Ann. Inst. Statist. Math.
Vol. 42, No. 4, 753–768 (1990)

BOOTSTRAP IN MOVING AVERAGE MODELS

ARUP BOSE*

Stat.-Math. Division, Indian Statistical Institute, 203, B. T. Road, Calcutta 700035, India and Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.

(Received December 24, 1987; revised February 13, 1990)

Abstract. We prove that the bootstrap principle works very well in moving average models, when the parameters satisfy the invertibility condition, by showing that the bootstrap approximation of the distribution of the parameter estimates is accurate to the order $o(n^{-1/2})$ a.s. Some simulation studies are also reported.

Key words and phrases: Moving average models, stationary autoregressions, Cramer's condition, Edgeworth expansions, empirical distribution function, bootstrap.

1. Introduction

The bootstrap procedure was introduced by Efron (1979, 1982). Since then there have been theoretical studies dealing with the accuracy of the bootstrap approximation in various senses (asymptotic normality, Edgeworth expansion, etc.). Some of the references are Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982), Babu and Singh (1984) and Hall (1988). One class of results show that in the i.i.d. situation where the normal approximation holds with an error of $O(n^{-1/2})$, if we replace the normal distribution by a sample dependent bootstrap distribution, then the error rate is $o(n^{-1/2})$ a.s.

The bootstrap does not give correct answers in general dependent models. However, some dependent models do allow for an appropriate resampling so that the bootstrap works. Freedman (1984) has shown that the bootstrap gives the correct asymptotic result for two stage least squares estimates in linear autoregressions with possible exogenous variables orthogonal to errors. Basawa *et al.* (1989) have proven the validity of bootstrap in explosive first order autoregressions. Bose (1988*a*) has shown that the rate result alluded to in the i.i.d. situation holds for stationary autoregressions.

*Now at Indian Statistical Institute.

Here we deal with moving average models. The moment estimators of the parameters have an asymptotic normal distribution and the approximation error can be shown to be $O(n^{-1/2})$. The structure of the process enables appropriate resampling. We show that the bootstrap distribution approximates the distribution of the parameter estimates with an accuracy of $o(n^{-1/2})$ a.s. The idea is to develop one term Edgeworth expansion for the distribution of the parameter estimates and their bootstrapped version. The leading terms of these expansions match and the difference of the second terms is $o(n^{-1/2})$, yielding the desired result.

2. Preliminaries

Let (Y_t) be a process satisfying $Y_t = \varepsilon_t + \sum_{i=1}^{l} \alpha_i \varepsilon_{t-i}$, where

(A1) (ε_t) is i.i.d. $\sim F_0$, $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = 1$, and $E\varepsilon_t^{2(s+1)} < \infty$ for some $s \geq 3$.

(A2) $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramer's condition, i.e. for every $d > 0$, $\exists \delta > 0$ such that

$$\sup_{\|t\| \geq d} | E \exp (it'(\varepsilon_1, \varepsilon_1^2))| < 1 - \delta .$$

$\alpha_1, \alpha_2, \ldots, \alpha_l$ are unknown parameters which can be estimated by moment estimates.

Remark 2.1. The assumption that the mean and variance of ε_t are known has been made to keep the proofs simple. See Remark 3.2 for a discussion of how this assumption can be dropped. The Cramer's condition is required to obtain Edgeworth expansions.

Remark 2.2. The minimum moment assumption we need is $E\varepsilon_1^8 < \infty$, which may seem too strong. However, the estimates of α_i's involve quadratic functions of ε_t and we need the $(s+1)$-th moment of ε_t^2 with s at least 3. This is in contrast to the situation of i.i.d. observations where the s-th moment suffices to derive the $o(n^{-(s-2)/2})$ expansion.

We first assume that $l = 1$, i.e. $Y_t = \varepsilon_t + \alpha \varepsilon_{t-1}$. The moment estimate of α, given the observations $Y_0, Y_1, \ldots, Y_n$, is $\alpha_n = n^{-1} \sum_{t=1}^{n} Y_t Y_{t-1}$. Under our assumptions, α_n is strongly consistent and $n^{1/2}(\alpha_n - \alpha)$ has an asymptotic normal distribution.

Define $\tilde{\varepsilon}_i = \sum_{j=0}^{i-1} (-1)^j \alpha^j Y_{i-j}$, and $\tilde{\varepsilon}_1 = Y_1$. Using the structure of the process,

$$\tilde{\varepsilon}_i = \varepsilon_i - (-\alpha)^i \varepsilon_0 . \tag{2.1}$$

Hence $\tilde{\varepsilon}_i$ and ε_i are close for all large i if $|\alpha| < 1$, which shows that resampling is proper in this situation. (For $l > 1$, this condition should be replaced by the invertibility condition, see Hannan (1970)). So motivated by (2.1), we compute the pseudo errors as

$$\hat{\varepsilon}_{in} = \sum_{j=0}^{i-1} (-1)^j \alpha_n^j Y_{i-j}, \quad i = 2,\ldots,n, \qquad \hat{\varepsilon}_{1n} = Y_1 .$$

For ease of notations we will often drop the suffix n. Let G_n denote the empirical distribution function which puts mass n^{-1} at each $\hat{\varepsilon}_{in}$, $i = 1, 2, \ldots, n$. Let $\hat{F}_n(x) = G_n(x + \bar{\varepsilon}_n)$ where $\bar{\varepsilon}_n = n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{in}$. It is expected that $\hat{F}_n$ will be close to F_0 with increasing n. Take an i.i.d. sample (ε_{in}^*) from $\hat{F}_n$ and define

$$Y_i^* = \begin{cases} \varepsilon_{in}^* + \alpha_n \varepsilon_{i-1,n}^*, & i = 1,\ldots,n , \\ \varepsilon_i^* + \alpha_n \varepsilon_{i-1}^*, & \text{dropping the suffix } n . \end{cases}$$

Pretend that α_n is unknown and obtain its moment estimate by $\alpha_n^* = n^{-1} \sum_{t=1}^{n} Y_t^* Y_{t-1}^*$. So the bootstrapped quantity corresponding to $n^{1/2}(\alpha_n - \alpha)$ is $n^{1/2}(\alpha_n^* - \alpha_n)$. In the next section we will see how accurate the distribution of $n^{1/2}(\alpha_n^* - \alpha_n)$ is (given $Y_0, Y_1, \ldots, Y_n$) in estimating the distribution of $n^{1/2}(\alpha_n - \alpha)$ as $n \to \infty$.

Before discussing the main results, we introduce a few notations. C stands for a generic constant, which in probability arguments may depend on the particular point w under consideration in the basic probability space. For a sequence of random vectors X_t, $S_n = n^{-1/2} \sum_{t=1}^{n} X_t$. $G_n \Rightarrow G$ denotes that the distribution G_n converges weakly to G (G_n may be random). The function $\psi_{n,s}$ represents the first $(s-1)$ terms of the Edgeworth expansion of the distribution of S_n whenever such an expansion is valid. See Bhattacharya and Ranga Rao (1976, p. 145) for the definition of $\psi_{n,s}$ when X_t are i.i.d. Götze and Hipp (1983) may be consulted for a definition of $\psi_{n,s}$ when X_t are dependent. For any random vector X, $D(X)$ denotes the dispersion matrix of X. $\beta = (\beta_1 \cdots \beta_k)$ denotes a vector where each β_i is a nonnegative integer and for $f: \mathbb{R}^k \to \mathbb{R}$,

$$D^\beta f(x) = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial x_k^{\beta_k}} f(x_1, \ldots, x_k) \qquad \text{and} \qquad |\beta| = \beta_1 + \beta_2 + \cdots + \beta_k .$$

3. Main results

We first need some auxiliary results. Let $\tilde{F}_n$ denote the empirical distribution function of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$.

LEMMA 3.1. *Under* (A1), *we have for all* $k \le 2(s+1)$,

(a) $n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_i^k \xrightarrow{a.s.} E_{F_0}(\varepsilon_1^k)$ *and* $n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{in}^k \xrightarrow{a.s.} E_{F_0}(\varepsilon_1^k)$,

(b) $\tilde{F}_n \Rightarrow F_0$ *a.s. and* $\hat{F}_n \Rightarrow F_0$ *a.s.*

PROOF. Throughout the proof, arguments are for a fixed w in the basic probability space and hence all bounds, etc. depend on w in general.

(a) To prove the first part, it is enough to show that $n^{-1} \sum_{i=1}^{n} (\varepsilon_i^k - \tilde{\varepsilon}_i^k) \to 0$ a.s. But $n^{-1}(\varepsilon_1^k - \tilde{\varepsilon}_1^k) \to 0$ trivially. Furthermore,

$$\left| n^{-1} \sum_{i=2}^{n} (\varepsilon_i^k - \tilde{\varepsilon}_i^k) \right| \le n^{-1} \sum_{j=0}^{k-1} \binom{k}{j} |\varepsilon_0|^{k-j} \sum_{i=2}^{n} |\varepsilon_i|^j |\alpha|^i .$$

It easily follows from Theorem 2.18 of Hall and Heyde (1980) that $n^{-1} \sum_{i=1}^{n} |\varepsilon_i|^j |\alpha|^i \xrightarrow{\text{a.s.}} 0 \ \forall j \le k-1$. To prove the second part, it suffices to show that $n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i^k - \tilde{\varepsilon}_i^k) \xrightarrow{\text{a.s.}} 0$. Note that $\alpha_n \xrightarrow{\text{a.s.}} \alpha$ and $|\alpha| < 1$. Hence for all large n and for all $j \ge 1$,

$$(3.1) \quad |\alpha| + |\alpha_n - \alpha| \le \tilde{\beta} < 1 \qquad \text{a.s.}$$

and

$$(3.2) \quad |\alpha_n^j - \alpha^j| = |(\alpha_n - \alpha + \alpha)^j - \alpha^j| \le C|\alpha_n - \alpha|\delta^j \qquad \text{for some } \delta < 1 .$$

Hence

$$\begin{aligned}
&\left| n^{-1} \sum_{i=1}^{n} (\tilde{\varepsilon}_i^k - \hat{\varepsilon}_i^k) \right| \\
&= \left| n^{-1} \sum_{i=2}^{n} \left[\left(\sum_{j=0}^{i-1} (-1)^j \alpha^j Y_{i-j} \right)^k - \left(\sum_{j=0}^{i-1} (-1)^j \alpha_n^j Y_{i-j} \right)^k \right] \right| \\
&\le n^{-1} \sum_{i=2}^{n} \left[2^{k-1} \sum_{j=0}^{i-1} |\alpha_n^j - \alpha^j| \, |Y_{i-j}| \left\{ \left(\left| \sum_{t=0}^{i-1} (-1)^t \alpha_n^t Y_{i-j} \right| \right)^{k-1} \right. \right. \\
&\qquad \left. \left. + \left(\left| \sum_{t=0}^{i-1} (-1)^t \alpha^t Y_{i-j} \right| \right)^{k-1} \right\} \right]
\end{aligned}$$

$$\leq Cn^{-1}|\alpha_n - \alpha| \sum_{i=2}^{n} \left(\sum_{j=0}^{i-1} \delta^j |Y_{i-j}| \right)^k \quad \text{(by (3.1) and (3.2))} .$$

Since $Y_i = \varepsilon_i + \alpha\varepsilon_{i-1}$, it is enough to show that $n^{-1} \sum_{i=2}^{n} \left(\sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}| \right)^k$ is bounded a.s. The sequence $Z_i = \sum_{j=0}^{\infty} \delta^j |\varepsilon_{i-j}|$, $i \geq 1$ is a stationary autoregressive process of order one and hence is ergodic (see Hannan (1970), p. 204). So the second part of (a) follows from the observation

$$n^{-1} \sum_{i=2}^{n} \left(\sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}| \right)^k \leq n^{-1} \sum_{i=1}^{n} Z_i^k \overset{\text{a.s.}}{\to} E(Z_1^k) < \infty .$$

(b) Since (ε_i) is i.i.d. F_0, the first part readily follows from (2.1). The second part follows from the following observation: if F_n and G_n are empirical distributions based on n tuples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, then for all f such that f' is bounded,

$$|E_{F_n}(f) - E_{G_n}(f)| \leq \frac{1}{n} \sum_{i=1}^{n} |f(x_i) - f(y_i)| \leq \|f'\|_\infty \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| . \quad \square$$

To study the bootstrap approximation, we need an Edgeworth expansion for α_n. For this, we use a result of Götze and Hipp (1983) (henceforth referred to as GH).

Let (X_t) be $\mathbb{R}^k$ valued random variables on $(\Omega, \mathscr{F}, P)$. Let there be σ-fields $\mathscr{D}_j$ $\left(\text{write } \sigma\left(\bigcup_{j=a}^{b} \mathscr{D}_j \right) = \mathscr{D}_a^b \right)$ and $\alpha^* > 0$ such that

C(1) $EX_t = 0 \; \forall t$.

C(2) $E\|X_t\|^{s+1} \leq M_{s+1} < \infty \; \forall t$ for some $s \geq 3$.

C(3) $\exists Y_{nm} \in \mathscr{D}_{n-m}^{n+m} \ni E\|X_n - Y_{nm}\| \leq C \exp(-\alpha^* m)$.

C(4) $\forall A \in \mathscr{D}_{-\infty}^{n}, B \in \mathscr{D}_{n+m}^{\infty}, |P(A \cap B) - P(A)P(B)| \leq C \exp(-\alpha^* m)$.

C(5) $\exists d, \delta > 0 \ni \forall \|t\| \geq d, E \left| E \exp\left(it' \sum_{j=n-m}^{n+m} X_j \right) \middle| \mathscr{D}_j, j \neq n \right| < 1 - \delta < 1$.

C(6) $\forall A \in \mathscr{D}_{n-p}^{n+p}$, $\forall n, p, m, E|P(A|\mathscr{D}_j,\ j \neq n) - P(A|\mathscr{D}_j, 0 < |j-n| \leq m+p)| \leq C \exp(-\alpha^* m)$.

C(7) $\lim_{n\to\infty} D\left(n^{-1/2} \sum_{t=1}^{n} X_t \right) = \Sigma$ exists and is positive definite.

Let s_0 be s or $(s-1)$ according as s is even or odd, ϕ_Σ be the normal density with mean 0 and dispersion matrix Σ. The following results are due to Götze and Hipp (1983).

THEOREM 3.1. *Let f: $\mathbb{R}^k \to \mathbb{R}$ be a measurable function such that $|f(x)| \leq M(1 + \|x\|^{s_0})$ for every $x \in \mathbb{R}^k$. Assume that* C(1)–C(7) *hold. Then there exists a positive constant δ_0 not depending on f and M, and for any* *arbitrary $k > 0$ there exists a positive constant C depending on M but not on f such that*

$$\left| Ef(S_n) - \int f d\psi_{n,s} \right| \le Cw(f, n^{-k}) + o(n^{-(s-2+\delta_0)/2})$$

where $w(f, n^{-k}) = \int \sup (|f(x+y) - f(x)|: |y| \le n^{-k})\phi_\Sigma(x)dx$. *The term* $o(\cdot)$ *depends on f through M only.*

COROLLARY 3.1. *Assume* C(1)–C(7). *Then the following approximation holds uniformly over convex measurable* $C \subseteq \mathbb{R}^k$.

$$P(S_n \in C) = \psi_{n,s}(C) + o(n^{-(s-2)/2}) .$$

Let $X_t = Y_t Y_{t-1} - \alpha$ and $\mathscr{D}_j$ = sigma field generated by ε_j. It can then be easily shown that X_t satisfies the conditions of the above theorem under (A1) and (A2). We omit the details (see Bose (1988*b*)). Thus we have the following proposition.

PROPOSITION 3.1. *Assume that* (A1) *and* (A2) *hold. Let* $S_n = n^{-1/2} \sum_{t=1}^{n} (Y_t Y_{t-1} - \alpha)$. *Then*

(a) *Theorem* 3.1 *holds with the above* S_n; *consequently*,

(b) $P(S_n \in C) = \psi_{n,s}(C) + o(n^{-(s-2)/2})$, *uniformly over convex subsets of* $\mathbb{R}$.

We now develop an Edgeworth expansion for the bootstrapped version of the above S_n. In what follows we make the convention that the presence of (*) indicates that we are dealing with a bootstrapped quantity; as a result, expectations, etc. are taken w.r.t. (ε_i^*) i.i.d. $\hat{F}_n$ given $Y_0, Y_1, \ldots, Y_n$.

Define

$$X_j^* = Y_j^* Y_{j-1}^* - \alpha_n, \quad j \ge 1 \qquad \text{and} \qquad H_n^*(t) = E\left[e^{itn^{-1/2}} \sum_{j=1}^{n} X_j^* \right].$$

We have the following lemmas. The proofs are only sketched and the details can be completed by arguments similar to those used by GH.

LEMMA 3.2. $\forall |t| \le Cn^{\varepsilon_0}$ *and* $|\beta| \le s + 2$, *we have*

$$|D^\beta(H_n^*(t) - \hat{\psi}_{n,s}^*(t))|$$
$$\le C(1 + m_{s+1,n}^*)(1 + |t|^{3(s-1)+|\beta|}) \exp(-C|t|^2) n^{-(s-2+\varepsilon_0)/2}$$

for some $\varepsilon_0 < 1/2$ *and* C *depends on the bounds of* $m^*_{s+1,n} = (s+1)$*-th moment of* X_j^*. *Here* $\hat{\psi}^*_{n,s}(t)$ *is the Fourier transform of* $\psi^*_{n,s}$.

The proof is exactly as the proof of Lemma 3.33 of GH and we omit it.

Let

$$I_1 = \{t\colon Cn^{\varepsilon_0} \le |t| \le C_1 n^{1/2}\}, \qquad I_2 = \{t\colon C_1 n^{1/2} \le |t| \le \tilde{\varepsilon}^{-1} n^{1/2}\}$$

where C_1 is to be chosen and $0 < \tilde{\varepsilon} < 1$ is fixed.

LEMMA 3.3. *Under* (A1) *and* (A2), *we have for almost every sequence* (Y_i),

$$\int_{t \in I_2} |D^\alpha H_n^*(t)|\, dt = o(n^{-(s-2)/2}) .$$

PROOF. A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that $E^*|E^* A_p^* | \mathscr{D}_j^*, j \ne j_p| < 1$ uniformly in $t \in I_2$ and $p = 1, 2, \ldots, J$ where $\mathscr{D}_j^* = \sigma(\varepsilon_j^*)$, $A_p^* = \exp\left(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} Z_j^* \right)$; see GH for the definition of j_p and m. We omit the details of these definitions since they are not explicitly used in the sequel. It suffices to note that j_p is fixed and that the above expectation is independent of m (see below). This expectation equals

$$\delta^*_{nm} = E^* \left| E^* \exp\left(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} X_j^* \right) \right| \varepsilon_j^*, \quad j \ne j_p \left| \right. .$$

Note that $\sum_{j=j_p-m}^{j_p+m} X_j^* = \varepsilon^*_{j_p}(Y^*_{j_p-1} + \alpha_n Y^*_{j_p+2} + \varepsilon^*_{j_p+1} + \alpha_n^2 \varepsilon^*_{j_p-1}) + \alpha_n \varepsilon^{*2}_{j_p} + V$ where V is independent of $\varepsilon^*_{j_p}$.

Let K_n^* denote the distribution function of $Y^*_{j_p-1} + \alpha_n Y^*_{j_p+2} + \varepsilon^*_{j_p+1} + \alpha_n^2 \varepsilon^*_{j_p-1}$. Then $\delta^*_{nm} = \int \left| \int \exp(itn^{-1/2}xy + itn^{-1/2}\alpha_n x^2) d\hat{F}_n(x) \right| dK_n^*(y)$.

As t varies in I_2, $(tn^{-1/2}, tn^{-1/2}\alpha_n)$ varies in a compact set bounded away from zero. Let D denote any such set in $\mathbb{R}^2$. Let $b_1, b_2 > 0$ (to be chosen). Then

$$\begin{aligned} \delta^*_{nm} &\le \sup_{(d_1,d_2) \in D} \int \left| \int \exp(id_1 xy + id_2 x^2) d\hat{F}_n(x) \right| dK_n^*(y) \\ &\le K_n^*(b_1 \le |Y| \le b_2) I_{1n} + K_n^*(|Y| < b_1) + K_n^*(|Y| > b_2) \end{aligned}$$

where

$$I_{1n} \leq \sup_{b_1 \leq |y| \leq b_2} \sup_{(d_1, d_2) \in D} \left| \int \exp(id_1 xy + id_2 x^2) d\hat{F}_n(x) \right| .$$

By Lemma 3.1, $K_n^* \Rightarrow K$ a.s. where K is the distribution of $Y_{j-1} + \alpha Y_{j+2} + \varepsilon_{j+1} + \alpha^2 \varepsilon_{j-1}$, which is non-degenerate. Thus b_1 and b_2 can be chosen such that for all large n,

$$K_n^*(|Y| < b_1) + K_n^*(|Y| > b_2) < \alpha_0 < 1 .$$

Note that $F_n^* \Rightarrow F_0$ a.s. and F_0 satisfies Cramer's condition. Since the convergence of characteristic functions to the limit is uniform over compact sets, we have $I_{1n} < 1 - \gamma < 1$ for all large n. Thus $\delta_{nm}^* \leq (1-\gamma) + \alpha_0 \gamma < 1$, proving the lemma. □

LEMMA 3.4. *Under* (A1) *and* (A2), *for a sufficiently small* C_1, *we have for almost every sequence* (Y_i), $|\beta| \leq s + 2$,

$$\int_{t \in I_1} |D^\beta H_n^*(t)| dt = o(n^{-(s-2)/2}) .$$

PROOF. We proceed as for Lemma 3.3 but use a different estimate for $E^*|E^* A_p^* | \varepsilon_j^*, j \neq j_p|$. We have to deal with $\delta_{nm}^* = E^*|E^* \exp(itn^{-1/2} \cdot (\varepsilon_n^* A_n^* + \alpha_n \varepsilon_n^{*2})) | \mathscr{D}_j^*, j \neq n|$ where $A_n^* = Y_{n-1}^* + \alpha_n Y_{n+2}^* + \varepsilon_{n+1}^* + \alpha_n^2 \varepsilon_{n-1}^*$.

Note that $\delta_{nm}^* = E^*|1 - (t_n'/2n) D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n + (y/6)(\|t_n\|^3 / n^{3/2}) E^* \cdot \|(\varepsilon_n^*, \varepsilon_n^{*2})\|^3|$ where $t_n' = (tA_n^*, t\alpha_n)$, $|y| \leq 1$. Thus

$$\delta_{nm}^* \leq E^* \left| 1 - \frac{t_n'}{2n} D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n \right| + \frac{E^*(\|t_n\|^3)}{6n^{3/2}} \mu_{3n}^* ,$$

where

$$\mu_{3n}^* = E^* \|(\varepsilon_n^*, \varepsilon_n^{*2})\|^3 \to E\|(\varepsilon_1, \varepsilon_1^2)\|^3 \quad \text{a.s,}$$
$$E^*(\|t_n\|^3) \leq |t|^3 [E^*(A_n^{*2} + \alpha_n^2)^3]^{1/2} .$$

Note that $E^*(A_n^{*2} + \alpha_n^2)^3 \to E(A^2 + \alpha^2)^3$ a.s. where $A = Y_1 + \alpha Y_3 + Y_2 + \alpha^2 Y_1$. Thus

$$\frac{E^*(\|t_n\|^3)}{6n^{3/2}} \mu_{3n}^* \leq \frac{C\|t\|^3}{n^{3/2}} \leq CC_1 \frac{\|t\|^2}{n} \quad \text{a.s.}$$

Let $\bar{\lambda}(A)$ and $\underline{\lambda}(A)$ denote, respectively, the maximum and minimum eigenvalues of A. Denote $\tilde{\Sigma} = D(\varepsilon_1, \varepsilon_1^2)$, $\tilde{\Sigma}_n = D(\varepsilon_1^*, \varepsilon_1^{*2})$. Note that $\bar{\lambda}(\tilde{\Sigma}_n) \to \bar{\lambda}(\tilde{\Sigma}) > 0$ a.s. and $\underline{\lambda}(\tilde{\Sigma}_n) \to \underline{\lambda}(\tilde{\Sigma}) > 0$ a.s. (by Lemma 3.1). Then we have the following relations.

$$E^*\left|1-\frac{t_n'}{2n}D(\varepsilon_n^*,\varepsilon_n^*)^2t_n\right|$$

$$\leq\left[E^*\left\{1-\frac{t_n'}{2n}D(\varepsilon_n^*,\varepsilon_n^{*2})+\left(\frac{t_n'D(\varepsilon_n^*,\varepsilon_n^{*2})}{2n}\right)^2\right\}\right]^{1/2},$$

$$E^*\left(\frac{t_n'\tilde{\Sigma}_nt_n}{2n}\right)^2\leq\bar{\lambda}^2(\tilde{\Sigma}_n)E^*\left(\frac{\|t_n\|^4}{4n^2}\right)\leq CC_1\frac{\|t\|^2}{n}\quad\text{a.s.}$$

$$E^*\left(\frac{t_n'\tilde{\Sigma}_nt_n}{2n}\right)\geq\underline{\lambda}(\tilde{\Sigma}_n)E^*\left(\frac{\|t_n\|^2}{2n}\right)\geq C\underline{\lambda}(\tilde{\Sigma}_n)\frac{\|t\|^2}{n}\quad\text{a.s.}$$

By combining these estimates and choosing a C_1 sufficiently small,

$$\delta_{nm}\leq1-\gamma\frac{\|t\|^2}{n}\leq\exp\left(-\frac{\gamma\|t\|^2}{n}\right)\quad\text{for some}\quad\gamma>0\quad\text{a.s.}$$

A look at the proof of Lemma 3.43 of GH shows that this proves the lemma. □

From Lemmas 3.2–3.4 and Lemma 1 of Babu and Singh (1984) (henceforth referred to as BS), we have the following theorems for the cases $l=1$ and $l>1$, respectively.

THEOREM 3.2. *Assume* (A1), (A2) *and* $|\alpha|<1$. *Suppose* $f\colon\mathbb{R}\to\mathbb{R}$ *is such that* $|f(x)|\leq M(1+|x|^2)$. *Let* $\sigma_n^{*2}=E^*(S_n^{*2})$. *For a.e.* Y_0, $Y_1,\ldots$ *and uniformly over* $x\in\mathbb{R}$,

(a) $\left|E^*f(S_n^*)-\int fd\psi_{n,3}^*\right|\leq Cw(f,n^{-k},\sigma_n^{*2})+o(n^{-1/2})$.

(b) $P^*(\sigma_n^{*-1}S_n^*\leq x)=\int_{-\infty}^x d\psi_{n,3}^*(\sigma_n^*y)+o(n^{-1/2})=P(\sigma^{-1}S_n\leq x)+o(n^{-1/2})$.

We omit the proof. See BS for a proof in the i.i.d. case.

THEOREM 3.3. *Let* H *be a function from* $\mathbb{R}^l\to\mathbb{R}$ *which is thrice continuously differentiable in a neighbourhood of* 0. *Let* h *denote the vector of first order partial derivatives of* H *at* 0. *Assume* $h\neq0$ *and that* (α_i) *satisfies the invertibility condition. Let*

$$\Sigma=\lim_{n\to\infty}D\left(n^{-1/2}\sum_{k=1}^nY_kY_{k-i},\ 1\leq i\leq l\right),$$

$$\Sigma_n^* = D^*\left(n^{-1/2}\sum_{k=1}^{n} Y_k^* Y_{k-i}^*,\ 1 \le i \le l\right),$$

$$\tilde{\beta}_i = E(Y_k Y_{k-i}), \qquad \tilde{\beta}_{in} = E^*(Y_k^* Y_{k-i}^*), \qquad i = 1,\dots,l.$$

$$T(F) = n^{1/2}\left[H\left(n^{-1}\sum_{k=1}^{n}(Y_k Y_{k-i} - \tilde{\beta}_i),\ i = 1,\dots,l\right) - H(0)\right],$$

$$\sigma^2 = l'\Sigma l.$$

$$T(F_n^*) = n^{1/2}\left[H\left(n^{-1}\sum_{k=1}^{n}(Y_k^* Y_{k-i}^* - \tilde{\beta}_{in}^*),\ i = 1,\dots,l\right) - H(0)\right],$$

$$\sigma_n^{*2} = l'\Sigma_n^* l.$$

Then $\sup_x |P(\sigma^{-1}T(F) \le x) - P^*(\sigma_n^{*-1}T(F_n^*) \le x)| = o(n^{-1/2})$ *a.s.*

PROOF. Proposition 3.1 and Lemmas 3.2–3.4 remain valid, respectively, for

$$\left(n^{-1/2}\sum_{k=1}^{n}(Y_k Y_{k-i} - \tilde{\beta}_i),\ i = 1,\dots,l\right) \quad \text{and}$$
$$\left(n^{-1/2}\sum_{k=1}^{n}(Y_k^* Y_{k-i}^* - \tilde{\beta}_{in}^*),\ i = 1,\dots,l\right).$$

Arguments analogous to Theorem 3 and Corollary 2 of BS yield the theorem. We omit the details which involve a Taylor expansion of H and a change of variable formula. □

The above result is true for vector valued H with proper modifications since Theorem 3 and Corollary 2 of BS remain true for such functions. The estimates of $\alpha_1,\dots,\alpha_l$ in a general MA model are smooth functions of $n^{-1}\sum_{j=1}^{n} Y_j Y_{j-i}$, $i = 1,\dots,l$. Hence Theorem 3.9 can be utilized to prove the results for these parameter estimates.

THEOREM 3.4. *Under the assumptions* (A1) *and* (A2) *for a.e.* Y_0, $Y_1,\dots$,

(a) *Let* $l = 1$, *and* $|\alpha| < 1$. *Let* σ^2 *and* σ_n^2 *be respectively the limiting variance of* $n^{1/2}(\alpha_n - \alpha)$ *and the variance of* $n^{1/2}(\alpha_n^* - \alpha_n)$ (*given* $Y_0, Y_1,\dots, Y_n$). *Then*

$$\sup_x |P(n^{1/2}(\alpha_n - \alpha)/\sigma_n \le x) - P(n^{1/2}(\alpha_n^* - \alpha_n)/\sigma_n \le x)| = o(n^{-1/2}).$$

(b) *Let* $l \ge 2$ *and* (α_i) *satisfy the invertibility condition. Let* G_n *be the distribution function of* $\Sigma^{-1/2}n^{1/2}(\alpha_n - \alpha_i,\dots,\alpha_{ln} - \alpha_l)$, *where* Σ *is the limiting* *covariance matrix of* $n^{1/2}(\alpha_{1n}-\alpha_1,\ldots,\alpha_{ln}-\alpha_l)$. *Let* G_n^* *be the corresponding bootstrapped version. Then*

$$\sup_{x\in\mathbf{R}^l} |G_n(x)-G_n^*(x)| = o(n^{-1/2}).$$

PROOF. The case $l=1$ is Theorem 3.2. For $l=2$, the moment equations are

$$\alpha_{2n}=n^{-1}\sum_{t=1}^{n} Y_t Y_{t-2} \quad\text{and}\quad \alpha_{1n}(1+\alpha_{2n})=n^{-1}\sum_{t=1}^{n} Y_t Y_{t-1}.$$

Thus

$$(\alpha_{1n}-\alpha_1,\ \alpha_{2n}-\alpha_2)=\left(\frac{\bar{Z}_{1n}+\alpha_1(1+\alpha_2)}{1+\alpha_2+\bar{Z}_{2n}}-\alpha_1,\ \bar{Z}_{2n}\right),$$

where

$$n^{-1}\sum_{t=1}^{n}(Y_tY_{t-1}-\tilde{\beta}_1)=\bar{Z}_{1n} \quad\text{and}\quad n^{-1}\sum_{t=1}^{n}(Y_tY_{t-2}-\tilde{\beta}_2)=\bar{Z}_{2n}.$$

Now the result follows from the multidimensional version of Theorem 3.3. The idea of proof for a general l is clear from what we have shown. However, solving for the estimates $\alpha_{1n},\ldots,\alpha_{ln}$ becomes increasingly difficult with an increase in l. □

Remark 3.1. For i.i.d. observations, Hall (1988) has shown that error rates of $O(n^{-1})$ can be achieved for quantile estimates. This is based on a $O(n^{-1})$ expansion of the bootstrap statistic. Abramovitch and Singh (1985) have shown that an error rate of $o(n^{-(s-2)/2})$, $s\geq 3$ can be obtained for the cdf of a modified bootstrap statistic provided that a sufficiently high order Edgeworth expansion is valid for the bootstrap statistic. Our attempts to derive $O(n^{-1})$ results in the present context have not been successful since we have not yet been able to prove a higher order Edgeworth expansion for the bootstrap distribution.

Remark 3.2. The assumption that (ε_t) has mean 0 and variance 1 was imposed to keep the proofs simpler. We sketch below how the case $E\varepsilon_t=\mu$, $E\varepsilon_t^2=\sigma^2$ (both μ and σ^2 unknown) can be tackled. We illustrate the case $l=1$ only.

The model in this case is $Y_t=\mu+\varepsilon_t+\alpha\varepsilon_{t-1}$ where (A1) and (A2) hold but $E\varepsilon_t^2=\sigma^2>0$. Under the assumptions (A1) and (A2), the Edgeworth expansion is valid for the distribution of

$$(3.3) \qquad n^{-1/2}\left(\sum_{t=1}^{n}(Y_t-\gamma_1),\ \sum_{t=1}^{n}(Y_tY_{t-1}-\gamma_2),\ \sum_{t=1}^{n}(Y_t^2-\gamma_3)\right)$$

where $\gamma_1 = EY_t$, $\gamma_2 = EY_tY_{t-1}$ and $\gamma_3 = EY_t^2$.

The estimates μ_n, α_n and σ_n^2 of μ, α and σ^2 are obtained by solving the following moment equations:

$$n^{-1}\sum_{t=1}^{n} Y_t = \mu_n, \qquad n^{-1}\sum_{t=1}^{n} Y_tY_{t-1} = \mu_n^2 + \alpha_n\sigma_n^2 \qquad \text{and}$$

$$n^{-1}\sum_{t=1}^{n} Y_t^2 = \mu_n^2 + \alpha_n^2(1+\sigma_n^2)\,.$$

Hence

$$\mu_n = n^{-1}\sum_{t=1}^{n} Y_t, \qquad \sigma_n^2 = [y_2^2 + (y_2^4 + 4y_2^2y_3)^{1/2}]/2y_3 \qquad \text{and} \qquad \alpha_n = y_2/\sigma_n$$

where

$$y_2 = n^{-1}\sum_{t=1}^{n} Y_tY_{t-1} - \mu_n^2 \qquad \text{and} \qquad y_3 = n^{-1}\sum_{t=1}^{n} Y_t^2 - \mu_n^2\,.$$

Thus all these estimates are smooth functions of $\sum_{t=1}^{n} Y_t$, $\sum_{t=1}^{n} Y_tY_{t-1}$ and $\sum_{t=1}^{n} Y_t^2$.

Hence for a normalizing factor β_0, the distribution of $n^{1/2}\beta_0(\alpha_n - \alpha)$ admits an Edgeworth expansion of order $o(n^{-1/2})$, with the leading term as $\Phi(x)$ and the coefficients in the second term (which is $O(n^{-1/2})$) being smooth functions of α, μ and σ^2 and of moments of Y_t, Y_tY_{t-1} and Y_{t-1}^2 of order less or equal to three. β_0 can be explicitly calculated and depends on α, μ and moments of ε_1. The empirical distribution is computed as before, the only difference is that Y_i's are now replaced by $Y_i - \mu_n$. As in the case $\mu = 0$, $\sigma^2 = 1$, an asymptotic expansion is valid for the bootstrapped version of (3.3), which yields an expansion of order $o(n^{-1/2})$ for the distribution of $n^{1/2}\beta_n(\alpha_n^* - \alpha_n)$ where β_n is the bootstrap equivalent of β_0. The leading term in this expansion is also $\Phi(x)$ and the polynomial involved in the second term is of the same form as that in the expansion of $n^{1/2}\beta_0(\alpha_n - \alpha)$. By the ergodic theorem, the empirical moments of Y_t, Y_tY_{t-1} and Y_t^2 converge to the true moments a.s. and hence α_n, μ_n and σ_n are strongly consistent estimates of α, μ and σ, respectively. Thus the difference between the two expansions is $o(n^{-1/2})$ a.s.

4. Simulations

It is interesting to see how the bootstrap performs in small samples. The accuracy is expected to decrease as the parameter values move towards the boundary (for $l = 1$, as $|\alpha| \to 1$). A small simulation study for the moving average model with $l = 1$ was done. We also simulated the autoregressive process

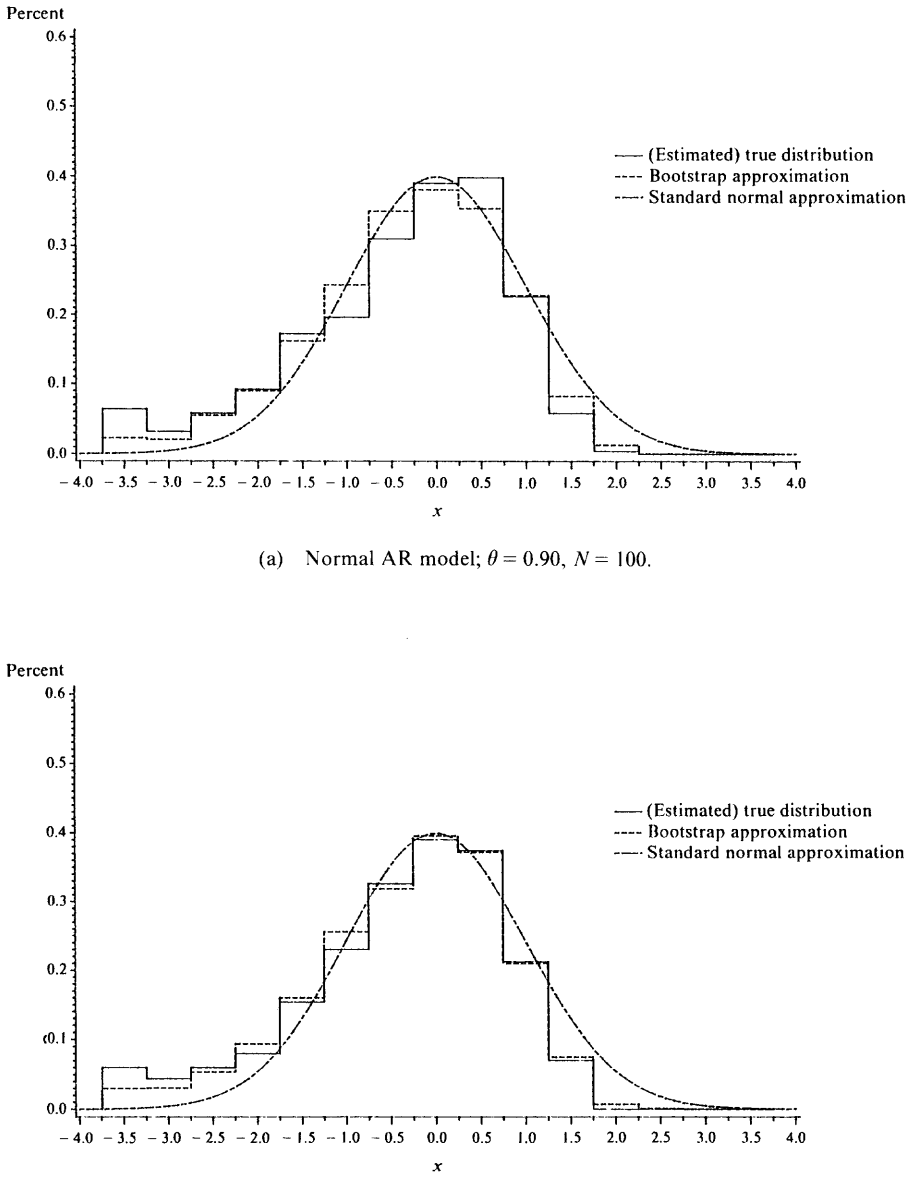
$$Y_t = \theta Y_{t-1} + \varepsilon_t, \qquad |\theta| < 1$$

when θ is estimated by the least squares method. The rate of $o(n^{-1/2})$ is also valid for this situation as was shown in Bose (1988*a*). See Bose (1988*a*) for the details of bootstrapping the distribution of the least squares estimates.

For both the MA and AR models, we generated ε_i's from $N(0, 1)$ and centered exp (1) densities. The parameter values were set at $\alpha = 0.9$ and $\theta = 0.9$ and a series of size $n = 100$ was generated. The distribution of the estimator, standardized by its true mean and true limiting variance, was approximated by using 1000 replications of the series. The first set of $n = 100$ observations was used to estimate the residuals and generate the bootstrap distribution. The bootstrap distribution was approximated by using 5000 repetitions for the AR case and 10,000 repetitions for the MA case.

The true (approximate) distribution, the bootstrap distribution and the standard normal distribution have been shown in each case in Figs. 1(a)–(d). It is evident that the bootstrap works very well in the AR case and reasonably well in the MA case. Similar results were seen to hold for other parameter values. In fact the bootstrap does better as we move away from the boundary values of ± 1. In the AR case, the sampling distribution for the normal and the exponential are close. This robustness is absent in the MA model. The bootstrap captures this deviation to some extent. However, it is not clear why the bootstrap performs better for the exponential than the normal. This is an interesting topic for further research. Apparently the skewness of the underlying distribution together with the structure of the model is affecting the performance of the bootstrap in some way.

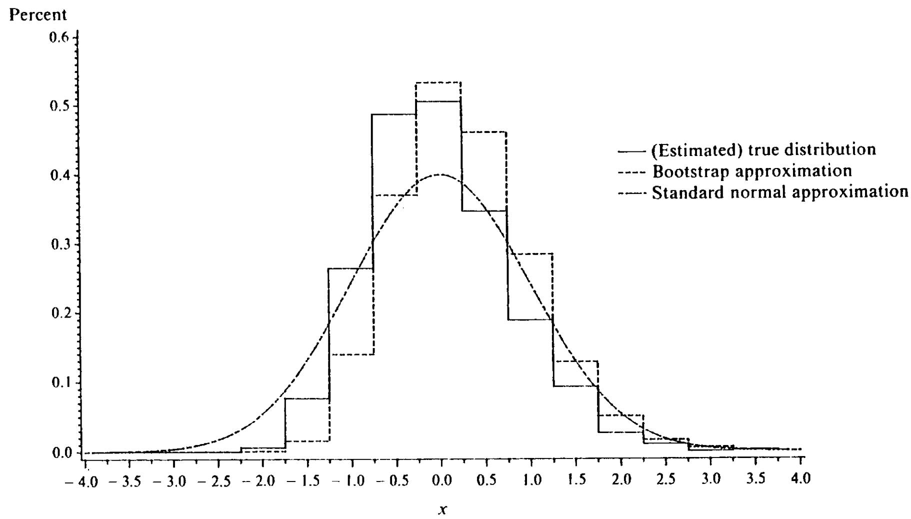
See Chatterjee (1985) for some more simulation studies. The study of the behavior of the bootstrap in other complicated time series models is still open. The author is currently working on the bootstrap in the class of nonlinear autoregressive models.



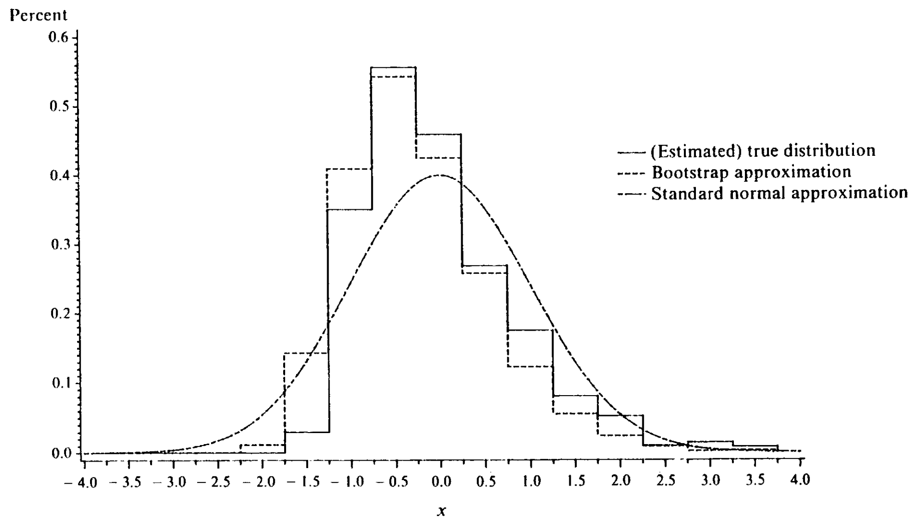
(a) Normal AR model; $\theta = 0.90$, $N = 100$.

(b) Exponential AR model; $\theta = 0.90$, $N = 100$.

Fig. 1. Histogram of estimator in Bootstrap study.



(c) Normal MA model; $\alpha = 0.90$, $N = 100$.



(d) Exponential MA model; $\alpha = 0.90$, $N = 100$.

Fig. 1. (continued).

Acknowledgements

The author is grateful to Professor G. J. Babu of Pennsylvania State University for his help and suggestions. The contents of this paper is part of the author's Ph.D. thesis written under his guidance at the Indian Statistical Institute. He is also grateful to anonymous referees who read an earlier version of the paper. The three referees of the current version have all made very valuable suggestions. These suggestions have greatly benefited this paper. The author is also extremely grateful to Dr. Tai Fang Lu of Purdue University for her generous help in the simulation study.

REFERENCES

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap, *Ann. Statist.*, **13**, 116–132.

Babu, G. J. and Singh, K. (1984). On one term Edgeworth correction by Efron's bootstrap, *Sankhyā Ser. A*, **40**, 219–232.

Basawa, I. V., Mallik, A. K., McCormick, W. P. and Taylor, R. L. (1989). Bootstrapping explosive autoregressive processes, *Ann. Statist.*, **17**, 1479–1486.

Beran, R. (1982). Estimated sampling distribution: the bootstrap and competitors, *Ann. Statist.*, **10**, 212–225.

Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions*, Wiley, New York.

Bickel, P. J. and Freedman, D. (1980). On Edgeworth expansion for the bootstrap, Preprint, University of California, Berkeley.

Bickel, P. J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, 1196–1217.

Bose, A. (1988*a*). Edgeworth correction by bootstrap in autoregressions, *Ann. Statist.*, **16**, 1709–1722.

Bose, A. (1988*b*). Higher order approximations for auto-covariances from linear processes with applications, *Statistics*, **19**, 259–269.

Chatterjee, S. (1985). Bootstrapping ARMA models: some simulations, Preprint, College of Business Administration, Northeastern University, Boston.

Efron, B. (1979). Bootstrap methods: another look at Jackknife, *Ann. Statist.*, **7**, 1–26.

Efron, B. (1982). The Jackknife, the bootstrap and other resampling plans, *CBMS-NSF Regional Conference Series in Applied Maths. Monograph*, **38**, SIAM, Philadelphia.

Freedman, D. (1984). On bootstrapping two stage least squares estimates in stationary linear models, *Ann. Statist.*, **12**, 827–842.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors, *Z. Wahrsch. Verw. Gebiete*, **64**, 211–239.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *Ann. Statist.*, **16**, 927–952.

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

Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.

Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap, *Ann. Statist.*, **9**, 1187–1195.