Journal of the Korean Statistical Society (2002), **31**: *1, pp 121–128*

A New Proof of Efficiency of LAD Estimation in an Autoregressive Process

Key-Il Shin[1], Hee-Jeong Kang[2] and Songyong Sim[3]

ABSTRACT

In this paper we provide a new proof of the asymptotic distributions of LAD estimators using the martingale limit theorem and show the efficiency of LAD estimators in a stationary AR(1) model setting.

Keywords: AR(1), LAD estimator, asymptotic normality, efficiency.

1. Introduction

Let $\{X_t\}$ be a sequence of the first-order autoregressive process given by

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

where β is a parameter of the model with $|\beta| < 1$ and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (*iid*) random errors with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2 < \infty$ and unknown distribution function F. In data analysis, the primary concern is to estimate the unknown parameter β in (1.1) from the data and a typical estimation procedure is to use the least squares method. One of the advantages of the least squares(LS) estimator is that the form of the LS estimator is well known and easy to compute. The asymptotic theory of the LS estimator is everywhere as in Theorem 8.2.1 of Fuller (1996). However, one of the disadvantages of the LS estimator is that when the error distribution is non-normal, the LS estimator is asymptotically less efficient than other robust estimators.

The revolutionary computer technologies make it possible for robust estimation procedures to play a more important role in data analysis. The least absolute deviations(LAD) estimator $\hat{\beta}_{LAD}$ which is a solution of the following

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

[1]Department of Statistics, Hankuk University of Foreign Studies, Kyungki 449-791, Korea
[2]Department of Statistics, Chonbuk National University, Chonju, Chonbuk 561-756, Korea
[3]Department of Statistics, Hallym University, Chuncheon, Kangwon 200-702, Korea

can be a good alternative to the LS estimator. The conceptual simplicity and the less sensitivity to extreme errors of LAD estimates make them worth being considered. In fact, when the errors follow the double exponential distribution, the LAD estimates are maximum likelihood estimates and hence asymptotically more efficient than LS estimates. A cost for the robustness of LAD estimates is computational difficulties. The form of LAD estimates is not easy to obtain and the computation of them is more expensive than that of LS estimates. To cope with this non-explicit problem of LAD estimates in (1.2), Koul and Zhu (1995) developed a *Bahadur-Kiefer* type representations for LAD estimators in autoregression models. Kang and Shin (1998) also obtained a strong representation for LAD estimators in (1.2) under some conditions such as the followings :

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

Kang and Shin (1998) have shown that under the conditions **(A)-(D)**, the LAD estimator $\hat{\beta}_{LAD}$ in (1.2) can be written as

$$\sqrt{n}\,(\,\hat{\beta}_{LAD} - \beta\,)\,2f(0)\,\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \;=\; \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}\,\text{sign}(\varepsilon_t) \;+\; R_n \qquad (1.3)$$

where $\text{sign}(\varepsilon_t) = 1$ or -1 if $\varepsilon_t > 0$ or $\varepsilon_t < 0$, respectively, and $R_n = O\left(n^{-\delta/(4(2+\delta))}(\ln n)^{3/4}\right)$ with probability 1.

The main purpose of this study is to derive the asymptotic normality of LAD estimates using the representation form in (1.3) and to show the efficiency of them in situations when the error terms do not follow the normal distribution. We have noticed that the asymptotic normality of LAD estimates has been shown in Dunsmuir and Spencer (1991). The contribution of this paper is to give a new way to derive the results using the *Bahadur-Kiefer* type representation in (1.3) and the central limit theorem for martingale differences.

The remainder of this paper is organized as follows. In Section 2, we present the asymptotic normality and the efficiency of LAD estimates in AR(1) models. The martingale limit theory will play an important role in this paper. See Hall and Heyde (1980) for more details on the martingale limit theory. Section 3 contains some Monte Carlo simulation results to support the efficiency of LAD estimates obtained in Section 2.

2. Asymptotic Normality

We begin with the main result of this paper to describe the limiting distribution of LAD estimates $\hat{\beta}_{LAD}$ in (1.3).

Theorem 2.1. *Under the conditions* (A)-(D) *in Section 1, as* $n \to \infty$,

$$\sqrt{n}\left(\hat{\beta}_{LAD} - \beta\right) \xrightarrow{d} N\left(0\,,\ \frac{1}{(2f(0))^2} \cdot \frac{1-\beta^2}{\sigma^2}\right) \tag{2.1}$$

where $E(\varepsilon_t^2) = \sigma^2 < \infty$ *and* $f(0)$ *is the value of the density function at 0.*

Proof. First, for each $n \geq 1$ and $1 \leq t \leq n$, let $W_{n,n} = \sum_{t=1}^{n} Z_{n,t} = \sum_{t=1}^{n} \frac{1}{\sqrt{n}} X_{t-1}\ \mathrm{sign}(\varepsilon_t)$ and $\mathcal{F}_{n,t}$ be a σ-field generated by $\{\varepsilon_1, \ldots, \varepsilon_t\}$. Then it is obvious that $Z_{n,t}$ is $\mathcal{F}_{n,t}$-measurable and the σ-fields are nested ; $\mathcal{F}_{n,t-1} \subset \mathcal{F}_{n,t}$. Also, with the fact that $Z_{n,n} = W_{n,n} - W_{n,n-1}$ and the following

$$E\left(Z_{n,n} \mid \mathcal{F}_{n,n-1}\right) = \frac{1}{\sqrt{n}} X_{n-1} E\left(\mathrm{sign}(\varepsilon_n)\right) = \frac{1}{\sqrt{n}} X_{n-1}\left(1 - 2F(0)\right) = 0$$

because $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are *iid* random errors and the distribution function F has a unique median at 0, we can say that for each $n \geq 1$, $\{Z_{n,t}\}$ is a sequence of martingale differences and $\{W_{n,i}\}_{i=1}^{n}$ is a martingale sequence. For more details on martingale, we refer the reader to Billingsley (1986). With the above arguments, in order to obtain the limiting normal distribution in (2.1), we shall use the central limit theorem for martingale differences. That is, we need to investigate the conditions (ii) and (iii) of Theorem 5.3.4 in Fuller (1996) or the conditions of Corollary 3.1 in Hall and Heyde (1980). Therefore we need to show the followings

$$\sum_{t=1}^{n} E\left(Z_{n,t}^2 \,|\mathcal{F}_{n,t-1}\right) = \sum_{t=1}^{n} E\left(\frac{X_{t-1}^2}{n} \cdot \mathrm{sign}(\varepsilon_t)^2 \,|\mathcal{F}_{n,t-1}\right) \xrightarrow{p} \frac{\sigma^2}{1-\beta^2} \tag{2.2}$$

$$\text{for all } \epsilon > 0, \quad \sum_{t=1}^{n} E\left(Z_{n,t}^2 \cdot I\left(|Z_{n,t}| > \epsilon\right) |\mathcal{F}_{n,t-1}\right) \xrightarrow{p} 0 \tag{2.3}$$

where $I(A)$ denotes the indicator function of a set A.

It is easy to prove (2.2) from the facts that $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are *iid* random errors and the distribution function F has a unique median at 0 and the following result in Theorem 4.2 of Anderson (1959)

$$\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \quad \xrightarrow{p} \quad \frac{\sigma^2}{1-\beta^2}\ . \tag{2.4}$$

For (2.3), we got the following, for all $\epsilon > 0$,

$$\begin{aligned}
&\sum_{t=1}^{n} E\left(Z_{n,t}^2 \cdot I\left(|Z_{n,t}| > \epsilon\right) \mid \mathcal{F}_{n,t-1}\right) \\
&\quad = \sum_{t=1}^{n} \frac{X_{t-1}^2}{n} \cdot E\left\{\operatorname{sign}(\varepsilon_t)^2 \cdot I\left(|\operatorname{sign}(\varepsilon_t)| > \frac{\sqrt{n}\epsilon}{|X_{t-1}|}\right) \mid \mathcal{F}_{n,t-1}\right\} \\
&\quad \le \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \cdot E\left\{\operatorname{sign}(\varepsilon_t)^2 \cdot I\left(|\operatorname{sign}(\varepsilon_t)| > \frac{\sqrt{n}\epsilon}{\max_{1\le t\le n}|X_{t-1}|}\right)\right\} \\
&\quad = \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \cdot E\left\{\operatorname{sign}(\varepsilon_1)^2 \cdot I\left(|\operatorname{sign}(\varepsilon_1)| > \frac{\sqrt{n}\epsilon}{\max_{1\le t\le n}|X_{t-1}|}\right)\right\} \\
&\quad \xrightarrow{p} \frac{\sigma^2}{1-\beta^2} \cdot 0 \;=\; 0\,. \qquad (2.5)
\end{aligned}$$

The last result in (2.5) comes from (2.4) and

$$\operatorname{sign}(\varepsilon_1)^2 \cdot I\left(|\operatorname{sign}(\varepsilon_1)| > \sqrt{n}\epsilon / \max_{1\le t\le n}|X_{t-1}|\right) \le 1$$

and the dominated convergence theorem.

Therefore, with the results of (2.4), (2.5) and the central limit theorem for martingale differences in Corollary 3.1 of Hall and Heyde (1980), we obtain that as $n \to \infty$,

$$W_{n,n} = \sum_{t=1}^{n} Z_{n,t} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}\operatorname{sign}(\varepsilon_t) \xrightarrow{d} N\left(0\,,\, \frac{\sigma^2}{1-\beta^2}\right). \qquad (2.6)$$

Hence, the limiting normal distribution for LAD estimates $\hat{\beta}_{LAD}$ in (2.1) follows from (2.4) and (2.6). □

With the result of Theorem 2.1 and the limiting normal distribution theory of least squares estimates (denoted by $\hat{\beta}_{LS}$) as is given in Theorem 8.2.1 of Fuller (1996) or Theorem 4.3 of Anderson (1959)

$$\sqrt{n}\left(\hat{\beta}_{LS} - \beta\right) \;=\; \sqrt{n}\,\frac{\sum_{t=1}^{n} X_{t-1}\,\varepsilon_t}{\sum_{t=1}^{n} X_{t-1}^2} \xrightarrow{d} N\left(0\,,\,1-\beta^2\right), \qquad (2.7)$$

it is easy to obtain the asymptotic relative efficiency of LAD estimates $\hat{\beta}_{LAD}$ relative to least squares estimates. We state the asymptotic relative efficiency results of LAD estimates in the following corollary.

Corollary 2.2. *Let the asymptotic relative efficiency of estimator $\hat{\theta}_B$ relative to estimator $\hat{\theta}_A$, $(ARE(\hat{\theta}_B, \hat{\theta}_A))$ be defined as $\frac{Asym.\ Var(\hat{\theta}_A)}{Asym.\ Var(\hat{\theta}_B)}$. Then, with the conditions in Theorem 2.1, the $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ of the LAD estimator in (1.3) is $4f(0)^2\sigma^2$.*

Proof. The result comes from (2.1) and (2.7). □

Example 2.3 When the random errors $\{\varepsilon_t\}$ in (1.1) follow $N(0, \sigma^2)$, then, the $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ is $2/\pi$. For the logistic distribution cases with $E(\varepsilon_t^2) = \sigma^2 = \pi^2 a^2/3$ for an arbitrary scale parameter $a > 0$, the $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ is $\pi^2/12$. For the double exponential distribution cases with $E(\varepsilon_t^2) = \sigma^2 = 2/b^2$ for an arbitrary scale parameter $b > 0$, the $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ is 2.

3. Simulation Studies

This section contains some Monte Carlo studies to support the theoretical results on the asymptotic relative efficiency of LAD estimates. For this purpose, we simulated the AR(1) model in (1.1) with $|\beta| < 1$ and n from 50 to 500. With given values of β and n, we generated the sample $\{X_t\}$, $t = 1, \ldots, n$, under the conditions that $\{\varepsilon_t\}$ is a sequence of *iid* normal distribution $N(0, \sigma^2)$ or double exponential distribution with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$. For each sample of size n, we calculated the LS squares estimator and the LAD estimator using the form in (2.7) and (1.3), respectively. The above process was iterated for 20,000 times and we recorded the empirical sample variance of LS estimates and LAD estimates. The ratio of the empirical sample variance of LAD estimates over that of LS estimates was compared with the theoretical asymptotic relative efficiency in Example 2.3. These Monte Carlo experiments were performed in Visual Fortran 5.0 with IMSL library 3.0 on Pentium II PC. The simulation results are summarized in Table 3.1 and Table 3.2, and show that the empirical results are almost equal to the theoretical values, as expected. This fact appears clear as the sample size n increases. Therefore, we can see that LAD estimates are more efficient than LS estimates for small samples as well when the random errors follow the double exponential distribution.

TABLE 3.1 *The empirical* $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ *for normal distributions*

β	σ^2	n							
		50	100	150	200	250	300	400	500
-0.7	1	0.61	0.62	0.63	0.62	0.62	0.63	0.63	0.63
	4	0.61	0.62	0.63	0.63	0.62	0.63	0.64	0.64
	9	0.61	0.63	0.63	0.63	0.63	0.64	0.63	0.63
	16	0.62	0.62	0.62	0.62	0.64	0.62	0.64	0.64
	25	0.60	0.62	0.63	0.63	0.64	0.63	0.64	0.63
-0.5	1	0.59	0.62	0.62	0.62	0.62	0.62	0.63	0.64
	4	0.59	0.62	0.63	0.63	0.63	0.62	0.63	0.63
	9	0.60	0.61	0.62	0.62	0.63	0.62	0.63	0.63
	16	0.59	0.61	0.62	0.62	0.63	0.63	0.63	0.63
	25	0.60	0.62	0.63	0.62	0.64	0.63	0.63	0.63
-0.2	1	0.58	0.62	0.62	0.63	0.63	0.63	0.63	0.63
	4	0.60	0.61	0.62	0.62	0.63	0.63	0.63	0.63
	9	0.59	0.61	0.63	0.63	0.63	0.63	0.64	0.64
	16	0.59	0.61	0.62	0.61	0.63	0.63	0.64	0.63
	25	0.59	0.62	0.62	0.62	0.63	0.63	0.62	0.63
0.2	1	0.60	0.61	0.61	0.62	0.61	0.64	0.64	0.63
	4	0.60	0.60	0.62	0.62	0.63	0.62	0.64	0.63
	9	0.59	0.61	0.62	0.62	0.61	0.62	0.63	0.63
	16	0.60	0.61	0.63	0.62	0.63	0.64	0.63	0.63
	25	0.59	0.61	0.62	0.62	0.62	0.64	0.64	0.63
0.5	1	0.58	0.61	0.62	0.63	0.62	0.63	0.62	0.63
	4	0.59	0.62	0.62	0.63	0.62	0.63	0.63	0.63
	9	0.59	0.61	0.62	0.62	0.63	0.63	0.63	0.62
	16	0.59	0.62	0.62	0.63	0.63	0.62	0.64	0.64
	25	0.60	0.62	0.62	0.62	0.63	0.62	0.62	0.63
0.7	1	0.61	0.61	0.63	0.62	0.63	0.63	0.64	0.63
	4	0.61	0.62	0.63	0.63	0.63	0.62	0.63	0.64
	9	0.61	0.62	0.63	0.62	0.63	0.63	0.63	0.63
	16	0.61	0.63	0.64	0.62	0.63	0.63	0.62	0.63
	25	0.60	0.63	0.63	0.63	0.63	0.63	0.63	0.63

TABLE 3.2 *The empirical $ARE(\hat{\beta}_{LAD}, \hat{\beta}_{LS})$ for double exponential distributions*

β	σ^2	*n*							
		50	*100*	*150*	*200*	*250*	*300*	*400*	*500*
-0.7	1	1.76	1.86	1.91	1.95	1.92	1.94	1.95	2.00
	4	1.77	1.86	1.92	1.91	1.93	1.97	1.93	1.95
	9	1.80	1.86	1.88	1.91	1.96	2.01	2.01	1.94
	16	1.75	1.82	1.88	1.91	1.93	1.92	1.96	1.96
	25	1.78	1.91	1.89	1.93	1.92	1.97	1.94	1.95
-0.5	1	1.73	1.85	1.87	1.91	1.92	1.95	1.99	1.97
	4	1.71	1.84	1.88	1.93	1.94	1.95	1.96	1.97
	9	1.73	1.83	1.89	1.90	1.91	1.96	1.98	1.98
	16	1.70	1.80	1.87	1.92	1.90	1.95	1.95	1.97
	25	1.69	1.84	1.91	1.93	1.93	1.93	1.95	2.00
-0.2	1	1.71	1.82	1.89	1.89	1.96	1.92	1.95	1.98
	4	1.68	1.81	1.87	1.90	1.92	1.96	1.92	1.96
	9	1.67	1.84	1.89	1.92	1.91	1.94	1.96	1.98
	16	1.70	1.82	1.90	1.93	1.95	1.91	1.96	1.95
	25	1.66	1.80	1.90	1.94	1.92	1.93	1.96	1.94
0.2	1	1.69	1.82	1.90	1.91	1.89	1.94	1.92	1.94
	4	1.68	1.84	1.85	1.93	1.92	1.93	1.97	1.96
	9	1.68	1.80	1.88	1.93	1.93	1.95	1.97	1.98
	16	1.66	1.80	1.87	1.90	1.92	1.95	1.95	1.94
	25	1.70	1.84	1.90	1.91	1.91	1.95	2.00	1.95
0.5	1	1.71	1.85	1.86	1.94	1.90	1.94	1.97	1.98
	4	1.70	1.83	1.91	1.91	1.95	1.94	1.94	2.00
	9	1.67	1.85	1.89	1.92	1.93	1.93	1.93	1.96
	16	1.68	1.83	1.92	1.90	1.97	1.92	1.90	1.96
	25	1.71	1.84	1.88	1.90	1.93	1.99	1.95	1.96
0.7	1	1.76	1.87	1.92	1.93	1.96	1.93	1.96	1.96
	4	1.75	1.86	1.89	1.98	1.94	1.95	1.95	1.96
	9	1.74	1.86	1.94	1.92	1.93	1.93	1.96	1.96
	16	1.76	1.88	1.90	1.94	1.95	1.94	1.96	1.96
	25	1.74	1.87	1.87	1.91	1.93	1.93	1.97	1.95

References

Anderson, T. W. (1959). "On asymptotic distributions of estimates of parameters of stochastic difference equations", *The Annals of Mathematical Statistics*, **30**, 676-687.

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

Dunsmuir, W. T. M. and Spencer, N. M. (1991). "Strong consistency and asymptotic normality of *l1* estimates of the autoregressive moving-average model", *Journal of Time Series Analysis*, **12**, 95-104.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed., John Wiley & Sons, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press, New York.

Kang, H. -J. and Shin, K. -I. (1998). "Strong representations for LAD estimators in AR(1) models", *Journal of the Korean Statistical Society*, **27**, 349-358.

Koul, H. L. and Zhu, Zhiwei (1995). "Bahadur-Kiefer representations for GM-estimators in autoregression models", *Stochastic Processes and Their Applications*, **57**, 167-189.