Journal of the Korean Statistical Society (1998), **27**: *3, pp 349–358*

Strong Representations for LAD Estimators in AR(1) Models

Hee-Jeong Kang[1] and Key-Il Shin[2]

ABSTRACT

Consider the AR(1) model $X_t = \beta X_{t-1} + \varepsilon_t$ where $|\beta| < 1$ is an unknown parameter to be estimated and $\{\varepsilon_t\}$ denotes the independent and identically distributed error terms with unknown common distribution function F. In this paper, a strong representation for the least absolute deviation (LAD) estimate of β in AR(1) models is obtained under some mild conditions on F.

Keywords: Stationary AR(1); LAD estimator; Strong Representation

1. INTRODUCTION

The high speed computer technologies have brought back considerable interest in the robust procedures which have been ignored for a long time in statistical literature. A typical example is the LAD(least absolute deviation) method that was abandoned in favor of the LS(least squares) method, because of the computational difficulties. One of the computational difficulties using the LAD method rather than the LS method comes from the fact that it is not easy to obtain the explicit form for the LAD estimator, which is usually needed for the statistical inferences. Since Bahadur (1966) has obtained a strong representation for the sample p-th quantile, a lot of studies have been made on this kind of strong representation for the robust statistical methods. See, for example, Kiefer (1967) for the sample quantiles and also see Babu (1989) for the LAD estimator in linear models. In the times series settings, Koul and Zhu (1995) have obtained the *Bahadur-Kiefer* type representation for the class of generalized M-estimators for the parameter vector in AR(p) models. However, the conditions such as $E|\varepsilon_t|^\tau < \infty$, for some $\tau > 8$ they used for LAD estimator is too strict so that the representation formular is little useful. In this paper, we are also concerned with

[1]Dept. of Statistics, Chonbuk National University, Chonju, Chonbuk 561-756, Korea.
[2]Dept. of Statistics, Hankuk University of Foreign Studies, Kyunggi-do 449-791, Korea.

the same representation problem for the LAD estimator in AR(1) models under much milder conditions than in Koul and Zhu (1995). For this purpose, let $\{X_t\}$ be a sequence of the AR(1) model defined by

$$X_t = \beta X_{t-1} + \varepsilon_t \,, \qquad X_0 = 0 \,, \qquad t = 1, \dots, n \tag{1.1}$$

where β is a parameter of the model with $|\beta| < 1$ and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random errors with unknown distribution function F which satisfies the following conditions:

(A) $E|\varepsilon_t|^{2+\delta} < \infty$, for some $\delta > 0$.

(B) The *d.f.* F has a unique median at 0.

(C) F has a continuous density f in a neighborhood of 0 and $f(0) > 0$.

(D) For some $c > 0$, $|f(h) - f(0)| \le c\,|h|^{1/2}$ for all h in a neighborhood of 0.

In order to estimate the unknown parameter β in (1.1), we are primarily concerned with the LAD estimator $\hat{\beta}$ which is a solution of

$$\sum_{t=1}^{n} \left| X_t - \hat{\beta} X_{t-1} \right| = \inf_{\beta \in \mathcal{R}^1} \sum_{t=1}^{n} |X_t - \beta X_{t-1}| \,. \tag{1.2}$$

The main aim of this study is to understand the large sample behavior of $\{X_t\}$ in the stochastic process (1.1) under the conditions **(A)**-**(D)** and to find the explicit representation for the LAD estimator $\hat{\beta}$ in (1.2) which is also very useful in sequential analysis. Thus under the conditions **(A)**-**(D)**, we have the following result.

Theorem 1.1. Under $H_0 : \beta = \beta_0$, with probability 1,

$$\sqrt{n}\,(\hat{\beta} - \beta_0)\, 2f(0)\, \frac{1}{n} \sum_{t=1}^{n} X_{t-1}^2 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}\, \mathrm{sign}(\varepsilon_t) + R_n \tag{1.3}$$

where $R_n = O\left(n^{-\delta/(4(2+\delta))} (\ln n)^{3/4}\right)$ as $n \to \infty$.

Remark 1.1: Except for the condition **(A)** and the order of the remainder term, the strong representation for $\hat{\beta}$ in (1.3) is the same as that in Koul and Zhu (1995). See Theorem 5 in Koul and Zhu (1995) for the details.

The remaining of this paper is organized as follows. In Section 2, we present some technical results including exponential inequalities for sums of martingale difference sequences which will play key roles in this paper. See Billingsley (1986) for more details on martingales. Section 3 contains the proofs of the *Bahadur-Kiefer* type representation of the LAD estimator $\hat{\beta}$. We essentially follow the ideas of Babu (1989) to derive Theorem 1.1.

2. PRELIMINARY RESULTS

To begin with, we state a result to describe the large sample behavior of $\{X_t\}$ under the conditions **(A)-(D)**.

Lemma 2.1. *With probability 1,*

$$n^{-1/(2+\delta)} \max_{1\le t\le n} |X_t| \longrightarrow 0 \qquad as \;\; n \to \infty\,. \tag{2.1}$$

Proof: First notice that by the relation (1.1),

$$\max_{1\le t\le n} |X_t| \le \max_{1\le t\le n} |\varepsilon_t| \,/(1-|\beta|)\,.$$

For arbitray $\epsilon > 0$, let $\varepsilon_t^{'} = \varepsilon_t \; I\left(|\varepsilon_t| \le \epsilon\, t^{1/(2+\delta)}\right)$. Then, $Pr(\varepsilon_t \ne \varepsilon_t^{'}\; i.o.) = 0$ because

$$\sum_{t=1}^{\infty} Pr(\,|\varepsilon_t| > \epsilon\, t^{1/(2+\delta)}\,) \le 1 + E\,|\varepsilon_t/\epsilon|^{2+\delta} < \infty\,.$$

Denote $A_\epsilon = \{\omega : |\varepsilon_t| \le \epsilon\, t^{1/(2+\delta)}$ for all large t $\}$. The above argument implies that $Pr(A_\epsilon) = 1$. Here, let $\epsilon = 1/k$ for arbitrary large integer k and denote $A = \cap_{k=1}^{\infty} A_{1/k}$. Then, $Pr(A) = 1$ and we can say that on such A,

$$\max_{1\le t\le n} |X_t| \le \frac{1}{1-|\beta|} \left(C_{(k,\omega)} + \frac{1}{k}\, n^{1/(2+\delta)} \right)$$

where $C_{(k,\omega)}$ is an arbitrary large constant which depends on not n but k and ω and ω is an element of the σ-field generated by $\{\varepsilon_1, \dots, \varepsilon_n\}$. Therefore,

$$\limsup_{n\to\infty} n^{-1/(2+\delta)} \max_{1\le t\le n} |X_t| \le \frac{1}{k}$$

and the result in (2.1) follows by letting $k \to \infty$. □

The following lemma is the basic scheme to derive the main result of this paper.

Lemma 2.2. *Suppose that* $Pr\left(\limsup(B_n \cap A_n)\right) = 0$ *and* $Pr\left(\liminf A_n\right) = 1$. *Then,* $Pr\left(\limsup B_n\right) = 0$.

Proof: Note that $B_n \subset (A_n^c \cup (B_n \cap A_n))$. Thus, we can get

$$\bigcup_{n \geq k}^{\infty} B_n \subset \left(\bigcup_{n \geq k}^{\infty} A_n^c\right) \bigcup \left(\bigcup_{n \geq k}^{\infty} (B_n \cap A_n)\right) .$$

So, with the assumptions and the fact that $\cup_{n \geq k}^{\infty} B_n$ is decreasing in k, it is easy to show that

$$Pr\left(\limsup B_n\right) = Pr\left(\bigcap_{k \geq 1}^{\infty} \bigcup_{n \geq k}^{\infty} B_n\right) = \lim_{k \to \infty} Pr\left(\bigcup_{n \geq k}^{\infty} B_n\right) = 0 .$$

□

In order to use Lemma 2.2, by the first Borel-Cantelli lemma, it suffices to show that $Pr(B_n \cap A_n) = O\left(n^{-r}\right)$ for some $r > 1$. The next lemma is an exponential inequalities for sum of martingale difference sequences which will play key roles in our proofs of Theorem 1.1. See Billingsley (1986) for more details on martingales.

Lemma 2.3. *Let* $\{Z_i\}$ *be martingale difference sequences with* $|Z_i| \leq b$ *a.e. for some* $b > 0$ *and let* $\sum_{i=1}^n E\left(Z_i^2 \mid \mathcal{F}_{i-1}\right) = \sum_{i=1}^n \sigma_i^2$ *with* $\mathcal{F}_{i-1}$ *being the* σ*-field generated by* $Z_1, \ldots, Z_{i-1}$. *Then for some* $v > 0$ *and* $0 < a < v/b$,

$$Pr\left(\sum_{i=1}^{n} Z_i > 2a \ , \ \sum_{i=1}^{n} \sigma_i^2 \leq v\right) \leq exp(-a^2/v) .$$

Proof: Since $e^y \leq 1 + y + y^2$ for $|y| \leq 1$, we have for any $0 < \mu \leq 1/b$,

$$
\begin{aligned}
& Pr\left(\sum_{i=1}^{n} Z_i > a + \mu v \ , \ \sum_{i=1}^{n} \sigma_i^2 \le v\right) \\
\le\ & \exp(-\mu a) E\left\{\exp\left(\mu \sum_{i=1}^{n}(Z_i - \mu\sigma_i^2)\right)\right\} \\
\le\ & \exp(-\mu a) E\left\{\exp\left(\mu \sum_{i=1}^{n-1}(Z_i - \mu\sigma_i^2)\right) \exp(-\mu^2\sigma_n^2) E\big(1 + \mu Z_n + \mu^2 Z_n^2 | \mathcal{F}_{n-1}\big)\right\} \\
\le\ & \exp(-\mu a) E\left\{\exp\left(\mu \sum_{i=1}^{n-1}(Z_i - \mu\sigma_i^2)\right) \exp(-\mu^2\sigma_n^2)\exp(\mu^2\sigma_n^2)\right\} \\
\le\ & \exp(-\mu a) E\left\{\exp\left(\mu \sum_{i=1}^{n-1}(Z_i - \mu\sigma_i^2)\right)\right\} \\
\le\ & \exp(-\mu a)
\end{aligned}
$$

because $\{Z_i\}$'s are martingale difference sequences. Then, the proof will be completed by taking $\mu = a/v$. □

We are now ready to obtain the strong representation for the LAD estimator $\hat{\beta}$ in Theorem 1.1. From now on, we assume that the null hypothesis $H_0 : \beta = \beta_0$ holds throughout the paper.

3. PROOF OF THEOREM 1.1

The proof is divided into several lemmas.

Lemma 3.1. *Under the conditions* **(A)**-**(D)**, *we have with probability 1,*

$$\sum_{t=1}^{n} Z_t = O\left(n^{-\delta/(4(2+\delta))}(\ln n)^{3/4}\right)$$

uniformly for $|\beta - \beta_0| \le B\,(\ln n/n)^{1/2}$ *and for any* $B > 0$ *where*

$$Z_t = \frac{1}{\sqrt{n}} X_{t-1}\left\{sign\left(\varepsilon_t - (\beta - \beta_0)X_{t-1}\right) - sign(\varepsilon_t) + 2F\left((\beta - \beta_0)X_{t-1}\right) - 1\right\}$$

and $sign(\varepsilon_t) = 1$ *or* -1 *if* $\varepsilon_t > 0$ *or* $\varepsilon_t < 0$, *respectively.*

Proof: Let $Z_t' = Z_t\, I\left(|X_{t-1}|/\sqrt{n} \le n^{-\delta/(2(2+\delta))}\right)$ and let

$$H_1 = \left\{\omega : \exists\ n_0(\omega) \text{ such that } Z_t' = Z_t \text{ for all } t \ge n_0(\omega)\right\}. \tag{3.1}$$

Then, by Lemma 2.1, $Pr(H_1) = 1$. Notice that $\{Z_t'\}$'s are martingale difference sequences. Thus, with the conditions and the fact that $|Z_t'| \le 2n^{-\delta/(2(2+\delta))}$ for $1 \le t \le n$, it is not difficult to prove that $\sum_{t=1}^n E\left(Z_t'^2 \mid \mathcal{F}_{t-1}\right) \le C \cdot B \cdot (\ln n)^{1/2} n^{-\delta/(2(2+\delta))}$ for some $C > 0$ and $B > 0$ where $\mathcal{F}_{t-1}$ is the σ-field generated by $\varepsilon_1, \cdots, \varepsilon_{t-1}$. Therefore, by Lemma 2.3 with taking $A = 2(CBr)^{1/2}$ for any $C > 0,\ B > 0,\ r > 1$ we can show that

$$Pr\left\{\left|\sum_{t=1}^n Z_t'\right| > An^{-\delta/(4(2+\delta))}(\ln n)^{3/4}, \sum_{t=1}^n E\left(Z_t'^2 \mid \mathcal{F}_{t-1}\right) \le C\, B\, (\ln n)^{1/2} n^{-\delta/(2(2+\delta))}\right\}$$
$$= O\left(n^{-r}\right)$$

Thus by Lemma 2.2, with probability 1, $\sum_{t=1}^n Z_t' = O\left(n^{-\delta/(4(2+\delta))}(\ln n)^{3/4}\right)$ for all large n. Hence on H_1 in (3.1), since

$$\left|\sum_{t=1}^n Z_t(\omega) - \sum_{t=1}^n Z_t'(\omega)\right| \le \sum_{t=1}^{n_0(\omega)} 2\, \frac{|X_{t-1}(\omega)|}{\sqrt{n}} = O(n^{-1/2}) \tag{3.2}$$

for all large n, the result now follows. □

Lemma 3.2. *Under the conditions* **(A)-(D)**, *with probability 1,*

$$\limsup_{n\to\infty} \sup_{|\beta-\beta_0| = B(\ln n/n)^{1/2}} \frac{|\, g_n(\beta)\,|}{(\beta-\beta_0)^2\, n} \le \frac{f(0)}{2}\,\tau^2$$

where $\tau^2 = \gamma\,\sigma^2/(1-|\beta_0|)^2,\ 0 < \gamma < (\,(1-|\beta_0|)/(1+|\beta_0|)\,)^2,\ \sigma^2 = E(\varepsilon_t^2),$ $B = 8(1-|\beta_0|)/(f(0)\gamma\sigma)$ *and*

$$g_n(\beta) = \sum_{t=1}^n \{\, |\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)| - E\,(\,|\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)|\ \mid \mathcal{F}_{t-1})\,\}$$

with $\mathcal{F}_{t-1}$ *being the* σ*-field generated by* $\varepsilon_1, \cdots, \varepsilon_{t-1}$.

Proof: With similar arguments in the proof of Lemma 3.1, let $W_t = |\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)| - E\left(\,|\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)| \mid \mathcal{F}_{t-1}\right)$ and $W_t' = W_t\, I(|X_{t-1}| \le n^{1/(2+\delta)})$ and denote

$$H_2 = \left\{\omega :\ \exists\ n_0(\omega)\ \textit{such that}\ W_t' = W_t\ \textit{for all}\ t \ge n_0(\omega)\right\}. \tag{3.3}$$

Then, by Lemma 2.1, $Pr(H_2) = 1$ and since for any y and z,

$$|y| - |y-z| = \int_0^z (1 - 2I(y \le x))\, dx\ , \tag{3.4}$$

we can get with the conditions **(A)**-**(D)** that for $1 \le t \le n \quad |W_t'| \le 2\cdot B\cdot (\ln n)^{1/2} n^{-\delta/(2(2+\delta))}$ and $\sum_{t=1}^n E(W_t'^2 \mid \mathcal{F}_{t-1}) \le C\cdot B^2\cdot(\ln n)$ uniformly for $|\beta-\beta_0| = B(\ln n/n)^{1/2}$ and for some $C>0$. Notice that $\{W_t'\}$'s are martingale difference sequences. Thus on H_2 in (3.3) by applying Lemma 2.3 and Lemma 2.2 to $(\left|\sum_{t=1}^n W_t'\right|)/((\beta-\beta_0)^2 n)$ with $a = f(0)\tau^2/4$ and $v = CB^2(\ln n)$, we can show that with probability 1,

$$\frac{\left|\sum_{t=1}^n W_t'\right|}{(\beta-\beta_0)^2\, n} \le \frac{f(0)\tau^2}{2}$$

uniformly for $|\beta-\beta_0| = B(\ln n/n)^{1/2}$ and for all large n. Therefore with the similar arguments in (3.2), we can prove that with probability 1,

$$\sup_{|\beta-\beta_0| = B(\ln n/n)^{1/2}} \frac{|\, g_n(\beta)\,|}{(\beta-\beta_0)^2\, n} \le O\left((n\ln n)^{-1/2}\right) + \frac{f(0)}{2}\,\tau^2$$

for all large n. Hence the result follows by letting $n \to \infty$. □

Lemma 3.3. *Under the conditions* **(A)**-**(D)**, *with probability 1,*

$$\left|\,\hat\beta - \beta_0\,\right| \le B\,(\ln n/n)^{1/2}$$

for all large n where $\hat\beta$ is a LAD estimator in (1.2) and B is in the conditions of Lemma 3.2.

Proof: With the fact that F is differentiable at 0 and the relation in (3.4), we can get

$$\sum_{t=1}^n E\left(\,|\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)| \mid \mathcal{F}_{t-1}\,\right) = -f(0)(1+o(1))(\beta-\beta_0)^2 \sum_{t=1}^n X_{t-1}^2$$

uniformly for $|\beta - \beta_0| = B(\ln n/n)^{1/2}$. So by the following relation

$$\frac{1}{(1+|\beta_0|)^2}\,\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 \;\le\; \frac{1}{n}\sum_{t=1}^{n}X_t^2 \;\le\; \frac{1}{(1-|\beta_0|)^2}\,\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\,,$$

we obtain with probability 1,

$$\limsup_{n\to\infty}\ \sup_{|\beta-\beta_0|=B(\ln n/n)^{1/2}} \frac{1}{(\beta-\beta_0)^2}\,\frac{1}{n}\sum_{t=1}^{n}E\big(\,|\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)|\ \big|\ \mathcal{F}_{t-1}\,\big) \le \frac{-f(0)\,\sigma^2}{(1+|\beta_0|)^2}\ .$$

Thus, by the result of Lemma 3.2, it is easy to show that with probability 1,

$$\limsup_{n\to\infty}\ \sup_{|\beta-\beta_0|=B(\ln n/n)^{1/2}} \frac{1}{(\beta-\beta_0)^2}\,\frac{1}{n}\sum_{t=1}^{n}\big(\,|\varepsilon_t| - |\varepsilon_t - X_{t-1}(\beta-\beta_0)|\,\big) \le \frac{-f(0)\,\sigma^2}{2\,(1+|\beta_0|)^2} < 0 \tag{3.5}$$

Notice that $h(\beta) = \sum_{t=1}^{n}|\varepsilon_t - X_{t-1}(\beta-\beta_0)|$ is convex in β. Therefore, this and the result in (3.5) imply that with probability 1, for all large n,

$$\sum_{t=1}^{n}|\varepsilon_t| < \inf_{|\beta-\beta_0|\ge B(\ln n/n)^{1/2}} \sum_{t=1}^{n}|\varepsilon_t - X_{t-1}(\beta-\beta_0)|$$

and this completes the proof. □

Lemma 3.4. *Under the conditions* **(A)-(D)**, *with probability 1,*

$$\left|\,\frac{1}{\sqrt{n}}\,y_n(\hat\beta)\,\right| = o\left(n^{-\delta/(2(2+\delta))}\right)$$

where $y_n(\beta) = \sum_{t=1}^{n} X_{t-1}\, sign\,(\varepsilon_t - X_{t-1}(\beta-\beta_0))$ *and* $\hat\beta$ *is a LAD estimator in (1.2).*

Proof: A LAD estimator is a solution of

$$\inf\left\{\sum_{t=1}^{n}|X_t - \beta X_{t-1}| : \beta\in\mathcal{R}^1\right\} = \inf\left\{\sum_{t=1}^{n}|\,\varepsilon_t - (\beta-\beta_0)X_{t-1}| : \beta\in\mathcal{R}^1\right\}.$$

Then, with the directional derivative arguments as in Babu (1989), all the non-zero directional derivatives of $\sum_{t=1}^{n} | \varepsilon_t - (\beta - \beta_0)X_{t-1}|$ at $\beta = \hat{\beta}$ must be non-negative. This implies that

$$\left| \sum_{t=1}^{n} X_{t-1} \operatorname{sign}\left(\varepsilon_t - X_{t-1}(\hat{\beta} - \beta_0)\right) \right| \le \sum_{t=1}^{n} |X_{t-1}| \; I(X_t = \hat{\beta} X_{t-1}) \,. \quad (3.6)$$

Thus, Since $\{\varepsilon_t\}$'s are independent and the distribution of $\{\varepsilon_t\}$ is continuous we can get

$$\sum_{t=1}^{n} |X_{t-1}| \; I(X_t = \hat{\beta} X_{t-1}) \le \max_{1 \le t \le n} |X_t| \,. \quad (3.7)$$

Therefore, the result now follows from (3.6), (3.7) and Lemma 2.1. □

Proof of Theorem 1.1. With the conditions **(A)-(D)** and the independence of $\{\varepsilon\}$'s, we can ontain that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1} \, E\left(\operatorname{sign}(\varepsilon_t - (\beta - \beta_0)X_{t-1}) \mid \mathcal{F}_{t-1}\right) \\ &= \frac{-2}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1} \left(F(\, (\beta - \beta_0)X_{t-1}) - F(0) \,\right) \\ &= -2\, f(0)\, (\beta - \beta_0) \, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}^2 \; + \; O\left(n^{-\delta/(4(2+\delta))} \, (\ln n)^{3/4}\right) \end{aligned} \quad (3.8)$$

uniformly for $|\beta - \beta_0| = B(\ln n/n)^{1/2}$ and for B in Lemma 3.2. Thus, from Lemma 3.3 and (3.8), it is esay to say that with probability 1,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1} \, E\left(\operatorname{sign}\left(\varepsilon_t - (\hat{\beta} - \beta_0)X_{t-1}\right) \mid \mathcal{F}_{t-1}\right) \\ &= -2\, f(0)\, (\hat{\beta} - \beta_0) \, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}^2 \; + \; O\left(n^{-\delta/(4(2+\delta))} \, (\ln n)^{3/4}\right) \end{aligned} \quad (3.9)$$

for all large n. Hence, from Lemma 3.1 and Lemma 3.3 , we can see that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \, y_n(\hat{\beta}) - \frac{1}{\sqrt{n}} \, y_n(\beta_0) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1} \, E\left(\operatorname{sign}\left(\varepsilon_t - (\hat{\beta} - \beta_0)X_{t-1}\right) \mid \mathcal{F}_{t-1}\right) \\ &= O\left(n^{-\delta/(4(2+\delta))} \, (\ln n)^{3/4}\right) \end{aligned}$$

where $y_n(\beta)$ is in Lemma 3.4 and this completes the proof with (3.9) and the result of Lemma 3.4. □

REFERENCES

Babu, G. J. (1989). "Strong Representations for LAD Estimators in Linear Models," *Probability Theory and Related Fields* **83**, 547-558.

Bahadur, R. R. (1966). "A Note on Quantiles in Large Samples," *The Annals of Mathematical Statistics* **37**, 577-580.

Billingsley, P. (1986). *Probability and Measure* 2nd ed., John Wiley & Sons, New York.

Kiefer, J. (1967). "On Bahadur's Representation of Sample Quantiles," *The Annals of Mathematical Statistics* **38**, No. 5, 1323-1342.

Koul, H. L. and Zhu, Zhiwei (1995). " Bahadur-Kiefer Representations for GM-estimators in Autoregression Models," *Stochastic Processes and their Applications* **57**, 167-189.