. Pollard, Asymptotics for least absolute deviation regression estimators, Econometric Theory 7 (1991) 186–199.
[7] N.L. Hjort, D. Pollard, Asymptotics for minimizers of convex processes, unpublished manuscript, 1993.
[8] C.J. Geyer, On the asymptotics of convex stochastic optimization, unpublished manuscript, 1996.
[9] K. Knight, Limiting distributions for L_1 regression estimators under general conditions, Ann. Statist. 26 (1998) 755–770.
[10] K. Knight, W. Fu, Asymptotics for Lasso-type estimators, Ann. Statist. 28 (2000) 1356–1378.

[11] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.
[12] D. Pollard, Empirical Processes: Theory and Applications, in: NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2, Institute of Mathematical Statistics, Hayword, California, 1990.
[13] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes: With Applications to Statistics, Springer-Verlag, New York, 1996.
[14] P. Billingsley, Convergence of Probability Measure, 2nd ed., Wiley, New York, 1999.
[15] R.T. Rockafellar, Convex Analysis, Princeton Univ. Press, Princeton, 1970.
[16] W. Niemiro, Asymptotics for *M*-estimators defined by convex minimization, Ann. Statist. 20 (1992) 1514–1533.
[17] R.M. Dudley, An extended Wichura theorem, definitions of Donsker classes, and weighted empirical distributions, in: Probability in Banach Spaces V, in: Lecture Notes in Mathematics, vol. 1153, Springer, New York, 1985, pp. 141–178.
[18] J. Kim, D. Pollard, Cube root asymptotics, Ann. Statist. 18 (1990) 191–219.
[19] R.A. Davis, K. Knight, J. Liu, *M*-estimation for autoregressions with infinite variance, Stochastic Process. Appl. 40 (1992) 145–180.
[20] D. Pollard, Convergence of Stochastic Processes, Springer-Verlag, New York, 1984.
[21] R. Koenker, Quantile Regression, Oxford Univ. Press, Oxford, 2005.
[22] C. Gutenbrunner, J. Jurečková, Regression rank scores and regression quantiles, Ann. Statist. 20 (1992) 305–330.
[23] R. Koenker, S. Portnoy, *L*-estimation for linear models, J. Amer. Statist. Assoc. 82 (1987) 851–857.
[24] S. Portnoy, Asymptotic behavior of regression quantiles in nonstationary, dependent cases, J. Multivariate Anal. 38 (1991) 100–113.
[25] C. Gutenbrunner, J. Jurečková, R. Koenker, S. Portnoy, Tests of linear hypotheses based on regression rank scores, J. Nonparametr. Stat. 2 (1993) 307–331.
[26] H.L. Koul, M.E. Saleh, Autoregression quantiles and related rank score processes, Ann. Statist. 23 (1995) 670–689.
[27] V.I. Koltchinskii, Differentiability of inverse operators and limit theorems for inverse functions, J. Theoret. Probab. 11 (1998) 645–699.
[28] J. Angrist, V. Chernozhukov, I. Fernandéz-Val, Quantile regression under mis-specification, with an application to the U.S. wage structure, Econometrica 74 (2006) 539–563.
[29] H.L. Koul, Some convergence theorems for ranks and weighted empirical cumulatives, Ann. Math. Statist. 41 (1970) 1768–1773.
[30] H.L. Koul, Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression, Ann. Math. Statist. 40 (1969) 1950–1979.
[31] H.L. Koul, A weak convergence result useful in robust autoregression, J. Statist. Plann. Inference 29 (1991) 291–308.
[32] C. Berge, Topological Spaces (E.M. Patterson, Trans.), MacMillan, New York, 1963.
[33] K.S. Chan, Testing for threshold autoregression, Ann. Statist. 18 (1990) 1886–1894.
[34] M. Caner, A note on LAD estimation of a threshold model, Econometric Theory 18 (2002) 800–814.