The Korean Communications in Statistics
Vol. 13 No. 3, 2006, pp.745-754

Comparison of Bootstrap Methods for LAD Estimator in AR(1) Model[1)]

Kee-Hoon Kang[2)] and Key-Il Shin[3)]

Abstract

It has been shown that LAD estimates are more efficient than LS estimates when the error distribution is double exponential in AR(1) model. In order to explore the performance of LAD estimates one can use bootstrap approaches. In this paper we consider the efficiencies of bootstrap methods when we apply LAD estimates with highly variable data. Monte Carlo simulation results are given for comparing generalized bootstrap, stationary bootstrap and threshold bootstrap methods.

Keywords : Least square estimator; stationary bootstrap; threshold bootstrap.

1. 서론

분석의 대상이 되는 시계열 자료 가운데 분산이 매우 큰 경우가 최근에는 많아지고 있다. 이러한 자료의 분석은 모수의 추정을 위해 일반적으로 쓰이는 최소제곱추정법과 같은 방법을 사용할 경우 효과적인 결과를 얻을 수 없다. 오차항의 분포가 코시분포 또는 이중지수분포와 같이 꼬리가 두꺼워서 분산이 크게 되는 분포를 따르는 경우에는 최소절대편차(Least Absolute Deviation: LAD) 기준을 사용하는 것이 효율적임이 알려져 있다. Shin et al. (2002)은 오차가 이중지수분포를 따르는 경우 LAD 추정량이 최소제곱추정량(Lease Square Estimator: LSE)에 비해 점근적 효율이 두 배가 됨을 보였다.

평균이 알려진 다음의 AR(1) 모형을 고려하자.

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad t \geq 1 \tag{1}$$

여기서 오차 ε_t는 $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$인 임의의 분포 F를 따르는 서로 독립인 확률변수라고 가정하자. 만약 (1)이 정상성(stationarity)을 만족한다면 $|\phi| < 1$이고

---

1) The first author was supported by grant No. (R01-2006-000-11061-0, 2006) from the Basic Research Program of the Korea Science & Engineering Foundation and the second author was supported by the research fund 2006 of Hankuk University of Foreign Studies.

2) Associate Professor, Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Korea.

3) Professor, Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Korea.
Correspondence : keyshin@hufs.ac.kr

$\sigma^2 < \infty$이 될 것이다. 정상 시계열인 경우 (1)에서 모수 ϕ의 최소제곱추정량은 다음과 같다.

$$\hat{\phi}_{LSE} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2} \tag{2}$$

최소절대편차 기준을 사용하여 추정된 모수를 $\hat{\phi}_{LAD}$라 하면 다음의 식 (3)을 만족하게 된다.

$$\sum_{t=2}^{n} |X_t - \hat{\phi}_{LAD} X_{t-1}| = \infty_{\phi} \sum_{t=2}^{n} |X_t - \phi X_{t-1}| \tag{3}$$

앞의 (2)와 (3)에 제시된 두 추정량 $\hat{\phi}_{LAD}$와 $\hat{\phi}_{LSE}$의 효율성은 오차항의 분포 F에 의해 결정된다. 예를 들어, 오차의 분포가 정규분포를 따르면 LAD 추정량의 효율은 점근적으로 최소제곱추정량의 64%에 지나지 않는다. 반면에 오차의 분포가 이중지수분포를 따를 경우 LAD 추정량의 효율은 점근적으로 최소제곱추정량에 비해 200%에 달한다. 이에 관한 자세한 내용은 Shin et al. (2002)을 살펴보기 바란다. 정규분포에서는 최소제곱추정량이 최대가능도추정량(Maximum Likelihood Estimator: MLE)이 되고, 이중지수분포에서는 LAD 추정량이 최대가능도추정량이 되므로 이는 충분히 예상할 수 있는 결과이다.

모형 (1)에서 평균을 μ라 할 때, 다음의 AR(1) 모형을 살펴보자.

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \varepsilon_t, \quad t \geq 1$$

여기서 $\phi_0 = \mu(1-\phi_1)$라 하면

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t, \quad t \geqq 1 \tag{4}$$

이 된다. 이제 모형 (4)에서 모수를 LAD를 이용하여 추정하는 식은

$$\sum_{t=2}^{n} |X_t - \hat{\phi}_{0,LAD} - \hat{\phi}_{1,LAD} X_{t-1}| = \infty_{\phi_0, \phi_1} \sum_{t=2}^{n} |X_t - \phi_0 - \phi_1 X_{t-1}| \tag{5}$$

와 같이 표현되고 여기서 $\hat{\phi}_{0,LAD}$, $\hat{\phi}_{1,LAD}$는 각각 모수 ϕ_0, ϕ_1의 LAD 추정량을 의미한다.

시계열 자료에서 오차의 분산이 큰 경우를 흔히 접할 수 있는데 이러한 경우에는 LAD 추정량을 사용하는 것이 더욱 효율적이다. LAD는 최소제곱추정량보다 먼저 제안된 방법이지만 계산상의 어려움으로 과거에는 그 사용이 제한적이었다. 그러나 컴퓨터의 발달은 이러한 문제를 해결해 주고 있다. 최근의 대부분 통계 패키지는 LAD 추정량을 계산해 주기 때문에 아무런 문제가 되지 않고 있다. 시계열 자료 분석에서 LAD 추정량의 사용은 그 형태를 명확하게 표현할 수 없기 때문에 분산 계산이나 신뢰구간을 구하는 측면에서 매우 복잡해진다. 따라서 LAD 추정량의 분산을 계산하는 것은 다른 방법을 사용해야 한다.

이러한 문제를 해결할 수 있는 한 가지 방법이 붓스트랩이며 시계열 분석에서 사용되는 붓스트랩 방법은 여러 가지가 있다. 시계열 자료 분석에서 사용되는 붓스트랩

방법에 관한 최근 연구들로는 Li and Maddala (1996), Cao (1999), Berkowitz and Kilian (2000)과 Härdle et al. (2003) 등이 있다.

본 논문에서는 오차가 정규분포, 코시분포 그리고 이중지수분포를 따르는 AR(1) 모형에서 LAD 추정량을 사용할 때 기존에 사용되고 있는 여러 붓스트랩 방법을 비교하여 효율적인 붓스트랩 방법이 어떤 것인지 모의실험을 통하여 살펴보았다. 본 논문에서 비교될 붓스트랩 방법들은 2장에서 간단히 설명하였고 3장에서는 평균이 있는 모형인 (4)식에 대한 모의실험을 통하여 여러 붓스트랩 방법의 효율성을 살펴보았고 4장에서 간단히 결론을 요약하였다.

2. 붓스트랩 방법

이 절에서는 본 논문에서 고려하는 붓스트랩 방법들에 관하여 살펴보자. 시계열 자료 분석에서 사용되는 붓스트랩 방법은 다른 분야와 마찬가지로 여러 가지가 있는데 이중에서 일반적인 붓스트랩(General Bootstrap: GB)은 주어진 자료에서 모형을 식별한 후 이 모형을 이용하여 모수를 추정하고, 이를 이용하여 잔차를 생성한 후 이 잔차를 이용하는 방법이다. 이 방법은 간단하여 사용하기에 편한 반면에 단점으로는 먼저 모형을 식별해야 한다는 것이다. 다음으로 사용되는 방법은 이동블럭 붓스트랩(Moving Block Bootstrap: MBB)이다. MBB 방법은 먼저 블럭의 크기를 정하고 이 크기를 이용하여 블럭을 생성한다. 생성된 블럭을 복원 추출함으로써 붓스트랩 표본이 이루어진다. 이 방법은 고정된 블럭의 크기를 결정하는 것이 문제점으로 대두되고 있다. 이에 관한 내용은 Efron and Tibshirani (1993)의 8.6절을 참조하기 바란다. 이 문제를 해결하는 방법으로 블럭의 크기를 기하분포에 의해 생성된 난수로 랜덤하게 결정하는 방법이 정상 붓스트랩(Stationary Bootstrap: SB)이다. 이에 관한 내용은 Politis and Romano (1994)를 참조하기 바란다. 최근 Park et al. (2001)이 제안한 방법인 문지방 붓스트랩(Threshold Bootstrap: TB)은 Block의 크기를 기하분포를 따르는 변수를 이용하지 않고 어떤 문지방(Threshold)이라는 기준에 의해 결정하게 된다. 이 방법은 블럭의 크기를 결정하는 다른 방법에 비해 더 효과적인 것으로 알려져 있다. 이제 위에서 언급한 네 가지 붓스트랩 방법 중에서 MBB는 효율이 많이 떨어지는 것으로 알려져 있으므로 본 비교에서는 제외하기로 한다. 즉 GB, SB 그리고 TB 방법을 비교하고자 한다. 이제 GB와 SB 그리고 TB에 관한 내용을 단계별로 간단히 살펴보기로 하자.

2.1 General Bootstrap(GB)

먼저 식 (4)의 모형에 있는 모수를 OLS와 LAD를 이용하여 추정하고 이를 이용하여 다음과 같이 잔차 e_t^G를 구한다.

$$e_t^G = X_t - \hat{\phi}_0 - \hat{\phi}_1 X_{t-1}, \quad t \geq 1$$

구해진 잔차 e_t^G를 복원 추출하여 $e_1^{G*}, e_2^{G*}, \cdots, e_n^{G*}$를 얻고, 이를 이용하여 다음의 식과

같이 붓스트랩 표본을 만든다.

$$X_t^{G*} = \hat{\phi}_0 + \hat{\phi}_1 X_{t-1}^{G*} + e_t^{G*}, \quad t \geq 1$$

여기서 $X_0^{G*} = \hat{\mu}$이다. 이렇게 만들어진 시계열 $X_1^{G*}, X_2^{G*}, \cdots, X_n^{G*}$이 붓스트랩 표본이 되며 이 표본으로부터 LSE와 LAD를 이용하여 붓스트랩 추정량을 구하고 평균과 분산을 구한다.

2.2 Stationary Bootstrap(SB)

SB는 MBB가 정상성을 만족하지 못하는 단점을 극복하기 위해 Politis and Romano (1994)에 의해 제안된 것으로 다음과 같이 구해진다.

[Step 1] $[0,1]$에서 임의의 값 p를 정한다. 본 논문에서는 $p = 0.1$을 이용한다.

[Step 2] 먼저 기하분포(geometric distribution) $f(x) = \Pr X = x = (1-p)^{m-1} p$을 이용하여 랜덤 넘버 $L_1, L_2, \cdots$ 을 생성한다.

[Step 3] $1, 2, \cdots, n$에서 이산형 균일분포(Discrete Uniform)를 이용하여 랜덤 넘버 $I_1, I_2, \cdots$ 를 생성한다.

[Step 4] $B_{i,b} = X_i, X_{i+1}, \cdots, X_{i+b-1}$이라 하자. 즉, $B_{i,b}$는 X_i에서 시작하여 b개를 모은 블록을 나타낸다. [Step 2]와 [Step 3]에서 얻어진 L_i와 I_i를 이용하여 B_{I_i, L_i}를 구한다. 예를 들어, $I_1 = 7$, $L_1 = 5$이면 $X_7, X_8, \cdots, X_{11}$이 첫 번째 블럭이 된다. 다음으로, $I_2 = 30$, $L_2 = 15$이면 $X_{30}, X_{31}, \cdots, X_{44}$가 두 번째 블럭이 된다. 이렇게 생성된 블록들 B_{I_1, L_1}, B_{I_2, L_2}, B_{I_3, L_3} 등을 연결하여 시계열을 만든다. 자료수가 n을 넘게 되면 자르고 버린다.

[Step 5] 만들어진 시계열 $X_1^{S*}, X_2^{S*}, \cdots, X_n^{S*}$이 붓스트랩 표본이 되며 LSE와 LAD를 이용하여 추정치를 구하고 이것들로부터 평균과 분산을 구한다.

2.3 Threshold Bootstrap(TB)

TB는 Kim et al. (1993)에 의해 자기상관이 있는 이진 데이터의 분석을 위한 이항 붓스트랩(binary bootstrap)의 확장으로 제안되었는데, 다음의 절차에 따라 적용할 수 있다.

[Step 1] 먼저 관측된 자료 $X_1, X_2, \cdots, X_n$에서 문지방 수준(threshold level)을 정하고 정해진 값에 의해 순환기(cycle)를 구한다. 예를 들어, 문지방 값(threshold value)이 자료의 평균을 기준으로 정해졌을 경우에는 평균보다 작은 자료에서 평균보다 큰 자료로 바뀔 때까지를 하나의 순환기(cycle)로 정할 수 있다.

[Step 2] 한 개 또는 여러 개의 순환기를 하나로 묶어서 하나의 덩어리(chunk)를 만든다. 다른 붓스트랩 방법에서는 블록(block)이라는 표현을 쓰나 TB에서는

덩어리(chunk)라는 표현을 사용한다.

[Step 3] 덩어리를 복원 추출한 후 추출된 덩어리들을 뽑힌 순서로 연결하여 하나의 시계열 자료로 만들고 만약 만들어진 자료가 원 자료의 수 보다 많아지면 끝 부분을 잘라 같은 길이가 되도록 한다.

[Step 4] 만들어진 시계열 $X_1^{T*}, X_2^{T*}, \cdots, X_n^{T*}$이 붓스트랩 표본이 되며 LSE와 LAD를 이용하여 추정치를 구하고 평균과 분산을 구한다.

3. 모의실험

이 절에서는 2절에서 살펴본 몇 가지 붓스트랩 방법의 효율을 비교하기 위한 모의실험 수행 결과를 살펴보고자 한다. 절편이 없는 AR(1) 모형과 절편이 있는 AR(1) 모형에서의 LAD 결과를 모두 비교해 보았다. 모의실험을 위하여 고려한 모형은 식 (4)에서 $\phi_0 = 0, 1$, $\phi_1 = 0.5, 0.8$인 경우의 조합인 네 가지 경우를 대상으로 하였다. 사용한 자료의 수는 $n = 100, 400$ 두 가지이다. 오차항 ε_t의 분포로는 표준정규분포 $N(0,1)$, 코시분포 $C(0,1)$과 이중지수분포 $DE(0,1)$ 등을 사용하였다. 얻어진 각 표본에 대하여 자료에서 250번의 붓스트랩 반복이 이루어졌으며 두 개의 모수, ϕ_0, ϕ_1를 LSE와 LAD를 사용하여 추정하였다. 250번의 붓스트랩 반복에서 얻어진 추정치들의 평균을 붓스트랩 추정량으로 하였다. 이러한 절차는 500번 반복되었으며 붓스트랩 추정량의 분산을 구하였다. 이에 관한 결과로 각 붓스트랩 방법에 기초한 추정량이 다음의 <표 1>부터 <표 4>에 제시되어 있다.

<표 1> 모형에 따라 붓스트랩에 기초한 ϕ_1의 추정치 ($n = 100$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	LAD	LSE	LAD	LSE	LAD	LSE
$\phi_0 = 0.0$ $\phi_1 = 0.5$	GB	0.455	0.450	0.495	0.459	0.472	0.455
	SB	0.417	0.408	0.475	0.419	0.442	0.414
	TB	0.448	0.441	0.491	0.464	0.467	0.453
$\phi_0 = 1.0$ $\phi_1 = 0.5$	GB	0.453	0.449	0.496	0.461	0.475	0.451
	SB	0.417	0.406	0.475	0.419	0.444	0.411
	TB	0.448	0.438	0.492	0.464	0.467	0.446
$\phi_0 = 0.0$ $\phi_1 = 0.8$	GB	0.735	0.733	0.793	0.742	0.759	0.731
	SB	0.698	0.664	0.777	0.679	0.726	0.665
	TB	0.731	0.729	0.788	0.741	0.755	0.730
$\phi_0 = 1.0$ $\phi_1 = 0.8$	GB	0.747	0.745	0.793	0.743	0.764	0.739
	SB	0.714	0.673	0.777	0.680	0.735	0.671
	TB	0.741	0.731	0.788	0.742	0.758	0.729

<표 1>은 $n = 100$인 경우에 네 가지 모수값의 모형에서 ϕ_1의 추정치를 계산한 결

과이다. 오차가 정규분포에서 생성된 경우 LAD와 LSE 간에 큰 차이를 보이지 않고 있다. 그러나 코시 또는 이중지수분포인 경우에서는 LAD가 LSE에 비해 훨씬 더 실제 모수값에 가깝게 추정되는 것을 확인할 수 있다. 따라서 오차가 어떤 분포에서 생성되었든지 편의(bias)를 기준으로 하였을 경우 LAD가 더 우수한 것을 알 수 있다. 또한 붓스트랩 방법들 간의 비교에서는 자료가 어떤 분포에 의해 생성되었는지 상관없이 GB와 TB간에는 별 차이가 없는 것으로 판단할 수 있으며 SB가 가장 좋지 않게 나타남을 확인할 수 있다.

<표 2> 모형에 따라 붓스트랩에 기초한 ϕ_0의 추정치 ($n=100$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	LAD	LSE	LAD	LSE	LAD	LSE
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.004	-0.001	0.000	0.839	0.000	0.004
	SB	0.002	-0.002	0.001	5.320	0.001	0.006
	TB	0.002	-0.002	-0.004	-0.329	0.000	0.005
$\phi_0=1.0$ $\phi_1=0.5$	GB	1.094	1.103	0.999	-6.726	1.051	1.090
	SB	1.159	1.179	1.039	-5.062	1.106	1.159
	TB	1.101	1.115	1.001	-7.466	1.064	1.091
$\phi_0=0.0$ $\phi_1=0.8$	GB	-0.003	-0.006	0.001	-9.903	0.000	-0.005
	SB	-0.007	-0.010	0.003	-11.108	0.000	-0.007
	TB	-0.006	-0.007	-0.003	-6.927	0.000	-0.005
$\phi_0=1.0$ $\phi_1=0.8$	GB	1.261	1.266	1.032	-8.633	1.162	1.281
	SB	1.407	1.579	1.103	-9.568	1.289	1.573
	TB	1.286	1.305	1.052	-5.668	1.189	1.302

<표 3> 모형에 따라 붓스트랩에 기초한 ϕ_1의 추정치 ($n=400$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	LAD	LSE	LAD	LSE	LAD	LSE
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.492	0.490	0.499	0.489	0.493	0.487
	SB	0.450	0.441	0.490	0.441	0.465	0.439
	TB	0.489	0.484	0.499	0.492	0.491	0.485
$\phi_0=1.0$ $\phi_1=0.5$	GB	0.487	0.488	0.499	0.491	0.492	0.486
	SB	0.449	0.440	0.486	0.439	0.465	0.438
	TB	0.485	0.480	0.499	0.493	0.492	0.485
$\phi_0=0.0$ $\phi_1=0.8$	GB	0.785	0.785	0.799	0.784	0.791	0.782
	SB	0.745	0.707	0.793	0.694	0.765	0.707
	TB	0.783	0.781	0.799	0.786	0.790	0.781
$\phi_0=1.0$ $\phi_1=0.8$	GB	0.783	0.785	0.799	0.783	0.790	0.784
	SB	0.744	0.705	0.792	0.699	0.764	0.706
	TB	0.782	0.779	0.798	0.785	0.789	0.782

<표 2>는 ϕ_0의 추정치에 관한 결과이다. <표 1>에서와 마찬가지로 정규분포의 경우에는 두 가지 추정치에 별 차이가 없으나 이중지수분포의 경우에는 LAD가 조금 좋고, 코시분포의 경우에는 LSE가 아주 안좋게 나타나고 있다. 붓스트랩 방법들 간의 비교에서도 <표 1>에서와 비슷한 결론을 얻는다.

<표 4> 모형에 따라 붓스트랩에 기초한 ϕ_0의 추정치 ($n=400$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	LAD	LSE	LAD	LSE	LAD	LSE
$\phi_0=0.0$ $\phi_1=0.5$	GB	-0.001	-0.003	-0.002	0.267	-0.007	-0.013
	SB	-0.004	-0.004	0.000	0.550	-0.006	-0.014
	TB	-0.002	-0.003	-0.002	0.397	-0.008	-0.013
$\phi_0=1.0$ $\phi_1=0.5$	GB	1.025	1.024	1.000	-1.381	1.015	1.026
	SB	1.098	1.118	1.028	-1.851	1.069	1.119
	TB	1.027	1.037	1.001	-1.381	1.015	1.025
$\phi_0=0.0$ $\phi_1=0.8$	GB	-0.004	-0.001	0.002	-0.744	0.003	-0.002
	SB	-0.004	-0.001	0.003	-1.234	0.002	-0.002
	TB	-0.004	-0.001	-0.001	-0.884	0.003	-0.001
$\phi_0=1.0$ $\phi_1=0.8$	GB	1.082	1.074	1.008	0.793	1.054	1.086
	SB	1.270	1.462	1.046	1.003	1.181	1.463
	TB	1.087	1.095	1.012	0.605	1.057	1.089

<표 5> 각 추정량의 분산비교($Var(\hat{\phi}_{LSE})/Var(\hat{\phi}_{LAD})$, $n=100$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.6517	0.7198	8.3781	3596475.93	1.4012	1.8552
	SB	0.7128	0.7081	4.0138	1611025.62	1.2923	1.7932
	TB	0.7440	0.7067	5.0998	2386510.38	1.4705	1.7974
$\phi_0=1.0$ $\phi_1=0.5$	GB	0.6581	0.7381	7.9752	1508570.80	1.4614	1.7068
	SB	0.6915	0.7661	2.8981	699310.09	1.2773	1.6177
	TB	0.7513	0.8239	4.5531	1407475.45	1.4759	1.7443
$\phi_0=0.0$ $\phi_1=0.8$	GB	0.6968	0.8504	8.7732	1558669.69	1.3879	1.9817
	SB	0.7524	1.0584	40.4338	1239388.55	1.4604	2.2670
	TB	0.7854	0.8198	2.8756	476107.39	1.3644	1.8894
$\phi_0=1.0$ $\phi_1=0.8$	GB	0.6173	0.6359	8.6495	1278337.88	1.3944	1.5569
	SB	0.7686	0.8170	40.2348	723167.23	1.4144	1.6384
	TB	0.7848	0.7692	2.8987	322527.57	1.4631	1.5485

<표 6> 각 추정량의 분산비교($Var(\hat{\phi}_{LSE})\,/\,Var(\hat{\phi}_{LAD})$, $n=400$)

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.7361	0.7240	42.0106	21785.79	1.5320	2.3376
	SB	0.7685	0.7676	6.1973	28213.43	1.5425	2.2822
	TB	0.8168	0.7371	27.5776	20658.37	1.6424	2.3144
$\phi_0=1.0$ $\phi_1=0.5$	GB	0.6432	0.6047	47.9415	21053.65	1.8245	2.0511
	SB	0.6953	0.6780	2.6299	14918.37	1.6354	1.9066
	TB	0.7744	0.7115	25.0069	21801.29	1.8394	2.0383
$\phi_0=0.0$ $\phi_1=0.8$	GB	0.6597	0.7557	47.1885	14637.79	1.7130	1.7070
	SB	0.8627	1.0491	6.8763	25119.59	1.8960	2.4178
	TB	0.7591	0.7556	39.7001	16186.38	1.7316	1.7269
$\phi_0=1.0$ $\phi_1=0.8$	GB	0.6800	0.7294	75.6893	10269.17	1.8015	1.8297
	SB	0.8979	0.9867	10.2269	4738.10	2.0477	2.1969
	TB	0.7809	0.8443	2.1641	9228.19	1.9454	1.8976

다음으로 $n=400$일 때 ϕ_1. ϕ_0의 추정 결과에 해당하는 <표 3>과 <표 4>를 살펴보자. 추정치의 편의가 좀 작아진 것을 제외하면 ϕ_1에 대한 전체적인 판단은 <표 1>의 경우와 같다. <표 4>의 결과도 <표 2>의 경우와 유사한 결론을 내리게 한다.

절편이 없는 경우에 이론적으로는 정규분포에서 LSE에 비해 LAD의 효율이 64%이고 이중지수분포인 경우에는 반대로 효율이 200%인 것이 알려져 있다. 추정량의 분산을 비교한 <표 5>의 결과에서 보면 정규분포에서 얻어진 자료에서는 LSE가 LAD에 비해 효율적인 것을 확인할 수 있다. 반대로, 자료가 이중지수분포에서 얻어진 경우에는 LAD가 LSE에 비해 효율적이며 <표 6>과 비교해 보면 n이 커지면서 더욱 좋아짐을 알 수 있다. 코시분포의 경우에는 LSE의 효율이 LAD에 비해 매우 떨어짐을 알 수 있고 이러한 사항은 ϕ_1은 물론 특히 ϕ_0의 추정에서는 심각하게 두드러진다.

이제 붓스트랩 방법 GB, SB 그리고 TB 간의 효율을 LAD 추정량에 대해서 비교해 보기로 하자. <표 7>과 <표 8>은 붓스트랩 LAD 추정량의 분산 값을 제시하고 있다. 표본의 크기에 따라, 오차항의 분포에 따라 그리고 모수에 따라 전체적으로 보면 GB와 TB가 유사하거나 상대적으로 좋을 때가 비슷하게 있는 반면에 SB의 경우는 좀 떨어지고 있음을 확인할 수 있다. GB의 경우 실제 모형인 AR(1)을 이용하여 얻어진 결과이므로 다른 비교 대상 가운데 효율이 나쁘지 않게 나오는 것이 당연할 것이다. 모형에 대한 설정이 잘못 될 경우를 고려한다면 TB를 사용하는 것이 권장될 수 있을 것이다. 특히, 꼬리가 두꺼운 코시분포의 경우에는 SB가 두드러지게 나쁘게 나타나고 있음을 유의해야 할 것이다.

본 연구에서 표본의 크기가 아주 작은 경우는 다루지 않았으나 저자에게 확인할 수 있는 $n=50$인 경우의 모의실험 결과를 보면 앞에서와 동일하게 결론을 내릴 수 있

다. 다만, 표본의 크기가 아주 작아질 때는 SB의 경우 [Step 1]에서 사용되는 p값을 조금 크게 할 필요가 있다. 블록의 평균 길이가 $1/p$에 해당되므로 p값이 작으면 표본의 크기에 비해 상대적으로 큰 블록의 길이가 생성되어 이에 따른 결과로 SB가 많이 안좋게 나타나기 때문이다.

<표 7> $n=100$인 경우 붓스트랩 LAD 추정량의 분산

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.0115	0.0170	0.0005	0.0262	0.0053	0.0136
	SB	0.0108	0.0209	0.0017	0.0335	0.0059	0.0161
	TB	0.0106	0.0188	0.0007	0.0299	0.0051	0.0140
$\phi_0=1.0$ $\phi_1=0.5$	GB	0.0113	0.0543	0.0005	0.0283	0.0054	0.0313
	SB	0.0113	0.0562	0.0022	0.0430	0.0062	0.0352
	TB	0.0107	0.0530	0.0007	0.0326	0.0054	0.0314
$\phi_0=0.0$ $\phi_1=0.8$	GB	0.0063	0.0214	0.0003	0.0291	0.0031	0.0187
	SB	0.0075	0.0269	0.0016	0.0481	0.0043	0.0265
	TB	0.0065	0.0237	0.0013	0.0439	0.0034	0.0210
$\phi_0=1.0$ $\phi_1=0.8$	GB	0.0051	0.1449	0.0003	0.0355	0.0030	0.0848
	SB	0.0059	0.1698	0.0016	0.0832	0.0041	0.1180
	TB	0.0052	0.1545	0.0013	0.0648	0.0032	0.0946

<표 8> $n=400$인 경우 붓스트랩 LAD 추정량의 분산

오차항 분포		$N(0,1)$		$C(0,1)$		$DE(0,1)$	
계수	방법	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$
$\phi_0=0.0$ $\phi_1=0.5$	GB	0.0026	0.0036	0.0000	0.0055	0.0012	0.0029
	SB	0.0025	0.0041	0.0007	0.0069	0.0014	0.0035
	TB	0.0024	0.0037	0.0000	0.0054	0.0012	0.0029
$\phi_0=1.0$ $\phi_1=0.5$	GB	0.0024	0.0134	0.0000	0.0057	0.0010	0.0066
	SB	0.0023	0.0133	0.0023	0.0131	0.0012	0.0080
	TB	0.0021	0.0120	0.0000	0.0057	0.0010	0.0066
$\phi_0=0.0$ $\phi_1=0.8$	GB	0.0014	0.0043	0.0000	0.0065	0.0005	0.0032
	SB	0.0016	0.0057	0.0010	0.0085	0.0007	0.0042
	TB	0.0013	0.0046	0.0000	0.0071	0.0006	0.0033
$\phi_0=1.0$ $\phi_1=0.8$	GB	0.0012	0.0303	0.0000	0.0064	0.0005	0.0150
	SB	0.0014	0.0365	0.0006	0.0254	0.0006	0.0189
	TB	0.0012	0.0308	0.0007	0.0121	0.0005	0.0145

4. 결론

3장 모의실험의 결과를 살펴보면 추정량은 오차가 정규분포인 경우 LSE가 우수하

고 다른 두 꼬리가 두꺼운 분포에서는 LAD가 우수한 것을 확인 할 수 있었다. 붓스트랩 방법의 비교에서는 일반적으로 GB와 TB가 우수한 것을 알 수 있다. 그러나 GB를 사용하기 위해서는 먼저 모형 식별이 이루어 져야하고 얻어진 모형을 이용하여 잔차를 구한 후 사용해야 한다. 본 논문에서는 실제 모형(true model) 즉, 자료가 AR(1) 모형에서 생성되었고 또한 AR(1) 모형을 이용하여 잔차를 구하였다. 그러나 실제 자료를 가지고 분석하는 경우에는 모형이 어떤 것인지 추정해야 하기 때문에 모형의 설정에 따른 문제가 있을 수 있다. 이에 반하여 SB와 TB는 모형 식별 과정 없이 사용할 수 있는 방법이며, 두 붓스트랩 방법의 비교 결과로는 SB보다 TB의 사용을 권장할 수 있으리라 판단된다.

참고문헌

[1] Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometrics Reviews*, Vol. 19, 1-48.

[2] Cao, R. (1999). An overview of bootstrap mothods for estimating and predict-ing in time series. *Test*, Vol. 8, 95-116.

[3] Efron, B and Tibshirani, R.J. (1993). *An introduction to the Bootstrap*, Chapman & Hall, New York.

[4] Härdle, W., Horowitz, J. and Kreiss, J.P. (2003). Bootstrap methods for time series. *International Statistical Review*, Vol. 71, 435-460.

[5] Kim, Y., Willemain, T., Haddock, J. and Runger, G. (1993). The threshold bootstrap: A new approach to simulation output analysis. In: Evans, G.W., Mollaghasemi, M., Russel, E.C., Biles, W.E. (Eds.), *Proceedings: 1993 Winter Simulation Conference*, 498-502.

[6] Li, H. and Maddala, G. (1996). Bootstrapping time series models. *Econometric Theory*, Vol. 15, 115-195.

[7] Shin, K.-I., Kang, H.-J. and Sim, H. (2002). A new proof of efficient of LAD estimation in an autoregressive process. *Journal of the Korean Statistical Society*, Vol. 31, 121-128.

[8] Park, D.S., Kim, Y.B., Shin, K.-I. and Willemain, T.R. (2001). Simulation output analysis using the threshold bootstrap. *European Journal of Operational Research*, Vol. 134, 17-28.

[9] Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, Vol. 89, 1303-1313.

[Received August 2006, Accepted October 2006]