EDGEWORTH CORRECTION BY BOOTSTRAP IN AUTOREGRESSIONS

by

Arup Bose
Indian Statistical Institute, Calcutta
and
Purdue University

EDGEWORTH CORRECTION BY BOOTSTRAP IN AUTOREGRESSIONS

by

Arup Bose
Indian Statistical Institute, Calcutta
and
Purdue University

Technical Report #87-37

Department of Statistics
Purdue University

August 1987

★

EDGEWORTH CORRECTION BY BOOTSTRAP IN AUTOREGRESSIONS

by

Arup Bose

Indian Statistical Institute, Calcutta

and

Purdue University, Indiana

Summary: 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 $0(n^{-1/2})$.

AMS 1985 Subject Classification: Primary 62 F 12, 62 M 10; Secondary 60 F 99, 60 G 10

Key Words and Phrases: Autoregressions, Cramer's condition, Edgeworth expansion, bootstrap.

1. <u>Introduction</u>: 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 (1984), Freedman and Peters (1984a,b,c). 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 accuracy of 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 lease 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 pseudo-data can be generated. We show that the distribution of the parameter estimates can be bootstrapped with accuracy $o(n^{-1/2})$, thereby improving the normal approximation. The idea is to develop one term Edgeworth expansion for the distribution of the parameter estimates and its bootstrapped version, and then compare these two.

2. <u>Preliminaries</u>

Let Y_t be a stationary autoregressive process satisfying

$$Y_t = \sum_{i=1}^{p} \theta_i Y_{t-i} + \varepsilon_t \qquad (2.1)$$

where we assume that

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

(A2) $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramer's condition, i.e. for every $d > 0$, there exists $\delta > 0$ such that
$$\sup_{\|t\| > d} |E \exp (it'(\varepsilon_1, \varepsilon_1^2))| \leq \exp (-\delta).$$

(A3) Roots of $\sum\limits_{j=0}^{p} \theta_j z^{p-j} = 0$ lie within the unit circle. Here $\theta_0 = 1$.

Remarks on the assumptions

2.1. The stationarity of the autoregression is assumed at this stage to keep the computations simple. This assumption can be dropped. See Remark 3.11.

2.2. 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.

2.3. 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 $\varepsilon_i (Y_t Y_{t-1}$ involves $\varepsilon_i^2)$ 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 existence of sth moment suffices. The $(s + 1)^{th}$ moment is needed because Y_t's are dependent. (See Gotze and Hipp (1983)). We have stated our assumption in terms of $s \geq 3$ since under the assumption we can obtain Edgeworth expansion for the distribution of the l.s.e. upto 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.

2.4. 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 notations.

3

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

$$S_n \begin{pmatrix} \theta_{1n} \\ \vdots \\ \theta_{pn} \end{pmatrix} = \begin{pmatrix} \Sigma Y_{t-1} Y_t \\ \vdots \\ \Sigma Y_{t-p} Y_t \end{pmatrix} \tag{2.2}$$

where $S_n = \begin{pmatrix} \Sigma Y_{t-1}^2 & \Sigma Y_{t-1} Y_{t-2} & \ldots & \Sigma Y_{t-1} Y_{t-p} \\ \Sigma Y_{t-2} Y_{t-1} & \Sigma Y_{t-2}^2 & \ldots & \Sigma Y_{t-2} Y_{t-p} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \Sigma Y_{t-p} Y_{t-1} & \ldots & \ldots & \Sigma Y_{t-p}^2 \end{pmatrix}$

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

The errors $\hat{\varepsilon}_t$ are "recovered" by

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

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 $(\varepsilon_i^*), i = 0, \pm 1, \pm 2$ be i.i.d. $F_n^*(\cdot)$. (Strictly speaking we should write ε_{in}^*, but we shall drop the suffix n to ease notations).

Given (ε_i^*), generate Y_i^* by

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

Let $\theta_n' = (\theta_{1n}, \ldots, \theta_{pn})$. 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 $Y_0, Y_1, \ldots, Y_n$.

Let $\sigma_i = \text{Cov}(Y_o, Y_i), i = 0, \ldots, (p-1)$.

4

It is well known that

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

Let Σ_n^* denote its bootstrap version, i.e. $\sigma_i^* = \text{Cov}(Y_o^*, Y_i^*)$.

$n^{1/2}\Sigma^{1/2}(\theta_n - \theta)$ 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 $n^{1/2}\Sigma_n^*(\theta_n^* - \theta_n)$. This will help us to study the accuracy of the bootstrap approximation. In practice the (conditional) distribution of $n^{1/2}\Sigma_n^*(\theta_n^* - \theta_n)$ 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

$$S_n \begin{pmatrix} \theta_{1n} - \theta_1 \\ \vdots \\ \theta_{pn} - \theta_p \end{pmatrix} = \begin{pmatrix} \Sigma Y_{t-1}\varepsilon_t \\ \vdots \\ \Sigma Y_{t-p}\varepsilon_t \end{pmatrix}$$

Define $\quad X_{it} = Y_{t-i}\varepsilon_t, i = 1, \ldots, p$

$$X_{p+1,t} = \varepsilon_t^2 - 1$$

$$X_t' = (X_{1t}, \ldots, X_{p+1,t}), \mathcal{D}_j = \sigma(\varepsilon_j).$$

3. <u>Main results</u>: 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 $(\Omega, \mathcal{F}, P)$ and 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) $EX_t = 0 \quad \forall t$

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

5

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

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

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

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

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 $s_o \le s$ by

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

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 Gotze and Hipp (1983) (henceforth referred as GH).

Theorem 3.1. Let $f : \mathbb{R}^k \to \mathbb{R}$ denote a measurable function such that $|f(x)| \le M(1 + \|x\|^{s_o})$ for every $x \varepsilon \mathbb{R}^k$. 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

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

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

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 $C \subseteq \mathbb{R}^k$,

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

Under our assumptions,

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

6

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

$$\text{Clearly} \qquad Y_{j-i}\varepsilon_j = \varepsilon_n \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r} + \beta \text{ if } j = n$$

$$= \varepsilon_n \delta_{j-i-n} \varepsilon_j + \beta \text{ if } j - i \geq n$$

$$= \beta \text{ if } j - i < n \text{ and } j > n.$$

where β is a random variable independent of ε_n.

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

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

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

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

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

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

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

$$\text{Thus} \qquad E|E \exp(it' \sum_{j=n-m}^{n+m} X_t)|\mathcal{D}_j, j \neq n|$$

$$= E|E \exp(i(t_1, \ldots, t_p)\varepsilon_n Z_{nm} + it_{p+1}\varepsilon_n^2)|\varepsilon_j, j \neq n| \qquad (3.1)$$

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

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

Suppose that $\|t\|^2 = \sum_{i=1}^{p} t_i^2 + t_{p+1}^2 \geq d_1^2 = d^2/\ell^2$ where $0 < \ell < 1$ is to be chosen.

Then $P(\|t_{nm}\| \geq d) \geq P((a'Z_{nm})^2 \geq \ell^2)$ where $\|a\| = 1$.

7

Let $b_i, i = 1, \ldots, (p+1)$ be points in $\mathbb{R}^p$ and $r > 0$ be such that $P(Z_{j1} + Z_{j2} \; \varepsilon B(b_i, r)$ $\forall j = 1, \ldots p) > 0, \forall i = 1, \ldots, (p+1)$ and $B(b_i, r)$ are such that not all $(b_1, \ldots, b_{p+1})$ lie in any given hyperplane of dimension $(p-1)$. This is possible since the dispersion matrix of $(Z_{i1} + Z_{i2}, i = 1, \ldots, p)$ is positive definite.

Chose ℓ sufficiently small. Then for any a with $\|a\| = 1$, $\{x : (a'x)^2 < \ell^2\}$ does not intersect at least one of the balls $B(b_i, r)$.

Hence $P((a'Z_{nm})^2 \geq \ell^2) \geq \min_{i=1,\ldots,(p+1)} P(Z_{nm} \varepsilon B(b_i, r))$ and lim inf of the right hand side is positive. Thus there exists n_o, m_o large enough and $\varepsilon > 0$ such that

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

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

Further $\lim_{n \to \infty} D(n^{-1/2} \sum_{t=1}^n X_t) = \begin{pmatrix} \Sigma & 0 \\ 0 & V(\varepsilon_1^2) \end{pmatrix}$ is positive definite verifying C(7).

Thus Theorem 3.1 and Corollary 3.2 are valid for

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

Remark 3.3

The above arguments also show that if conditions (A1) and (A3) hold and ε_1 satisfies Cramer's condition, then the distribution of $n^{-1/2} S_n(\theta_n - \theta)$ admits an Edgeworth expansion of order $o(n^{-(s-2)/2})$.

Our next task is to derive 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: $\forall \|t\| \leq C.n^{\varepsilon^o}$ we have

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

8

for some $\varepsilon^o < 1/2$, and C depends on the bounds of β_{4n}^* ($=$ 4th 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 $I_1 = \left\{ t : Cn^{\varepsilon^o} \leq \|t\| \leq C_1 n^{1/2} \right\}$

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

<u>Lemma 3.5</u>: Under (A1), (A2) we have, for almost every sequence $Y_o, Y_1, \ldots$ and $|\alpha| \leq 6$,

$$\int\limits_{t \varepsilon I_2} |D^\alpha H_n^*(t)| dt = o(n^{-1/2}).$$

<u>Proof</u>: A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that

$$E^* |E^* A_p^* | \mathcal{D}_j^*, j \neq j_p| < 1 \text{ uniformly in } t \varepsilon I_2 \text{ and } p = 1, \ldots, \ell$$

where $A_p^* = \exp(it' n^{-1/2} \sum\limits_{j=j_p-m}^{j_p+m} Z_j^*)$. For definition 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

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

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

$$\delta_{nm} = \int |\int \exp(it_1' xy + it_2' x^2) dF_n^*(x)| dF_{nm}^*(y) \quad \text{(see (3.1))}$$

Note that a.s., $F_{nm}^* \Rightarrow F$ where F is the distribution of $Z_1 + Z_2$, Z_1, Z_2 iid $Z_1 \overset{\mathcal{L}}{=} \sum\limits_{t=0}^{\infty} \theta^t \varepsilon_t$ and by Levy's theorem F is continuous. Further $F_n^* \Rightarrow F_o$ a.s.. Since the convergence of F_{nm}^* to F is uniform we have

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

9

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

<u>Lemma 3.6</u>: Assume (A1), (A2) hold. For sufficiently small C_1, we have for almost every sequence $Y_o, Y_1, \ldots,$

$$\int\limits_{t\varepsilon I_1} |D^\alpha H_n^*(t)| dt = o(n^{-1/2}).$$

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

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

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

$$\delta_{nm}^* \leq P^*(|A_n^* + B_{nm}^*| \geq b)$$

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

$$\left. I(|A_n^* + B_{nm}^*| \leq b) \right]$$

For large b, the first term is $< 1/2$.

In the second term, the inner expectation equals

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

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

10

In the above expression the last term is less than or equal to

$$\frac{1}{6}b^3\frac{\|t\|^3}{n^{3/2}}\mu_{3n}^* \quad \text{where } \mu_{3n}^* = E^*\|(\varepsilon_n^*,\varepsilon_n^{*2})\|^3$$

$$\leq \beta\frac{b^3 C_1}{6}\frac{\|t\|^2}{n} \quad \text{a.s. noting that } \mu_{3n}^* \overset{a.s.}{\to} E\|(\varepsilon_1,\varepsilon_1^2)\|^3$$

$$\leq \alpha\frac{\|t\|^2}{n}, \quad \text{where } \alpha \text{ is as small as we please by taking } C_1 \text{ sufficiently small.}$$

Note that $D(\varepsilon_n^*,\varepsilon_n^{*2}) \overset{a.s.}{\to} D(\varepsilon_1,\varepsilon_1^2)$, which is positive definite, and on $\{|A_n^* + B_{nm}^*| \leq b\}$ we have,

$$|\frac{t_n'}{2n}D(\varepsilon_n^*,\varepsilon_n^{*2})t_n| \leq |D(\varepsilon_n^*,\varepsilon_n^{*2})|\frac{b^2}{2n}\|t\|^2 \leq 1 \text{ if } C_1 \text{ is small.}$$

Hence second term of δ_{nm}^* is bounded by

$$\delta_{2nm} = \alpha\frac{\|t\|^2}{n} + E^*\left[(1 - \frac{1}{2n}t_n'D(\varepsilon_n^*,\varepsilon_n^{*2})t_n)I(|A_n^* + B_{nm}^*| \leq b)\right]$$

$$\leq \alpha\frac{\|t\|^2}{n} + E^*(1 - \frac{1}{2n}I(|A_n^* + B_{nm}^*| \leq b)\lambda_1(\Sigma_n^*)\|t_n\|^2)$$

where $\lambda_1(\Sigma_n^*) = $ smallest eigen value of $D(\varepsilon_n^*,\varepsilon_n^{*2})$ and hence $\lambda_1(\Sigma_n^*) \overset{a.s.}{\to} \lambda_1(D(\varepsilon_1,\varepsilon_1^2)) > 0$.

If b is large enough so that $P^*\{|A_n^* + B_{nm}^*| \leq b\} > \frac{1}{2}$, we have

$$\left[E^*\left[(A_n^* + B_{nm}^*)^2 I(|A_n^* + B_{nm}^*| \leq b)\right] - E^*(A_n^* + B_{nm}^*)^2\right]$$

$$\leq \left[E^*(A_n^* + B_{nm}^*)^4 P^*(|A_n^* + B_{nm}^*| > b)\right]^{1/2}$$

$$< \eta, \quad \text{where } \eta \text{ can be made arbitrarily small by choosing } b \text{ large enough.}$$

(Note that $E^*(A_n^* + B_{nm}^*)^4 \overset{a.s.}{\to} E(Z_1 + Z_2)^4$, where Z_1 and Z_2 are iid, $Z_1 \overset{\mathcal{L}}{=} \sum_{t=0}^{\infty}\theta^t\varepsilon_t$).

And thus

$$\delta_{2nm} \leq \alpha\frac{\|t\|^2}{n} + E^*\left(1 - \frac{\lambda_1(\Sigma_n^*)}{2n}I(|A_n^* + B_{nm}^*| \leq b)(t_2^2 + t_1^2(A_n^* + B_{nm}^*)^2)\right)$$

This term in turn is dominated by

$$\alpha\frac{\|t\|^2}{n} + 1 - \gamma\frac{\|t\|^2}{n} \quad \text{(where } \gamma > \alpha)$$

$$\leq \exp(-\delta\frac{\|t\|^2}{n}).$$

11

This shows that $\delta^*_{nm} \le \rho + \exp\left(-\delta \frac{\|t\|^2}{n}\right)$ where $\rho < \frac{1}{2}$. 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).

<u>Lemma 3.7</u>: 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 $s \ge 2$. Further let $\alpha = K(x : \|x\| \le 1) > \frac{1}{2}$ and $\beta = \int \|x\|^{s+2} K(dx) < \infty$. Then for any $0 < \varepsilon < 1$,

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

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

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

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

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

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

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

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

Combining Lemmas 3.4-3.7 we can say that Theorem 3.1 and Corollary 3.2 hold for $S_n = n^{-1/2}\Sigma X^*_t$ for $s = 3$. $\qquad$ (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 the above result to the multidimension case is true and shall also be used.

<u>Lemma 3.8</u>: Let $\ell = (\ell_1, \ldots, \ell_k)$ be a vector $L = (L_{ij})$ be a $k \times k$ matrix and Q be a polynomial in k variables. Let $M \ge max(|V_{ij}|, |u_{ij}|, |\ell_i|, |L_{ij}|, |a_\lambda|)$ where $(V_{ij}) = V, (u_{ij}) = V^{-1}$

and a_λ are coefficients of Q. Let $|\ell_1| > \ell_o > 0$ and $b_n = (\ell_1 n^{1/2})^{-1}$. Then there exists a polynomial p in one variable, whose coefficients are continuous functions of $\ell_i, L_{ij}, V_{ij}, u_{ij}$ and a_λ such that

$$\int\limits_{\{z: \ell \cdot z + n^{-1/2} z' L z < u(\ell' V \ell)^{1/2}\}} (1 + n^{-1/2} Q(z)) \varphi_V(z) dz = \int_{-\infty}^{u} (1 + b_n p(y)) \varphi(y) dy + o(n^{-1/2}).$$

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

We now state and prove our main result.

Theorem 3.9: Under assumptions (A1) - (A3) for a.e. (Y_i),

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

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 equation (3.4)).

Hence for a.e. (Y_i),

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

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

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

Also note that by (3.3)

$$\sup_x |P(S_n \le x) - \int_{-\infty}^{x} (1 + n^{-1/2} p(x)) d\Phi_\Gamma(x)| = o(n^{-1/2}) \text{ where } \Gamma = \begin{pmatrix} \Sigma & 0 \\ 0 & V(\varepsilon_1^2) \end{pmatrix}$$

where $p(x)$ denotes a polynomial of the same form as $p(x, n)$.

13

Note that for $p = 1, \Sigma = (1 - \theta^2)^{-1}, \Sigma_n^* = (1 - \theta_n^2)^{-1}$.

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

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

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

$$A_n^* = n^{-1}(Y_o^{*2} - Y_n^{*2})$$

$$\text{Let} \quad B_i^* = \left\{|X_i^*| \geq c \log n\right\} \quad i = 1, 2$$

$$B_3^* = \left\{n^{3/4}|A_n^*| \geq c \log n\right\}$$

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 $B_1^{*c} \cap B_2^{*c} \cap B_3^{*c}$,

$$(1 - \theta_n^2)^{-1/2}n^{1/2}(\theta_n^* - \theta_n) = (\ell' X^* + n^{-1/2}X^{*\prime}A^* X^*)(\ell' \Gamma_n^* \ell)^{-1/2} + o(n^{-1/2}) \quad (3.6)$$

$$\text{where} \quad \ell' = (1,0), X^{*\prime} = (X_1^*, X_2^*), A^* = \begin{pmatrix} -2\theta_n, & -1/2 \\ -1/2 & 0 \end{pmatrix} \text{ and}$$

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

Analogous representation holds for $(1 - \theta^2)^{-1}n^{1/2}(\theta_n - \theta)$. Now note that the moments of $(X_{j-1}^* \varepsilon_j^*, \varepsilon_j^{*2} - \sigma_n^{*2})$ (under F_n^*) converges almost surely to those of $(X_{j-1}\varepsilon_j^*, \varepsilon_j^2 - 1)$ by 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.

$$\text{Define} \quad \tilde{Y}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}_{p \times 1}, \tilde{Z}_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p \times 1}$$

$$B = \begin{pmatrix} \theta_1 \ldots \theta_p \\ I \end{pmatrix}_{p \times p}$$

14

Clearly (2.1) is equivalent to $\tilde{Y}_t = B\tilde{Y}_{t-1} + \tilde{Z}_t$ $\qquad$ (3.7)

Let $A_n = n^{-1} \sum\limits_{t=1}^{n} \tilde{Y}_t \tilde{Y}_t'$, $B_n = n^{-1} \sum\limits_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Y}_{t-1}'$.

This gives $B_n - BB_nB' = n^{-1} \sum\limits_{t=1}^{n} \tilde{Z}_t \tilde{Z}_t' + 2Bn^{-1} \sum\limits_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t' + 0_p(n^{-1})$ $\qquad$ (3.8)

Note that $\Sigma = E\tilde{Y}_t \tilde{Y}_t'$ and satisfies $\Sigma = B\Sigma B + I^*$ where $I^* = \begin{pmatrix} 1 & 0\ldots0 \\ 0 & 0 \end{pmatrix}$

Equation (3.8) yields

$$B_n - \Sigma + B(B_n - \Sigma)B' = n^{-1} \sum\limits_{t=1}^{n} (\tilde{Z}_t \tilde{Z}_t' - I^*) + 2Bn^{-1} \sum\limits_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t' + 0_p(n^{-1})$$

which gives

$$B_n - \Sigma = G_1 \left[n^{-1} \sum\limits_{t=1}^{n} (\tilde{Z}_t \tilde{Z}_t' - I^*) + 2Bn^{-1} \sum\limits_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t' \right] G_2 + 0_p(n^{-1})$$

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

Let $\qquad V_n' = \left(\sum\limits_{t=1}^{n} X_{1t}, \ldots, \sum\limits_{t=1}^{n} X_{pt} \right).$

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

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

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

Let $\qquad B_i = \left\{ n^{-1/2} | \sum\limits_{t=1}^{n} X_{it}| \geq c \log n \right\}, i = 1, \ldots, (p+1).$

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

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

15

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

$$n^{1/2}(\theta_n - \theta) = n^{-1/2}\Gamma^{-1}\begin{pmatrix} \Sigma X_{1t} \\ \vdots \\ \Sigma X_{p+1,t} \end{pmatrix} + n^{-3/2}\begin{pmatrix} V_n'L_1V_n \\ \vdots \\ V_n'L_pV_n \end{pmatrix} + o(n^{-1/2}) \qquad (3.10)$$

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 sketch below 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 is,

$Y_t = \theta Y_{t-1} + \varepsilon_t + \mu$ where (ε_t) satisfies (A1) - (A3) but $E\varepsilon_t^2 = \sigma^2 > 0$ and μ and σ^2 are unknown.

Under assumptions (A1) - (A3), Edgeworth expansion is valid for the distribution of

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

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

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

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

An estimate σ_n^2 of σ^2 is given by

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

16

Thus, the estimates θ_n, μ_n and σ_n are all smooth functions of $\sum\limits_{t=1}^{n} Y_t, \sum\limits_{t=1}^{n} Y_t Y_{t-1}$ and $\sum\limits_{t=1}^{n} Y_t^2$ 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 upto $o(n^{-1/2})$, with the leading term as $\Phi(x)$, and the coefficients involved in the polynomial in the second term (which is $0(n^{-1/2})$) are smooth functions of θ, μ, σ^2 and of moments of $Y_t, Y_t Y_{t-1}$ and Y_t^2 of order less or equal to three. β can be explicitly calculated and depends on θ, μ and moments of ε_1.

The empirical distribution is computed by putting mass $1/n$ at each $\hat{\varepsilon}_i = Y_i - \theta_n Y_{i-1} - \mu_n, i = 1, \ldots, n$. 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 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 is 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.

References

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

Beran, R. (1982): Estimated sampling distribution: 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. (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.

Bickel, P. J. and Freedman, D. (1983): Bootstrapping regression models with a large number of parameters. In the *Lehmann Festschrift* ed. P. Bickel, J. L. Hodges and K. Doksum, Wadsworth, Belmont, California.

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

Daggett, R. and Freedman, D. A. (1984): Econometrics and the law: a case study in the proof of anti trust damages. *Proceedings of the Berkeley Conference in honor of J. Neyman and J. Kiefer*, ed. LeCam and Olshen, Vol. 1, 123–172. Wadsworth, Inc., Belmont, California.

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. *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.

Freedman, D. and Peters, S. (1984a): Bootstrapping a regression equations; some empirical results. *J. Amer. Statist. Assoc.* **79**, 97–106.

Freedman, D. and Peters, S. (1984b): Bootstrapping an econometric model; some empirical results. *J. Bus. Econ. Statist.* **2**, 150–158.

Freedman, D. A. and Peters, S. C. (1984c): Using the bootstrap to evaluate forecasting equations. *Tech. Rep. No.* **20**, Univ. of California, Berkeley, Dept. of Statistics. To appear in J. of Forecasting.

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. Statist.* **9**, 1187–1195.

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

Arup Bose
Department of Statistics
Mathematical Sciences Building
Purdue University
West Lafayette, IN   47907