*The Korean Journal of Applied Statistics* (2019) DOI: https://doi.org/10.5351/KJAS.2019.32.1.099
**32**(1), 99–109

# Robust confidence interval for random coefficient autoregressive model with bootstrap method

Na Rae Jo[a] · Do Sang Lim[b] · Sung Duck Lee[a,1]

[a]Department of Information and Statistics, Chungbuk National University;
[b]Division of Chronic Disease Control Prevention, Korea Centers for Disease Control & Prevention

(Received December 3, 2018; Revised December 11, 2018; Accepted December 11, 2018)

**Abstract**

We compared the confidence intervals of estimators using various bootstrap methods for a Random Coefficient Autoregressive(RCA) model. We consider a Quasi score estimator and M-Quasi score estimator using Huber, Tukey, Andrew and Hempel functions as bounded functions, that do not have required assumption of distribution. A standard bootstrap method, percentile bootstrap method, studentized bootstrap method and hybrid bootstrap method were proposed for the estimations, respectively. In a simulation study, we compared the asymptotic confidence intervals of the Quasi score and M-Quasi score estimator with the bootstrap confidence intervals using the four bootstrap methods when the underlying distribution of the error term of the RCA model follows the normal distribution, the contaminated normal distribution and the double exponential distribution, respectively.

Keywords: random coefficient autoregressive model, standard bootstrap method, percentile bootstrap method, studentized bootstrap method, hybrid bootstrap method

## 1. 서론

시계열 자료가 강한 비대칭성을 가지고 있거나 시간의 가역성을 만족하지 않는 자료들에 대해서는 선형 시계열 모형의 적합은 적당하지 않다고 알려져 있다 (Tong, 1990). 비선형 시계열 모형에는 중선형 모형, 확률계수 자기회귀 모형(random coefficient autoregressive; RCA), 조건부 이분산 자기회귀 모형 등이 있는데, 이러한 모형들은 일반적으로 환율변화 또는 옵션의 가격분석 등 경제 지표의 시계열 예측모형을 위해서 연구되고 있다. 실상에서 접할 수 있는 시계열 자료는 분포에 대한 가정을 할 수 없는 경우가 많이 발생하고, 더군다나 시계열 자료 중 강한 비대칭성을 갖거나 시간의 가역성을 만족하지 않는 비선형 시계열 자료에서는 최우추정법과 같은 일반적인 모수추정법을 사용하기 어렵다. 이에 Wedderburn (1974)은 분포의 가정 없이 단지 평균과 분산만을 알고 있는 경우에도 최우추정법과 같은 성질을 얻을 수 있는 Quasi score 추정함수를 이용한 추정법을 제안하였다. 또한 시계열 자료가 정규성의 가정을 만족하지 않을 때 정규성 이론에 근거한 추정법들은 더 많은 오류의 가능성을 내포하고 있

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A3B03932557).
[1]Corresponding author: Department of Information and Statistics, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju, Chungbuk 28644, Korea. E-mail: sdlee@chungbuk.ac.kr

다. Huber (1981)는 이러한 이상치의 영향을 감지할 수 있도록 유계함수(bounded function)를 적용하여 로버스트한 추정량을 얻을 수 있는 M-Quasi score 추정법을 제안하였다. 한편, 컴퓨터의 계산능력의 발전과 함께 이러한 추정량을 바탕으로 하여 여러가지 붓스트랩 예측구간을 구할 수 있다. 모수적 예측구간은 Beran (1990)에 의해서 처음 논의되었고, Thombs와 Scbucany (1990)에 의해 AR 과정을 위한 비모수적 붓스트랩 예측구간이 연구되었다. Lee와 Kim (2004)은 ARCH 모형의 Quasi score 추정량과 Huber 유계함수를 이용한 M-Quasi score 추정량에 대하여 PB 붓스트랩 신뢰구간을 모의실험 하였다.

본 논문에서는 시간의 가역성을 만족하지 않으면서 이상치가 존재하는 비선형 시계열 자료에 대한 모형인 비선형 모형인 확률계수 자기회귀 모형을 다룬다. 자료의 분포를 가정하지 않아도 되는 Quasi score 추정함수를 이용한 추정량과 Huber, Tukey, Andrew, Hempal 4가지 유계함수를 이용한 M-Quasi 추정량을 제시하였다. 또한 Quasi 추정량과 M-Quasi 추정량의 근사적 신뢰구간과 붓스트랩 추정량들로부터 신뢰구간을 구하는 표준 붓스트랩 방법, 백분위수 붓스트랩 방법, 스튜던트화 붓스트랩 방법, 하이브리드 붓스트랩 방법을 이용한 붓스트랩 신뢰구간을 비교하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 일반적인 비선형 시계열 모형 중 확률계수 자기회귀 모형을 설명하고, 3장에서는 Quasi score 추정함수를 이용한 추정법과 로버스트 추정함수인 M-Quasi score 추정함수를 이용한 추정법을 소개하였다. 4장에서는 4가지 대표적인 붓스트랩 방법 및 절차를 소개하였다. 5장에서는 확률계수 자기회귀 모형의 오차항의 분포가 정규분포, 오염된 정규분포, 이중지수분포를 따를 때, 근사적인 방법을 이용한 신뢰구간과 붓스트랩을 이용한 신뢰구간을 모의실험을 실시하여 비교분석하였다. 마지막으로 6장에서는 본 논문의 결론 및 향후 연구에 대하여 언급하였다.

## 2. 확률계수 자기회귀 모형

시계열 $\{X_t,\, t = 1, 2, 3, \ldots\}$가 다음과 같은 $p$차 자기회귀식을 만족하는 비선형 시계열 모형이라 한다.

$$X_t = H\left(\underline{X_{t-1}}, Z_t; \underline{\theta}\right) + \varepsilon_t, \quad t = 1, 2, 3, \ldots, \tag{2.1}$$

여기서 $\varepsilon_t$는 평균이 0이고 분산이 $\sigma_\varepsilon^2$인 독립이고 동일한 분포를 갖는(independent and identically distributed; iid) 확률변수이고, 함수 $H$는 관찰값과 모수로 이루어진 알려진 함수이다. 관찰값 벡터는 $\underline{X_{t-1}} = (X_{t-1}, X_{t-2}, \ldots, X_{t-p})$로 $\theta$는 $p \times 1$ 모수 벡터이다. 한편 $Z_t$는 평균이 0이고 분산이 $\sigma_z^2$인 독립이며 동일한 분포를 갖는 미관측된 $p \times 1$ 확률벡터로서 $\varepsilon_t$와 독립이다. $Z_t$는 과거의 측정치 $\underline{X_{t-1}}$에 의하여 모수 $\theta$와 함수 $H$에 관한 불확실성으로 나타낸다. 조건부 기대치와 조건부 분산은 다음과 같이 정의된다.

$$\mu_t\left(\underline{X_{t-1}}; \theta\right) = E(X_t F_{t-1}), \tag{2.2}$$

$$v_t(\theta) = \mathrm{Var}(X_t F_{t-1}). \tag{2.3}$$

여기서 $F_{t-1}$는 $\sigma$-field로 $F_{t-1} = \sigma(X_{t-1}, X_{t-2}, \ldots, X_{t-p})$이다.

Nicholls와 Quinn (1982)은 일반화된 $p$차 확률계수 자기회귀 모형을 제안하였으며 다음과 같은 식을 갖는다.

$$X_t = (\theta_1 + Z_{t1})X_{t-1} + \cdots + (\theta_t + Z_{tp})X_{t-p} + \varepsilon_t. \tag{2.4}$$

여기서 $\varepsilon_t$는 평균이 0이고 분산이 $\sigma_\varepsilon^2$인 iid 확률변수이며, $Z_{tj}$는 평균이 0이고 분산이 $\sigma_z^2$으로 유한한

iid 확률변수이다. 또한 이 두 확률변수 $Z_{tj}$와 $\varepsilon_t$는 독립이다. $p$차 RCA 모형의 특수한 경우인 1차 RCA 모형은 다음과 같고, RCA(1)이라 한다.

$$X_t = (\theta + Z_t)X_{t-1} + \varepsilon_t, \quad t = 1, 2, 3, \ldots. \tag{2.5}$$

Nicholls와 Quinn (1982)은 RCA(1) 모형에서 정상성을 만족하기 위한 조건이 $\theta^2 + EZ_t^2 < 1$임을 밝혔다. 식 (2.2)와 식 (2.3)에 따른 $X_t$에 대한 조건부 기대치와 분산은 각각 다음과 같다.

$$E(X_t F_{t-1}) = \theta X_{t-1}, \tag{2.6}$$

$$\mathrm{Var}(X_t X_{t-1}) = \sigma_z^2 X_{t-1}^2 + \sigma_\varepsilon^2. \tag{2.7}$$

## 3. RCA(1) 모형의 추정

### 3.1. RCA(1) 모형에서 조건부 최소제곱추정법

일반적으로 조건부 최소제곱추정량은 다음을 만족하는 추정량을 말한다.

$$\min_\theta \sum_{t=1}^{n} \left\{X_t - \mu_t\left(\underline{X_{t-1}}; \theta\right)\right\}^2. \tag{3.1}$$

RCA(1) 모형에서 조건부 최소제곱함수는 다음과 같이 정의한다.

$$\begin{aligned} S_{n1}(\theta) &= \sum_{t=1}^{n} \left(X_t - \mu_t\left(\underline{X_{t-1}}; \theta\right)\right) \frac{\partial \mu_t\left(\underline{X_{t-1}}; \theta\right)}{\partial \theta} \\ &= \sum_{t=1}^{n} \left(X_t - \theta X_{t-1}\right) X_{t-1}. \end{aligned} \tag{3.2}$$

여기서 식 (3.2)의 $S_{n1}(\theta) = 0$을 만족하는 추정량이 $\theta$의 조건부 최소제곱추정량 $\hat{\theta}_{n1}$이다. Kim 등 (2003)은 조건부 최소제곱추정량에 대한 극한분포를 얻기 위해 정규조건을 제시하였으며 [정규조건 1] 하에서 [정리 2.1]을 이용하여 조건부 최소제곱추정량의 극한분포가 다음과 같다는 것을 밝혔다.

$$\sqrt{n}\left(\hat{\theta}_{n1} - \theta\right) \xrightarrow{d} N\left(0, \frac{\sigma_z^2 E\left(X_{t-1}^4\right) - \sigma_\varepsilon^2 E\left(X_{t-1}^2\right)}{\left(E\left(X_{t-1}^2\right)\right)^2}\right). \tag{3.3}$$

### 3.2. RCA(1) 모형에서 Quasi score 추정법

일반적으로 우도함수(likelihood function)를 정의하기 위해서는 관찰값들의 분포형태가 정규분포라는 가정을 하는 것이 일반적이다. 하지만 관찰값들의 분포가 정규분포라는 가정은 추정법에서 강한 제약조건이다. Wedderburn (1974)는 분포를 가정하지 않고 오직 관찰값들의 평균과 분산만이 알려져 있을 경우 최우추정량(maximum likelihood estimate)과 같은 성질을 얻을 수 있는 Quasi score 추정을 제안하였다. Quasi score 추정(Quasi score estimation)은 일반화선형모형에서 분산이 평균의 함수로 정의됨을 가정하고 Quasi score 우도함수를 정의하여 추정하는 방법이다. 확률변수 $X$의 평균이 $\mu$이고 분산이 $V(\mu)$일 때, Quasi score 우도함수 $Q$는 다음과 같이 정의된다.

$$\frac{\partial Q}{\partial \mu} = \frac{X - \mu}{V(\mu)}. \tag{3.4}$$

Quasi score 우도함수는 다음의 성질을 만족한다.

$$E\left(\frac{\partial Q}{\partial \mu}\right)=0,$$
$$E\left(\frac{\partial Q}{\partial \mu}\right)^2=-E\left(\frac{\partial^2 Q}{\partial \mu^2}\right)=\frac{1}{V(\mu)}.$$

RCA(1) 모형에서 Quasi score 함수는 다음과 같이 정의된다.

$$\begin{aligned} S_{n2}(\theta) &= \sum_{t=1}^{n}\left(X_t-\mu_t\left(\underline{X_{t-1}};\theta\right)\right)v_t^{-1}(\theta)\frac{\partial \mu_t(\underline{X_{t-1}};\theta)}{\partial \theta} \\ &= \sum_{t=1}^{n}\frac{\varepsilon_t+Z_tX_{t-1}}{\sigma_\varepsilon^2+\sigma_z^2\varepsilon_{t-1}^2}X_{t-1}. \end{aligned} \tag{3.5}$$

여기서 식 (3.7)의 $S_{n2}(\theta)=0$을 만족하는 추정량을 Quasi score 추정량이라 하는데, RCA(1) 모형에서 Quasi score 추정량 $\hat{\theta}_{n2}$은 다음과 같다.

$$\hat{\theta}_{n2}=\sum_{t=1}^{n}\left(\frac{X_tX_{t-1}}{\sigma_\varepsilon^2+\sigma_z^2\varepsilon_{t-1}^2}\right)\Big/\sum_{t=1}^{n}\left(\frac{X_{t-1}^2}{\sigma_\varepsilon^2+\sigma_z^2\varepsilon_{t-1}^2}\right). \tag{3.6}$$

Kim 등 (2003)은 [정규조건 2] 하에서 [정리 2.2]를 이용하여 Quasi score 추정량의 극한분포가 다음과 같다는 것을 밝혔다.

$$\sqrt{n}\left(\hat{\theta}_{n2}-\theta\right)\xrightarrow{d}N\left(0,1/E\left(\frac{X_{t-1}^2}{\sigma_z^2X_{t-1}^2+\sigma_\varepsilon^2}\right)\right). \tag{3.7}$$

### 3.3. RCA(1) 모형에서 M-Quasi score 추정법

정규성 가정이 만족되지 않을 경우 정규분포에 근거한 추정법은 이론적인 오차보다 더 큰 오류를 내포할 가능성이 있다. 그리고 최우추정법 또는 최소제곱추정법이 이상치의 영향을 많이 받는다. 이러한 문제점을 보완하기 위하여 Huber (1981)는 추정방정식에 이상치를 감지할 수 있도록 유계(bounded) 함수를 적용하여 로버스트(Robust)한 추정량을 얻을 수 있는 M 추정법(maximum likelihood type estimation)을 제안하였다. 로버스트한 추정법으로 많이 이용되고 있는 유계함수로는 Huber (1964)가 제안한 Huber 유계함수, Beaton 등 (1974)가 제안한 Tukey 유계함수, Andrews 등 (1972)가 제안한 Andrews 유계함수와 Hampel 유계함수 등이 있다. 유계함수에서 상수 $k, l, m$을 조율상수(tuning constant)라 한다.

- Huber 유계함수

$$\psi(x)=\begin{cases} x, & |x|\le k,\\ k\,\mathrm{sign}(x), & |x|>k. \end{cases}$$

Huber의 유계함수에서 $k=1.5$는 합리적인 상수로 추천된다.

- Tukey 유계함수

$$\psi(x)=\begin{cases} x\left(1-\dfrac{x^2}{k^2}\right)^2 b, & |x|\le k,\\ 0, & |x|>k. \end{cases}$$

Tukey의 유계함수는 흔히 Tukey의 biweight 함수라고 부른다. Tukey의 biweight 함수에서 조율상

수로는 $k = 6.0$이 합리적이라고 알려져 있다.

- Andrews 유계함수

$$\psi(x) = \begin{cases} \sin\left(\dfrac{x}{k}\right), & |x| \le k\pi, \\ 0, & |x| > k\pi. \end{cases}$$

Andrews의 유계함수에서 조율상수는 $k = 1.5$ 또는 $1.8$이 추천된다.

- Hampel 유계함수

$$\psi(x) = \begin{cases} x, & |x| \le k, \\ k\,\mathrm{sign}(x), & k < |x| \le l, \\ \dfrac{k\,\mathrm{sign}(x)(m - |x|)}{m - l}, & l < |x| \le m, \\ 0, & |x| > m. \end{cases}$$

Hampel의 유계함수는 재감소(redescending)하는 함수이다. 조율상수로는 $k = 1.7$, $l = 3.4$, $m = 8.5$가 추천된다.

이러한 유계함수를 Quasi score 함수에 적용하여 보다 로버스트한 추정함수를 얻기 위하여 Cha 등 (1999)은 M-Quasi score 추정함수 및 극한분포를 정의하였다. 이에 따라 RCA(1) 모형에서 M-Quasi score 함수는 다음과 같다.

$$\begin{aligned} S_{n3}(\theta) &= \sum_{t=1}^{n} \psi_1\left(X_t - \mu_t\left(\underline{X_{t-1}}; \theta\right) v_t^{-1}(\theta)\right) \psi_2\left(\frac{\partial \mu_t(\underline{X_{t-1}}; \theta)}{\partial \theta}\right) \\ &= \sum_{t=1}^{n} \psi_i\left(\frac{\varepsilon_t(\theta)}{\sigma^{2_\varepsilon} + \sigma_z^2 X_{t-1}^2}\right) \psi_2\left(X_{t-1}\right). \end{aligned} \tag{3.8}$$

여기서 $\psi_1$과 $\psi_2$는 유계함수로 $\theta$에 대해 미분가능하며, $\varepsilon_t(\theta) = X_t - \theta X_{t-1}$이다. 식 (3.10)의 $S_{n3}(\theta) = 0$을 만족하는 $\theta$를 M-Quasi score 우도추정량 $\hat{\theta}_{n3}$이라 한다. 또한 M-Quasi score 우도추청량에 대한 극한분포는 다음과 같다.

$$\sqrt{n}\left(\hat{\theta}_{n3} - \theta\right) \xrightarrow{d} N\left(0, \frac{E\left[\psi_1\left(\varepsilon_t(\theta) / \left(X_{t-1}^2 \sigma_z^2 + \sigma_\varepsilon^2\right)\right) \psi_2\left(X_{t-1}\right)\right]^2}{E\left[\frac{\partial \psi_1\left(\varepsilon_t(\theta) / \left(X_{t-1}^2 \sigma_z^2 + \sigma_\varepsilon^2\right)\right)}{\partial \theta} \psi_2\left(X_{t-1}\right)\right]}\right). \tag{3.9}$$

## 4. RCA 모형에서의 붓스트랩 연구

붓스트랩 방법은 Efron (1979)에 의해 처음으로 소개되었는데, 각 관측값들에 동일한 확률 $1/n$을 주어서 구한 경험적 누적 확률분포 $F_n(x)$를 미지의 분포에 대체시켜서 통계량의 성질을 추정하는 기법이다. 원래의 모집단에 대한 분포를 가정하지 않고 자료로부터 직접 통계량을 구할 수 있다는 장점 때문에 시계열 자료에서는 붓스트랩을 이용한 신뢰구간을 구하는 모의실험을 통해 예측의 정확성을 측정할 수 있다. 붓스트랩 추정량들로부터 신뢰구간을 구하는 대표적인 방법으로 표준 붓스트랩 방법, 백분위수 붓스트랩 방법, 스튜던트화 붓스트랩 방법, 하이브리드 붓스트랩 방법이 있다. 오차항 $\varepsilon_t$의 분포에 따라 각 방법에 의한 붓스트랩 신뢰구간에 모수가 포함되는 비율을 알아본다. 붓스트랩의 절차는 다음과 같다.

Step 1. 오차항 $\{\varepsilon_t, t = 1, 2, 3, \ldots, T\}$를 난수발생하고 RCA(1) 모형을 따르는 시계열 난수 $\{X_t, t = 1, 2, 3, \ldots, T\}$를 발생한다. 발생된 시계열 $\{X_t, t = 1, 2, 3, \ldots, T\}$로부터 Quasi score 추정량과 M-Quasi score 추정량을 구한다.

Step 2. 시계열 난수로부터 추정된 Quasi score 추정량과 M-Quasi score 추정량으로부터 잔차 $\{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_T\}$를 구하고 이를 표준화 한다.

$$\tilde{r}_t = \frac{\hat{r}_t - \hat{r}(\bullet)}{\widehat{\mathrm{se}}_r}, \tag{4.1}$$

여기서 $\hat{r}(\bullet) = \sum_{t=1}^{n} \hat{r}_t / n$이고, $\widehat{\mathrm{se}}_r = \sqrt{\sum_{t=1}^{T} (\hat{r}_t - \hat{r}(\bullet))^2 / (T-1)}$이다.

Step 3. 표준화된 잔차 $\{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_T\}$를 바탕으로 붓스트랩 표본 $\{X_1^*, X_2^*, \ldots, X_T^*\}$를 생성한다. 그리고 붓스트랩 표본으로부터 붓스트랩 추정량 $\hat{\theta}^*$를 계산한다.

Step 4. Step 3을 $B$번 반복하여 붓스트랩 추정량 $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\}$를 구한다.

Step 5. $B$개의 붓스트랩 추정량 $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\}$으로부터 각각의 방법을 적용하여 $(1-2\alpha) \times 100\%$ 신뢰구간을 구한다.

(1) 표준 붓스트랩 방법(standard bootstrap method; SB)

가장 보편화된 $(1-2\alpha) \times 100\%$ SB 신뢰구간은 다음과 같다.

$$\left(\hat{\theta} - z_\alpha \widehat{\mathrm{se}}_B, \hat{\theta} + z_\alpha \widehat{\mathrm{se}}_B\right), \tag{4.2}$$

여기서 $\widehat{\mathrm{se}}_B = \sqrt{\sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}^*)^2 / (B-1)}$이고 $\bar{\theta}^* = \sum_{b=1}^{B} \hat{\theta}_b^* / B$이다.

(2) 백분위수 붓스트랩 방법(percentile bootstrap method; PB)

$B$개의 붓스트랩 추정량 $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\}$를 크기순 $\hat{\theta}_{(1)}^* \le \hat{\theta}_{(2)}^* \le \cdots \le \hat{\theta}_{(B)}^*$으로 나열하여 구한 $\theta$에 대한 $(1-2\alpha) \times 100\%$ PB 신뢰구간을 다음과 같다.

$$\left(\hat{\theta}_{(\alpha \times B)}^*, \hat{\theta}_{((1-\alpha) \times B)}^*\right). \tag{4.3}$$

(3) 스튜던트화 붓스트랩 방법(studentized bootstrap method; STUB)

STUB 방법은 $\sqrt{n}(\hat{\theta}^* - \hat{\theta}) / \widehat{\mathrm{se}}^*$의 분포로부터 백문위 지점을 찾아 신뢰구간을 결정하는 방법으로 $\theta$에 대한 $(1-2\alpha) \times 100\%$ STUB 신뢰구간은 다음과 같다.

$$\left(\hat{\theta} - \hat{t}_{(1-\alpha)} \widehat{\mathrm{se}}, \hat{\theta} - \hat{t}_\alpha \widehat{\mathrm{se}}\right). \tag{4.4}$$

여기서 $\hat{t}_\alpha$는 $\Pr(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) / \widehat{\mathrm{se}}^* \le \hat{t}_\alpha) = \alpha$를 만족하는 값이다. 또한 $\widehat{\mathrm{se}}^*$는 붓스트랩 표본으로부터 구한 붓스트랩 추정량 $\hat{\theta}^*$의 표준오차이다.

(4) 하이브리드 붓스트랩 방법(hybrid bootstrap method; HYB)

$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) / \widehat{\mathrm{se}}$의 분포로부터 백분위 지점을 찾아 신뢰구간을 설정하는 방법으로 $\theta$에 대한 $(1-2\alpha) \times 100\%$ HYB 신뢰구간은 다음과 같다.

$$\left(\hat{\theta} - \hat{h}_{(1-\alpha)} \widehat{\mathrm{se}}, \hat{\theta} - \hat{h}_\alpha \widehat{\mathrm{se}}\right). \tag{4.5}$$

여기서 $\hat{h}_\alpha$는 $\Pr(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) / \widehat{\mathrm{se}} \le \hat{h}_\alpha) = \alpha$를 만족하는 값이다.

**Table 5.1.** Comparison of the simulated proportion that $\varepsilon_t$ is standard normal distribution

| Estimator | | Method | Confidence level | | |
|---|---|---|---|---|---|
| | | | 95% | 90% | 80% |
| Quasi score | | Asymptotics | 0.947 | 0.902 | 0.789 |
| | | SB | 0.946 | 0.907 | 0.800 |
| | | PB | 0.934 | 0.872 | 0.772 |
| | | STUD | 0.940 | 0.884 | 0.775 |
| | | HYB | 0.954 | 0.896 | 0.800 |
| M-Quasi score | Huber | Asymptotics | 0.940 | 0.882 | 0.790 |
| | | SB | 0.947 | 0.899 | 0.801 |
| | | PB | 0.960 | 0.904 | 0.801 |
| | | STUD | 0.952 | 0.897 | 0.794 |
| | | HYB | 0.955 | 0.905 | 0.813 |
| | Tukey | Asymptotics | 0.937 | 0.876 | 0.788 |
| | | SB | 0.946 | 0.903 | 0.796 |
| | | PB | 0.955 | 0.901 | 0.806 |
| | | STUD | 0.950 | 0.901 | 0.782 |
| | | HYB | 0.941 | 0.891 | 0.790 |
| | Andrews | Asymptotics | 0.959 | 0.908 | 0.807 |
| | | SB | 0.953 | 0.894 | 0.787 |
| | | PB | 0.952 | 0.904 | 0.797 |
| | | STUD | 0.949 | 0.888 | 0.793 |
| | | HYB | 0.956 | 0.918 | 0.787 |
| | Hampel | Asymptotics | 0.951 | 0.901 | 0.801 |
| | | SB | 0.954 | 0.894 | 0.800 |
| | | PB | 0.949 | 0.903 | 0.800 |
| | | STUD | 0.946 | 0.893 | 0.793 |
| | | HYB | 0.955 | 0.910 | 0.781 |

SB = standard bootstrap method; PB = percentile bootstrap method; STUB = studentized bootstrap method; HYB = hybrid bootstrap method.

## 5. 모의실험

비선형 시계열 모형인 확률계수 자기회귀 모형에서 Quasi score 우도추정량과 M-Quasi score 우도추정량의 근사적인 신뢰구간을 얻기 위해 두 추정량의 극한분포를 이용하여 신뢰구간을 구한다. 즉, $n =$ 200개의 난수로부터 95%, 90%, 80%의 근사적인 신뢰구간을 구하고 이 신뢰구간이 모수값을 포함하는지 확인하고, 이러한 과정을 $N =$ 1,000번 반복실시하여 신뢰구간이 모수값에 포함되는 비율을 계산한다. 또한 붓스트랩 신뢰구간을 얻기 위하여 앞에서 정의한 SB, PB, STUD, HYB 방법을 이용한다. RCA 시계열 난수로부터 크기가 200인 $B(N = 1{,}000)$개의 붓스트랩 표본을 생성하고 붓스트랩 표본을 이용하여 $B$개의 추정량을 구한 다음 95%, 90%, 80% 신뢰구간을 설정한다. 그리고 설정된 신뢰구간에 모수의 참값이 포함되는지 확인한다. 이런 과정을 $N =$ 1,000번 반복실시하여 신뢰구간에 모수의 참값이 포함되는 비율을 계산한다. 정상조건을 만족하기 위해 RCA(1) 모형에서 모수 $\theta = 0.5$, $Z_t$의 분산을 $\sigma_z^2 = 0.03$으로 가정하였다. 확률변수 $\varepsilon_t$의 분포를 표준정규분포, 이중지수분포 그리고 오염정규분포로 나누어 시뮬레이션을 수행하였다.

RCA(1)모형에서 근사적인 신뢰구간과 붓스트랩 신뢰구간에 대한 포함비율 결과는 조건에 따라 Tables 5.1–5.3에 나타내었다. Table 5.1은 $\varepsilon_t$가 표준정규분포인 RCA(1) 모형의 결과로 Quasi score 추정량

**Table 5.2.** Comparison of the simulated proportion that $\varepsilon_t$ is double exponential distribution

| Estimator | | Method | Confidence level | | |
|---|---|---|---|---|---|
| | | | 95% | 90% | 80% |
| Quasi score | | Asymptotics | 0.890 | 0.819 | 0.701 |
| | | SB | 0.936 | 0.851 | 0.751 |
| | | PB | 0.912 | 0.832 | 0.730 |
| | | STUD | 0.916 | 0.840 | 0.746 |
| | | HYB | 0.934 | 0.833 | 0.732 |
| M-Quasi score | Huber | Asymptotics | 0.940 | 0.882 | 0.790 |
| | | SB | 0.947 | 0.899 | 0.801 |
| | | PB | 0.960 | 0.904 | 0.801 |
| | | STUD | 0.952 | 0.897 | 0.794 |
| | | HYB | 0.955 | 0.905 | 0.813 |
| | Tukey | Asymptotics | 0.937 | 0.876 | 0.788 |
| | | SB | 0.946 | 0.903 | 0.796 |
| | | PB | 0.955 | 0.901 | 0.806 |
| | | STUD | 0.950 | 0.901 | 0.782 |
| | | HYB | 0.941 | 0.891 | 0.790 |
| | Andrews | Asymptotics | 0.959 | 0.908 | 0.807 |
| | | SB | 0.953 | 0.894 | 0.787 |
| | | PB | 0.952 | 0.904 | 0.797 |
| | | STUD | 0.949 | 0.888 | 0.793 |
| | | HYB | 0.956 | 0.918 | 0.787 |
| | Hampel | Asymptotics | 0.951 | 0.901 | 0.801 |
| | | SB | 0.954 | 0.894 | 0.800 |
| | | PB | 0.949 | 0.903 | 0.800 |
| | | STUD | 0.946 | 0.893 | 0.793 |
| | | HYB | 0.955 | 0.910 | 0.781 |

SB = standard bootstrap method; PB = percentile bootstrap method; STUB = studentized bootstrap method; HYB = hybrid bootstrap method.

과 M-Quasi score 추정량에서 근사적인 신뢰구간과 붓스트랩 신뢰구간의 포함비율은 유사하였다. 그리고 Table 5.2의 확률변수 $\varepsilon_t$가 이중지수분포인 경우와 Table 5.3의 확률변수 $\varepsilon_t$가 오염정규분포인 경우를 살펴보면 Quasi score 추정량과 M-Quasi score추정량에서 붓스트랩 신뢰구간의 포함비율이 근사적인 신뢰구간보다 더 신뢰도에 근접함을 알 수 있다. 붓스트랩 방법 중 SB와 STUD이 다른 붓스트랩 방법보다 더 신뢰도에 근접함을 알 수 있다.

## 6. 결론

자료가 시간의 가역성을 따르지 않으면서 분포를 가정할 수 없는 비선형 시계열 자료에서 평균과 분산이 알려져 있는 RCA모형을 고려하였다. 이때 Quasi score 함수를 이용한 Quasi score 추정량과 유계함수를 적용한 M-quasi score 함수를 이용한 M-Quasi score 추정량을 이용하여 근사 신뢰구간을 구하였다. 근사 신뢰구간과 SB, PB, STUD, HYB의 네 가지의 붓스트랩 신뢰구간을 구하여 참 추정량의 포함비율을 구하였다. RCA(1) 모형에서 확률변수 $\varepsilon_t$가 정규분포를 따르는 경우, Quasi score 추정량과 M-Quasi score 추정량에 대한 근사적인 신뢰구간에는 별다른 차이를 보이지 않았다, 그리고 확률변

**Table 5.3.** Comparison of the simulated proportion that $\varepsilon_t$ is contaminated normal distribution with 10% level of contamination and 10 magnitude of contamination

| Estimator | | Method | Confidence level | | |
|---|---|---|---|---|---|
| | | | 95% | 90% | 80% |
| Quasi score | | Asymptotics | 0.882 | 0.810 | 0.675 |
| | | SB | 0.922 | 0.820 | 0.742 |
| | | PB | 0.914 | 0.832 | 0.780 |
| | | STUD | 0.927 | 0.806 | 0.762 |
| | | HYB | 0.904 | 0.826 | 0.773 |
| M-Quasi score | Huber | Asymptotics | 0.925 | 0.877 | 0.759 |
| | | SB | 0.921 | 0.823 | 0.714 |
| | | PB | 0.932 | 0.812 | 0.712 |
| | | STUD | 0.944 | 0.864 | 0.739 |
| | | HYB | 0.920 | 0.832 | 0.723 |
| | Tukey | Asymptotics | 0.908 | 0.864 | 0.753 |
| | | SB | 0.926 | 0.855 | 0.764 |
| | | PB | 0.930 | 0.845 | 0.735 |
| | | STUD | 0.940 | 0.866 | 0.780 |
| | | HYB | 0.930 | 0.824 | 0.736 |
| | Andrews | Asymptotics | 0.912 | 0.864 | 0.765 |
| | | SB | 0.942 | 0.860 | 0.757 |
| | | PB | 0.931 | 0.888 | 0.760 |
| | | STUD | 0.905 | 0.853 | 0.771 |
| | | HYB | 0.936 | 0.865 | 0.769 |
| | Hampel | Asymptotics | 0.919 | 0.859 | 0.755 |
| | | SB | 0.936 | 0.848 | 0.743 |
| | | PB | 0.921 | 0.855 | 0.721 |
| | | STUD | 0.931 | 0.892 | 0.763 |
| | | HYB | 0.932 | 0.878 | 0.776 |

SB = standard bootstrap method; PB = percentile bootstrap method; STUB = studentized bootstrap method; HYB = hybrid bootstrap method.

수 $\varepsilon_t$가 이중지수분포이거나 오염정규분포인 경우 근사적인 방법보다 붓스트랩 방법이 더 효율적이었다. RCA(1) 모형에서는 붓스트랩 방법중 SB와 STUD가 PB와 HYB 보다 효율적이었다.

## References


Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press, New Jersey.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, *Technometrics*, **16**, 147–185.

Beran, R. (1990). Calibrating prediction regions, *Journal of the American Statistical Association*, **85**, 715–723.

Cha, K. U., Kim, S. Y., and Lee, S. D. (1999). Robust estimation using estimating functions for time series models, *The Journal of Applied Statistics*, **12**, 479–490.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics*, **7**, 1–26

Huber, P. J. (1964). Robust estimation of a location Parameter, *The Annals of Mathematical Statistics*, **35**, 73–101.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Kim, S., Cha, K., and Lee, S. (2003). Efficient Quasi-likelihood estimation for nonlinear time series models and its application, *Communications for Statistical Applications and Methods*, **10**, 101–113.

Lee, S. D. and Kim, J. S. (2004). Prediction intervals for nonlinear time series models using the bootstrap method, *The Korean Journal of Applied Statistics*, **17**, 219–228.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, New York.

Thombs, L. A. and Scbucany, W. R. (1990). Bootstrap prediction intervals for autoregression, *Journal of the American Statistical Association*, **85**, 486–492.

Tong, H. (1990). *Nonlinear Time Series*, Oxford University Press, Oxford.

Wedderburn, R. P. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method, *Biometrika*, **61**, 439–447.

# 붓스트랩 방법을 적용한 확률계수 자기회귀 모형에 대한 로버스트 구간추정

조나래[a] · 임도상[b] · 이성덕[a,1]

[a]충북대학교 정보통계학과, [b]질병관리본부 만성질환관리과

(2018년 12월 3일 접수, 2018년 12월 11일 수정, 2018년 12월 11일 채택)

**요 약**

비선형 시계열인 확률계수 자기회귀(random coefficient autoregressive; RCA) 모형에 대하여 여러 가지 방법을 이용한 추정량의 신뢰구간 비교하였다. RCA 모형에 대하여 자료의 분포를 가정하지 않아도 되는 Quasi 스코어 추정량과 Huber, Tukey, Andrew, Hempal 4가지 유계함수를 이용한 M-Quasi 스코어 추정량을 제시하였다. 이러한 추정량에 대하여 표준 붓스트랩 방법, 백분위수 붓스트랩 방법, 스튜던트화 붓스트랩 방법, 하이브리드 붓스트랩 방법을 이용한 신뢰구간을 구하였다. 모의실험을 통하여 RCA 모형의 오차항의 분포가 정규분포, 오염정규분포, 이중지수분포를 따를 때 Quasi 스코어 추정량과 M-Quasi 스코어 추정량들의 근사적 신뢰구간과 네가지 붓스트랩 방법을 이용한 신뢰구간을 비교하였다.

주요용어: 확률계수 자기회귀 모형, 표준 붓스트랩 방법, 백분위수 붓스트랩 방법, 스튜던트화 붓스트랩 방법, 하이브리드 붓스트랩 방법

이 논문은 2016년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 이공학 개인기초연구지원사업 연구비 지원에 의하여 연구되었음 (NRF-2016R1D1A3B03932557).

[1]교신저자: (28644) 충북 청주시 서원구 충대로 1, 충북대학교 정보통계학과. E-mail: sdlee@chungbuk.ac.kr