Bootstrap in Moving Average Models

by

Arup Bose
Indian Statistical Institute, Calcutta
and
Purdue University, West Lafayette

Technical Report #87-55

Department of Statistics
Purdue University

December 1987
First revision July 1989
Second revision September 1989

*

Bootstrap in Moving Average Models

by

Arup Bose
Indian Statistical Institute, Calcutta
and
Purdue University, West Lafayette

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 are accurate upto 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.

AMS 1980 Subject Classification: Primary 62F12, 62M10

Secondary 60F99, 60G10

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 (e.g. asymptotic normality, Edgeworth expansion etc.). A few of the references are Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982), Babu and Singh (1984) and Hall (1988). One class of results say that in i.i.d. situation where the normal approximation holds with error $0(n^{-1/2})$, if we replace the normal distribution function by the sample dependent bootstrap distribution, then the error rate is $o(n^{-1/2})$ a.s.

The bootstrap does not give the correct answers in general dependent models. However, some dependent models do allow for an appropriate resampling so that the bootstrap works. Freedman (1984) shows 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. (1987) prove the validity of bootstrap in unstable and explosive first order autoregressions. Bose (1988) shows that the rate result alluded to in the i.i.d. situation holds for stationary autoregressions.

In this paper we deal with moving average models. The moment estimators of the parameters have an asymptotic normal distribution and the error of approximation can be shown to be $0(n^{-1/2})$. The structure of the process enables appropriate resampling. We show that the bootstrap distribution approximates the distribution of the parameter estimates with accuracy $o(n^{-1/2})$ a.s. The idea is to develop one term Edgeworth expansion for the distribution of the parameter estimates and its 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

$$Y_t = \varepsilon_t + \sum_{i=1}^{\ell} \alpha_i \varepsilon_{t-i},$$

where we assume that

(A1) (ε_t) are 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

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

$\alpha_1,\alpha_2,\ldots\alpha_\ell$ are unknown parameters, which will be estimated by moment estimates.

Remark 2.1. The assumption that the mean and variance of ε_t are known have been made to keep the proofs simple. See Remark 2.12 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 $(s+1)$th moment of ε_t^2 with s at least 3. This is in contrast to the situation of i.i.d. observations where sth moment suffices to derive an expansion of $o(n^{-(s-2)/2})$.

We first assume that $\ell=1$, 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,

$$\alpha_n=n^{-1}\sum_{t=1}^{n}Y_tY_{t-1}.$$

It is well known (see Hall and Heyde (1980), pp. 197–198) that under our assumptions, α_n is a strongly consistent estimate of α. Moreover $n^{1/2}(\alpha_n-\alpha)$ has an asymptotic normal distribution.

Define $\tilde{\varepsilon}_i=\sum_{j=0}^{i-1}(-1)^j\alpha^jY_{i-j}, i=2,\ldots,n$, and $\tilde{\varepsilon}_1=Y_1$. Using the structure of the process,

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

This shows that $\tilde{\varepsilon}_i$ and ε_i are close enough for large i, only if $|\alpha|<1$, which in turn shows that resampling is proper only in this situation. (For $p>1$, this condition should be replaced by the invertibility condition (see Hannan (1970)).

So motivated by (2.1), we compute the pseudo errors as

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

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 $\hat{F}_n(x) = G_n(x - \overline{\varepsilon}_n)$ where $\overline{\varepsilon}_n = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{in}$. 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

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

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 $n \to \infty$ in the next section.

Before we discuss the main results, we wish to introduce a few notations.

C will stand for a generic constant, and 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,

$$S_n = n^{-1/2}\sum_{t=1}^{n} X_t.$$

The symbol $G_n \Rightarrow G$ will denote that the distribution G_n converges weakly G. (G_n may be random).

The function $\psi_{n,s}$ will denote the usual function associated with Edgeworth expansions. This function 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, page 145) for the definition of $\psi_{n,s}$ when X_t are i.i.d. Gotze and Hipp (1983) may be consulted for a definition when X_t are dependent.

$\beta = (\beta_1 \ldots \beta_k)$ denotes a vector where each β_i is a nonnegative negative integer and for any β, and $f: \mathrm{R}^k \to \mathrm{R}$

$$D^{\beta} f(x) = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \ldots \partial x_k^{\beta_k}} f(x_1 \ldots x_k)$$

where

$$|\beta| = \beta_1 + \beta_2 + \ldots + \beta_k.$$

For any random vector X, $D(X)$ will denote the dispersion matrix of X.

3. The 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

(a) $n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_i^k \xrightarrow{a.s.} E_{F_0}(\varepsilon_1^k) \qquad \forall k \leq 2(s+1)$

(b) $n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{in}^k \xrightarrow{a.s.} E_{F_0}(\varepsilon_1^k) \qquad \forall k \leq 2(s+1)$

(c) $\tilde{F}_n \Rightarrow F_0$ a.s.

(d) $\hat{F}_n \Rightarrow F_0$ a.s.

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

(a) By strong law of large numbers, 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. Further

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

It follows easily that $n^{-1}\sum_{i=1}^{n}|\varepsilon_i|^j|\alpha|^i \overset{a.s.}{\to} 0\ \forall j \le k-1$. (Use e.g. Theorem 2.18 of Hall and Heyde (1980)). This proves (a).

(b) By (a), it suffices to show that $n^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_i^k - \tilde{\varepsilon}_i^k) \overset{a.s.}{\to} 0$. Note that $\alpha_n \overset{a.s.}{\to} \alpha$ and $|\alpha| < 1$. Hence for large n,

$$|\alpha| + |\alpha_n - \alpha| \le \tilde{\beta} < 1 \text{ a.s.} \dots\dots\dots \tag{2.2}$$

Also note that $\forall_j \ge 1$,

$$\begin{aligned} |\alpha_n^j - \alpha^j| &= |(\alpha_n - \alpha + \alpha)^j - \alpha^j| \\ &\le Cj|\alpha_n - \alpha|\tilde{\beta}^{j-1} \\ &\le C|\alpha_n - \alpha|\delta^j \text{ for some } \delta < 1\dots \end{aligned} \tag{2.3}$$

Hence $|n^{-1}\sum_{i=1}^{n}(\tilde{\varepsilon}_i^k - \hat{\varepsilon}_i^k)|$

$$= |n^{-1}\sum_{i=2}^{n}\left[\left(\sum_{j=0}^{i-1}(-1)^j\alpha^jY_{i-j}\right)^k - \left(\sum_{j=0}^{i-1}(-1)^j\alpha_n^jY_{i-j}\right)^k\right]|$$

$$\le n^{-1}\sum_{i=2}^{n}\left[2^{k-1}\sum_{j=0}^{i-1}|\alpha_n^j - \alpha^j||Y_{i-j}|\left\{\left(|\sum_{t=0}^{i-1}(-1)^t\alpha_n^tY_{i-j}|\right)^{k-1} + \left(|\sum_{t=0}^{i-1}(-1)^t\alpha^tY_{i-j}|\right)^{k-1}\right\}\right]$$

$$\le Cn^{-1}|\alpha_n - \alpha|\sum_{i=2}^{n}\left(\sum_{j=0}^{i-1}\delta^j|Y_{i-j}|\right)^k \quad \text{(by (2.2) and (2.3))}$$

Thus it is sufficient to show that

$$n^{-1}\sum_{i=2}^{n}\left(\sum_{j=0}^{i-1}\delta^j|Y_{i-j}|\right)^k \quad \text{is bounded a.s.}$$

But note that $Y_i = \varepsilon_i + \alpha\varepsilon_{i-1}$. Hence 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.

Define
$$Z_i = \sum_{j=0}^{\infty}\delta^j|\varepsilon_{i-j}|\dots\dots$$

Then the sequence (Z_i) is a stationary autoregressive process of order one and hence ergodic. (See Hannan (1970), page 204).

$$\text{Thus} \quad n^{-1}\sum_{i=1}^{n} Z_i^k \overset{a.s.}{\rightarrow} E(Z_1^k) < \infty.$$

$$\text{But} \quad \left(\sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}|\right)^k \leq Z_i^k.$$

This proves (b).

(c) Since (ε_i) are i.i.d. F_0, this readily follows from (2.1).

(d) Note that 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,

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

(d) follows easily from this observation.

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

Let (X_t) be $\mathbb{R}^k$ valued random variables on $(\Omega, \mathcal{F}, P)$. Introduce the following conditions.

Let there be σ-fields $\mathcal{D}_j$ (write $\sigma(\bigcup_{j=a}^{b} \mathcal{D}_j) = \mathcal{D}_a^b$) 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) $\exists Y_{nm} \varepsilon \mathcal{D}_{n-m}^{n+m} \ni E||X_n - Y_{nm}|| \leq$ c.exp $(-\alpha^* m)$

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

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

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

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

Define the integer $s_o \leq s$ by

$$s_0 = \begin{cases} s & \text{if s is even} \\ s-1 & \text{if s is odd.} \end{cases}$$

Recall that $\psi_{n,s}$ is the usual function associated with Edgeworth expansions and $S_n = n^{-1/2} \sum_{t=1}^{n} X_t$. Let ϕ_Σ be the normal density with mean 0 and dispersion matrix Σ.

The following results are due to Gotze and Hipp (1983).

Theorem 2.2. Let $f:\ \mathbb{R}^k \to \mathbb{R}$ denote a measurable function such that $|f(x)| \leq M(1 + ||x||^{s_0})$ for every $x \varepsilon \mathbb{R}^k$. Assume that C(1)–C(7) hold. Then there exists a positive constant δ_0 not depending on f and M, and for arbitrary $k > 0$ there exists a positive constant C depending on M but not on f such that

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

where $w(f, n^{-k}) = \int sup(|f(x+y) - f(x)| : |y| \leq n^{-k})\phi_\Sigma(x)dx.$

The term $o(\cdot)$ depends on f through M only.

Corollary 2.3: Under assumptions C(1)–C(7) we have uniformly for convex measurable $C \subseteq \mathbb{R}^k$,

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

Let $X_t = Y_t Y_{t-1} - \alpha$ and $\mathcal{D}_j$ = 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. Thus we have the following proposition.

Proposition 2.4: 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 2.2 holds with the above S_n and as a consequence,

b) $P(S_n \varepsilon C) = \psi_{n,s}(C) + o(n^{-(s-2)/2})$ uniformly over convex subsets of 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 the bootstrapped quantity and hence expectation etc. are taken w.r.t. (ε_i^*) i.i.d. $\hat{F}_n$ given $Y_0, Y_1, \ldots, Y_n$.

Define $X_j^* = Y_j^* Y_{j-1}^* - \alpha_n, j \geq 1$

$H_n^*(t) =$ the characteristic function of $n^{-1/2} \sum_{j=1}^{n} X_j^*$.

We have the following lemmas. The proofs are only sketched and the details can be filled in from GH.

Lemma 2.5: $\forall |t| \leq C.n^{\varepsilon_0}$, we have

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

for some $\varepsilon_0 < 1/2$ and C depends on the bounds of $m_{s+1,n}^* = (s+1)$th moment of X_j^*. D^β is the usual differential operator, $\hat{\psi}_{n,s}^*(t)$ the Fourier transform of $\psi_{n,s}^*$, the usual function associated with Edgeworth expansions and $|\beta| \leq s+2$.

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

Let $I_1 = \{t : \; Cn^{\tilde{\varepsilon}_0} \leq |t| \leq C_1 n^{1/2}\}$

$I_2 = \{t : \; C_1 n^{1/2} \leq |t| \leq \tilde{\varepsilon}^{-1} n^{1/2}\}$ where C_1 is to be chosen and $0 < \tilde{\varepsilon} < 1$ is fixed.

Lemma 2.6: Under (A1) and (A2), we have for almost every sequence (Y_i),

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

Proof: A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that $E^*|E^*A_p^*|\mathcal{D}_j^*, j \neq j_p| < 1$ uniformly in $t \varepsilon I_2$ and $p = 1, 2, \ldots, J$ where

$A_p^* = \exp(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} Z_j^*), \mathcal{D}_j^* = \sigma(\varepsilon_j^*)$ and for definition of j_p and m see GH. We omit the details of the definitions since they are not used explicitly in the sequel. Suffices to note that j_p is fixed and the above expectation is independent of m (see below).

The above expectation equals

$$\delta_{nm}^* = E^*|E^* \exp(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} X_j^*)|\varepsilon_j^*,\ j \neq j_p|.$$

Note that $\sum_{j=j_p-m}^{j_p+m} X_j^* = \varepsilon_{j_p}^*(Y_{j_p-1}^* + \alpha_n Y_{j_p+2}^* + \varepsilon_{j_p+1}^* + \alpha_n^2 \varepsilon_{j_p-1}^*) + \alpha_n \varepsilon_{j_p}^{*2} + V$

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 $\delta_{nm}^* = \int | \int \exp(itn^{-1/2}xy + itn^{-1/2}\alpha_n x^2) d\hat{F}_n(x)| dK_n^*(y)$.

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 $\mathbb{R}^2$.

$$\delta_{nm}^* \leq \sup_{(\vartheta_1,\vartheta_2)\varepsilon D} \int | \int \exp(i\vartheta_1 xy + i\vartheta_2 x^2) d\hat{F}_n(x)| dK_n^*(y).$$

Let $b_1, b_2 > 0$ (to be chosen). Then

$$\delta_{nm}^* \leq K_n^*(b_1 \leq |Y| \leq b_2) I_{1n} + K_n^*(|Y| < b_1) + K_n^*(|Y| > b_2)$$

where $I_{1n} \leq \sup_{b_1 \leq |y| \leq b_2} \sup_{(\vartheta_1,\vartheta_2)\varepsilon D} |\int \exp(i\vartheta_1 xy + i\vartheta_2 x^2) d\hat{F}_n(x)|$.

Note that by Lemma 2.1, $K_n^* \Rightarrow K$ a.s. where K is the distribution function 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 large n,

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

Note that $F_n^* \Rightarrow F_o$ a.s. and we have Cramer's condition for $(\varepsilon_1, \varepsilon_1^2)$, $(\varepsilon_1, \varepsilon_1^2) \sim F_o$. Using the fact that the convergence of a sequence of characteristic functions to its limit is uniform over compact sets, we have $I_{1n} < 1 - \gamma < 1$ for large n. Thus $\delta_{nm}^* \leq (1-\gamma) + \alpha_o \gamma < 1$ proving the lemma.

Lemma 2.7: Under (A1) and (A2), for sufficiently small C_1, we have for almost every sequence (Y_i), $|\beta| \leq s+2$,

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

Proof: As in Lemma 2.6 it is sufficient to deal with the original variables instead of truncations. As before we proceed as in Lemma 2.6 following GH but use a different estimate for $E^*|E^* A_p^* | \varepsilon_j^*, j \neq j_p|$.

We have to deal with

$$\delta_{nm}^* = E^*|E^* \exp(it\ n^{-1/2}(\varepsilon_n^* A_n^* + \alpha_n \varepsilon_n^{*2}))|\mathcal{D}_j^*, j \neq n|$$

where $A_n^* = Y_{n-1}^* + \alpha_n Y_{n+2}^* + \varepsilon_{n+1}^* + \alpha_n^2 \varepsilon_{n-1}^*$.

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

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

where

$$\mu_{3n}^* = E^*||(\varepsilon_n^*, \varepsilon_n^{*2}||^3 \to E||(\varepsilon_1, \varepsilon_1^2)||^3 a.s.$$

$$E^*(||t_n||^3) \leq |t|^3 [E^*(A_n^{*2} + \alpha_n^2)^3]^{1/2}$$

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

Hence for some constant C,

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

On the other hand,

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

Let $\overline{\lambda}(A)$ 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 $\overline{\lambda}(\tilde{\Sigma}_n) \to \overline{\lambda}(\tilde{\Sigma}) > 0$ a.s. and $\underline{\lambda}(\tilde{\Sigma}_n) \to \underline{\lambda}(\tilde{\Sigma}) > 0$ a.s. (by Lemma 2.1).

$$E^*\left(\frac{t_n'\tilde{\Sigma}_n t_n}{2n}\right)^2 \leq \overline{\lambda}^2(\tilde{\Sigma}_n)E^*\left(\frac{\|t_n\|^4}{4n^2}\right)$$

and arguing as before, the proceeding quantity

$$\leq CC_1\frac{\|t\|^2}{n}a.s.$$

On the other hand,

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

Combining these estimates and choosing C_1 sufficiently small,

$$\delta_{nm} \leq 1 - \gamma\frac{\|t\|^2}{n}, \text{ for some } \gamma > 0 \ a.s.$$
$$\leq \exp\left(-\frac{\gamma\|t\|^2}{n}\right).$$

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

The following lemma is stated in Babu and Singh (1984) (henceforth referred as BS) and is a modified version of a lemma in Sweeting (1977).

Lemma 2.8: Let P and K be probability measures and Q be a signed measure on $\mathbb{R}^k$. Let f be a measurable function such that $|f(x)| \leq M(1 + |x|^s)$ for some $s \geq 2$. Further let

$\alpha = K(x : ||x|| \leq 1) > \frac{1}{2}$ and $M_0 = \int ||x||^{s+2} K(dx) < \infty$. Then for any $0 < \tilde{\varepsilon} < 1$,

$$\left|\int f d(P-Q)\right| \leq (2\alpha - 1)^{-1} [B(1-\alpha)/\alpha]^{-1+\tilde{\varepsilon}^{-1/4}} + \beta \varepsilon B$$
$$+ B \int (1 + ||x||^s) |K^*_{\tilde{\varepsilon}}(P-Q)|(dx)$$
$$+ \sup_{||x|| < \tilde{\varepsilon}^{1/4}} \int w(f, 2\tilde{\varepsilon}, x-y)|Q|dy$$

where

$$K_{\tilde{\varepsilon}}(dx) = K(\tilde{\varepsilon}^{-1}dx) \text{ and } B = 9^s M_s(f) \int (1 + ||x||^s)(P + |Q|)dx,$$
$$M_s(f) = sup(1 + ||x||^s)^{-1} |f(x)|.$$

Further we have for any $0 < ||x|| < 1,\ 0 < \delta_0 < 1$,

$$\int w(f, \delta_0, x-y)\phi(y)dy \leq 3 \int w(f, \delta, y)\phi(y)dy$$
$$+ C_o M_s(f) ||x||^{-k-s+1} \exp(-\frac{1}{8}||x||^{-2}).$$

From Lemmas 2.5 – 2.8, we have the following theorem.

Theorem 2.9: Assume (A1), (A2) and $|\alpha| < 1$. Suppose $f : \mathbb{R} \to \mathbb{R}$ is such that $|f(x)| \leq M(1 + |x|^2)$. Let $\sigma_n^{*2} = E^*(S_n^{*2})$. For a.e. $Y_o, Y_1, \ldots$ and uniformly over $x \varepsilon \mathbb{R}$,

(a) $|E^* f(S_n^*) - \int f d\psi^*_{n,3}| \leq C.w(f, n^{-k}, \sigma_n^{*2}) + o(n^{-1/2})$

(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. For a proof in the i.i.d. case see BS.

For $p > 1$, we have correspondingly the following results.

Theorem 2.10: Let H be a function from $\mathbb{R}^p \to \mathbb{R}$ which is thrice continuously differentiable in a neighbourhood of 0. Let ℓ denote the vector of first order partial derivatives of H at 0. Assume $\ell \neq 0$ and that (α_i) satisfies the invertibility condition.

Let

$$T(F) = n^{1/2}\left[H(n^{-1}\sum_{k=1}^{n}(Y_kY_{k-i} - \tilde{\beta}_i),\ i = 1,\ldots,\ell) - H(0)\right],\ \sigma^2 = \ell'\Sigma\ell$$

$$T(F_n^*) = n^{1/2}\left[H(n^{-1}\sum_{k=1}^{n}(Y_k^*Y_{k-i}^* - \tilde{\beta}_{in}^*),\ i = 1,\ldots,\ell) - H(0)\right],\ \sigma_n^{*2} = \ell'\Sigma_n^*\ell$$

where

$$\Sigma = \lim_{n\to\infty} D(n^{-1/2}\sum_{k=1}^{n} Y_kY_{k-i},\ i = 1,\ldots,\ell)$$

$$\Sigma_n^* = D^*(n^{-1/2}\sum_{k=1}^{n} Y_k^*Y_{k-i}^*,\ i = 1,\ldots,\ell)$$

$$\tilde{\beta}_i = E(Y_kY_{k-i}),\ \tilde{\beta}_{in} = E^*(Y_k^*Y_{k-i}^*)\ i = 1,\ldots,\ell.$$

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 2.4 and Lemma 2.5 – 2.7 remain valid for

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

respectively. Thus arguments analogous to Theorem 3 and Corollary 2 of BS yields the theorem. We omit the details which involve Taylor expansion of H and a kind of change of variable formula.

The above result is true with vector valued H with proper modifications. This is because Theorem 3 and Corollary 2 of BS remain true for such functions. The estimates of $\alpha_1,\ldots,\alpha_\ell$ in general moving average model are smooth functions of $n^{-1}\sum_{j=1}^{n} Y_jY_{j-i}$ $i = 1,\ \ldots,\ell$.

Hence Theorem 2.10 can be utilized to prove results for these parameter estimates.

Theorem 2.11: Under assumptions (A1) and (A2) for a.e. $Y_o, Y_1, \ldots,$

(a) If $\ell = 1$, and $|\alpha| < 1$,

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

where

$$\sigma^2 = \text{ limiting variance of } n^{1/2}(\alpha_n - \alpha)$$
$$\sigma_n^2 = \text{ Variance of } n^{1/2}(\alpha_n^* - \alpha_n) \text{ (given } Y_o, Y_1, \ldots, Y_n).$$

(b) Let $\ell \geq 2$ and (α_i) satisfy the invertibility condition.

Let G_n denote the distribution function of $\Sigma^{-1/2} n^{1/2}(\alpha_n - \alpha_i, \ldots, \alpha_{pn} - \alpha_p)$, where Σ is the limiting variance–covariance matrix of $n^{1/2}(\alpha_{1n} - \alpha_1, \ldots, \alpha_{pn} - \alpha_p)$. Let G_n^* denote the corresponding bootstrapped distribution function. Then

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

Proof: (a) The case $\ell = 1$ is Theorem 2.9.

(b) For $\ell = 2$, the moment equations are

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

Thus

$$\alpha_{2n} - \alpha_2 = n^{-1} \sum_{t=1}^{n} (Y_t Y_{t-2} - \tilde{\beta}_2) = \overline{Z}_{2n} \text{ say}$$
$$\alpha_{1n}(1 + \alpha_{2n}) - \alpha_1(1 + \alpha_2) = n^{-1} \sum_{t=1}^{n} (Y_t Y_{t-1} - \tilde{\beta}_1) = \overline{Z}_{1n} \text{ say}$$
$$\text{Thus } (\alpha_{1n} - \alpha_1, \alpha_{2n} - \alpha_2) = \left(\frac{\overline{Z}_{1n} + \alpha_1(1 + \alpha_2)}{1 + \alpha_2 + \overline{Z}_{2n}} - \alpha_1, \ \overline{Z}_{2n} \right).$$

Now the result follows from the multidimensional version of Theorem 2.10.

The idea of proof for general p is clear from what we have shown. However, solving for the estimates $\alpha_{1n}, \ldots, \alpha_{\ell n}$ becomes increasingly difficult with increase in ℓ.

Remark 2.12. In the situation of 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 a sufficiently high order Edgeworth expansion is valid for the bootstrap statistic. Our attempts to derive $O(n^{-1})$ results in the present context has not been successful since we have not yet been able to prove a higher order Edgeworth expansion for the bootstrap distribution.

Remarks 2.13: The assumption that (ε_t) have 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 $p = 1$ only.

The model in this case is,

$$Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1} \text{ where (A1), (A2) hold but } E\ \varepsilon_t^2 = \sigma^2 > 0$$

Under assumptions (A1), (A2), Edgeworth expansion is valid for the distribution of

$$n^{-1/2}\left(\sum_{t=1}^{n}(Y_t - \gamma_1),\ \sum_{t=1}^{n}(Y_tY_{t-1} - \gamma_2),\ \sum_{t=1}^{n}(Y_t^2 - \gamma_3)\right) \cdots \cdots \tag{2.5}$$

where $\gamma_1 = E\ Y_t$, $\gamma_2 = E\ Y_tY_{t-1}$ and $\gamma_3 = E\ Y_t^2$.

Estimates μ_n, α_n and σ_n^2 of μ, α and σ^2 are obtained by solving

$$n^{-1}\sum_{t=1}^{n} Y_t = \mu_n$$

$$n^{-1}\sum_{t=1}^{n} Y_tY_{t-1} = \mu_n^2 + \alpha_n\sigma_n^2$$

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

These equations give the moment estimates.

Hence

$$\mu_n = n^{-1} \sum_{t=1}^{n} Y_t$$

$$\sigma_n^2 = \left[y_2^2 + (y_2^4 + 4y_2^2 y_3)^{1/2} \right] / 2y_3$$

$$\text{where } y_2 = n^{-1} \sum_{t=1}^{n} Y_t Y_{t-1} - \mu_n^2$$

$$y_3 = n^{-1} \sum_{t=1}^{n} Y_t^2 - \mu_n^2.$$

Note that the positive square root has to be taken since as $n \to \infty$, a.s. $y_3 > 0$.

$$\alpha_n = y_2 / \sigma_n.$$

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

Hence for a suitable normalizing factor β_0, the distribution of $n^{1/2}\beta_0(\alpha_n - \alpha)$ admits an Edgeworth expansion upto $o(n^{-1/2})$, with the leading term as $\Phi(x)$, and the coefficients involved in the second term (which is $0(n^{-1/2})$) are smooth functions of α, μ and σ^2 and of moments of $Y_t, Y_t Y_{t-1}$ 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 by proceeding as before, the only difference is that Y_i's are now replaced by $Y_i - \mu_n$.

Proceeding as in the case $\mu = 0, \sigma^2 = 1$, an asymptotic expansion is valid for the bootstrapped version of (2.5), 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_ϑ. 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 ergodic theorem the empirical moments of $Y_t, Y_t Y_{t-1}$ 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 $\ell = 1$, as $|\alpha| \to 1$). A small simulation study for the moving average model with $\ell = 1$ was done. We also simulated the autoregressive process

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

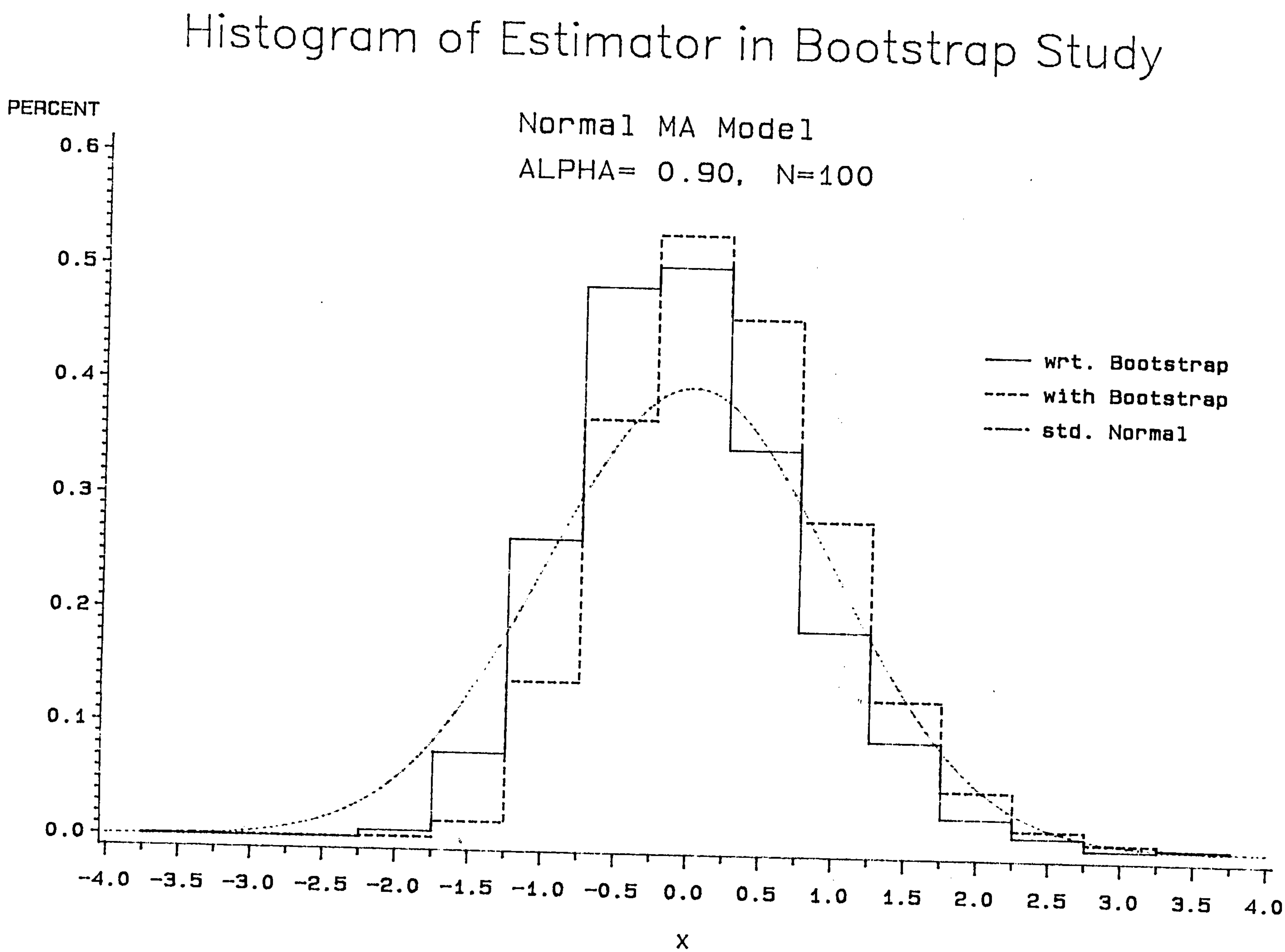
when θ is estimated by the least squares method. Rate of $o(n^{-1/2})$ is also valid for this situation as was shown in Bose (1988). See Bose (1988) for the details of bootstrapping the distribution of the least squares estimates.

For both the MA and AR models, we generated ε_i's from either $N(0,1)$ or from 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.

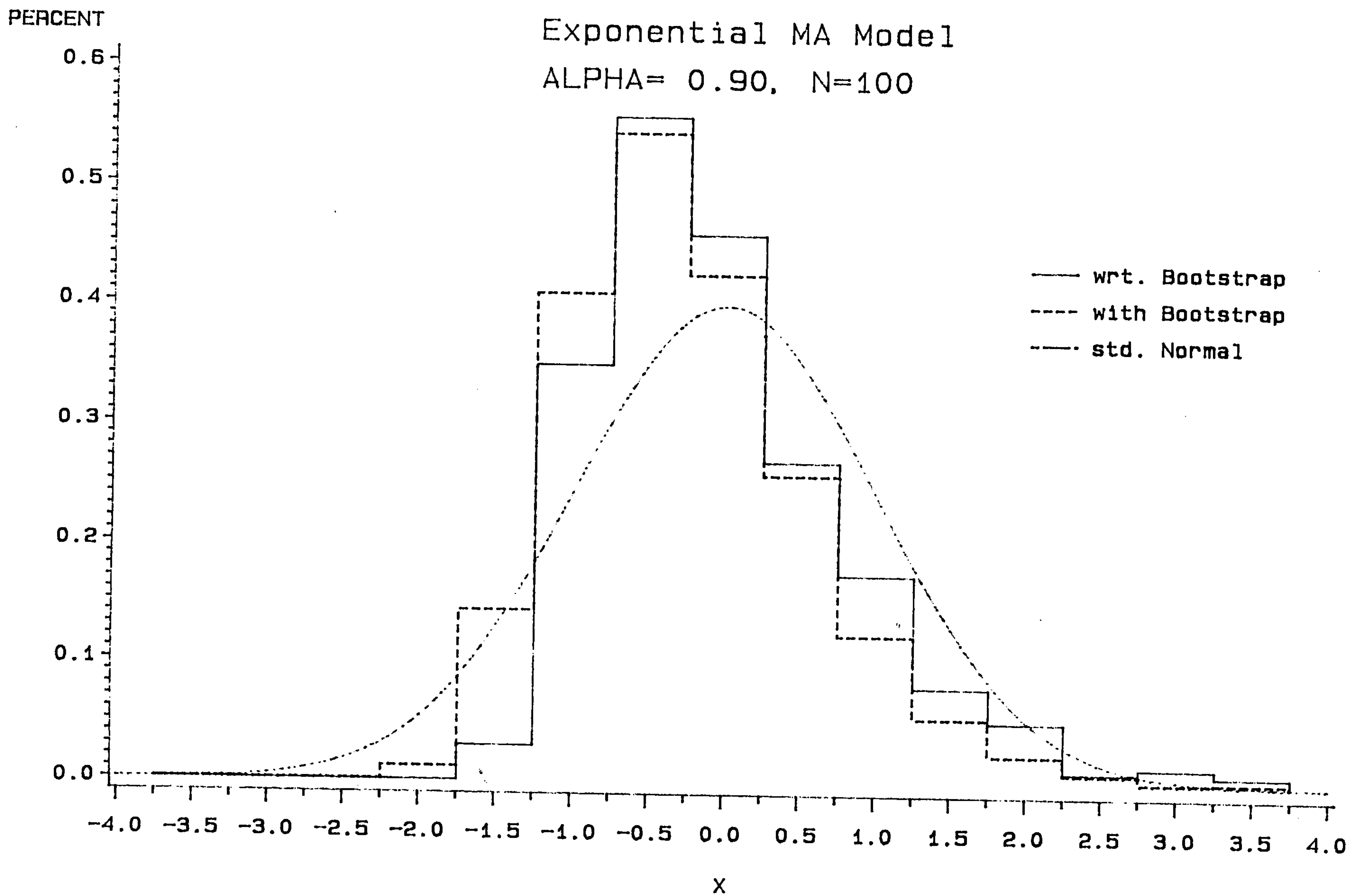
The true (approximate) distribution, the bootstrap distribution and the standard normal distribution have been shown in each case in the graphs. It is evident that the bootstrap works very well in the AR case and reasonably well in the MA case. Similar results were seen to hold for other parameter values. In fact the bootstrap does better as we move away from boundary values of ± 1.

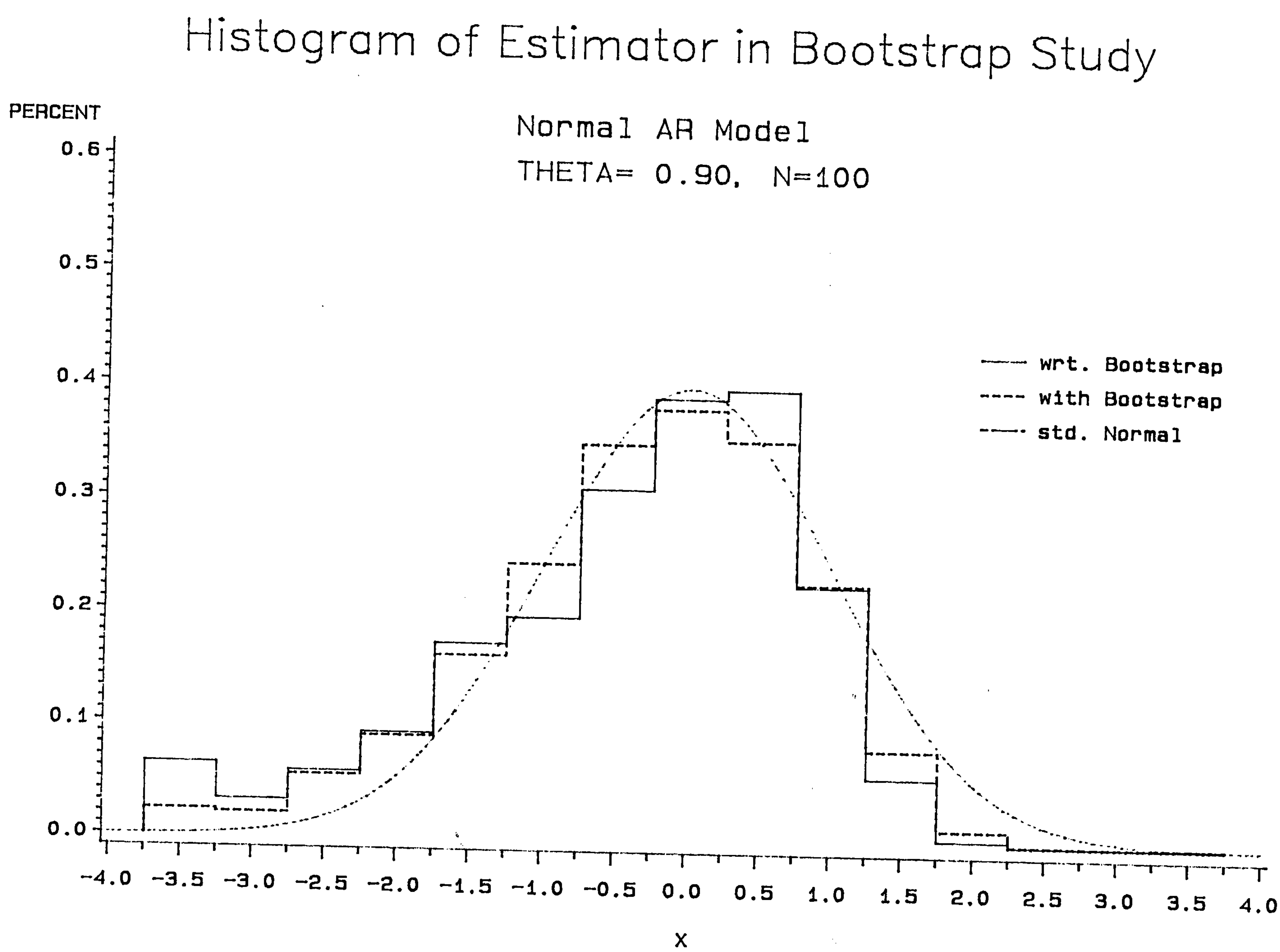
See Chatterjee (1985) for some more simulation studies. The study of the behavior of the bootstrap in other complicated time series models is still open. The author is currently working on the bootstrap in the class of nonlinear autoregressive models.

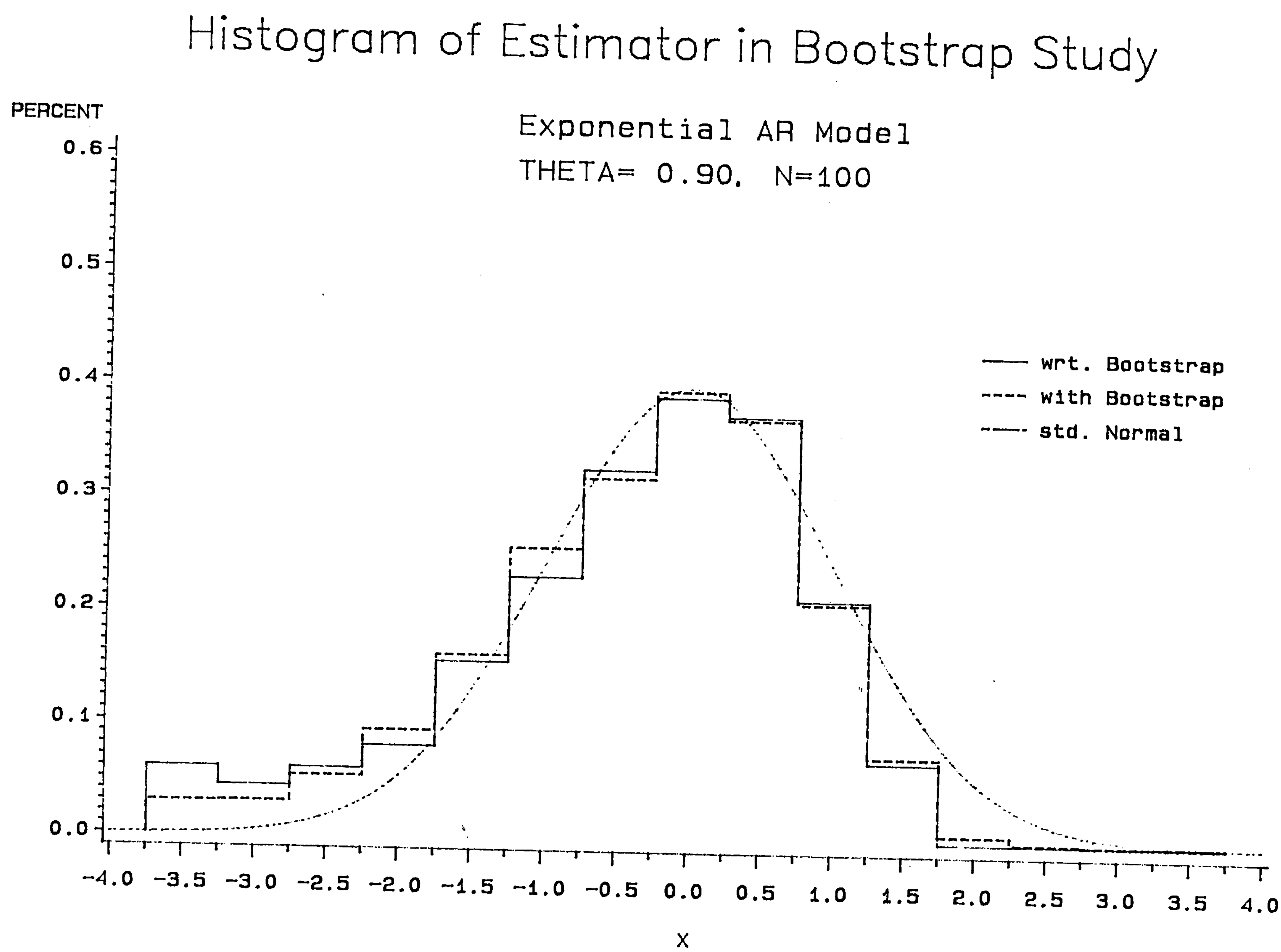
Acknowledgements. The author is grateful to Professor G. J. Babu of Pennsylvania State University for his help and suggestions. The contents of this paper is a part of the author's Ph.D. thesis written under his guidance at Indian Statistical Institute. He is also grateful


Histogram of Estimator in Bootstrap Study
PERCENT
Normal MA Model
ALPHA= 0.90, N=100
0.6
0.5
0.4
0.3
0.2
0.1
0.0
wrt. Bootstrap
with Bootstrap
std. Normal
-4.0 -3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
X

Histogram of Estimator in Bootstrap Study




Histogram of Estimator in Bootstrap Study
PERCENT
Normal AR Model
THETA= 0.90, N=100
wrt. Bootstrap
with Bootstrap
std. Normal
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-4.0
-3.5
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
X


Histogram of Estimator in Bootstrap Study
Exponential AR Model
THETA= 0.90. N=100
PERCENT
0.6
0.5
0.4
0.3
0.2
0.1
0.0
wrt. Bootstrap
with Bootstrap
std. Normal
-4.0 -3.5 -3.0 -2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
X

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

Babu, G.J. and Singh, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā* **40**, Ser. A, Pt. 2, 219–232.

Basawa, I.V., Mallik, A.K., McCormick, W.P. and Taylor, R.L. (1987). Bootstrapping nonstationary autoregressive processes. University of Georgia, Tech. Rep. #**STA64**.

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.* John Wiley & Sons, New York.

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

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

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16(4)**, 1709–1722.

Bose, A. (1987). Higher order approximations for auto-covariances from linear processes with applications. To appear in *Statistics.*

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

Efron, B. (1979). Bootstrap methods: another look at Jackknife. *Ann. Statistics*, **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.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16(3)**, 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. John Wiley and Sons.

Gotze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zeit Wahr. verw. Gebiete* **64**, 211–239.

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

Sweeting, T.J. (1977). Speeds of convergence for the multidimensional central limit theorems. *Ann. Prob.* **5**, 28–41.