The Annals of Statistics
1988, Vol. 16, No. 4, 1709–1722

EDGEWORTH CORRECTION BY BOOTSTRAP IN AUTOREGRESSIONS

BY ARUP BOSE

Indian Statistical Institute and Purdue University

We prove that the distribution of least-squares estimates in autoregressions can be bootstrapped with accuracy $o(n^{-1/2})$ a.s., thereby improving the normal approximation error of $O(n^{-1/2})$.

1. Introduction. Since the introduction of the bootstrap procedure by Efron (1979, 1982), there has been a fast-growing literature on the topic. Empirical evidence has suggested that the bootstrap performs usually very well. See, e.g., Efron (1979, 1982), Bickel and Freedman (1983), Daggett and Freedman (1985) and Freedman and Peters (1984a,b, 1985). Simultaneously, there have been attempts to provide theoretical justification as to why this method performs well. These results provide an insight into the working of the bootstrap procedure. We would like to mention the papers by Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982) and Babu and Singh (1984). They deal with the accuracy of the bootstrap approximation in various senses (e.g., asymptotic normality, Edgeworth expansions, etc.) mainly for sample mean type statistics (or their functionals), quantiles etc., in the i.i.d. situation.

The bootstrap cannot, in general, work for dependent processes; Singh (1981) provides an example. However, it was anticipated that it would work if the dependence is taken care of while resampling. Freedman (1984) confirms this by showing that it does work for certain linear dynamic models (e.g., for two-stage least-squares estimates in linear autoregressions with possible exogeneous variables orthogonal to errors). To the author's knowledge, this is the only theoretical work available for bootstrap in dependent models.

In the absence of distributional assumptions on the errors, the autoregressive parameters are estimated by the least-squares method. The structure of the process enables us to resample the errors and then pseudodata can be generated. We show that the distribution of the parameter estimates can be bootstrapped with accuracy $o(n^{-1/2})$ a.s., thereby improving the normal approximation. The idea is to develop one-term Edgeworth expansions for the distribution of the parameter estimates and its bootstrapped version and then compare these two.

2. Preliminaries. Let Y_t be a stationary autoregressive process satisfying

$$Y_t = \sum_{i=1}^{p} \theta_i Y_{t-i} + \varepsilon_t, \tag{2.1}$$

Received September 1985; revised February 1988.
AMS 1980 *subject classifications*. Primary 62F12, 62M10; secondary 60F99, 60G10.
Key words and phrases. Autoregressions, Cramér's condition, Edgeworth expansion, bootstrap.

where we assume that:

(A.1) (ε_t) are i.i.d. $\sim F_0$, $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, $E\varepsilon_t^{2(s+1)} < \infty$ for some $s \geq 3$.
(A.2) $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramér's condition, i.e., for every $d > 0$, there exists $\delta > 0$ such that $\sup_{\|t\| > d} |E \exp(it'(\varepsilon_1, \varepsilon_1^2))| \leq \exp(-\delta)$.
(A.3) Roots of $\sum_{j=0}^{p} \theta_j z^{p-j} = 0$ lie within the unit circle. Here $\theta_0 = 1$.

REMARKS ON THE ASSUMPTIONS.

1. The stationarity of the autoregression is assumed at this stage to keep the computations simple. This assumption can be dropped. See Remark 3.11.
2. The assumption that (ε_t) have mean 0 and variance 1 seems too rigid. It is possible to drop this assumption and allow arbitrary unknown mean and variance. The resulting proofs shall be more messy since we have to tackle the estimates of these parameters too. However, the results do go through, since these estimates are nice functions of the observations (Y_t). See Remark 3.10.
3. The minimum moment assumption we need is $E\varepsilon_1^8 < \infty$, which at first glance might seem too strong. However, note that the l.s.e. involve quadratic functions of ε_i ($Y_t Y_{t-1}$ involves ε_t^2) and we need $(s+1)$th moment of these with s at least 3. This is in contrast to the i.i.d. situation, where the existence of sth moment suffices. The $(s+1)$th moment is needed because Y_t's are dependent [see Götze and Hipp (1983)]. We have stated our assumption in terms of $s \geq 3$ since under this assumption we can obtain an Edgeworth expansion for the distribution of the l.s.e. up to the order $o(n^{-(s-2)/2})$ which is of independent interest. However, there seems to be a problem in obtaining such an expansion for the bootstrapped distribution.
4. Our results are stated and proved for real-valued processes. The results continue to hold for vector-valued processes. The proofs are similar with added complexity in the notation.

Given $(Y_{1-p}, \ldots, Y_n)$, the least-squares estimates $\theta_{1n}, \ldots, \theta_{pn}$ of $\theta_1, \ldots, \theta_p$ are obtained by solving

$$\tilde{S}_n \begin{pmatrix} \theta_{1n} \\ \vdots \\ \theta_{pn} \end{pmatrix} = \begin{pmatrix} \sum Y_{t-1} Y_t \\ \vdots \\ \sum Y_{t-p} Y_t \end{pmatrix}, \tag{2.2}$$

where

$$\tilde{S}_n = \begin{pmatrix} \sum Y_{t-1}^2 & \sum Y_{t-1} Y_{t-2} & \cdots & \sum Y_{t-1} Y_{t-p} \\ \sum Y_{t-2} Y_{t-1} & \sum Y_{t-2}^2 & \cdots & \sum Y_{t-2} Y_{t-p} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sum Y_{t-p} Y_{t-1} & \cdots & \cdots & \sum Y_{t-p}^2 \end{pmatrix}.$$

The bootstrap distribution of $(\theta_{1n}, \ldots, \theta_{pn})$ is obtained as follows. The errors $\hat{\varepsilon}_t$

are "recovered" by

$$\hat{\varepsilon}_t = Y_t - \sum_{i=1}^{p} \theta_{in} Y_{t-i}, \qquad t = 1, \dots, n.$$

Let $G_n(\cdot)$ denote the distribution function which puts mass $1/n$ at each $\hat{\varepsilon}_i$. Let $F_n^*(x) = G_n(x + \bar{\varepsilon}_n)$, $\bar{\varepsilon}_n = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i$. Let (ε_i^*), $i = 0, \pm 1, \pm 2, \dots,$ be i.i.d. $F_n^*(\cdot)$. (Strictly speaking we should write ε_{in}^*, but we shall drop the suffix n to ease the notation.)

Given (ε_i^*), generate Y_i^* by

$$Y_i^* = \sum_{j=1}^{p} \theta_{jn} Y_{i-j}^* + \varepsilon_i^*, \qquad i = 0, \pm 1, \pm 2, \dots .$$

Let $\theta_n' = (\theta_{1n}, \dots, \theta_{pn})$. Pretend that θ_n is unknown and obtain its l.s.e. θ_n^*. In general, the presence of (*) will denote that we are dealing with the bootstrap quantity and hence expectation etc., are taken under (ε_i^*) i.i.d. F_n^* given $Y_0, Y_1, \dots, Y_n$. Let $\sigma_i = \operatorname{Cov}(Y_0, Y_i)$, $i = 0, \dots, p-1$. It is well known that

$$\Sigma = \begin{pmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{p-1} \\ & \sigma_0 & \cdots & \sigma_{p-2} \\ & & \vdots & \\ & & & \sigma_0 \end{pmatrix} \text{ is positive definite.}$$

Let Σ_n^* denote its bootstrap version, i.e., $\sigma_i^* = \operatorname{Cov}(Y_0^*, Y_i^*)$. $n^{1/2}\Sigma^{1/2}(\theta_n - \theta)$ has an asymptotic normal $(0, I)$ distribution under our assumptions. In fact, we will show that an Edgeworth expansion can be developed for its distribution function. Then we will show that an analogous expansion is valid for its bootstrapped version $n^{1/2}\Sigma_n^{*1/2}(\theta_n^* - \theta_n)$. This will help us to study the accuracy of the bootstrap approximation. In practice the (conditional) distribution of $n^{1/2}\Sigma_n^{*1/2}(\theta_n^* - \theta_n)$ is approximated to any desired degree of accuracy by drawing repeated sets of observations and forming the histogram.

Note that the least-squares equation can be written as

$$\tilde{S}_n \begin{pmatrix} \theta_{1n} - \theta_1 \\ \vdots \\ \theta_{pn} - \theta_p \end{pmatrix} = \begin{pmatrix} \sum Y_{t-1}\varepsilon_t \\ \vdots \\ \sum Y_{t-p}\varepsilon_t \end{pmatrix}.$$

Define

$$\begin{aligned} X_{it} &= Y_{t-i}\varepsilon_t, \qquad i = 1, \dots, p, \\ X_{p+1,t} &= \varepsilon_t^2 - 1, \qquad\qquad (2.3) \\ X_t' &= \left(X_{1t}, \dots, X_{p+1,t}\right), \qquad \mathscr{D}_j = \sigma(\varepsilon_j). \end{aligned}$$

3. Main results. We first obtain an asymptotic expansion for the distribution of $n^{-1/2}\sum_{t=1}^n X_t$ by using the following results due to Götze and Hipp (1983).

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

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

(C.2) $$E\|X_t\|^{s+1} \le \beta_{s+1} < \infty, \quad \forall\, t \text{ for some } s \ge 3,$$

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

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

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

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

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

Define the integer $s_0 \le s$ by

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

Let $\psi_{n,s}$ be the usual function associated with Edgeworth expansions [see Götze and Hipp (1983)]. Let φ_Σ be the normal density with mean 0 and dispersion matrix Σ.

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

The following two results are due to Götze and Hipp (1983) (henceforth referred to as GH).

THEOREM 3.1. *Let f: $\mathbb{R}^k \to \mathbb{R}$ denote a measurable function such that $|f(x)| \le M(1 + \|x\|^{s_0})$ for every $x \in \mathbb{R}^k$. Assume that* (C.1)–(C.7) *hold. Then there exists a positive constant δ 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*

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

where

$$w(f, n^{-K}) = \int \sup(|f(x + y) - f(x)|: |y| \le n^{-K})\varphi_\Sigma(x)\, dx.$$

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

COROLLARY 3.2. *Under assumptions* (C.1)–(C.7) *we have uniformly for convex measurable $C \subseteq \mathbb{R}^k$,*

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

Under our assumptions,

$$Y_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}, \qquad t = 0 \pm 1, \pm 2, \dots,$$

where $\exists\ c, \alpha > 0$ and an integer N_0 such that $\forall\ N \geq N_0, \sum_{r=N}^{\infty} |\delta_r| \leq c \exp(-\alpha N)$. Conditions (C.1)–(C.4) and (C.6) can be easily checked for X_t's defined by (2.3). We now verify (C.5) for X_t's. Let β denote a random variable independent of ε_n.

Clearly,

$$Y_{j-i}\varepsilon_j = \begin{cases} \varepsilon_n \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r} + \beta & \text{if } j = n, \\ \varepsilon_n \delta_{j-i-n} \varepsilon_j + \beta & \text{if } j - i \geq n, \\ \beta & \text{if } j - i < n \text{ and } j > n. \end{cases}$$

Hence

$$\sum_{j=n}^{n+m} \sum_{i=1}^{p} t_i Y_{j-i} \varepsilon_j = \varepsilon_n \sum_{i=1}^{p} t_i (A_{in} + B_{inm}) + \beta,$$

where

$$A_{in} = \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r}, \qquad B_{inm} = \sum_{r=0}^{m-i} \delta_r \varepsilon_{r+i+n}, \qquad i = 1, \dots, p.$$

Note that for all $i, i' = 1, \dots, p$, A_{in} and $B_{i'nm}$ are independent and

$$Z'_{nm} = (A_{in} + B_{inm}, i = 1, \dots, p) \to_{\mathscr{L}} (Z_{i1} + Z_{i2}, i = 1, \dots, p),$$

where Z_{i1} and Z_{i2} are independent and

$$\operatorname{Cov}(Z_{i1}, Z_{j1}) = \sigma_{i-j} = \operatorname{Cov}(Z_{i2}, Z_{j2}).$$

Hence the limiting dispersion matrix of Z_{nm} is positive definite. Thus

$$E\left|E \exp\left(it' \sum_{j=n-m}^{n+m} X_t\right) | \mathscr{D}_j,\ j \neq n\right| \tag{3.1}$$

$$= E|E \exp(i(t_1, \dots, t_p)\varepsilon_n Z_{nm} + it_{p+1}\varepsilon_n^2)|\varepsilon_j,\ j \neq n|$$

$$\leq \exp(-\delta) P(\|t_{nm}\| \geq d) + P(\|t_{nm}\| \leq d), \tag{3.2}$$

where $t_{nm} = ((t_1, \dots, t_p)Z_{nm}, t_{p+1})$.

Suppose that $\|t\|^2 = \sum_{i=1}^{p} t_i^2 + t_{p+1}^2 \geq d_1^2 = d^2/l^2$, where $0 < l < 1$ is to be chosen. Then $P(\|t_{nm}\| \geq d) \geq P((a'Z_{nm})^2 \geq l^2)$, where $\|a\| = 1$. Let b_i, $i = 1, \dots, (p+1)$, be points in $\mathbb{R}^p$ and $r > 0$ be such that $P(Z_{j1} + Z_{j2} \in B(b_i, r), \forall\ j = 1, \dots, p) > 0$, $\forall\ i = 1, \dots, (p+1)$ and $B(b_i, r)$ are such that not all $(b_1, \dots, b_{p+1})$ lie in any given hyperplane of dimension $(p-1)$. This is possible since the dispersion matrix of $(Z_{i1} + Z_{i2}, i = 1, \dots, p)$ is positive definite. Choose l sufficiently small. Then for any a with $\|a\| = 1$, $\{x: (a'x)^2 < l^2\}$ does not intersect at least one of the balls $B(b_i, r)$. Hence $P((a'Z_{nm})^2 \geq l^2) \geq \min_{i=1,\dots,(p+1)} P(Z_{nm} \in B(b_i, r))$ and lim inf of the right side is positive. Thus there exist n_0, m_0 large enough and $\varepsilon > 0$ such that

$$P(\|t_{nm}\| \geq d^2) \geq \varepsilon > 0, \quad \forall \; \|t\|^2 \geq d_1^2,\; n \geq n_0,\; m \geq m_0.$$

Combining this with (3.2) verifies condition (C.5) for X_t's defined in (2.3).

Further,

$$\lim_{n\to\infty} D\left(n^{-1/2}\sum_{t=1}^{n} X_t\right) = \begin{pmatrix} \Sigma & 0 \\ 0 & V(\varepsilon_1^2) \end{pmatrix}$$

is positive definite verifying (C.7). Thus Theorem 3.1 and Corollary 3.2 are valid for

$$n^{-1/2}\sum_{t=1}^{n} X_t, \quad \text{with } X_t\text{'s as defined in (2.3).} \tag{3.3}$$

REMARK 3.3. The preceding arguments also show that if conditions (A.1) and (A.3) hold and ε_1 satisfies Cramér's condition, then the distribution of $n^{-1/2}\tilde{S}_n(\theta_n - \theta)$ admits an Edgeworth expansion of order $o(n^{-(s-2)/2})$.

Our next task is to derive an Edgeworth expansion for the bootstrapped version of X_t's. We will derive this expansion only for the case $p = 1$ to avoid notational complexity.

Let $H_n^*(\cdot)$ denote the characteristic function of $n^{-1/2}\sum_{j=1}^{n} Z_j^*$, where Z_j^* is a certain truncation (as in GH) of $X_j^* = (Y_{j-1}^*\varepsilon_j^*, \varepsilon_j^{*2} - \sigma_n^{*2})$. Here $\sigma_n^{*2} = E^*(\varepsilon_n^{*2})$. We omit the explicit definition of Z_j^* since this shall not be used in subsequent calculations.

LEMMA 3.4. *$\forall\; \|t\| \leq Cn^{\varepsilon^0}$ we have*

$$|D^{\alpha}(H_n^*(t) - \hat{\psi}_{n,3}^*(t))| \leq C(1 + \beta_{4n}^*)(1 + \|t\|^{6+|\alpha|})\exp(-C\|t\|^2)n^{-\varepsilon^0 - 1/2},$$

for some $\varepsilon^0 < \frac{1}{2}$, and C depends on the bounds of β_{4n}^* ($=$ fourth moment of X_j^*). $\hat{\psi}_{n,3}^*(t)$ denotes the Fourier transform of $\psi_{n,3}^*$, the signed measure associated with the Edgeworth expansion of X_j^*. D^{α} is the usual differential operator with $|\alpha| \leq 6$.

This same lemma is proved in GH and hence we skip the proof.

Let

$$I_1 = \{t\colon Cn^{\varepsilon^0} \leq \|t\| \leq C_1 n^{1/2}\},$$
$$I_2 = \{t\colon C_1 n^{1/2} \leq \|t\| \leq \varepsilon^{-1} n^{1/2}\},$$

where C_1 is to be chosen later and $0 < \varepsilon < 1$ is fixed.

LEMMA 3.5. *Under* (A.1) *and* (A.2) *we have, for almost every sequence* $Y_0, Y_1, \ldots$ *and* $|\alpha| \leq 6$,

$$\int_{t\in I_2} |D^{\alpha}H_n^*(t)|\, dt = o(n^{-1/2}).$$

PROOF. A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that

$$E^*|E^*A_p^*|\mathscr{D}_j^*,\ j \neq j_p| < 1 \quad \text{uniformly in } t \in I_2 \text{ and } p = 1,\dots,l,$$

where $A_p^* = \exp(it'n^{-1/2}\sum_{j=j_p-m}^{j_p+m} Z_j^*)$. For the definitions of l and j_p one can consult GH. We omit these definitions since they are not used explicitly in our calculation. But note that the effect of truncation is negligible and it suffices to deal with

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

If F_{nm}^* is the distribution of $\sum_{t=0}^{\infty}\theta_n^t\varepsilon_{n-1-t}^* + \sum_{j=n+1}^{n+m}\theta_n^{j-1-n}\varepsilon_j^*$, then writing $t_1' = t_1 n^{-1/2}$, $t_2' = t_2 n^{-1/2}$, we have [see (3.1)]

$$\delta_{nm} = \int\left|\int \exp\left(it_1'xy + it_2'x^2\right) dF_n^*(x)\right| dF_{nm}^*(y).$$

Note that a.s. $F_{nm}^* \Rightarrow F$, where F is the distribution of $Z_1 + Z_2$, Z_1, Z_2 i.i.d. $Z_1 =_{\mathscr{L}} \sum_{t=0}^{\infty}\theta^t\varepsilon_t$ and by Lévy's theorem F is continuous. Further, $F_n^* \Rightarrow F_0$ a.s. Since the convergence of F_{nm}^* to F is uniform we have

$$\delta_{nm} \to \delta = \int\left|\int \exp\left(it_1'xy + it_2'x^2\right) dF_0(x)\right| dF(y) \quad \text{a.s.}$$

uniformly on compact sets of (t_1', t_2'), i.e., uniformly over $t \in I_2$, and by Cramér's condition $\delta < 1$. This proves the lemma. □

LEMMA 3.6. *Assume* (A.1) *and* (A.2) *hold. For sufficiently small* C_1, *we have for almost every sequence* $Y_0, Y_1, \dots,$

$$\int_{t \in I_1} |D^\alpha H_n^*(t)|\, dt = o(n^{-1/2}).$$

PROOF. As in Lemma 3.5, it is sufficient to deal with the original variables instead of truncations. We proceed as in Lemma 3.5 following GH but use a different estimate for $E^*|E^*A_p^*|\mathscr{D}_j^*,\ j \neq j_p|$ (see Lemma 3.5 for the definition of A_p^*). We have to deal with

$$\delta_{nm}^* = E^*|E^*\exp\left(it_1 n^{-1/2}\varepsilon_n^*(A_n^* + B_{nm}^*) + it_2 n^{-1/2}\varepsilon_n^{*2}\right)|\varepsilon_j^*,\ j \neq n|,$$

where

$$A_n^* = \sum_{t=0}^{\infty}\theta_n^t\varepsilon_{n-1-t}^* \quad \text{and} \quad B_{nm}^* = \sum_{j=n+1}^{n+m}\theta_n^{j-1-n}\varepsilon_j^*,$$

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

where $t_n' = (t_1(A_n^* + B_{nm}^*), t_2)$ and $|\gamma| \le 1$. Thus

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

where

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

$$E^*(\|t_n\|^3) \le \left[E^*\left(t_1^2(A_n^* + B_{nm}^*)^2 + t_2^2\right)^3\right]^{1/2}$$

$$\le C\left[t_1^6 E^*(A_n^* + B_{nm}^*)^6 + t_2^6\right]^{1/2}$$

Note that $E^*(A_n^* + B_{nm}^*)^6 \to E(Z_1 + Z_2)^6$ a.s., where Z_1 and Z_2 are i.i.d. $Z_1 =_{\mathscr{L}} \sum_{t=0}^{\infty} \theta^t \varepsilon_t$. Hence, for some constant C,

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

$$\le CC_1 \frac{\|t\|^2}{n} \quad \text{a.s.}$$

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

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

Let $\bar{\lambda}(A)$ and $\underline{\lambda}(A)$ denote the maximum and minimum eigenvalues of a matrix A. Note that $\bar{\lambda}(\Sigma_n) \to \bar{\lambda}(\Sigma) > 0$ a.s. and $\underline{\lambda}(\Sigma_n) \to \underline{\lambda}(\Sigma) > 0$ a.s.

$$E^*\left(\frac{t_n' D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n}{2n}\right)^2 \le \bar{\lambda}^2(\Sigma_n) \frac{E^*(\|t_n\|^4)}{4n^2}$$

and arguing as before, the preceding quantity

$$\le CC_1 \frac{\|t\|^2}{n} \quad \text{a.s.}$$

$$E^*\left(\frac{t_n' D(\varepsilon_n^*, \varepsilon_n^*)^2 t_n}{2n}\right) \ge \underline{\lambda}(\Sigma_n) E^*\left(\frac{\|t_n\|^2}{2n}\right)$$

$$\ge \underline{\lambda}(\Sigma_n) C \frac{\|t\|^2}{n} \quad \text{a.s.}$$

Combining these estimates and choosing C_1 sufficiently small,

$$\delta_{nm}^* \le 1 - \gamma \frac{\|t\|^2}{n}, \quad \text{for some } \gamma > 0 \text{ a.s.}$$

$$\le \exp\left(-\frac{\gamma\|t\|^2}{n}\right) \quad \text{a.s.}$$

A look at the proof of Lemma 3.43 of GH shows that this proves Lemma 3.6. □

Our next lemma is stated in Babu and Singh (1984) and is a modified version of a lemma in Sweeting (1977).

LEMMA 3.7. *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* $M_s(f) < \infty$ *for some* $s \geq 2$. *Further, let* $\alpha = K(x\colon \|x\| \leq 1) > \frac{1}{2}$ *and* $\beta = \int \|x\|^{s+2} K(dx) < \infty$. *Then for any* $0 < \varepsilon < 1$,

$$\left|\int f\, d(P-Q)\right| \leq (2\alpha - 1)^{-1}[B(1-\alpha)/\alpha]^{-1+\varepsilon^{-1/4}} + \beta\varepsilon B$$

$$+B\int (1 + \|x\|^s)|K_\varepsilon * (P-Q)|\, dx$$

$$+ \sup_{\|x\| < \varepsilon^{1/4}} \int w(f, 2\varepsilon, x - y)|Q\, dy,$$

where

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

$$M_s(f) = \sup_x (1 + \|x\|^s)^{-1}|f(x)|.$$

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

$$\int w(f, \delta, x - y)\varphi(y)\, dy \leq 3\int w(f, \delta, y)\varphi(y)\, dy$$

$$+ C_0 M_s(f)\|x\|^{-k-s+1}\exp\left(-\tfrac{1}{8}\|x\|^{-2}\right).$$

Combining Lemmas 3.4–3.7, we can say that Theorem 3.1 and Corollary 3.2 hold a.s. for

$$S_n = n^{-1/2}\sum X_t^*, \quad \text{for } s = 3. \tag{3.4}$$

To study the accuracy of the bootstrap approximation, we need to compare two Edgeworth expansions and for which the following lemma is needed. This lemma is due to Babu and Singh and is essentially a modified version of a lemma of Bhattacharya and Ghosh (1978). A natural extension of this result to the multidimensional case is true and will also be used.

LEMMA 3.8. *Let* $l = (l_1, \ldots, l_k)$ *be a vector* $L = (L_{ij})$ *be a* $k \times k$ *matrix and Q be a polynomial in k variables. Let* $M \geq \max(|V_{ij}|, |u_{ij}|, |l_i|, |L_{ij}|, |a_\lambda|)$, *where* $(V_{ij}) = V$, $(u_{ij}) = V^{-1}$ *and* a_λ *are coefficients of Q. Let* $|l_1| > l_0 > 0$ *and* $b_n = (l_1 n^{1/2})^{-1}$. *Then there exists a polynomial p in one variable, whose coefficients are continuous functions of* l_i, L_{ij}, V_{ij}, u_{ij} *and* a_λ *such that*

$$\int_{\{z\colon lz + n^{-1/2}z'Lz < u(l'Vl)^{1/2}\}} (1 + n^{-1/2}Q(z))\varphi_V(z)\, dz$$

$$= \int_{-\infty}^{u} (1 + b_n p(y))\varphi(y)\, dy + o(n^{-1/2}).$$

The $o(\cdot)$ *term depends on M and* l_0.

We now state and prove our main result.

THEOREM 3.9. *Under assumptions* (A.1)–(A.3) *for a.e.* (Y_i),

$$\sup_x |P^*\big(n^{1/2}\Sigma_n^{*1/2}(\theta_n^* - \theta_n) \le x\big) - P\big(n^{1/2}\Sigma^{1/2}(\theta_n - \theta) \le x\big)| = o(n^{-1/2}).$$

PROOF. We first give the proof for the case $p = 1$. We have an Edgeworth expansion of $n^{-1/2}(\sum_{j=1}^n Y_{j-1}^* \varepsilon_j^*, \sum_{j=1}^n(\varepsilon_j^{*2} - \sigma_n^{*2}))$ [see (3.4)]. Hence, for a.e. (Y_i),

$$\sup_x \left| P^*(S_n^* \le x) - \int_{-\infty}^x \big(1 + n^{-1/2}p(y,n)\big)\, d\Phi_{\Gamma_n^*}(y)\right| = o(n^{-1/2}), \tag{3.5}$$

where $p(x,n)$ denotes a polynomial in x whose coefficients are continuous functions of the moments of $Y_{j-1}^*\varepsilon_j^*$ and $(\varepsilon_j^{*2} - \sigma_n^{*2})$ of order 3 or less and

$$\Gamma_n^* = D^*\big(Y_{j-1}^*\varepsilon_j^*, \varepsilon_j^{*2}\big) = \begin{pmatrix} \Sigma_n^* & 0 \\ 0 & V\big(\varepsilon_1^{*2}\big) \end{pmatrix}.$$

Also note that by (3.3),

$$\sup_x \left| P(S_n \le x) - \int_{-\infty}^x \big(1 + n^{-1/2}p(y)\big)\, d\Phi_{\Gamma}(y)\right| = o(n^{-1/2}),$$

where

$$\Gamma = \begin{pmatrix} \Sigma & 0 \\ 0 & V\big(\varepsilon_1^2\big) \end{pmatrix}$$

and $p(y)$ denotes a polynomial of the same form as $p(y,n)$. Note that for $p = 1$, $\Sigma = (1 - \theta^2)^{-1}$, $\Sigma_n^* = (1 - \theta_n^2)^{-1}$,

$$n^{1/2}\big(1 - \theta_n^2\big)^{-1/2}\big(\theta_n^* - \theta_n\big) = \frac{X_1^*\big(1 - \theta_n^2\big)^{-1/2}}{1 + n^{-1/2}(2\theta X_1^* + X_2^*) + A_n^*},$$

where

$$X_1^* = n^{-1/2}\sum_{t=1}^n Y_{t-1}^*\varepsilon_t^*,$$

$$X_2^* = n^{-1/2}\sum_{t=1}^n \big(\varepsilon_t^{*2} - \sigma_n^{*2}\big),$$

$$A_n^* = n^{-1}\big(Y_0^{*2} - Y_n^{*2}\big).$$

Let

$$B_i^* = \{|X_i^*| \ge c\log n\}, \qquad i = 1,2,$$

$$B_3^* = \big\{n^{3/4}|A_n^*| \ge c\log n\big\}.$$

By (3.4), $P(B_i^*) = o(n^{-1/2})$, $i = 1,2$, a.s. Also $P(B_3^*) = o(n^{-1/2})$ a.s. On $B_1^{*c} \cap B_2^{*c} \cap B_3^{*c}$,

$$\big(1 - \theta_n^2\big)^{-1/2} n^{1/2}\big(\theta_n^* - \theta_n\big) = \big(l'X^* + n^{-1/2}X^{*\prime}A^*X^*\big)\big(l'\Gamma_n^* l\big)^{-1/2} + o(n^{-1/2}), \tag{3.6}$$

where

$$l' = (1,0), \qquad X^{*\prime} = (X_1^*, X_2^*), \qquad A^* = \begin{pmatrix} -2\theta_n & -1/2 \\ -1/2 & 0 \end{pmatrix}$$

and

$$\Gamma_n^* = \begin{pmatrix} \left(1-\theta_n^2\right)^{-1} & 0 \\ 0 & V\left(\varepsilon_1^{*2}\right) \end{pmatrix}.$$

Analogous representation holds for $(1-\theta^2)^{-1/2}n^{1/2}(\theta_n - \theta)$. Now note that the moments of $(X_{j-1}^*\varepsilon_j^*, \varepsilon_j^{*2} - \sigma_n^{*2})$ (under F_n^*) converge almost surely to those of $(X_{j-1}\varepsilon_j, \varepsilon_j^2 - 1)$ by the ergodic theorem and the fact that $\theta_n \to \theta$ a.s. The proof now follows from this observation and Lemma 3.8.

For $p > 1$, we need a representation of the form (3.6) and then we can apply (3.3), (3.4) and the multidimensional version of Lemma 3.8.

Define

$$\tilde{Y}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}_{p\times 1}, \qquad \tilde{Z}_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p\times 1}, \qquad B = \begin{pmatrix} \theta_1 & \cdots & \theta_p \\ & I & \end{pmatrix}_{p\times p}.$$

Clearly, (2.1) is equivalent to

$$\tilde{Y}_t = B\tilde{Y}_{t-1} + \tilde{Z}_t. \tag{3.7}$$

Let

$$A_n = n^{-1}\sum_{t=1}^{n} \tilde{Y}_t\tilde{Y}_t', \qquad B_n = n^{-1}\sum_{t=1}^{n} \tilde{Y}_{t-1}\tilde{Y}_{t-1}'.$$

This gives

$$A_n - BB_nB' = n^{-1}\sum_{t=1}^{n} \tilde{Z}_t\tilde{Z}_t' + 2Bn^{-1}\sum_{t=1}^{n} \tilde{Y}_{t-1}\tilde{Z}_t'. \tag{3.8}$$

Note that $\Sigma = E\tilde{Y}_t\tilde{Y}_t'$ and satisfies $\Sigma = B\Sigma B + I^*$, where $I^* = \begin{pmatrix} 1 & 0\cdots 0 \\ 0 & 0 \end{pmatrix}$. Equation (3.8) yields

$$B_n - \Sigma - B(B_n - \Sigma)B' = n^{-1}\sum_{t=1}^{n}\left(\tilde{Z}_t\tilde{Z}_t' - I^*\right) + 2Bn^{-1}\sum_{t=1}^{n}\tilde{Y}_{t-1}\tilde{Z}_t' + O_p(n^{-1}),$$

which gives

$$B_n - \Sigma = G_1\left[n^{-1}\sum_{t=1}^{n}\left(\tilde{Z}_t\tilde{Z}_t' - I^*\right) + 2Bn^{-1}\sum_{t=1}^{n}\tilde{Y}_{t-1}\tilde{Z}_t'\right]G_2 + O_p(n^{-1}),$$

where G_1, G_2 are independent of n but depend on $\theta_1, \dots, \theta_p$.

Let

$$V_n' = \left(\sum_{t=1}^{n} X_{1t}, \dots, \sum_{t=1}^{n} X_{pt}\right).$$

Then

$$n^{1/2}(\theta_n - \theta) = \left[\frac{\sum_{t=1}^{n}\tilde{Y}_{t-1}\tilde{Y}'_{t-1}}{n}\right]^{-1} n^{-1/2}V_n$$

$$(3.9)\qquad = \Sigma^{-1}\left[I + (B_n - \Sigma)\Sigma^{-1}\right]^{-1} n^{-1/2}V_n$$

$$= \Sigma^{-1}\left[I + G_1\left(n^{-1}\sum_{t=1}^{n}\left(\tilde{Z}_t\tilde{Z}'_t - I^*\right) + 2Bn^{-1}\sum_{t=1}^{n}\tilde{Y}_{t-1}\tilde{Z}'_t\right)G_3\Sigma^{-1}\right]^{-1} n^{-1/2}V_n.$$

Let

$$B_i = \left\{n^{-1/2}\left|\sum_{t=1}^{n} X_{it}\right| \geq c\log n\right\}, \qquad i = 1,\dots,p+1.$$

Since we have an asymptotic expansion for the distribution of $n^{-1/2}\sum_{t=1}^{n}X_t$, it easily follows that

$$P(B_i) = o(n^{-1/2}), \qquad i = 1,\dots,p+1.$$

On $\bigcap_{i=1}^{p}B_i^c$, by (3.9),

$$(3.10)\qquad n^{1/2}(\theta_n - \theta) = n^{-1/2}\Gamma^{-1}\begin{pmatrix}\Sigma X_{1t}\\ \vdots \\ \Sigma X_{p+1,t}\end{pmatrix} + n^{-3/2}\begin{pmatrix}V_n' L_1 V_n\\ \vdots \\ V_n' L_p V_n\end{pmatrix} + o(n^{-1/2}).$$

An analogous representation holds for the bootstrapped version. This completes the proof of Theorem 3.9. □

REMARK 3.10. The assumptions $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$ may seem to be too restrictive. Actually these restrictions were imposed to keep the proofs simpler. We will sketch how the case $E\varepsilon_t = \mu$, $E\varepsilon_t^2 = \sigma^2$ can be tackled. We illustrate the case $p = 1$ only.

The model in this case are

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

where (ε_t) satisfies (A.1)–(A.3) but $E\varepsilon_t^2 = \sigma^2 > 0$ and μ and σ^2 are unknown.

Under assumptions (A.1)–(A.3), Edgeworth expansion is valid for the distribution of

$$(3.11)\qquad n^{-1/2}\left(\sum_{t=1}^{n}(Y_t - \alpha_1), \sum_{t=1}^{n}(Y_tY_{t-1} - \alpha_2), \sum_{t=1}^{n}\left(Y_t^2 - \alpha_3\right)\right),$$

where $\alpha_1 = EY_t$, $\alpha_2 = EY_tY_{t-1}$ and $\alpha_3 = EY_t^2$.

Estimates θ_n and μ_n of θ and μ are obtained by solving

$$\sum_{t=1}^{n} (Y_t - \theta_n Y_{t-1} - \mu_n) = 0$$

and

$$\sum_{t=1}^{n} Y_{t-1}(Y_t - \theta_n Y_{t-1} - \mu_n) = 0.$$

An estimate σ_n^2 of σ^2 is given by

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

Thus the estimates θ_n, μ_n and σ_n are all smooth functions of $\sum_{t=1}^n Y_t$, $\sum_{t=1}^n Y_t Y_{t-1}$ and $\sum_{t=1}^n Y_t^2$ except for terms which can be neglected.

Thus for a suitable normalizing factor β, the distribution of $n^{1/2}\beta(\theta_n - \theta)$ admits an Edgeworth expansion up to $o(n^{-1/2})$, with the leading term as $\Phi(x)$, and the coefficients involved in the polynomial in the second term [which is $O(n^{-1/2})$] are smooth functions of θ, μ and σ^2 and of moments of Y_t, $Y_t Y_{t-1}$ and Y_t^2 of order less or equal to 3. β can be explicitly calculated and depends on θ and μ and moments of ε_1.

The empirical distribution is computed by putting mass $1/n$ at each $\hat{\varepsilon}_i = Y_i - \theta_n Y_{i-1} - \mu_n$, $i = 1, \dots, n$. Proceeding as in the case $\mu = 0$, $\sigma^2 = 1$, an asymptotic expansion is valid for the bootstrapped version of (3.11), which yields an expansion of order $o(n^{-1/2})$ for the distribution of $n^{1/2}\beta_n(\theta_n^* - \theta_n)$, where β_n is the variance-normalizing factor, the bootstrap equivalent of β. The leading term of 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 $n^{1/2}\beta(\theta_n - \theta)$. By the 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 θ_n, μ_n and σ_n are all strongly consistent estimators of θ, μ and σ, respectively. Thus, the difference between the two Edgeworth expansions is $o(n^{-1/2})$ a.s.

REMARK 3.11. The assumption of stationarity of (Y_t) was made since the calculations (e.g., of Σ) in this case are simpler. The results hold even if this assumption is dropped. This is fairly obvious, since the asymptotic structure does not change and the results of GH go through.

It would be interesting to see how the bootstrap performs in small samples. The accuracy is expected to decrease as the parameter values move toward the boundary. The absence of stationarity will also decrease the accuracy in small samples. See Chatterjee (1985) for some simulation studies.

It will also be interesting to study the bootstrap in other complicated time-series models.

Acknowledgments. The author is grateful to Professor G. J. Babu of the Indian Statistical Institute, Calcutta, for his help and suggestions. This paper is part of the author's Ph.D. thesis written under his guidance. Constructive criticism by the referees and the Editor has improved the paper beyond recognition. The author is extremely grateful to them. He is also indebted to one of the referees for pointing out an error in the original version.

REFERENCES

BABU, G. J. and SINGH, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46** 219–232.

BERAN, R. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BICKEL, P. J. and FREEDMAN, D. A. (1980). On Edgeworth expansion for the bootstrap. Preprint.

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

BICKEL, P. J. and FREEDMAN, D. A. (1983). Bootstrapping regression models with many parameters. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 28–48. Wadsworth, Belmont, Calif.

CHATTERJEE, S. (1985). Bootstrapping ARMA models: Some simulations. Preprint, College of Business Administration, Northeastern Univ.

DAGGETT, R. S. and FREEDMAN, D. A. (1985). Econometrics and the law: A case study in the proof of antitrust damages. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **1** 123–172. Wadsworth, Monterey, Calif.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

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

FREEDMAN, D. A. and PETERS, S. C. (1984a). Bootstrapping a regression equation: Some empirical results. *J. Amer. Statist. Assoc.* **79** 97–106.

FREEDMAN, D. A. and PETERS, S. C. (1984b). Bootstrapping an econometric model: Some empirical results. *J. Bus. Econom. Statist.* **2** 150–158.

FREEDMAN, D. A. and PETERS, S. C. (1985). Using the bootstrap to evaluate forecasting equations. *J. Forecasting* **4** 251–262.

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

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

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

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907