

A note on moment convergence of bootstrap M-estimators

Kengo Kato

Received: October 25, 2009; Accepted: October 27, 2010

Summary: This paper studies the consistency of bootstrap moment estimators for a general M-estimator. We establish a theorem on the uniform integrability of the bootstrap M-estimator, thereby giving sufficient conditions for the consistency of the bootstrap moment estimators. As an application of our theorem, we provide sufficient conditions for the consistency of the bootstrap variance estimator for the quantile regression estimator, which has been considered as an important unsolved problem in the literature. We also discuss a justification of a bootstrap information criterion.

1 Introduction

The bootstrap introduced by Efron (1979) is a convenient general method for making statistical inference. It is well known that under suitable regularity conditions, the bootstrap is consistent in estimating the distribution of a general M-estimator (see Arcones and Giné, 1992; Wellner and Zhan, 1996). The distributional consistency of the bootstrap, however, does not imply the consistency of the bootstrap moment estimators, where “the bootstrap moment estimators” mean the corresponding conditional moments of the bootstrap estimator given the sample. In this paper, we study the consistency of the bootstrap moment estimators for a general M-estimator. Our framework allows for non-smooth objective functions such as the absolute value function or more generally the “check” function used in quantile regression (Koenker and Bassett, 1978). We establish a theorem on the uniform integrability of the bootstrap M-estimator, thereby giving sufficient conditions for the consistency of the bootstrap moment estimators.

There is a vast literature on the consistency of bootstrap moment estimators. Shao and Tu (1995) reviewed some earlier results on this topic. More recently, Gonçalves and White (2005) proved the consistency of the bootstrap variance estimator for the least squares estimator in the time series context (more precisely, we should say “the bootstrap covariance matrix estimator” rather than “the bootstrap variance estimator”; however, we would use the latter term for convenience). For general M-estimation that allows for non-smooth objective functions, (primitive) conditions for the consistency of the bootstrap moment estimators do not appear to be available.

AMS 2010 subject classification: 62F40, 62E20

Key words and phrases: Bootstrap, M-estimator, moment convergence, quantile regression

Of particular interest is the consistency of the bootstrap variance estimator for the quantile regression estimator. Under suitable regularity conditions, the quantile regression estimator meets the $\sqrt{n}$ -asymptotic normality (see Koenker, 2005, Ch.4). The distinctive point of the quantile regression estimator is that the asymptotic covariance matrix depends on the unknown conditional density. Therefore, the bootstrap variance estimation is particularly convenient to the quantile regression case as it can avoid the non-parametric density estimation. Hahn (1995) proved the distributional consistency of the bootstrap quantile regression estimator but did not study the consistency of the bootstrap variance estimator. A motivation to study the consistency of the bootstrap variance estimator to the quantile regression case also comes from the observation of Buchinsky (1995) who compared several inference methods for quantile regression models based on the Monte Carlo study. Buchinsky (1995) reported that inference based on the bootstrap variance estimator performs quite well in his numerical examples. It is thus of interest to study some theoretical justification of the bootstrap variance estimation to the quantile regression case. Gonçalves and White (2005, p.972) remarked that “establishing theoretical results that justify the application of the bootstrap to variance estimation for the quantile regression estimator is an important area of future research.” This paper gives an answer to a more general problem than what Gonçalves and White (2005) posed, since it considers a general moment rather than the second moment and general M-estimation that includes quantile regression as a special case. We give in Section 3 (relatively) primitive sufficient conditions for the consistency of the bootstrap variance estimator for the quantile regression estimator.

Another important application is a justification of the extended information criterion (EIC) proposed by Ishiguro et al. (1997), in which the bias term of the information criterion is estimated by the bootstrap. We also give a brief discussion on conditions for the consistency of the bootstrap bias estimator in Section 3.

A closely related topic is the moment convergence of an M-estimator. Nishiyama (2010) established sufficient conditions for the moment convergence of a general M-estimator by using a connection to the convergence rate theorem of van der Vaart and Wellner (1996, Theorem 3.2.5). The approach taken by this paper is based on the technique used in the same convergence rate theorem. Thus, the present result may be viewed as a bootstrap version of Nishiyama’s (2010) result, although Nishiyama allows for a rate different from $\sqrt{n}$ while this paper focuses on the case where the estimator is $\sqrt{n}$ -consistent. However, it should be noted that we are dealing with the convergence of conditional moments of the bootstrap M-estimator, which we believe is sufficiently different from Nishiyama’s topic to make this paper non-trivial and to require a separate treatment. Yoshida (2010) also tackled the moment convergence problem of an M-estimator in a different approach.

The rest of the paper consists of two sections. In Section 2, we present the main result including the proofs. In Section 3, we consider an application of our theorem to the quantile regression case and EIC.

2 Main result

Let $X, X_1, \dots, X_n$ be an independent sample from a distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$. For each $\theta \in \Theta$, which is assumed to be a (Borel) measurable subset of $\mathbb{R}^d$, let $m_\theta : \mathcal{X} \rightarrow \mathbb{R}$ be a known function. We assume the joint measurability of the map $(x, \theta) \mapsto m_\theta$. We consider the M-estimator $\hat{\theta}_n := \arg \min_{\theta \in \Theta} \mathbb{M}_n(\theta)$ where $\mathbb{M}_n(\theta) := n^{-1} \sum_{i=1}^n m_\theta(X_i)$. We assume the existence of a measurable solution $\hat{\theta}_n$, which is satisfied, for instance, when Θ is compact and the map $\theta \mapsto m_\theta$ is continuous (see Jenrich, 1969, Lemma 2).

Suppose for a while that $M(\theta) := E[m_\theta(X)]$ is minimized at $\theta_0 \in \Theta$ and twice continuously differentiable at θ_0 with nonsingular second derivative matrix A , and that there exists a vector-valued measurable function $\dot{m}_{\theta_0} : \mathcal{X} \rightarrow \mathbb{R}^d$ such that $E[\|\dot{m}_{\theta_0}(X)\|^2] < \infty$ and $E[\{m_\theta(X) - m_{\theta_0}(X) - (\theta - \theta_0)' \dot{m}_{\theta_0}(X)\}^2] = o(\|\theta - \theta_0\|^2)$ as $\theta \rightarrow \theta_0$. Then, under some additional regularity conditions including the consistency of $\hat{\theta}_n$, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = -A^{-1}\{n^{-1/2} \sum_{i=1}^n \dot{m}_{\theta_0}(X_i)\} + o_p(1) \xrightarrow{d} N(0, A^{-1}BA^{-1})$ where $B := E[\dot{m}_{\theta_0}(X)\dot{m}_{\theta_0}(X)']$ (see, for instance van der Vaart and Wellner, 1996, Sec.3.2). Statistical inference on θ_0 based on the asymptotic distribution often requires a consistent estimator of the asymptotic covariance matrix $C := A^{-1}BA^{-1}$. However, in some cases such as the quantile regression case, the estimation of the asymptotic covariance matrix turns out to be a non-trivial issue. In such cases, the bootstrap gives a convenient alternative to the estimation of the asymptotic covariance matrix.

Let $X_1^*, \dots, X_n^*$ denote a bootstrap sample, i.e., an independent sample from the empirical distribution of $X_1, \dots, X_n$. We consider the bootstrap M-estimator $\hat{\theta}_n^* := \arg \min_{\theta \in \Theta} \mathbb{M}_n^*(\theta)$ where $\mathbb{M}_n^*(\theta) := n^{-1} \sum_{i=1}^n m_\theta(X_i^*)$. A bootstrap estimator of the asymptotic covariance matrix is given by $\hat{C}^* := E[n(\hat{\theta}_n^* - \hat{\theta}_n)(\hat{\theta}_n^* - \hat{\theta}_n)' \mid X_1, \dots, X_n]$. It is known that under suitable regularity conditions, the conditional distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ given the sample converges weakly to $N(0, C)$ in probability (see below the definition of the conditional weak convergence in probability). The distributional consistency, however, does not imply the consistency of $\hat{C}^*$. We study conditions under which conditional moments of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ given the sample converge in probability to those of the limiting distribution, given the distributional consistency of the bootstrap estimator.

We note that $\mathbb{M}_n^*(\theta)$ can be written as $\mathbb{M}_n^*(\theta) = n^{-1} \sum_{i=1}^n W_{ni} m_\theta(X_i)$ where W_{ni} is the number of times that X_i is redrawn from the original sample. The vector $(W_{n1}, \dots, W_{nn})'$ has multinomial distribution with parameters n and (probabilities) $n^{-1}, \dots, n^{-1}$. The randomness of bootstrap quantities (such as $\hat{\theta}_n^*$) comes from the randomness of both $X_1, \dots, X_n$ and $W_{n1}, \dots, W_{nn}$. As in van der Vaart and Wellner (1996, Sec. 3.6.1), we view $X_1, X_2, \dots$ as the coordinate projection on the first countably infinite coordinates of the product space $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty) \times (\mathcal{W}, \mathcal{C}, Q)$ and let the triangular sequence $\{W_{ni} : i = 1, \dots, n; n = 1, 2, \dots\}$ depends on the last factor only. We assume that $\hat{\theta}_n^*$ is chosen such that it is a measurable map from the product space to $\mathbb{R}^d$, which is often satisfied by suitable primitive regularity conditions. Let $E_W[\cdot]$ denote the expectation with respect to W_{ni} ($i = 1, \dots, n; n = 1, 2, \dots$) conditional on $X_1, X_2, \dots$.

In this paper, we presume the distributional consistency of $\hat{\theta}_n^*$ since there are several available results on that topic (see Arcones and Giné, 1992; Hahn, 1995; Wellner and

Zhan, 1996). For completeness, we clarify the concept of the conditional weak convergence in probability. Put $D_n := \{X_1, \dots, X_n\}$. Recall the bounded Lipschitz metric on the space of distributions (see van der Vaart and Wellner, 1996, p.73). Let T_n^* be some scalar statistic of $X_1, \dots, X_n$ and $W_{n1}, \dots, W_{nn}$. We say that the conditional distribution of T_n^* given D_n converges weakly to some fixed distribution (ν , say) in probability if the bounded Lipschitz metric between the two distributions converges in probability to zero, i.e.,

$$\sup_{g \in BL_1} \left| E_W[g(T_n^*)] - \int g d\nu \right| \xrightarrow{P} 0,$$

where BL_1 is the set of all functions on $\mathbb{R}$ with Lipschitz norm bounded by one. The next lemma gives a sufficient condition for the consistency of conditional moments of T_n^* given D_n when the conditional distribution of T_n^* given D_n converges weakly to ν in probability. The lemma might look obvious from the standard uniform integrability argument. However, we give a proof for it for clarity.

Lemma 2.1 *Let T_n^* be a scalar statistic of $X_1, \dots, X_n$ and $W_{n1}, \dots, W_{nn}$ such that the conditional distribution of T_n^* given D_n converges weakly to some fixed distribution (ν , say) in probability. If $E_W[|T_n^*|^q] = O_p(1)$ for some $q > 1$, then (a) ν has q -th absolute moment; (b) for any integer $1 \leq r < q$, we have $E_W[T_n^{*r}] \xrightarrow{P} \int t^r d\nu(t)$.*

Proof: Let T denote a random variable with distribution ν .

Part (a): Take a subsequence $\{n'\} \subset \{n\}$ such that conditionally on $X_1, X_2, \dots$, $T_{n'}^* \xrightarrow{d} \nu$ for almost every sequence $X_1, X_2, \dots$. By Fatou's lemma together with Skorohod's theorem, we have $E[|T|^q] \leq \liminf_{n'} E_W[|T_{n'}^*|^q]$, a.s. The fact that $E_W[|T_n^*|^q] = O_p(1)$ implies that the liminf is finite with positive probability. Since $E[|T|^q]$ is non-random, we obtain the first assertion.

Part (b): The proof is a modification of Lemma 4.5.2 in Chung (2001). Fix $\epsilon > 0$ and $\eta > 0$. Take a sufficiently large K such that $P(E_W[|T_n^*|^q] > K) \leq \eta$ for all $n \geq 1$ and $E[|T|^q] \leq K$. For a positive L such that $K/L^{(q-r)} \leq \epsilon$, define $g_L(t) := L^r$ if $t > L$; $:= t^r$ if $|t| \leq L$ and $:= (-L)^r$ if $t < -L$. Since g_L is bounded and Lipschitz continuous, we have $E_W[g_L(T_n^*)] \xrightarrow{P} E[g_L(T)]$. On the other hand,

$$\begin{aligned} |E_W[T_n^{*r}] - E_W[g_L(T_n^*)]| &\leq E_W[|T_n^*|^r I(|T_n^*| > L)] \\ &\leq \frac{E_W[|T_n^*|^q]}{L^{q-r}}, \end{aligned}$$

which is less than or equal to ϵ with probability greater than $1 - \eta$. We also have $|E[g_L(T)] - E[T^r]| \leq K/L^{q-r} \leq \epsilon$. Therefore,

$$\begin{aligned} P(|E_W[T_n^{*r}] - E[T^r]| > 3\epsilon) &\leq P(|E_W[T_n^{*r}] - E_W[g_L(T_n^*)]| > \epsilon) \\ &\quad + P(|E_W[g_L(T_n^*)] - E[g_L(T)]| > \epsilon) \\ &\leq P(|E_W[g_L(T_n^*)] - E[g_L(T)]| > \epsilon) + \eta. \end{aligned}$$

Taking the limit of both the sides, we obtain $\limsup_{n \rightarrow \infty} P(|E_W[T_n^{*r}] - E[T^r]| > 3\epsilon) \leq \eta$. Therefore, the proof is completed. $\square$

We now present the main result of the paper. In the statement of the theorem, we use the notation $J(1, \mathcal{F})$ to represent a uniform metric entropy integral (see van der Vaart and Wellner, 1996, p. 239).

Theorem 2.2 Suppose that: (i) There exist a $\theta_0 \in \Theta$ and a positive constant c such that $M(\theta) - M(\theta_0) \geq c\|\theta - \theta_0\|^2$ for all $\theta \in \Theta$. (ii) The class of functions $\mathcal{M}_\delta := \{m_\theta - m_{\theta_0} : \|\theta - \theta_0\| \leq \delta, \theta \in \Theta\}$ has envelope M_δ such that for some $p \geq 2$ and $\epsilon > 0$, $E[M_\delta^{p+\epsilon}] \leq \text{const.} \times \delta^{p+\epsilon}$ for all $\delta > 0$, and the class $\mathcal{M}_\delta$ with envelope M_δ satisfies the uniform metric entropy condition: $J(1, \mathcal{M}_\delta) \leq \text{const.}$ for all $\delta > 0$, where the constants are independent of δ . Then, we have $\sup_{n \geq 1} E[\|\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\|^{p+\epsilon'}] < \infty$ for any $\epsilon' \in (0, \epsilon)$.

Remark 2.3 A primitive sufficient condition for condition (ii) is: (ii)' There exists a measurable function $\dot{m} : \mathcal{X} \rightarrow \mathbb{R}$ such that $|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq \dot{m}(x)\|\theta_1 - \theta_2\|$ and $E[\dot{m}(X)^{p+\epsilon}] < \infty$ for some $p \geq 2$ and $\epsilon > 0$. Use Theorem 2.7.11 of van der Vaart and Wellner (1996) and the relation between covering numbers and bracketing numbers.

Before going to the proof, we explain an implication of the theorem. Suppose that $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, C)$ and the conditional distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ given D_n converges weakly to $N(0, C)$ in probability, where C is given in the previous discussion. Suppose also that p is a positive integer. Theorem 2.2 establishes sufficient conditions under which $E_W[g(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n))] \xrightarrow{P} E[g(Z)]$ with $Z \sim N(0, C)$ for a polynomial function g of degree less than or equal to p (Theorem 2.2 indeed ensures the L_1 -convergence). In particular, if the conditions of Theorem 2.2 holds with $p = 2$, the bootstrap variance estimator $\hat{C}^*$ will be consistent.

Proof of Theorem 2.2: We first show that $\sup_{n \geq 1} E[\|\sqrt{n}(\hat{\theta}_n^* - \theta_0)\|^{p+\epsilon'}] < \infty$. The proof consists of a combination of the proof of Theorem 3.2.5 in van der Vaart and Wellner (1996). Define $S_{j,n} := \{\theta \in \Theta : 2^{j-1} < \sqrt{n}\|\theta - \theta_0\| \leq 2^j\}$ for $j = 1, 2, \dots$. If $\sqrt{n}\|\hat{\theta}_n^* - \theta_0\| > 2^L$ for some integer L , then $\inf_{\theta \in S_{j,n}} \{\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0)\} \leq 0$ for some $j \geq L$. Therefore,

$$P\left(\sqrt{n}\|\hat{\theta}_n^* - \theta_0\| > 2^L\right) \leq \sum_{j \geq L} P\left(\inf_{\theta \in S_{j,n}} \{\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0)\} \leq 0\right).$$

Decompose $\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0)$ as

$$\begin{aligned} \mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0) &= [\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0) - \{\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0)\}] \\ &\quad + [\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0) - \{M(\theta) - M(\theta_0)\}] \\ &\quad + \{M(\theta) - M(\theta_0)\} \\ &=: I_{1n}(\theta) + I_{2n}(\theta) + I_3(\theta). \end{aligned}$$

177 By condition (i), for $\theta \in S_{j,n}$, $I_3(\theta) \geq c2^{2j-2}/n$. This implies that

$$\begin{aligned}
 178 \quad & \mathbb{P} \left(\inf_{\theta \in S_{j,n}} \{\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0)\} \leq 0 \right) \\
 179 \quad & \leq \mathbb{P} \left(\inf_{\theta \in S_{j,n}} \{I_{1n}(\theta) + I_{2n}(\theta)\} \leq -\frac{c2^{2j-2}}{n} \right) \\
 180 \quad & \leq \mathbb{P} \left(\sup_{\theta \in S_{j,n}} |I_{1n}(\theta) + I_{2n}(\theta)| \geq \frac{c2^{2j-2}}{n} \right) \\
 181 \quad & \leq \mathbb{P} \left(\sup_{\theta \in S_{j,n}} |I_{1n}(\theta)| \geq \frac{c2^{2j-2}}{2n} \right) + \mathbb{P} \left(\sup_{\theta \in S_{j,n}} |I_{2n}(\theta)| \geq \frac{c2^{2j-2}}{2n} \right). \quad (2.1)
 \end{aligned}$$

182 Recall the definition of $\mathcal{M}_\delta$ and M_δ . By the joint measurability of the map $(x, \theta) \mapsto m_\theta$,
 183 $\mathcal{M}_\delta$ is image admissible Suslin (see Dudley, 1999, Sec. 5.3). Put $\delta_{j,n} := 2^j/\sqrt{n}$. By
 184 Theorem 2.14.1 of van der Vaart and Wellner (1996), we have

$$\begin{aligned}
 185 \quad & \mathbb{E} \left[\sup_{\theta \in S_{j,n}} |I_{2n}(\theta)|^{p+\epsilon} \right] \leq \text{const.} \times n^{-(p+\epsilon)/2} J(1, \mathcal{M}_{\delta_{j,n}})^{p+\epsilon} \mathbb{E}[M_{\delta_{j,n}}(X)^{p+\epsilon}] \\
 186 \quad & \leq \text{const.} \times n^{-(p+\epsilon)} 2^{(p+\epsilon)j},
 \end{aligned}$$

187 where the constants are independent of (j, n) . Thus, by Markov's inequality, the second
 188 term on the right hand side of (2.1) is bounded by $\text{const.} \times 2^{-(p+\epsilon)j}$. To bound the first
 189 term, recall that $\mathbb{E}_W[\mathbb{M}_n^*(\theta)] = \mathbb{M}_n(\theta)$. By Theorem 2.14.1 of van der Vaart and Wellner
 190 (1996), we have

$$\begin{aligned}
 191 \quad & \mathbb{E}_W \left[\sup_{\theta \in S_{j,n}} |I_{1n}(\theta)|^{p+\epsilon} \right] \\
 192 \quad & \leq \text{const.} \times n^{-(p+\epsilon)/2} J(1, \mathcal{M}_{\delta_{j,n}})^{p+\epsilon} \{n^{-1} \sum_{i=1}^n M_{\delta_{j,n}}(X_i)^{p+\epsilon}\},
 \end{aligned}$$

193 where the constant is independent of (j, n) . The fact that $\mathcal{M}_{\delta_{j,n}}$ is image admissible
 194 Suslin ensures to apply Fubini's theorem to get

$$195 \quad \mathbb{E} \left[\sup_{\theta \in S_{j,n}} |I_{1n}(\theta)|^{p+\epsilon} \right] \leq \text{const.} \times n^{-(p+\epsilon)} 2^{(p+\epsilon)j}.$$

196 We have shown that there exists a constant D such that for any positive integer L ,

$$\begin{aligned}
 197 \quad & \mathbb{P} \left(\sqrt{n} \|\hat{\theta}_n^* - \theta_0\| > 2^L \right) \leq D \sum_{j \geq L} 2^{-(p+\epsilon)j} \\
 198 \quad & \leq 2D 2^{-(p+\epsilon)L}.
 \end{aligned}$$

199 Take $L = \lceil \log_2 t \rceil$ for $t \geq 2$ where $[a]$ denotes the integer part of a number a . Then, we
 200 can see that there exists another constant D' independent of t such that $\mathbb{P}(\sqrt{n} \|\hat{\theta}_n^* - \theta_0\| >$
 201 $t) \leq D' t^{-(p+\epsilon)}$. Because of the fact that for a non-negative random variable Z , $\mathbb{E}[Z^q] =$
 202 $q \int_0^\infty t^{q-1} \mathbb{P}(Z > t) dt$ for $q \geq 1$, we obtain: $\sup_{n \geq 1} \mathbb{E}[\|\sqrt{n}(\hat{\theta}_n^* - \theta_0)\|^{p+\epsilon}] < \infty$.

203 The analogous argument leads to that $\sup_{n \geq 1} E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{p+\epsilon'}] < \infty$. Combin-
 204 ing the previous result, we obtain: $\sup_{n \geq 1} E[\|\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\|^{p+\epsilon'}] < \infty$. $\square$

205 We give a brief discussion on the conditions of Theorem 2.2. Conditions (i) and (ii)
 206 (or (i) and (ii)') are adapted from conditions for the $\sqrt{n}$ -consistency of the M-estimator
 207 discussed in van der Vaart and Wellner (1996, p. 291). The different points are: (a) we put
 208 a global restriction on the behavior of $M(\theta)$ rather than a local one; (b) we put a higher
 209 moment restriction on M_δ (or $\dot{m}$). Part (b) is natural for the present purpose. Part (a) is
 210 essential for the present proof since we have to control the behavior of $P(\sqrt{n}\|\hat{\theta}_n^* - \theta_0\| >$
 211 $t)$ for large t and hence have to control the behavior of $\mathbb{M}_n^*(\theta) - \mathbb{M}_n^*(\theta_0)$ over all “shells”
 212 $S_{j,n}$ for large j . Not surprisingly, the conditions of Theorem 2.2 are analogous to those of
 213 Nishiyama's (2010) Theorem 1 that establishes the moment convergence (of any order)
 214 of an original M-estimator, as the proof strategies of both the theorems have the same
 215 root, Theorem 3.2.5 of van der Vaart and Wellner (1996).

216 It is worthwhile to remark that the proof uses the uniform integrability of the original
 217 M-estimator (i.e., $\sup_{n \geq 1} E[\|\sqrt{n}(\hat{\theta}_n - \theta_0)\|^{p+\epsilon'}] < \infty$). Thus, under the same set of
 218 conditions and the $\sqrt{n}$ -asymptotic normality of $\hat{\theta}_n$, the moment convergence of $\hat{\theta}_n$ also
 219 follows. In view of the previous discussion, there seems essentially no additional cost to
 220 ensure the uniform integrability of the bootstrap M-estimator, in comparison with that of
 221 the original one.

222 3 Applications

223 3.1 Quantile regression

224 In this section, we consider an application of our Theorem 2.2 to the quantile regression
 225 case. In particular, we are interested in the consistency of the bootstrap variance estimator
 226 for the quantile regression estimator. Let Y be a scalar dependent variable and let Z
 227 be a d -dimensional vector of explanatory variables. We consider the quantile regression
 228 model:

$$229 \quad Q_\tau(Y|Z) = Z'\beta_0,$$

230 where $\tau \in (0, 1)$ is a quantile of interest, which is assumed to be fixed, $Q_\tau(Y|Z)$ is the
 231 conditional τ -quantile of Y given Z and $\beta_0 \in \mathbb{R}^d$ is an unknown parameter vector. Sup-
 232 pose that we have n independent observations $(Y_1, Z_1), \dots, (Y_n, Z_n)$ of (Y, Z) . Koenker
 233 and Bassett (1978) proposed an estimator (“the quantile regression estimator”)

$$234 \quad \hat{\beta}_n := \arg \min_{\beta \in \mathcal{B}} \left[\sum_{i=1}^n \rho_\tau(Y_i - Z_i'\beta) \right]$$

235 where $\rho_\tau(u) := \{\tau - I(u \leq 0)\}u$. We restrict the parameter space to be a compact
 236 and convex subset $\mathcal{B}$ of $\mathbb{R}^d$ for some technical reason stated later. Let $f_{Y|Z}(y|z)$ denote
 237 the conditional density of Y given $Z = z$. Under suitable regularity conditions, it is
 238 shown that $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$, where $A := E[f_{Y|Z}(Z'\beta_0|Z)ZZ']$ and
 239 $B := \tau(1 - \tau)E[ZZ']$. We consider to estimate the asymptotic covariance matrix $C :=$

240 $A^{-1}BA^{-1}$ by the bootstrap. Let $(Y_1^*, Z_1^*), \dots, (Y_n^*, Z_n^*)$ denote a bootstrap sample from
 241 $(Y_1, Z_1), \dots, (Y_n, Z_n)$. Consider the bootstrap quantile regression estimator

$$242 \quad \hat{\beta}_n^* := \arg \min_{\beta \in \mathcal{B}} \left[\sum_{i=1}^n \rho_\tau(Y_i^* - Z_i^{*\prime} \beta) \right].$$

243 Put $D_n := \{(Y_1, Z_1), \dots, (Y_n, Z_n)\}$. A bootstrap estimator of C is given by

$$244 \quad \hat{C}^* := E[n(\hat{\beta}_n^* - \hat{\beta}_n)(\hat{\beta}_n^* - \hat{\beta}_n)' \mid D_n],$$

245 which can be calculated by a simulation method. As usual in the quantile regression
 246 literature, we use $m_\beta(y, z) := \rho_\tau(y - z'\beta) - \rho_\tau(y - z'\beta_0)$ as an objective function.

247 We investigate sufficient conditions for the consistency of $\hat{C}^*$. Possible sufficient
 248 conditions are:

249 **(Q0)** The conditional distribution of $\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n)$ given D_n converges weakly to $N(0, C)$
 250 in probability.

251 **(Q1)** The parameter space $\mathcal{B}$ is a compact and convex subset of $\mathbb{R}^d$.

252 **(Q2)** $E[\|Z\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$.

253 **(Q3)** The conditional density $f_{Y|Z}(y|z)$ is continuous in y and there exists a constant
 254 $C_f < \infty$ such that $f_{Y|Z}(y|z) \leq C_f$. The matrix $A_\beta := E[f_{Y|Z}(Z'\beta|Z)ZZ']$ is
 255 positive definite for all $\beta \in \mathcal{B}$.

256 Condition (Q0) is a high level condition. Primitive sufficient conditions for (Q0) are
 257 found in Hahn (1995). Conditions (Q1)-(Q3) guarantee that there exists a positive con-
 258 stant c such that $M(\beta) := E[m_\beta(Y, Z)] \geq c\|\beta - \beta_0\|^2$ for all $\beta \in \mathcal{B}$. On the other hand,
 259 it is not difficult to see that $|m_{\beta_1}(y, z) - m_{\beta_2}(y, z)| \leq \|z\| \cdot \|\beta_1 - \beta_2\|$ for all $\beta_1, \beta_2 \in \mathbb{R}^d$.
 260 Thus, given condition (Q2), condition (ii)' in Remark 2.3 is satisfied with $\dot{m}(y, z) = \|z\|$
 261 and with $p = 2$. In summary, we have shown that:

262 **Corollary 3.1** Under conditions (Q0)-(Q3), $\hat{C}^* \xrightarrow{p} C$.

263 The boundedness of the parameter space is in usual not assumed in the quantile re-
 264 gression literature, although it is standard in general asymptotic theory. The boundedness
 265 of the parameter space is indeed essential for the present purpose. To see this, we recall
 266 the result of Ghosh et al. (1984). Ghosh et al. (1984) showed that the bootstrap variance
 267 estimator of the sample quantile may not be consistent despite the distributional con-
 268 sistency of the bootstrap sample quantile. The inconsistency of the bootstrap variance
 269 estimator is caused by the fact that the bootstrap sample quantile may sometimes take
 270 an extremely large value when there is no moment restriction. Ghosh et al. (1984) also
 271 showed that the bootstrap variance estimator will be consistent when a mild moment re-
 272 striction is satisfied. In the present case, since there is no moment restriction on Y , if $\mathcal{B}$ is
 273 unbounded, $\hat{C}_n^*$ can be inconsistent (recall that the quantile regression estimator reduces
 274 to the sample quantile when $Z = 1$. In that case, A_β reduces to the density of Y , of which

the infimum over the entire real line must be zero). The role of the boundedness of the parameter space is to prevent $\hat{\beta}_n^*$ from taking an extremely large value, thereby ensuring the consistency of $\hat{C}^*$. An alternative possible way of ensuring the consistency of $\hat{C}^*$ is to put a suitable moment restriction on Y instead of restricting the parameter space, as Ghosh et al. (1984) did in the sample quantile case. We leave this extension as a future research topic.

It is worthwhile to remark that the bootstrap variance estimator is robust to misspecification. Suppose that the model (3.1) is misspecified but there exists a unique solution β_0 to the unconditional moment restriction:

$$E[\{\tau - I(Y \leq Z'\beta_0)\}Z] = 0.$$

Then, under suitable regularity conditions, it is shown that $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$ where A is the same as before but $B := E[\{\tau - I(Y \leq Z'\beta_0)\}^2 ZZ']$ (Angrist et al., 2006). It is not difficult to see that the conclusion of Corollary 3.1 is valid under the present situation with B being the present specification.

3.2 EIC

In this section, we give an informal discussion on the justification of EIC proposed by Ishiguro et al. (1997). We do not intend to make a full list of regularity conditions for EIC to make the paper succinct and focus on how to use our Theorem 2.2 to the justification of EIC.

Let $X, X_1, \dots, X_n$ be an independent sample from a distribution P . Consider a parametric model $\{f(x|\theta) : \theta \in \Theta \subset \mathbb{R}^d\}$, where for each $\theta \in \Theta$, $f(x|\theta)$ is a probability density with respect to some common base measure. We assume that the map $\theta \mapsto f(x|\theta)$ is sufficiently smooth. We allow for that the model does not contain the true distribution but assume that there exists a unique solution θ_0 to the equation:

$$E[\dot{\ell}(X, \theta_0)] = 0,$$

where $\ell(x, \theta) := \log f(x|\theta)$ and $\dot{\ell}(x, \theta) := \partial \ell(x, \theta) / \partial \theta$. Let $\hat{\theta}_n$ denote the maximum likelihood estimator (MLE) based on the sample $X_1, \dots, X_n$. Then, under suitable regularity conditions, it is shown that $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$, where $A := E[-\ddot{\ell}(X, \theta_0)]$, $\ddot{\ell}(x, \theta) := \partial^2 \ell(x, \theta) / \partial \theta \partial \theta'$ and $B := E[\dot{\ell}(X, \theta_0)\dot{\ell}(X, \theta_0)']$ (White, 1982). Akaike (1974) proposed to use minus of the expected log likelihood, $-E[\ell(X, \hat{\theta}_n)]$, to measure the adequacy of the estimated model. It is well known that $-n^{-1} \sum_{i=1}^n \ell(X_i, \hat{\theta}_n)$ has a bias of order n^{-1} . Put $b_n := E[n^{-1} \sum_{i=1}^n \{\ell(X_i, \hat{\theta}_n) - \ell(X, \hat{\theta}_n)\}]$. Takeuchi (1976) heuristically showed that $b_n = \text{tr}(BA^{-1})/n + o(n^{-1}) =: b/n + o(n^{-1})$, which can be formally justified by using Theorem 1 of Nishiyama (2010), and proposed an information criterion ("TIC"): $-n^{-1} \sum_{i=1}^n \ell(X_i, \hat{\theta}_n) + b/n$, which reduces to "AIC" (Akaike, 1974) when the model is correctly specified.

Ishiguro et al. (1997) proposed a bootstrap estimator of the bias term b . Let $X_1^*, \dots, X_n^*$ denote a bootstrap sample from $D_n := \{X_1, \dots, X_n\}$ and let $\hat{\theta}_n$ denote the bootstrap MLE. Ishiguro et al. (1997) proposed the estimator $\hat{b}^* := E[\sum_{i=1}^n \{\ell(X_i^*, \hat{\theta}_n^*) -$

314 $\ell(X_i, \hat{\theta}_n^*) \mid D_n]$. We argue the consistency of $\hat{b}^*$. Decompose $\hat{b}^*$ as

$$\begin{aligned}
 315 \quad \hat{b}^* &= \mathbb{E} \left[\sum_{i=1}^n \{ \ell(X_i^*, \hat{\theta}_n^*) - \ell(X_i^*, \hat{\theta}_n) \} \mid D_n \right] \\
 316 \quad &+ \mathbb{E} \left[\sum_{i=1}^n \{ \ell(X_i, \hat{\theta}_n) - \ell(X_i, \hat{\theta}_n^*) \} \mid D_n \right] \\
 317 \quad &=: \mathbb{E}[I \mid D_n] + \mathbb{E}[II \mid D_n].
 \end{aligned}$$

318 The Taylor expansion gives that $I = 2^{-1}n(\hat{\theta}_n^* - \hat{\theta}_n)' \hat{A}_n(\tilde{\theta}_n^*)(\hat{\theta}_n^* - \hat{\theta}_n)'$ and $II =$
 319 $2^{-1}n(\hat{\theta}_n^* - \hat{\theta}_n)' \hat{A}_n(\tilde{\theta}_n^*)(\hat{\theta}_n^* - \hat{\theta}_n)'$, where $\hat{A}_n(\theta) := -n^{-1} \sum_{i=1}^n \ddot{\ell}(X_i, \theta)$, $\hat{A}_n^*(\theta) :=$
 320 $-n^{-1} \sum_{i=1}^n \ddot{\ell}(X_i^*, \theta)$ and $\tilde{\theta}_n^*$ is on the line segment between $\hat{\theta}_n$ and $\hat{\theta}_n^*$ (I and II may
 321 have different $\tilde{\theta}_n^*$). Under suitable regularity conditions, the conditional distributions of I
 322 and II given D_n converge weakly to the distribution of $2^{-1}Z'AZ$ in probability where
 323 $Z \sim N(0, A^{-1}BA^{-1})$, and $\mathbb{E}[Z'AZ] = \text{tr}(BA^{-1}) = b$. Suppose that, for instance,
 324 there exists a function $H(x)$ such that $\|\ddot{\ell}(x, \theta)\| \leq H(x)$ for all $\theta \in \Theta$ for some suit-
 325 able norm $\|\cdot\|$. Then, $|I| \leq 2^{-1}\{n^{-1} \sum_{i=1}^n H(X_i^*)\} \cdot \|\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\|^2$ and $|II| \leq$
 326 $2^{-1}\{n^{-1} \sum_{i=1}^n H(X_i)\} \cdot \|\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\|^2$. In view of Lemma 2.1, sufficient conditions
 327 for the moment convergence are $\mathbb{E}[H(X)^{2(1+\epsilon)}] < \infty$ and $\mathbb{E}[\|\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)\|^{4(1+\epsilon)} \mid$
 328 $D_n] = O_p(1)$ for some $\epsilon > 0$. Theorem 2.2 gives primitive sufficient conditions for the
 329 latter condition (Theorem 2.2 indeed gives sufficient conditions for the stronger assertion
 330 that $\mathbb{E}[\|\hat{b}^* - b\| \rightarrow 0)$.

331 **Acknowledgments.** The author thanks Professor Tatsuya Kubokawa for his valuable
 332 comments. This research was supported by the Grant-in-Aid for Scientific research pro-
 333 vided by the JSPS.

334 References

- 335 Akaike, H. (1974). Information theory and an extension of the maximum likelihood prin-
 336 ciple. In: *2nd International Symposium on Information Theory*, ed. by B.N. Petrov and
 337 F. Csaki, pp. 267–281, Akademiai Kiado.
- 338 Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006). Quantile regression under
 339 misspecification, with an application to the US wage structure. *Econometrica* **74** 539–
 340 563.
- 341 Arcones, M. and Giné, E. (1992). On the bootstrap of M-estimators and other statistical
 342 functionals. In: *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard, pp.
 343 14–47, Wiley.
- 344 Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regres-
 345 sion models : A Monte Carlo study. *J. Econometrics* **68** 303–338.
- 346 Chung, K. L. (2001). *A Course in Probability Theory, 3rd edition*. Academic Press.
- 347 Dudley, R. M. (1999). *Uniform Central Limit Theorem*. Cambridge Univ. Press.

- 348 Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.
- 349 Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the
350 sample median. *Ann. Statist.* **12** 1130–1135.
- 351 Gonçalves, S. and White, H. (2005). Bootstrap standard error estimates for linear regres-
352 sion. *J. Amer. Stat. Assoc.* **100** 970–979.
- 353 Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* **11**
354 105–121.
- 355 Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and
356 EIC, an extension of AIC. *Ann. Inst. Stat. Math.* **49** 411–434.
- 357 Jenrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann.*
358 *Math. Stat.* **40** 633–643.
- 359 Koenker, R. (2005). *Quantile Regression*. Oxford Univ. Press.
- 360 Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.
- 361 Nishiyama, Y. (2010). Moment convergence of M-estimators. *Statist. Neerlandica*, **64**
362 505–507.
- 363 Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag.
- 364 Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of
365 models. *Mathematical Sciences* **153** 12–18 (in Japanese).
- 366 Wellner, J. A. and Zhan, Y. (1996). Bootstrapping Z-estimators. Unpublished manuscript.
- 367 White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*
368 **50** 1–25.
- 369 van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Pro-*
370 *cesses: With Applications to Statistics*. Springer-Verlag.
- 371 Yoshida, N. (2010). Polynomial type large deviation inequalities and quasi-likelihood
372 analysis for stochastic differential equations. *Ann. Inst. Stat. Math.*, to appear, Online
373 first: May 20, 2010, DOI: 10.1007/s10463-009-0263-z.

374 Kengo Kato
375 Department of Mathematics
376 Graduate School of Science
377 Hiroshima University
378 1-3-1 Kagamiyama
379 Higashi-Hiroshima
380 Hiroshima 739-8526
381 Japan
382 kkato@hiroshima-u.ac.jp