The Annals of Statistics
1988, Vol. 16, No. 3, 1283–1314

ON THE STRONG APPROXIMATION OF THE DISTRIBUTIONS OF ESTIMATORS IN LINEAR STOCHASTIC MODELS, I AND II: STATIONARY AND EXPLOSIVE AR MODELS[1]

BY P. JEGANATHAN

The University of Michigan

It is shown that in the stationary autoregressive case (Part I [Sections 1–5)] the distributions of least squares estimators and their close relatives such as sample autocovariances and the recently introduced general M-estimators converge to suitable Gaussian distributions uniformly over all Borel sets and uniformly over suitable neighborhoods of the parameter. Specifically, the notion of strongly asymptotically shift equivariance is introduced and it is shown that the distributions of any estimators satisfying this asymptotic equivariance condition converge in the preceding strong sense whenever they converge weakly (in law), provided the likelihood function of the sample is appropriately smooth. This smoothness of the likelihood is verified under mild conditions. Then, restricted to a broad class of models which include autoregressive models, a more easily verifiable condition implying the aforementioned asymptotic equivariance is derived and is shown to be satisfied by the estimators mentioned earlier. The methods used in the present paper are different from the usual method of characteristic functions; some indications of their possible wider scope are given.

In Part II (Sections 6–8) the explosive autoregressive model is considered and a simple extension of the preceding result is applied to show that the least squares estimators converge in the preceding strong sense under suitable random or nonrandom normalization.

1. Stationary autoregressive case: Introduction. Consider the pth order univariate autoregressive model

$$Y_n = \beta_0 + \beta_1 Y_{n-1} + \cdots + \beta_p Y_{n-p} + \varepsilon_n,$$

$n = 1, 2, \ldots,$ where ε_i, $i \geq 1$, are i.i.d., independent of $(Y_0, \ldots, Y_{1-p})$, $E(\varepsilon_1^2) = \sigma^2 < \infty$ and the roots of the characteristic polynomial in m,

$$m^p - \beta_1 m^{p-1} - \cdots - \beta_p = 0,$$

are all less than 1 in absolute value (stationary case). $\beta' = (\beta_0, \ldots, \beta_p)$ and σ^2 are parameters to be estimated.

One of the most familiar estimators of β and σ^2 is the least squares (L.S.) estimator $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ defined by

$$\hat{\beta}_n = \left(\hat{\beta}_{0n}, \ldots, \hat{\beta}_{pn}\right)' = A_n^{-1} \sum_{j=1}^{n} \tilde{Y}_{j-1} Y_j$$

Received December 1986; revised December 1987.
[1]Supported by NSF Grant DMS-85-09837.
AMS 1980 *subject classifications*. Primary 62F12; secondary 62G20.
Key words and phrases. Strong convergence, local asymptotic normality, local asymptotic mixed normality, asymptotic shift equivariance, stationary autoregression, purely and partially explosive autoregression, least squares, M-estimators.

and

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^{n} \left(Y_j - \hat{\beta}_{0n} - \sum_{i=1}^{p} \hat{\beta}_{in} Y_{j-i} \right)^2,$$

where

$$\tilde{Y}_j = (1, Y_j, Y_{j-1}, \dots, Y_{j-p+1})'$$

and

$$A_n = \sum_{j=1}^{n} \tilde{Y}_{j-1} \tilde{Y}'_{j-1}.$$

It was shown by Mann and Wald (1943) under the further restriction that all moments of ε_1 exist and are finite that $n^{-1}A_n$ converges in probability to a p.d. matrix A and that $\sqrt{n}(\hat{\beta}_n - \beta)$ converges in law to a $(p+1)$-variate Gaussian distribution with mean vector 0 and covariance matrix $\sigma^2 A^{-1}$. A detailed treatment of this result with improvements and generalizations can be found, e.g., in Anderson (1971) and Fuller (1976). More general estimators of β and σ^2 called M-estimators have also been recently introduced in connection with robust estimation and have been shown to converge in law to appropriate Gaussian distributions, see, e.g., Bustos (1982) and Martin and Yohai (1985). In the present paper it is shown that the Gaussian approximation to the L.S. estimators (and their close relatives such as sample autocovariances) and more general M-estimators holds uniformly over all Borel subsets and uniformly over suitable neighborhoods of the parameter. (Such convergence will loosely be referred to later as strong convergence.) A similar result is also proved for the L.S. estimators in the explosive case, that is, the case where the characteristic polynomial has at least one root outside the unit circle but has no roots on the unit circle. The unstable case, that is, the case where the characteristic polynomial has no roots outside the unit circle, but has at least one root on the unit circle, is treated in Jeganathan (1987a). Estimators in other stochastic models such as, e.g., regression with lagged dependent variables and with time series errors and ARMA models will be treated in a subsequent part.

Some indications of the statistical motivation for considering such a strong convergence can be found in Jeganathan (1987b). [This paper will henceforth be referred to in short by J(1987b)]. Briefly, the information contained in the estimators is asymptotically reflected in the limiting distributions only when the likelihood ratios of the estimators converge in law appropriately to the likelihoods of the limiting distributions, as is explained in, e.g., Le Cam (1986), and such convergence of likelihood ratios becomes equivalent to the strong convergence of distributions of estimators under the asymptotic equivariant restriction of the present paper.

We now would like to explain some of the differences between the methods that will be used in the present paper and the usual method of characteristic functions (Ch.f.) used in the case of sums of independent random variables to prove strong convergence. The method of Ch.f.'s is very powerful when the summands are independent, though it depends to some extent on the identical nature of the distributions of the summands when strong convergence and other higher order (Edgeworth) expansions are needed; see Bhattacharya and Ranga Rao (1976) for a thorough treatment of the subject. See also Lemma 6.2 of Le Cam (1960). It has also been shown by Götze and Hipp (1983) that this method works well even when the summands are not independent but have suitable asymptotic independence structures and the events or functions that are approximated are assumed to be appropriately smooth. In the preceding case of L.S. estimators the asymptotic independence structures and other required conditions are satisfied [cf. Bose (1985)], but our interest is approximation over all Borel sets, and we do not know if the methods of Götze and Hipp can be successfully employed for such a strong approximation. Another difficulty, which arises even when the sample is i.i.d. is the following. In most cases of interest the estimators, in particular M-estimators, can be approximated by sums of i.i.d. r.v.'s when the sample is i.i.d., but the remainder term will also be random. In such cases Bhattacharya and Ghosh (1978) have shown that it is possible to get (higher order) approximations if the events or functions that are approximated are appropriately smooth and if the (random) remainder terms converge to 0 at suitable rates, but again it is not clear if such methods work well when the strong (higher order) approximations are needed. But note that one of the fundamental results of Bhattacharya and Ghosh (1978) is that strong higher order approximation holds if the estimator itself is a smooth function, independent of the sample size, of the sums of independent r.v.'s, but, e.g., M-estimators are typically not of this form. The crucial idea of our previous paper J(1987b) and the present paper is that if one has the approximation even only for extremely smooth events or functions and if the likelihood ratios of the sample are sufficiently smooth, then one can get the approximation for all events provided the estimator is *strongly asymptotically shift equivariant* (SASE) (see Section 2 for a precise definition). In the present paper, as in our previous paper J(1987b), we consider only the first order approximation. An attempt to further develop this method to get higher order approximations in time series and nonlinear regression models is presently being made. A definite form of the idea of the present method occurred to us after having seen the paper by Boos (1985); see J(1987b) for details.

The present paper is actually a continuation of our previous paper J(1987b). The main result of Section 2 of that paper is that, under the restriction that the likelihoods of the sample are smooth in the sense that they are *locally asymptotically normal* (LAN) [cf. Le Cam (1960)] and under the equivariance restriction indicated previously, convergence in law entails strong convergence; this result with suitable modifications will be recalled together with some further essential results in Section 2. Further, the applications considered in that paper were estimates of location, scatter and regression parameters, and the previously mentioned restrictions were readily verified in those cases because of the exact equivariance of the estimators and the models considered there. The situation of the present paper is not that immediate and, in fact, most of Sections 3–7 are devoted to the verification of the SASE condition. Specifically, in Section 3, restricted to a broad class of models which include stationary AR models, a general result is obtained giving an easily verifiable condition implying the aforementioned SASE condition. For L.S. estimators and sample autocovariances the conditions of this result are verified in Section 4. In Section 5 we then consider a "contaminated AR model" and show that M-estimators satisfy the requirements of the preceding result. Sections 6–8, which constitute Part II of the present paper, are devoted to the L.S. estimators in the explosive case, but in this case approximations more general than LAN are needed. For instance, when the errors are Gaussian the likelihoods of the sample are *locally asymptotically mixed normal* (LAMN) [see, e.g., Davies (1985)], a situation which is more general than, but very similar to, that of Section 2. However, as will be noted, the results of Sections 2 and 3 extend to these more general cases also. In the explosive case when there are also some roots inside the unit circle, the existing results on the convergence in law of L.S. estimators do not seem to be satisfactory, especially on the choice of normalizing constants; we have taken some care on this point, see Section 8 for the details.

2. Likelihood of the sample and the strong convergence of the estimators. For each $n \geq 1$, let $\{P_{\theta,n}: \theta \in \Theta\}$ be a family of probability measures (defined on some measurable space), where the parameter space Θ is assumed to be an open subset of the k-dimensional Euclidean space R^k, $k \geq 1$. Usually, $P_{\theta,n}$ stands for the joint probability distribution of the sample.

DEFINITION 1. The family $\{P_{\theta,n}: \theta \in \Theta\}$ is said to be LAN at $\theta_0 \in \Theta$ if there exists a sequence $\{W_n(\theta_0)\}$ of random k-vectors and a p.d. matrix $B(\theta_0)$ such that the differences

$$\log \frac{dP_{\theta_0+\delta_n h_n, n}}{dP_{\theta_0, n}} - \left[h_n' W_n(\theta_0) - \frac{1}{2} h_n' B(\theta_0) h_n\right]$$

converge to 0 in $P_{\theta_0,n}$-probability for every bounded $\{h_n\}$ of R^k, where δ_n, which may depend on θ_0, $n \geq 1$, are suitable p.d. matrices and the sequence $\{W_n(\theta_0)\}$ converges in law under $P_{\theta_0,n}$ to the k-variate Gaussian distribution $N(0, B(\theta_0))$. (In Sections 4 and 5 of the present paper, δ_n will be taken to be $n^{-1/2}I$; I is the identity matrix.)

The following notation is needed for the next definition: μ denotes the Lebesgue measure in R^k, $D_\alpha = \{h \in R^k: |h| < \alpha\}$, $\alpha > 0$, and $\theta_n(h)$ stands for $\theta_0 + \delta_n h$. [Throughout the paper $(h_1, \dots, h_k)' = h \in R^k$ and $|h|$ stands for the supremum norm $|h| = \max\{|h_1|, \dots, |h_k|\}$.]

DEFINITION 2. A sequence of estimators $\{T_n\}$ of a k-vector $g(\theta)$ is said to be SASE at $\theta_0 \in \Theta$ if for some nonsingular matrix F_{θ_0}, the differences

$$E\Big[f\big(T_n^*(\theta_0) - F_{\theta_0} h_n\big) | P_{\theta_n(h_n), n}\Big] \\ - \frac{1}{\mu(D_\alpha)} \int_{D_\alpha} E\Big[f\big(T_n^*(\theta_0) - F_{\theta_0}(h_n + u)\big) | P_{\theta_n(h_n+u), n}\Big]\, du$$

converge to 0 for all bounded sequences $\{h_n\}$ of R^k, for all large $\alpha > 0$ and uniformly for all Borel measurable f such that $|f| \le 1$, where we let $T_n^*(\theta) = \delta_n^{-1}(T_n - g(\theta))$.

DEFINITION 3. A sequence of estimators $\{T_n\}$ of a k-vector $g(\theta)$ is said to *converge strongly* to a distribution Q_θ at $\theta_0 \in \Theta$ if, for some nonsingular matrix F_{θ_0}, the differences

$$E\Big[f\big(T_n^*(\theta_0) - F_{\theta_0}h_n\big)|P_{\theta_0(h_n), n}\Big] - \int f(x)\, dQ_{\theta_0}(x)$$

converge to 0 for all bounded $\{h_n\}$ of R^k and uniformly for all Borel measurable functions f such that $|f| \le 1$, where we let $T_n^*(\theta) = \delta_n^{-1}(T_n - g(\theta))$.

Note that the preceding convergence is equivalent to

$$\sup_{|h| \le \alpha} \sup_{B \in \mathscr{B}} |P_{\theta_n(h), n}\big[T_n^*(\theta_0) - F_{\theta_0}h \in B\big] - Q_{\theta_0}(B)| \to 0$$

for all $\alpha > 0$, where $\mathscr{B}$ is the Euclidean Borel field of R^k. A similar remark applies to the SASE condition. For the cases treated in the present paper, it is possible to get the convergence to be uniform in $\theta_0 \in K \subseteq \Theta$ also for every compact K.

We now recall the main result (Theorem 1) of J(1987b).

THEOREM 1. *Assume that* $\{P_{\theta, n}: \theta \in \Theta\}$, $n \ge 1$, *satisfies the LAN condition at* $\theta_0 \in \Theta$. *Let* $\{T_n\}$ *be a sequence of estimators of a k-vector* $g(\theta)$ *such that*

(i) *the SASE condition is satisfied at* θ_0 *and that*

(ii) *for some nonsingular matrix* F_{θ_0}, *the sequence* $\{\delta_n^{-1}(T_n - g(\theta_0) - \delta_n F_{\theta_0} h)\}$ *converges in law under* $P_{\theta_n(h), n}$ *to a distribution* Q_{θ_0} *for every* $h \in R^k$, *where* Q_{θ_0} *does not depend on h.*

Then the sequence $\{T_n\}$ *converges strongly to the distribution* Q_{θ_0} *at* θ_0.

REMARK. The preceding formulation of the result is slightly different from Theorem 1 of J(1987b) in that the preceding SASE condition is weaker than the corresponding condition in J(1987b), but note that the proof in J(1987b) is actually the proof of Theorem 1. Even though both the preceding SASE condition and the corresponding stronger one in J(1987b) are implied by the strong convergence, in the present situation we could succeed verifying directly only the weaker SASE condition. It may further be noted that the main reason to state the SASE condition in the preceding form with the uniform prior on D_α is that it suits the proof given in J(1987b). It is possible to replace the uniform prior density on D_α by any density of the form $f(u)/\nu(D_\alpha)$, where f is an appropriately smooth probability density on R^k with the corresponding probability measure ν.

The following result [cf. Le Cam (1960)] will be used in the verification of the SASE condition. This result also entails that the measurability of the functions $\theta \to P_{\theta, n}$ involved in the definition of the SASE condition is not a restriction.

THEOREM 2. *Assume that* $\{P_{\theta, n}: \theta \in \Theta\}$, $n \geq 1$, *satisfy the LAN condition at* $\theta_0 \in \Theta$. *Then one can construct a sequence* $\{W_n^*(\theta_0)\}$ *of random k-vectors satisfying the following properties*:

(i) $E[\exp[u'W_n^*(\theta_0)] | P_{\theta_n(h), n}] < \infty$ *for all* $n \geq 1$ *and all* $h, u \in R^k$.

(ii) *Let* $K_n(\theta_0, h, u)$ *be functions such that the measures* $Q_n(\theta_0, h, u)$, *defined by*

$$\frac{dQ_n(\theta_0, h, u)}{dP_{\theta_n(h), n}} = K_n(\theta_0, h, u)\exp\left[u'[W_n^*(\theta_0) - B(\theta_0)h] - \frac{1}{2}u'B(\theta_0)u\right],$$

are probability measures. Then

$$\sup_{|h| \leq \alpha, |u| \leq \alpha} |K_n(\theta_0, h, u) - 1| \to 0$$

and

$$\sup_{|h| \leq \alpha, |u| \leq \alpha} \|P_{\theta_n(h+u), n} - Q_n(\theta_0, h, u)\| \to 0$$

for all $\alpha > 0$. (*Here* $\|\nu\|$ *denotes the total variation of the signed measure* ν.)

Note that as a consequence of (ii), the differences $W_n(\theta_0) - W_n^*(\theta_0) \to 0$ in $P_{\theta_0, n}$-probability where $W_n(\theta_0)$ are the random k-vectors involved in the LAN condition.

We now consider the LAN of the AR model. Let $(Y_{1-p}, \dots, Y_0, Y_1, \dots, Y_n)$ be a sample from the stationary AR process defined in the Introduction. $P_{\theta, n}$, $\theta' = (\beta', \sigma^2)$, will henceforth stand for the conditional distribution of the preceding sample given that $(Y_{1-p}, \dots, Y_0)$ is a known constant vector. In what follows, notations and assumptions introduced in Section 1 will be used throughout. In particular, recall that:

(B1) The roots of the polynomial equation in m,

$$m^p - \beta_1 m^{p-1} - \cdots - \beta_p = 0,$$

are all less than 1 in absolute value.

Let $(1/\sigma)f(x/\sigma)$ be the common p.d.f. with respect to Lebesgue measure of the i.i.d. sequence ε_n, $n \geq 1$, recall that $\varepsilon_n = Y_n - \tilde{Y}'_{n-1}\beta$. Then the p.d.f. of $P_{\theta, n}$ is given by

$$\frac{1}{\sigma^{n/2}} \prod_{j=1}^{n} f(\varepsilon_j/\sigma). \tag{1}$$

To proceed further, we introduce the following assumption.

(B2) The density f is absolutely continuous with respect to Lebesgue measure,

$$0 < \int_{-\infty}^{\infty} \left[\frac{\dot{f}(x)}{f(x)} \right]^2 f(x)\,dx < \infty \tag{2}$$

and

$$0 < \int_{-\infty}^{\infty} \left[1 + x\frac{\dot{f}(x)}{f(x)} \right]^2 f(x)\,dx < \infty, \tag{3}$$

where $\dot{f}$ denotes the derivative, whenever it exists, of the density f.

Note that condition (3) will not be needed if the scale parameter σ^2 is assumed to be known.

THEOREM 3. (i) *Let $P_{\beta, n}$ be the distribution of the sample from the autoregressive model of Section 1, assuming that σ^2 is known. Let Θ be the set of all β's such that Assumption* (B1) *is satisfied. Further assume that condition* (2) *holds. Then the families $\{P_{\beta, n}: \beta \in \Theta\}$, $n \geq 1$, satisfy the LAN condition at all $\beta \in \Theta$.*

(ii) *Let $P_{\theta, n}$, $\theta' = (\beta', \sigma^2)$, be the distribution of the sample when both β and σ^2 are unknown. Let Θ be as before. Assume that conditions* (2) *and* (3) *are satisfied. Then the families $\{P_{\theta, n}: \theta \in \Theta \times (0, \infty)\}$, $n \geq 1$, satisfy the LAN condition at all $\theta \in \Theta \times (0, \infty)$.*

Proof of statement (i) can be obtained from Swensen (1985). A proof of both statements is also given in Jeganathan (1986).

3. A general result implying the SASE condition. The purpose of this section is to prove the following general result. The result is general in the sense that it is applicable to a wide class of econometric models.

THEOREM 4. *Let $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ be a sample of n observations having the joint distribution $P_{\theta, n}$, $\theta \in \Theta \subseteq R^k$, $k \geq 1$, where Θ is open. Let $\hat{\theta}_n(\mathbf{Y}_n)$, $n \geq 1$, be a sequence of estimators of a k-vector $g(\theta)$. Assume that the following three conditions are satisfied*:

(A) *The sequence $\{P_{\theta, n}: \theta \in \Theta\}$, $n \geq 1$, satisfies the LAN condition at $\theta \in \Theta$ with the normalizing matrices δ_n, $n \geq 1$.*

(B) *There are transformations $U_{n, \theta, h_n, u} = U_n$: $R^n \to R^n$, such that the differences* [*with* $\theta_n(h) = \theta + \delta_n h$],

$$E\left[f(\hat{\theta}_n(\mathbf{Y}_n)) | P_{\theta_n(h_n), n} \right] - E\left[f(\hat{\theta}_n(U_n(\mathbf{Y}_n))) | P_{\theta_n(h_n + u), n} \right],$$

converge to 0 as $n \to \infty$ for all bounded sequences $\{h_n\}$ of R^k, for all $u \in R^k$ and uniformly for all Borel measurable f such that $|f| \leq 1$.

(C) *Denote, suppressing* h_n *and* θ, $\hat{\theta}_n(U_n(\mathbf{Y}_n))$ *by* $\tilde{\theta}_n(u)$, *where* U_n *is the transformation of Condition* (B). *Then there is a nonsingular matrix A such that the quantities*

$$\sup\left\{|\delta_n^{-1}\left(\tilde{\theta}_n(u) - \tilde{\theta}_n(v)\right) - A(v-u)|/|u-v|: |u| \le \alpha, |v| \le \alpha\right\}$$

converge to 0 *in* $P_{\theta,n}$*-probability for all* $\alpha > 0$.

Then the SASE condition is satisfied.

We first separate certain technical parts of the proof and state them as lemmas. We need some additional notation. The left open and right closed cube $\{(x_1,\dots,x_k)' = x \in R^k: a_i - \delta < x_i \le a_i + \delta, i = 1,2,\dots,k\}$ will be denoted by $D(a,\delta)$, $a' = (a_1,\dots,a_k)$. $\bar{D}(a,\delta)$ denotes the closure of $D(a,\delta)$; note that $\bar{D}(a,\delta) = \{x \in R^k: |x-a| \le \delta\}$. $\bar{D}(0,\delta)$ will be denoted by $\bar{D}_\delta$.

The following important lemma will be used crucially. It can be found, e.g., in Rudin [(1974), page 185], but the proof is only sketched there with arguments more involved than the following ones.

Lemma 5. *Let* $g: \bar{D}_\delta \to R^k$ *be a continuous function such that* $|g(x) - x| \le \varepsilon$ *for all* $x \in \bar{D}_\delta$. *Then*

$$\bar{D}_{\delta-\varepsilon} \subseteq g(\bar{D}_\delta) \subseteq \bar{D}_{\delta+\varepsilon}.$$

Proof. Since $x \in \bar{D}_\delta$ implies $|g(x)| \le |x| + \varepsilon \le \delta + \varepsilon$, we have $g(\bar{D}_\delta) \subseteq \bar{D}_{\delta+\varepsilon}$. To prove the other part, denote the column k-vector $g(x)$ by $(g_1(x),\dots,g_k(x))'$ and the column k-vector x by $(x_1,\dots,x_k)'$. Let $x^{k-1} = (x_1,\dots,x_{k-1},0)'$. Then

$$\begin{aligned}&\left\{(g_1(x),\dots,g_{k-1}(x))': x \in \bar{D}_\delta\right\}\\ &\quad\supseteq \left\{\left(g_1(x^{k-1}),\dots,g_{k-1}(x^{k-1})\right)': x^{k-1} \in \bar{D}_\delta\right\} = C_\delta, \quad \text{say.}\end{aligned}$$

Further, for each fixed $x_i \in [-\delta,\delta]$, $i = 1,2,\dots,k-1$, the function $g_k^*(x_k) = g_k(x)$: $[-\delta,\delta] \to R$ is continuous in x_k so that the set $g_k^*([-\delta,\delta])$ is both compact and connected in the real line. Hence

$$\begin{aligned}g_k^*([-\delta,\delta]) &= \left[\inf_{|x_k|\le\delta} g_k^*(x_k), \sup_{|x_k|\le\delta} g_k^*(x_k)\right],\\ &\supseteq [-\alpha+\varepsilon, \alpha-\varepsilon],\end{aligned}$$

since by assumption $|g_k(x) - x_k| \le \varepsilon$ for all $x \in \bar{D}_\delta$. Thus the set $g(\bar{D}_\delta)$ contains the Cartesian product $C_\delta \times \{x_k: |x_k| \le \delta - \varepsilon\}$.

Further note that by assumption [with $x^{k-1} = (x_1,\dots,x_{k-1},0)'$] $|g_i(x^{k-1}) - x_i| \le \varepsilon$, $i = 1,2,\dots,k-1$, for all $x^{k-1} \in \bar{D}_\delta$ so that by repeating the preceding arguments one arrives at the required result $g(\bar{D}_\delta) \supseteq \times_{i=1}^k\{x_i: |x_i| \le \delta - \varepsilon\} = \bar{D}_{\delta-\varepsilon}$. This completes the proof. □

To proceed further, we note the following fact.

Remark. Every open subset of R^k can be written as a countable disjoint union of the left open and right closed cubes of the form $D(a,\delta)$.

Results of the following type are probably well known, but we obtain it as an easy consequence of Lemma 5. (Recall that $D_\alpha = \{x \in R^k: |x| < \alpha\}$.)

LEMMA 6. *Let* $g: \overline{D}_\alpha \to R^k$ *be a function such that, for some* $0 < \varepsilon < 1$,

$$|g(x) - g(y) - (x - y)| \le \varepsilon|x - y| \tag{4}$$

for all x *and* y *in* $\overline{D}_\alpha$. *Then* g *is* 1–1 *such that* $g(B \cap D_\alpha)$ *is Borel measurable for every Borel measurable subset* B *of* R^k. *Furthermore,*

$$(1-\varepsilon)^k \int_{D_\alpha} f(x)\,dx \le \int_{g(D_\alpha)} f(g^{-1}(x))\,dx \le (1+\varepsilon)^k \int_{D_\alpha} f(x)\,dx$$

for all nonnegative Borel measurable f.

PROOF. Condition (4) implies that

$$|x-y|(1-\varepsilon) \le |g(x) - g(y)| \le (1+\varepsilon)|x-y|$$

for all x and y in $\overline{D}_\alpha$, so that g is 1–1 and continuous on $\overline{D}_\alpha$. Continuity implies that for every closed subset C of R^k that is contained in D_α, $g(C)$ is compact and hence measurable. Hence, since g is 1–1 one can easily see that the sets $g(D(a, \delta))$ are measurable if $D(a,\delta) \subseteq D_\alpha$. Using the preceding remark, this implies that $g(U \cap D_\alpha)$ is measurable for every open subset U of R^k. Hence, since g is 1–1 it follows from the usual arguments that $g(B \cap D_\alpha)$ is measurable for every measurable subset B of R^k. This proves the first part.

To prove the second part, note that Lemma 5 implies that

$$\overline{D}(g(a), \delta(1-\varepsilon)) \subseteq g(\overline{D}(a,\delta)) \subseteq \overline{D}(g(a), \delta(1+\varepsilon)) \tag{5}$$

for every $\overline{D}(a,\delta) \subseteq \overline{D}_\alpha$. Further, if μ is the Lebesgue measure in R^k,

$$\mu(\overline{D}(g(a), \delta(1 \pm \varepsilon))) = (1 \pm \varepsilon)^k \mu(D(a,\delta)). \tag{6}$$

Hence (5) and (6) readily imply that

$$(1-\varepsilon)^k \mu(D(a,\delta)) \le \mu(g(D(a,\delta))) \le (1+\varepsilon)^k \mu(D(a,\delta))$$

for every $D(a,\delta) \subseteq D_\alpha$. Hence, in view of the preceding remark,

$$(1-\varepsilon)^k \mu(U \cap D_\alpha) \le \mu(g(U \cap D_\alpha)) \le (1+\varepsilon)^k \mu(U \cap D_\alpha) \tag{7}$$

for every open subset U of R^k. Now note that since g is 1–1, $\nu(B) = \mu(g(B \cap D_\alpha))$ is a Borel measure. Further note that, as is well known, every finite Borel measure ν on R^k is regular in the sense that $\nu(B) = \inf\{\nu(U): B \subseteq U, U \text{ open})$ for every Borel subset B. Thus (7) implies that

$$(1-\varepsilon)^k \mu(B \cap D_\alpha) \le \mu(g(B \cap D_\alpha)) \le (1+\varepsilon)^k \mu(B \cap D_\alpha)$$

for all Borel sets B. This can be rewritten as

$$(1-\varepsilon)^k \int_{D_\alpha} I_B(x)\,dx \le \int_{g(D_\alpha)} I_B(g^{-1}(x))\,dx \le (1+\varepsilon)^k \int_{D_\alpha} I_B(x)\,dx$$

for every Borel subset B, where I_B denotes the indicator function of the set B, so that the result is proved for the functions $f = I_B$ and hence for every nonnegative measurable f. This completes the proof. □

In what follows, let

$$S_n(\theta, u) = \frac{|\det B|^{1/2}}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\left(u - B^{-1}W_n\right)' B\left(u - B^{-1}W_n\right)\right\},$$

where (suppressing θ), W_n and B are the ones involved in the definition of LAN condition.

As a consequence of Lemmas 5 and 6 we now obtain Lemma 7.

LEMMA 7. *Assume that Condition* (C) *of Theorem* 4 *is satisfied. Let* $\tilde{\theta}_n(u)$ *and the matrix A be as defined in Condition* (C). *Then the supremum, taken over all* f *such that* $|f| \le 1$, *of the absolute value of the differences*

$$(8)\quad \int_{D_\alpha} f\left(\delta_n^{-1}\left(\tilde{\theta}_n(u) - \theta\right)\right) S_n(\theta, u)\, du - \int_{D_\alpha} f\left(\delta_n^{-1}\left(\hat{\theta}_n - \theta\right) - Au\right) S_n(\theta, u)\, du,$$

converges to 0 *in* $P_{\theta, n}$*-probability.*

PROOF. Let $g_n(u) = \delta_n^{-1}(\hat{\theta}_n - \tilde{\theta}_n(u))$. Then Condition (C) in particular implies that the quantities

$$\sup_{|u| \le \alpha} |g_n(u) - Au|$$

converge to 0 in probability. Hence it follows easily that the difference between the l.h.s. of (8) and the quantity

$$\int_{D_\alpha} J_f(g_n(u))\, du$$

converges to 0 in probability, uniformly in f, where we let $J_f(u) = f(\delta_n^{-1}(\hat{\theta}_n - \theta) - u) S_n(\theta, A^{-1}u)$. Now, according to Condition (C), there is a decreasing sequence $\{\varepsilon_n\}$ of positive numbers such that $\varepsilon_n \downarrow 0$ and $P_{\theta, n}(E_n) \to 1$, where

$$E_n = \left\{\left[\sup |A^{-1}(g_n(u) - g_n(v)) - (v - u)|/|v - u| \colon |u| \le \alpha, |v| \le \alpha\right] \le \varepsilon_n\right\}.$$

Therefore, according to Lemma 6, we have for every nonnegative f,

$$(9)\quad \begin{aligned} (1 - \varepsilon_n)^k \int_{D_\alpha} J_f(g_n(u))\, du &\le |A|^{-1} \int_{g_n(D_\alpha)} J_f(u)\, du \\ &\le \left(1 + \varepsilon_n^k\right) \int_{D_\alpha} J_f(g_n(u))\, du \end{aligned}$$

whenever the event E_n is true, where $|A|$ is the determinant of A. Further, Lemma 5 implies that whenever the event E_n is true and f is nonnegative,

$$(10)\quad \int_{AD_{\alpha(1-\varepsilon_n)}} J_f(u)\, du \le \int_{g_n(D_\alpha)} J_f(u)\, du \le \int_{AD_{\alpha(1+\varepsilon_n)}} J_f(u)\, du.$$

But denoting the symmetric difference of two sets by Δ,

$$\left| \int_{AD_{\alpha(1 \pm \varepsilon_n)}} J_f(u)\, du - \int_{AD_\alpha} J_f(u)\, du \right| \le |A| \int_{D_\alpha \Delta D_{\alpha(1 \pm \varepsilon_n)}} S_n(\theta, u)\, du, \tag{11}$$

and since $\varepsilon_n \downarrow 0$ and $P_{\theta, n}(E_n) \to 1$ the required result follows easily from (9)–(11). □

Proof of Theorem 4. We verify the SASE condition when $h_n \equiv 0$, $n \ge 1$; verification for the general bounded $\{h_n\}$ requires only the notational change of replacing θ by $\theta_n(h_n)$ throughout. Explicitly, we want to verify that for some nonsingular matrix F_θ, the differences

$$E_\theta\left[f\left(\delta_n^{-1}(\hat{\theta}_n - \theta)\right)\right] - \frac{1}{\mu(D_\alpha)} \int_{D_\alpha} E_{\theta_n(u)}\left[f\left(\delta_n^{-1}(\hat{\theta}_n - \theta) - F_\theta u\right)\right] du \tag{12}$$

converge to 0 uniformly in f, $|f| \le 1$. According to Condition (B), the differences between the l.h.s. of (12) and the quantity

$$\frac{1}{\mu(D_\alpha)} \int_{D_\alpha} E_{\theta_n(u)}\left[f\left(\delta_n^{-1}(\tilde{\theta}_n(u) - \theta)\right)\right] du \tag{13}$$

converge to 0 uniformly in f. In what follows, we assume without loss of any generality that the measures $P_{\theta_n(h+u), n}$ are identically the same as the ones constructed in Theorem 2. Now (13) can be rewritten as

$$\frac{1}{\mu(D_\alpha)} \int_{D_\alpha} \int \left\{ \frac{G_n \int_{D_\alpha} f\left(\delta_n^{-1}(\tilde{\theta}_n(u) - \theta)\right)\left(dP_{\theta_n(u), n}/dP_{\theta, n}\right) du}{G_n \int \left(dP_{\theta_n(u), n}/dP_{\theta, n}\right) du} \right\} \tag{14}$$
$$\times dP_{\theta_n(v), n}\, dv,$$

where

$$G_n = \frac{|\det B|^{1/2}}{(2\pi)^{k/2}} \exp\left[\frac{1}{2} W_n' B^{-1} W_n \right].$$

Now, using Theorem 2, the numerator of the ratio inside the bracket of (14) can be approximated with $P_{\theta, n}$-probability tending to 1 by

$$\int_{D_\alpha} f\left(\delta_n^{-1}(\tilde{\theta}_n(u) - \theta)\right) S_n(\theta, u)\, du. \tag{15}$$

By Lemma 7 and again using Theorem 2, (15) can be approximated with $P_{\theta, n}$-probability tending to 1 by

$$G_n \int_{D_\alpha} f\left(\delta_n^{-1}(\hat{\theta}_n - \theta) - Au\right) \frac{dP_{\theta_n(u), n}}{dP_{\theta, n}}\, du. \tag{16}$$

Further, all the approximations hold, in view of contiguity, with respect to $P_{\theta_n(v), n}$-probabilities uniformly in v, $|v| \leq \alpha$. Thus, if the numerator of the ratio inside the bracket of (14) is replaced by (16), the difference between (14) and the quantity resulting by the replacement converge to 0 uniformly in f. Since this resulting quantity is the same as the r.h.s. of (12), the proof of the theorem is complete. □

4. Strong convergence of L.S. estimators and sample autocovariances. Let $(Y_{1-p}, \dots, Y_0, Y_1, \dots, Y_p)$ be a sample from the stationary AR model of Section 1. In this section, Theorem 4 of Section 3 will be applied to estimators of $\theta = (\beta', \sigma^2)'$ that can be written as explicit functions of the observations of the sample. L.S. estimators defined in Section 1 form one such example. Another estimator, $\bar{\theta}_n = (\hat{\beta}_{0n}, \dots, \bar{\beta}_{pn}, \sigma_n^2)'$, which is sometimes preferred for computational reasons [see Anderson (1971), page 186)], is defined by the relations

$$\sum_{j=1}^{p} C_{i-j}\bar{\beta}_{jn} = -C_i, \qquad i = 1, 2, \dots, p,$$

$$\bar{\beta}_{0n} = \bar{Y}_n\left(1 - \sum_{j=1}^{p} \bar{\beta}_{jn}\right)$$

and

$$\bar{\sigma}_n^2 = n^{-1} \sum_{j=1}^{n} \left(Y_j - \bar{\beta}_{0n} - \sum_{i=1}^{p} \bar{\beta}_{in} Y_{j-i}\right)^2,$$

where

$$\bar{Y}_n = n^{-1} \sum_{j=1}^{n} Y_j,$$

$$C_h = (n-h)^{-1} \sum_{j=1}^{n-h} \left(Y_j - \bar{Y}_n\right)\left(Y_{j+h} - \bar{Y}_n\right), \qquad h = 0, 1, 2, \dots, n-1.$$

According to Theorem 3 of Section 2, Condition (A) of Theorem 4 is satisfied; we now verify the remaining Conditions (B) and (C) when $h_n \equiv 0$, $n \geq 1$; verification for the general bounded $\{h_n\}$ requires only the notational change of replacing θ by $\theta + h_n/\sqrt{n}$ throughout.

Recall that the vector $(Y_{1-p}, \dots, Y_0)$ is assumed to be a known constant vector. It will also be convenient to reparametrize so that the new parameter is $\theta = (\nu, \beta_1, \dots, \beta_p, \sigma)$, where $\nu = \beta_1/1 - \sum_{j=1}^{p} \beta_j$. This of course does not change anything. Now let

$$\sigma u_1 = Y_1 - \nu,$$

$$\sigma u_2 = (Y_2 - \nu) - \beta_1(Y_1 - \nu),$$

$$\vdots$$

$$\sigma u_{p+1} = \left(Y_{p+1} - \nu\right) - \beta_1\left(Y_p - \nu\right) - \cdots - \beta_p(Y_1 - \nu),$$

$$\sigma u_i = \left(Y_i - \nu\right) - \sum_{j=1}^{p} \beta_j\left(Y_{i-j} - \nu\right), \qquad i \geq p+2.$$

The joint p.d.f. of $u_1, u_2, \ldots, u_n$ is given by

$$g_n(u_1, \ldots, u_n) = \prod_{i=1}^{p} f\left[u_i - \frac{1}{\sigma}\sum_{j=1}^{p} \beta_j Y_{j-1} + \frac{\nu}{\sigma}\left(1 - \sum_{j=1}^{i-1} \beta_j\right)\right] \prod_{l=p+1}^{n} f(u_l).$$

Now set

$$B_n = B_n(\beta, \sigma) = \frac{1}{\sigma}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\beta_1 & 1 & 0 & \cdots & 0 \\ -\beta_2 & -\beta_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{pmatrix}_{n\times n}.$$

Since the determinant of this matrix is $1/\sigma^{n/2}$, the joint p.d.f. of $(Y_1, Y_2, \ldots, Y_n)$ is given by

$$\frac{1}{\sigma^{n/2}} g_n(B_n(\mathbf{Y}_n - \nu \mathbf{1}_n)), \tag{17}$$

where $\mathbf{Y}_n' = (Y_1, \ldots, Y_n)$ and $\mathbf{1}_n' = (1, 1, \ldots, 1)_{1\times n}$. Also, with the notation $\theta_n(h) = \theta + h/\sqrt{n}$, $h' = (u, \nu', w)$, $u \in R$, $v \in R^p$ and $w \in R$, one has for every measurable function T_n,

$$E_\theta[T_n(\mathbf{Y}_n)] = E_{\theta_n(h)}\left(T_n\left((B_n^{-1}B_n^*)\left(\mathbf{Y}_n - \frac{u}{\sqrt{n}}\mathbf{1}_n\right)\right)\right), \tag{18}$$

where $B_n^* = B_n(\beta + v/\sqrt{n}, \sigma + w/\sqrt{n})$ and $E_{\theta_n(h)}$ denotes the expectation with respect to the density

$$\left(\sigma + \frac{w}{\sqrt{n}}\right)^{-1} g_n\left(B_n^*\left(\mathbf{Y}_n - \left(\nu + \frac{u}{\sqrt{n}}\right)\mathbf{1}_n\right)\right). \tag{19}$$

Note that this density need not be equal to the density of $P_{\theta_n(h), n}$, but the differences between the ratios of (19) and (17) and the ratios $dP_{\theta_n(h), n}/dP_{\theta, n}$ converge to 0 in $P_{\theta, n}$-probability as can be seen using condition (2). Thus, according to Theorem 2, Condition (B) of Theorem 4 is verified with

$$U_{n,\theta,h}(\mathbf{Y}_n) = B_n^{-1}B_n^*(\mathbf{Y}_n - (u/\sqrt{n})\mathbf{1}_n) = \mathbf{Y}_n^*, \quad \text{say.} \tag{20}$$

In order to simplify the verification of Condition (C), we now consider in detail only the first order autoregressive process. One can easily check that the verification in the pth order case is almost the same by writing it in the matrix form, see, e.g., Anderson [(1971), Section 5.3]. We now have to deal with only three parameters (μ, β, σ) where for convenience σ now stands for the inverse of the original scale parameter. Then

$$B_n^{-1} = \frac{1}{\sigma}\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \beta & 1 & 0 & 0 & \cdots & 0 \\ \beta^2 & \beta & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \beta^{n-1} & \beta^{n-2} & \cdots & & & 1 \end{pmatrix},$$

so that letting $-v = v$,

$$B_n^{-1}B_n^* = \left(1 + \frac{w}{\sigma\sqrt{n}}\right)\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{v}{\sqrt{n}} & 1 & 0 & \cdots & 0 \\ \beta\frac{v}{\sqrt{n}} & \frac{v}{\sqrt{n}} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \beta^{n-2}\frac{v}{\sqrt{n}} & \cdots & \frac{v}{\sqrt{n}} & & 1 \end{pmatrix}.$$

Thus, the kth component of the vector $\mathbf{Y}_n^*$ defined previously in (20) is given by

$$(21)\qquad Y_k^* = \left(1 + \frac{w}{\sigma\sqrt{n}}\right)\left[Y_k + \frac{v}{\sqrt{n}}\sum_{i=1}^{k-1} Y_{k-i}\beta^{i-1} - \frac{u}{\sqrt{n}} - \frac{uv}{n}\sum_{i=1}^{k-1}\beta^{i-1}\right].$$

To proceed further, we first prove Lemma 8.

LEMMA 8. *Condition* (C) *of Theorem* 4 *is satisfied for the statistics* $\tilde{\theta}_n(\mathbf{Y}_n^*) = \tilde{\theta}_n(h)$, *where* $\mathbf{Y}_n^*$ *is as defined in* (20) *and*

$$\tilde{\theta}_n(\mathbf{Y}_n) = \left(n^{-1}\sum_{i=1}^n Y_i,\ n^{-1}\sum_{i=1}^n Y_{i-1}Y_i,\ n^{-1}\sum_{i=1}^n Y_i^2\right).$$

PROOF. First note that $\tilde{\theta}_n(h) = (\tilde{\theta}_{n1}(h), \tilde{\theta}_{n2}(h), \tilde{\theta}_{n3}(h))$, where

$$\tilde{\theta}_{n1}(h) = n^{-1}\sum_{i=1}^n Y_i^*,$$

$$\tilde{\theta}_{n2}(h) = n^{-1}\sum_{i=1}^n Y_{i-1}^*Y_i^*$$

and

$$\tilde{\theta}_{n3}(h) = n^{-1}\sum_{i-1}^n Y_i^{*2},$$

with Y_i^* as defined in (21). We first consider $\tilde{\theta}_{n2}(h)$. Using (21) we have

$$(22)\qquad \begin{aligned} \sqrt{n}\,\tilde{\theta}_{n2}(h) &= n^{-1/2}\sum_{i=1}^n Y_{i-1}Y_i + v\frac{A_{n1}}{n} + u\frac{A_{n2}}{n} + w\frac{A_{n3}}{n} \\ &\quad + v^2\frac{A_{n4}}{n^{3/2}} + uv\frac{(A_{n2} + A_{n5})}{n^{3/2}} + uv^2\frac{A_{n5}}{n^2} \\ &\quad + vw\frac{A_{n1}}{n^{3/2}} + wu\frac{A_{n2}}{n^{3/2}} + wv^2\frac{A_{n2}}{n^2} + uvw\frac{(A_{n2} + A_{n5})}{n^2} \\ &\quad + uv^2w\frac{A_{n5}}{n^{5/2}}, \end{aligned}$$

where the coefficients A_{ni}, $i = 1, \dots, 5$, are such that $n^{-1}A_{ni}$ converge in probability to some constants. Before illustrating this behavior of the A_{ni}'s, we observe that this entails

$$\sqrt{n}\left(\hat{\theta}_{n2}(h_2) - \tilde{\theta}_{n2}(h_1)\right) = a_1(u_2 - u_1) + a_2(v_2 - v_1) + a_3(w_2 - w_1) + Z_{n2}(h_2, h_1) \tag{23}$$

for some constants a_1, a_2 and a_3, where the r.v.'s Z_{n2}'s satisfy the property that the quantities

$$\sup\{|Z_{n2}(h_2, h_1)|/|h_1 - h_2| \colon |h_1| \le \alpha, |h_2| \le \alpha\} \tag{24}$$

converge to 0 in $P_{\theta, n}$-probability for every $\alpha > 0$.

To illustrate the behavior of the coefficients in (22), consider the coefficient of v,

$$\frac{A_{n1}}{n} = n^{-1} \sum_{j=1}^{n} \sum_{i=1}^{j-1} Y_{j-1}Y_{j-i}\beta^{i-1} + n^{-1} \sum_{j=1}^{n} \sum_{i=1}^{j-2} Y_jY_{j-1-i}\beta^{i-1}. \tag{25}$$

The first term on the r.h.s. can be written for each fixed $l > 0$ as

$$n^{-1} \sum_{j=1}^{l} \sum_{i=1}^{j-1} Y_{j-1}Y_{j-i}\beta^{i-1} + n^{-1} \sum_{j=l}^{n} \sum_{i=1}^{l-3} Y_{j-1}Y_{j-i}\beta^{i-1} + n^{-1} \sum_{j=l}^{n} \sum_{i=l-2}^{j-1} Y_{j-1}Y_{j-i}\beta^{i-1} = I_1 + I_2 + I_3, \quad \text{say.}$$

Then clearly $I_1 \to_p 0$ for each l. Also,

$$E|I_3| \le n^{-1} \sum_{j=l}^{n} \sum_{i=l-2}^{j-1} \sqrt{E\left(Y_{j-1}^2\right)E\left(Y_{j-i}^2\right)}\,\beta^{i-1} \le \frac{K}{n} \sum_{j=l}^{n} \sum_{i=l-2}^{j-1} \beta^{i-1} \le K'|\beta|^l \to 0 \quad \text{as } l \to \infty,$$

since $|\beta| < 1$ where K and K' are some constants. Thus I_3 converges to 0 in probability, first by letting $n \to \infty$ and then $l \to \infty$. Further, convergence of I_2 for each fixed l follows from standard results and arguments [see, e.g., Anderson (1971) and Lemma 11 of Section 5]. Hence the convergence of the first sum on the r.h.s. of (25) follows easily; convergence of the second sum also follows using similar arguments. Similarly, the behavior of other coefficients can be checked.

The same arguments apply also to other statistics $\tilde{\theta}_{n1}(h)$ and $\tilde{\theta}_{n2}(h)$ so that

$$\sqrt{n}\left(\tilde{\theta}_{n1}(h_2) - \tilde{\theta}_{n1}(h_1)\right) = b_1(u_2 - u_1) + b_2(v_2 - v_1) + b_3(w_2 - w_1) + Z_{n1}(h_2, h_1) \tag{26}$$

and

$$\sqrt{n}\left(\tilde{\theta}_{n3}(h_2) - \tilde{\theta}_{n3}(h_1)\right) = c_1(u_2 - u_1) + c_2(v_2 - v_1) + c_3(w_2 - w_1) + Z_{n3}(h_2, h_1), \tag{27}$$

where b_i's and c_i's, $i = 1, 2, 3$, are constants and Z_{n1} and Z_{n3} satisfy property (24). Hence the proof of Lemma 8 is complete. □

Now, using the relationship between the components of the statistic $\tilde{\theta}_n$ and those of L.S. estimate $\tilde{\theta}_n$ and the estimator $\bar{\theta}_n$ defined previously, one can easily check, using (23), (26) and (27), that Condition (C) of Theorem 4 is satisfied for the estimators $\hat{\theta}_n$ and $\bar{\theta}_n$. We thus have:

THEOREM 9. (i) *Assume that Assumption* (B1) *and condition* (2) *of Section* 2 *are satisfied. Then the* $(p+1)$ *vectors* $\sqrt{n}(\hat{\beta}_n - \beta)$ *and* $\sqrt{n}(\bar{\beta}_n - \beta)$ *converge strongly to Gaussian distributions* ($\bar{\beta}_n = (\bar{\beta}_{0n}, \dots, \bar{\beta}_{pn})'$).

(ii) *In addition to the restrictions of the preceding statement, assume that* $E(\varepsilon_1^4) < \infty$ *and that condition* (3) *of Section* 2 *is satisfied. Then the* $(p+2)$-*vectors* $\sqrt{n}(\hat{\theta}_n - \theta)$ *and* $\sqrt{n}(\hat{\theta}_n - \theta)$ *converge strongly to Gaussian distributions.*

Note that condition (ii) of Theorem 1 is satisfied because of (18) and Condition (C) and the convergence in law of $\sqrt{n}(\hat{\theta}_n - \theta)$ and $\sqrt{n}(\bar{\theta}_n - \theta)$ under $P_{\theta, n}$. It may also be noted that it is possible to get the preceding strong convergence to be uniform in $\theta \in K \subseteq \Theta$ also for very compact K; this can be obtained by replacing θ in the preceding verification by arbitrary $\theta_n \in K$ throughout.

5. Strong convergence of general *M*-estimators. In this section, Theorem 4 will be applied in a situation where the estimators cannot in general be written as explicit functions of the observations as in the previous section. Consider the observations X_n, $n \geq 1$, such that

$$X_n = (1 - V_n)Y_n + V_n Z_n, \qquad n \geq 1,$$

where Y_n is the pth order autoregressive process of Section 1, $\{Z_n\}$ and $\{V_n\}$ are sequences of i.i.d. r.v.'s and the three sequences $\{\varepsilon_n\}$, $\{Z_n\}$ and $\{V_n\}$ are mutually independent. Also it will be assumed throughout that $E(V_1^2) < \infty$ and $E(Z_1^2) < \infty$. We refer to, e.g., Bustos (1982) and Martin and Yohai (1985) for the details of the significance of this contaminated autoregressive process in the context of robust estimation. It is possible to relax [cf. Bustos (1982)] the independence of the r.v.'s of the sequence $\{Z_n\}$ and those of $\{V_n\}$, but we proceed with the preceding independence restrictions. It is further assumed that the distributions of the sequences $\{V_n\}$ and $\{Z_n\}$ do not involve any unknown parameters so that the only unknown parameters are β and σ^2 of the process $\{Y_n\}$. (An apparent exception to this is when $\{Z_n\}$ is an autoregressive process with the same β and with error distributions possibly different from those of $\{Y_n\}$, but then $\{X_n\}$ itself becomes an autoregressive process with the same β and σ.) This assumption in particular entails that the likelihood ratios of the observations $\{(Y_i, V_i, Z_i), 1 \leq i \leq n\}$ will be the same as those of $\{Y_i, 1 \leq i \leq n\}$. Therefore by Theorem 3, since $\{Y_i, 1 \leq i \leq n\}$ is independent of $\{(Z_i, V_i), 1 \leq i \leq n\}$, the LAN condition holds under conditions (2) and (3) of Section 2, verifying Condition (A) of

Theorem 4. This independence also entails that, according to (18) with $\theta = (\beta, \sigma)$,

$$E_\theta\left[f\left(\hat{\theta}_n(\mathbf{Y}_n, \mathbf{Z}_n, \mathbf{V}_n)\right)\right] = E_{\theta_n(h)}\left[f\left(\hat{\theta}_n(\mathbf{Y}_n^*, \mathbf{Z}_n, \mathbf{V}_n)\right)\right] \tag{28}$$

for all statistics $\hat{\theta}_n$, for all measurable f and for all $h \in R^{p+2}$, where $\mathbf{Y}_n^*$ is as defined in (20). $\mathbf{Z}_n = (Z_1, \dots, Z_n)$, $\mathbf{V}_n = (V_1, \dots, V_n)$ and the expectation $E_{\theta_n(h)}$ is with respect to the probability measure whose conditional density given $(\mathbf{V}_n, \mathbf{Z}_n)$ is given by (19). Thus, as in Section 4, Condition (B) of Theorem 4 is also satisfied for all statistics $\hat{\theta}_n$.

Now consider the general M-estimators [see, e.g., Bustos (1982) and Martin and Yohai (1985)], that is, consider $\hat{\theta}_n = (\hat{\beta}_n, \hat{\sigma}_n)$ such that

$$\sum_{i=1}^{n} R_{i-1}\Phi\left(R_{i-1}, \left(X_i - R'_{i-1}\hat{\beta}_n\right)/\hat{\sigma}_n\right) = 0 \tag{29}$$

and

$$\sum_{i=1}^{n} \chi\left(\left(\left(X_i - R'_{i-1}\hat{\beta}_n\right)/\hat{\sigma}_n\right)^2\right) = 0, \tag{30}$$

where $R_{i-1} = (1, X_{i-1}, X_{i-2}, \dots, X_{i-p})$ and Φ: $R^{p+1} \times R \to R$ and χ: $[0, \infty) \to R$ are conveniently chosen functions. Under certain regularity conditions on Φ and χ, Bustos (1982) has shown that for some $(p+2)$-vector $g(\theta)$, $\sqrt{n}\,(\hat{\theta}_n - g(\theta))$ converge in law to a nondegenerate Gaussian distribution with mean vector 0; it is possible to strengthen his result so that condition (ii) of Theorem 1 is satisfied for $T_n = \hat{\theta}_n$. Recently, several other variants of M-estimators have also been introduced in the context of robust estimation [see, e.g., Martin and Yohai (1985)]; these variants will be considered in a subsequent part in the context of ARMA models.

We now state the assumptions. Let $\mathbf{X}_{n,\theta}^h$ be the n-vector whose ith component is defined by

$$X_i^h = V_i Z_i + (1 - V_i) Y_i^*, \tag{31}$$

where Y_i^* is the ith component of $\mathbf{Y}_n^*$ defined in (20). Note that Y_i^*, and hence X_i^h, depends on the parameter θ. Set $\mathbf{X}_n = (X_1, \dots, X_n)'$.

(C1) For all $\alpha > 0$, the quantities

$$\sup_{|u| \le \alpha} |\hat{\theta}_n\left(\mathbf{X}_{n,\theta_n}^u\right) - \hat{\theta}_n(\mathbf{X}_n)|$$

converge to 0 in $P_{\theta,n}$-probability for all sequences $\{\theta_n\}$ with $\theta_n = \theta + h_n/\sqrt{n}$, $\{h_n\}$ bounded.

Remark. Note that this condition is a necessary one for Condition (C) and seems to be a very mild one. In fact, if the difference $\hat{\theta}_n(\mathbf{X}_n) - \theta \to 0$ in $P_{\theta,n}$-probability then, by contiguity and by (31), $\hat{\theta}_n(\mathbf{X}_{n,\theta_n}^u) - \theta \to 0$ in $P_{\theta,n}$-probability for every u if Condition (A) of Theorem 4 is satisfied. Thus the differences $\hat{\theta}_n(\mathbf{X}_{n,\theta_n}^u) - \hat{\theta}_n(\mathbf{X}_n)$ converge to 0 in $P_{\theta,n}$-probability, but (C1) is stronger than this. It is possible to derive (C1) from more basic assumptions on the preceding functions Φ and χ.

(C2) For each $y \in R^{p+1}$, the function

$$(\beta, \sigma) \to \Big((y'\Phi(y, (x - y'\beta)/\sigma)), \chi\big(((x - y'\beta)/\sigma)^2\big)\Big),$$

defined from $R^{p+2} \to R^{p+2}$, is continuously differentiable with respect to all β and σ. Denote the derivative by [the $(p+2) \times (p+2)$ matrix] $W_1(y, (x - y'\beta), \sigma)$. Further, the function $(y, x) \to (y\Phi(y, x), \chi(x^2))$ defined from $R^{p+2} \to R^{p+2}$, is continuously differentiable with respect to all y and x and denote the derivative by [the $(p+2) \times (p+2)$ matrix] $W_2(y,x)$.

In order to state the next condition, let $\{Y_n'\}$ be the process satisfying the equations,

$$Y_n' = \beta_0 + \beta_1 Y_{n-1}' + \cdots + \beta_p Y_{n-p}' + \varepsilon_n,$$

where n now ranges over all integers, i.e., $n = 0, \pm 1, \pm 2, \dots,$ and the ε_n's are i.i.d. for all these values of n coinciding with the ε_n's of Section 1 for positive integers n; in other words, Y_n' is the stationary version of Y_n. Let $X_i' = (1 - V_i)Y_i' + V_i Z_i$ and $\xi = (Z_j, V_j, Y_j'\ j = -p+1, \dots, 1)'$.

(C3) Denote the matrix $W_1(y, (z - \beta' y), \sigma)$ by $g_1(\xi, \beta, \sigma)$ and $W_2(y, x)$ by $g_2(\xi, \beta, \sigma)$ when $y = (X_{-p+1}', \dots, X_0')'$, $z = X_1'$ and $x = (X_1' - \beta_1 X_0' - \cdots - \beta_p X_{-p+1}')/\sigma$, where W_1 and W_2 are as defined in (C2). Then

$$E\left[\sup_{|\varepsilon| \le \delta} |g_i(\xi + \varepsilon, \beta + \varepsilon, \sigma + \varepsilon) - g_i(\xi, \beta, \sigma)|\right] \to 0, \qquad i = 1, 2, \text{ as } \delta \to 0.$$

(C4) The matrix $E[g_1(\xi, \beta, \sigma)]$ is nonsingular.

(C5) For some $\delta > 0$,

$$E\left\{\left[\sup_{|t| \le \delta} |g_2(\xi + t, \beta + t, \sigma + t)|\right]^2\right\} < \infty.$$

We now state the main result of this section.

THEOREM 10. *Let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\sigma}_n)$ be the M-estimators defined previously (based on the observations $\{X_i,\ i \le i \le n\}$). Assume that $E(V_1^2) < \infty$, $E(Z_1^2) < \infty$, Conditions* (C1)–(C5) *and conditions* (2) *and* (3) *of Section* (2) *are satisfied. Further assume that condition* (*ii*) *of Theorem* 1 *is satisfied for the sequence* $\{\hat{\theta}_n\}$. *Then the $(p+1)$-vector $\hat{\theta}_n$ converges strongly to a Gaussian distribution.*

In order to prove this theorem, it remains only to verify Condition (C) of Theorem 4. To this end, it is convenient to prove Lemma 11 first.

Lemma 11. *Assume that Condition* (C3) *is satisfied. Let the* $(p+1)$*-vector* $\xi_i = (Z_j, V_j, Y_j,\ j = i-p, \dots, i)'$. *Then with* ξ, g_1 *and* g_2 *as in* (C3), *for every* $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\left[\sup_{|t| \le \delta} \left| n^{-1} \sum_{i=1}^{n} \{ g_k(\xi_i + t, \beta + t, \sigma + t) - E[g_k(\xi, \beta, \sigma)] \} \right| > \varepsilon \right] = 0, \qquad k = 1, 2.$$

Proof of Theorem 10. First note that the lemma is true when the $(p+1)$-vectors ξ_i's are replaced by $(Z_j, V_j, Y_j',\ j = i-p, \dots, i)'$, as can be easily seen using ergodicity. Also, it can be shown that for every increasing sequence $k_n \uparrow \infty$, the quantities $\sup_{k_n \le i \le n} |Y_i - Y_i'|$ converge to 0 in probability. Hence the lemma follows easily. □

In order to simplify further computations, we now restrict our attention to the case where $\{Y_n\}$ is a first order stationary AR model with known β_0, say $\beta_0 = 0$, so that the two unknown parameters are β and σ, where, as in Section 4 for convenience, σ stands for the inverse of the original scale parameter. Then, with $h' = (v, w)$ and using (21) and (31),

$$X_i^h = X_i + (1 - V_i)\left[\frac{v}{\sqrt{n}} g_i(\mathbf{Y}_{i-1}) + \frac{w}{\sigma\sqrt{n}} Y_i + \frac{vw}{\sigma n} g_i(\mathbf{Y}_{i-1}) \right], \tag{32}$$

where

$$g_i(\mathbf{Y}_{i-1}) = \sum_{j=1}^{i-1} Y_{i-j} \beta^{j-1}.$$

Also, define

$$\hat{\theta}_n(h) = \hat{\theta}_n(\mathbf{X}_n^h),$$

where $\mathbf{X}_n^h = (X_1^h, \dots, X_n^h)$ with X_i^h as defined in (31). Then, in view of (29) and (30), we have the following vector equation using Taylor's expansion around $(\hat{\beta}_n(u), \hat{\sigma}_n(u))$: For every $u' = (u_1, u_2)$ and $z' = (z_1, z_2)$ in R^2,

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \Big(X_{i-1}^z \Phi\big(X_{i-1}^z, (X_i^z - \hat{\beta}_n(z) X_{i-1}^z)/\hat{\sigma}_n(z) \big), \\ &\qquad \chi^*\big((X_i^z - \hat{\beta}_n(z) X_{i-1}^z)/\hat{\sigma}_n(z) \big) \Big) \\ &= \sum_{i=1}^{n} \Big(X_{i-1}^z \Phi\big(X_{i-1}^z, (X_i^z - \hat{\beta}_n(u) X_{i-1}^z)/\hat{\sigma}_n(u) \big), \\ &\qquad \chi^*\big((X_i^z - \hat{\beta}_n(u) X_{i-1}^z)/\hat{\sigma}_n(u) \big) \Big) \\ &\quad + \big(\hat{\theta}_n(z) - \hat{\theta}_n(u) \big) \sum_{i=1}^{n} W_1\big(X_{i-1}^z, X_i^z - \beta_n^* X_{i-1}^z, \sigma_n^* \big) \\ &= I_1 + I_2, \quad \text{say}, \end{aligned} \tag{33}$$

where $|(\beta_n^*, \sigma_n^*) - \hat{\theta}_n(u)| \leq |\hat{\theta}_n(z) - \hat{\theta}_n(u)|$ and $\chi^*(x) = \chi(x^2)$. Further, using (29) and (30) for the observations X_i^u and using Taylor's expansion around X_{i-1}^u and $(X_i^u - \hat{\beta}_n(u)X_{i-1}^u)/\hat{\sigma}_n(u)$,

$$(34)\quad \begin{aligned} I_1 = \sum_{i=1}^{n} & (U_{1i}(u,z), U_{2i}(u,z)) \\ & \times W_2\big(X_{i-1}^u + U_{1i}^*(u,z), \big(X_i^u - \hat{\beta}_n(u)X_{i-1}^u\big)/\hat{\sigma}_n(u) + U_{2i}^*(u,z)\big), \end{aligned}$$

where

$$|(U_{1i}^*, U_{2i}^*)| \leq |(U_{1i}, U_{2i})|$$

and where

$$(35)\quad \begin{aligned} U_{1i} = U_{1i}(u,z) &= X_{i-1}^z - X_{i-1}^u \\ &= n^{-1/2}(z_1 - u_1)(1 - V_i)g_i(\mathbf{Y}_{i-1}) \\ &\quad + (\sigma\sqrt{n})^{-1/2}(z_2 - u_2)(1 - V_i)Y_i \\ &\quad + (\sigma n)^{-1}(z_1 z_2 - u_1 u_2)(1 - V_i)g(\mathbf{Y}_{i-1}) \end{aligned}$$

and

$$(36)\quad U_{2i} = U_{2i}(u,z) = \big((X_i^z - X_i^u) - \beta_n(u)(X_{i-1}^z - X_{i-1}^u)\big)/\hat{\sigma}_n(u).$$

Now note further that, as can be easily seen,

$$\sup_{|h| \leq \alpha} \sup_{1 \leq i \leq n} |X_i^h - X_i| \to_p 0.$$

Using this fact together with Lemma 11 and Conditions (C1)–(C3), it can be shown that with $G = E[g_1(\xi, \beta, \sigma)]$,

$$(37)\quad \sup_{|u| \leq \alpha,\ |z| \leq \alpha} \left| \frac{1}{n} \sum_{i=1}^{n} W_i(u,z) - G \right| \to_p 0$$

for all $\alpha > 0$, where the sum ΣW_i is the coefficient of $(\hat{\theta}_n(z) - \hat{\theta}_n(u))$ in (33). Similarly, in view of relations (34)–(36), one can check (see the illustration in the proof of Lemma 8) using Lemma 11 and Conditions (C1)–(C3) and (C5), that there is a matrix H such that the quantities

$$(38)\quad \sup_{|h| \leq \alpha,\ |z| \leq \alpha} \left| \frac{I(u,z)}{\sqrt{n}} - H(z - u) \right| \Big/ |z - u| \to_p 0$$

for all $\alpha > 0$, where I is the sum in (34). It follows from (33) that

$$(39)\quad 0 = (H + Z_{n1}(u,z))(z - u) + [G + Z_{n2}(u,z)]\sqrt{n}\big(\hat{\theta}_n(z) - \hat{\theta}_n(u)\big),$$

where

$$\sup_{|u| \leq 0,\ |z| \leq \alpha} |Z_{ni}(u,z)| \to_p 0, \qquad i = 1, 2,$$

for all $\alpha > 0$. Now since G is nonsingular [Condition (C4)], it follows easily from (39) that Condition (C) of Theorem 4 is satisfied with $A = F$, where F is the nonsingular matrix involved in condition (ii) of Theorem 1 which we have assumed to hold for the sequence $\{\hat{\theta}_n\}$. This completes the proof of Theorem 10. □

6. Strong convergence of L.S. estimators in the explosive case: Introduction. Recall that an AR model is called explosive if its characteristic polynomial has no roots on the unit circle, but has at least one root outside the unit circle. The purpose of this part is to prove the strong convergence of L.S. estimators in this case with suitable random or nonrandom normalization. Since in the explosive case in at least part of the process only a few variables play a dominant role, central limit considerations cannot be invoked entirely and since our method depends on the smoothness of the likelihoods of the sample, we assume that errors are Gaussian in order to achieve the required smoothness. Actually, the required smoothness, and hence the strong convergence, can be achieved under only condition (2) of Part I using the methods of Example 1 of Levit (1974) as was kindly pointed out by the referee, but this will not be done here in order not to overload the paper. Note, however, that we do not assume the Gaussian restriction when the convergence in law of the L.S. estimators is derived. This restriction will entail that the suitably normalized log-likelihood of the sample will be quadratic in the parameter, but the variables of the quadratic term will stay random even in the limit, thus violating the LAN condition of Section 2. However, it will turn out that this case corresponds to the LAMN likelihoods [see, e.g., Davies (1985)] and that the results of Sections 2 and 3 will generalize to this case in a simple and natural way.

Section 7 treats the case in which all roots are outside the unit circle, to be called *purely explosive* case. The limit distribution of L.S. estimators in the first order case with Gaussian errors was obtained by White (1958). This result of White was extended to the purely explosive case of any order by Anderson (1959).

The more difficult part of the explosive case in which not all the roots are outside the unit circle, to be called *partially explosive* case, was treated by Venkataraman (1967, 1968) [see also Stigum (1974)]. Among several other results, Venkataraman (1968) obtained that the L.S. estimators converge in law to a Gaussian distribution when normalized by $\sqrt{n}$. The limiting Gaussian distribution he obtained was degenerate since the normalizing constants $\sqrt{n}$ completely eliminate the influence of the explosive part of the process and retain only the influence of the stationary part. In Section 8, we first obtain the convergence in law of L.S. estimators with different normalizing quantities so that the influence of both explosive and stationary parts are retained and that the limit is nondegenerate and then obtain the strong convergence.

7. Purely explosive case. Consider the AR model of Section 1 with $\beta_0 = 0$, that is, $\{Y_n\}$, $n \geq 1 - p$, is a process such that

$$Y_n = \beta_1 Y_{n-1} + \cdots + \beta_p Y_{n-p} + \varepsilon_n, \tag{40}$$

$n = 1, 2, \ldots,$ where ε_i, $i \geq 1$, are i.i.d., independent of $(Y_0, \ldots, Y_{1-p})$ with mean 0 and variance 1. Without further mention, it will be assumed throughout this section that the roots of the characteristic polynomial of the preceding AR model are all greater than 1 in absolute value. Whenever we assume that the ε_i's are Gaussian, it will be stated explicitly. Since $\beta_0 = 0$, we now let $\tilde{Y}_j = (Y_j, \dots, Y_{j-p+1})'$. Then if $\hat{\beta}_n$ is the L.S. estimator,

$$\hat{\beta}_n - \beta = A_n^{-1} C_n,$$

where

$$A_n = \sum_{j=1}^{n} \tilde{Y}_{j-1} \tilde{Y}'_{j-1} \quad \text{and} \quad C_n = \sum_{j=1}^{n} \tilde{Y}_{j-1} \varepsilon_j.$$

We first recall Theorem 12 from Anderson (1959). To state it, let

$$\delta = \begin{pmatrix} \beta_1 & \beta_2 & & \cdots & \beta_p \\ 1 & 0 & & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & & & \cdots \\ 0 & \cdots & & 1 & 0 \end{pmatrix}_{p \times p}$$

and $\tilde{\varepsilon}_i' = (\varepsilon_i, 0, \dots, 0)_{1 \times p}$, $i \geq 1$. Then relation (40) is equivalent to

$$\tilde{Y}_n = \delta \tilde{Y}_{n-1} + \tilde{\varepsilon}_n, \qquad n \geq 1.$$

Further, let

$$Z_n = \delta^{-(n-2)} \tilde{Y}_{n-1} = \tilde{\varepsilon}_1 + \delta^{-1} \tilde{\varepsilon}_2 + \cdots + \delta^{-(n-2)} \tilde{\varepsilon}_{n-1} + \delta \tilde{Y}_0. \tag{41}$$

THEOREM 12. *Let* $\delta_n = \delta^{-(n-2)}$. *Then*:

(i) *The differences*

$$\delta_n A_n \delta_n' - \sum_{i=0}^{n} \delta^{-i} Z_n Z_n' \delta'^{-i}$$

and

$$\delta_n C_n - \sum_{i=0}^{n-1} \delta^{-i} Z_n \varepsilon_{n-i}$$

converge to 0 *in* $P_{\beta, n}$*-probability.*

(ii) *The sequence* $\{Z_n\}$ *converges in* $P_{\beta, n}$*-probability to a random vector* Z *such that* ZZ' *is an a.s. positive definite matrix.*

(iii) *There is a sequence* $\{\varepsilon_i^*\}$, $i \geq 1$, *identically distributed with and independent of the sequence* $\{\varepsilon_i\}$, $i \geq 1$, *such that the vectors* $\sum_{i=0}^{n-1} \delta^{-i} Z_n \varepsilon_{n-i}$ *converge in law to the vector* $\sum_{i=1}^{\infty} \delta^{-(i-1)} Z \varepsilon_i^*$, *where* Z *is as in statement* (ii) *and hence is independent of* $\{\varepsilon_i^*\}$.

(iv) *It follows from the preceding three statements that when the* ε_i*'s are Gaussian,* $(\delta_n A_n \delta_n', \delta_n C_n)$ *converges in law under* $P_{\beta, n}$ *to* $(A, A^{1/2} W)$, *where* $A = \sum_{i=0}^{\infty} \delta^{-i} Z Z' \delta'^{-i}$ *and* W *is a copy of the standard p-variate Gaussian distribution independent of* A, *with mean* 0 *and unit covariance matrix.*

The fact that ZZ' is a.s. positive definite in statement (ii) was not established in Anderson (1959), but can be found, e.g., in Lai and Wei (1983).

Now note that when the ε_i's are Gaussian,

$$\log \frac{dP_{\beta+\delta_n h, n}}{dP_{\beta, n}} = h'\delta_n C_n - \frac{1}{2} h'\delta_n A_n \delta_n' h$$

for all $n \geq 1$ and $h \in R^p$. Hence the preceding result implies that the likelihood ratios $dP_{\beta+\delta_n h, n}/dP_{\beta, n}$ converge in law under $P_{\beta, n}$ to the likelihood ratios of the family $\{Q_h\colon h \in R^p\}$ such that

$$\frac{dQ_h}{dQ_0} = \exp\left(h'A^{1/2}W - \frac{1}{2} h'Ah \right),$$

where A and W are as in statement (iv) of Theorem 12. Now given A, the family $\{G_h\colon h \in R^p\}$ is a Gaussian family. For this reason the family $\{P_{\theta, n}\colon \theta \in \Theta\}$, $n \geq 1$, is said to be LAMN at θ_0 if Definition 1 of Section 2 is satisfied with $B(\theta_0)$ replaced by a sequence $\{B_n(\theta_0)\}$ of a.s. positive definite matrices such that, for some a.s. positive definite matrix $B(\theta_0)$, the sequence $\{(W_n(\theta_0), B_n(\theta_0))\}$ converges in law under $P_{\theta, n}$ to $(B^{1/2}(\theta_0)W, B(\theta_0))$, where W is a copy of the standard k-variate Gaussian distribution independent of $B(\theta_0)$. We refer to, e.g., Davies (1985) for a detailed study of LAMN families. For our purpose we mention that Theorems 1 and 2 of Section 2 and Theorem 4 of Section 3 extend to the LAMN case, since the arguments that involved the normality of the limiting families $\{Q_h\}$ hold true also when $\{Q_h\}$ is mixed Gaussian as defined previously. The required changes in the statements are that the nonsingular matrix F_{θ_0} involved in the SASE condition and condition (ii) of Theorem 1 have to be replaced by a sequence $\{F_{\theta_0, n}\}$ of a.s. nonsingular matrices such that for some a.s. nonsingular matrix F_{θ_0}, the sequence $(F_{\theta_0, n}, W_n(\theta_0), B_n(\theta_0))$ converge in law under $P_{\theta, n}$ to $(F_{\theta_0}, B^{1/2}(\theta_0)W, B(\theta_0))$, where W is, as before, independent of $(F_{\theta_0}, B(\theta_0))$. Furthermore, the nonsingular matrix A of Condition (C) of Theorem 4 has to be replaced by a sequence of a.s. nonsingular matrices, but under condition (ii) with the preceding modification of Theorem 1, this sequence can be taken to be the preceding sequence $\{F_{\theta_0, n}\}$. Henceforth, whenever Theorems 1 and 4 are referred to, it is to be understood that they are referred to with the preceding modifications.

It follows from the foregoing remarks that it remains only to check Condition (C) of Theorem 4, since then condition (ii) of Theorem 1 also follows. This will be done only in the first order case, since the general case can be checked in a similar way. We first consider the case in which β, $|\beta| < 1$, is the true parameter and then briefly indicate the proof for the general case where the true parameter is $\beta + \delta_n h_n$, $\{h_n\}$ bounded, $|\beta| > 1$ and $\delta_n = \beta^{-(n-2)}$. As in Section 4, the transformed variables satisfying Condition (B) of Theorem 4 are given by

$$Y_j^* = Y_j + \delta_n h\left(Y_1\beta^{i-2} + Y_2\beta^{i-3} + \cdots + Y_{i-1}\right). \tag{42}$$

Our aim is to prove the strong convergence of $(\hat{\beta}_n - \beta)$ under the normalizing quantities δ_n and $(\sum_{j=1}^n Y_{j-1}^2)^{1/2}$. We thus have to check that the variables,

$$\delta_n^{-1}\left(\sum_{j=1}^n Y_{j-1}^{*2}\right)^{-1}\sum_{j=1}^n Y_{j-1}^*(Y_j^* - \beta Y_{j-1}^*)$$

and

$$\left(\sum_{j=1}^n Y_{j-1}^{*2}\right)^{-1/2}\sum_{j=1}^n Y_{j-1}^*(Y_j^* - \beta Y_{j-1}^*),$$

considered as functions of h, satisfy Condition (C) of Theorem 4. This is a consequence of Lemma 13.

LEMMA 13.

$$\delta_n^2 \sum_{j=1}^n Y_{j-1}^{*2} = \delta_n^2 \sum_{j=1}^n Y_{j-1}^2 + hU_n + h^2 V_n$$

and

$$\delta_n \sum_{j=1}^n Y_{j-1}^*(Y_j^* - \beta Y_{j-1}^*) = \delta_n \sum_{j=1}^n Y_{j-1}(Y_j - \beta Y_{j-1}) + hU_n^* + h^2 V_n^*,$$

where the variables U_n, V_n and V_n^* converge to 0 in $P_{\beta,n}$-probability and the difference between the variables U_n^* and $(1-\beta^{-2})^{-1}Z_n^2$ converge to 0 in probability, where Z_n is as defined in (41).

PROOF. Using Theorem 12, we shall proceed in an indirect way which seems to be easier than the direct way. First note that Theorem 12 entails that the differences

$$\delta_n^2 \sum_{j=1}^n Y_{j-1}^2 - (1-\beta^{-2})^{-1}\left[\sum_{j=1}^{n-1}(Y_j - \beta Y_{j-1})\beta^{-i+1} + \beta Y_0\right]^2$$

converge to 0 in $P_{\beta,n}$-probability. Hence, contiguity and (42) imply that the differences

$$\delta_n^2 \sum_{j=1}^n Y_{j-1}^{*2} - (1-\beta^{-2})^{-1}\left[\sum_{j=1}^{n-1}(Y_j^* - \beta Y_{j-1}^*)\beta^{-i+1} + \beta Y_0\right]^2 \tag{43}$$

converge to 0 in $P_{\beta,n}$-probability. Both the left- and right-hand sides of these differences are quadratic in h, so that the differences between the corresponding linear and quadratic terms also converge to 0 in probability. In view of (42),

$$Y_j^* - \beta Y_{j-1}^* = Y_j - \beta Y_{j-1} + \delta_n h Y_{j-1}, \tag{44}$$

so that the coefficient of h of the sum inside the bracket of (43) is given by

$$\delta_n \sum_{i=1}^{n-1} Y_{i-1}\beta^{-i+1} + (\beta Y_0)\delta_n.$$

Since $\delta_n = \beta^{-(n-2)}$, it can be easily checked that this converges to 0 in $P_{\beta, n}$-probability. This proves the first part of the lemma.

Using again the arguments used to prove (43), it follows from another part of the first statement of Theorem 12 that the differences

$$\delta_n \sum_{j=1}^{n} Y_{j-1}^*(Y_j^* - \beta Y_{j-1}^*) - \left[\sum_{j=1}^{n-1} (Y_j^* - \beta Y_{j-1}^*)\beta^{-i+1} + \beta Y_0\right] \times \left[\sum_{j=1}^{n} (Y_j^* - \beta Y_{j-1}^*)\beta^{-n+j}\right] \tag{45}$$

converge to 0 in probability. Using (44),

$$\sum_{j=1}^{n} (Y_j^* - \beta Y_{j-1}^*)\beta^{-n+j} = \sum_{j=1}^{n} (Y_j - \beta Y_{j-1})\beta^{-n+j} + h\delta_n \sum_{j=1}^{n} Y_{j-1}\beta^{-n+j}.$$

Since $\delta_n = \beta^{-(n-2)}$,

$$\delta_n \sum_{j=1}^{n} Y_{j-1}\beta^{-n+j} = \sum_{j=1}^{n} \beta^{-2(n-j)} Z_j = \sum_{j=0}^{n-1} \beta^{-2j} Z_{n-j}.$$

Hence, it follows easily that [see Anderson (1959)] the differences $\delta_n \sum_{j=1}^{n} Y_{j-1}\beta^{-n+j} - (1-\beta^{-2})^{-1} Z_n$ converge to 0 in probability. Since both left- and right-hand sides of (45) are quadratic in h and in view of the first part of this proof, the second part of the lemma now follows, completing the proof of the lemma. □

As to the verification when the parameter is of the form $\beta + \delta_n h_n$, $|\beta| > 1$, it can be easily checked that Theorem 12 and Lemma 13 remain valid when β is replaced by $\beta + \beta^{-(n-2)} h_n$ throughout [including that $\delta_n = \beta^{-(n-2)}$ is replaced by $\delta_n' = (\beta + \beta^{-(n-2)} h_n)^{-(n-2)}$]. Hence the verification follows easily using the fact that $\delta_n^{-1}\delta_n' \to 1$.

We thus have the following.

THEOREM 14. *Let $\hat{\beta}_n$ be the L.S. estimator of β of the purely explosive AR model* (40). *Assume that the error variables are Gaussian with mean* 0 *and variance* 1. *Let $\delta_n = \delta^{-(n-2)}$ and A_n be matrices as defined in Theorem* 12. *Then the random quantities $\delta_n^{-1}(\hat{\beta}_n - \beta)$ and $A_n^{1/2}(\hat{\beta}_n - \beta)$ converge strongly to the distributions of $A^{-1/2}W$ and W, respectively, where A and W are as in statement* (iv) *of Theorem* 12.

8. Partially explosive case. Let B be the backshift operator, i.e., $BY_n = Y_{n-1}$. Then the equation

$$Y_n = \beta_1 Y_{n-1} + \beta_2 Y_{n-2} + \cdots + \beta_p Y_{n-p} + \varepsilon_n, \qquad n \ge 1,$$

can be rewritten as

$$\left(1 - \beta_1 B - \beta_2 B^2 - \cdots - \beta_p B^p\right) Y_n = \varepsilon_n, \qquad n \ge 1.$$

Let $\rho_1, \rho_2, \dots, \rho_p$ be the p roots of the characteristic polynomial . Assume that for some $1 \le k < p$, $|\rho_1| \le |\rho_2| \le \cdots \le |\rho_k| < 1 < |\rho_{k+1}| \le \cdots \le |\rho_p|$. We may write

$$\begin{aligned}(1 - \beta_1 B - \cdots - \beta_p B^p) Y_n = {} & \left[(1 - \rho_1 B) \cdots (1 - \rho_k B)\right] \\ & \times \left[(1 - \rho_{k+1} B) \cdots (1 - \rho_p B)\right] Y_n.\end{aligned} \tag{46}$$

Let $\alpha_1, \dots, \alpha_k$ be such that

$$(1 - \rho_1 B) \cdots (1 - \rho_k B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_k B^k$$

and $\gamma_1, \dots, \gamma_{p-k}$ be such that

$$(1 - \rho_{k+1} B) \cdots (1 - \rho_p B) = 1 - \gamma_1 B - \cdots - \gamma_{p-k} B^{p-k}.$$

It then follows from (46) that

$$V_n = ((1 - \rho_1 B) \cdots (1 - \rho_k B)) Y_n$$

is a purely explosive AR model of order $p - k$ with parameters $\gamma_1, \dots, \gamma_{p-k}$ and with characteristics roots $\rho_{k+1}, \dots, \rho_p$. In other words, we have

$$\left(1 - \gamma_1 B - \cdots - \gamma_{p-k} B^{p-k}\right) V_n = \varepsilon_n, \qquad n \ge 1.$$

Similarly, it follows that

$$U_n = \left((1 - \rho_{k+1} B) \cdots (1 - \rho_p B)\right) Y_n$$

satisfies the relation

$$\left(1 - \alpha_1 B - \cdots - \alpha_k B^k\right) U_n = \varepsilon_n, \qquad n \ge 1,$$

so that $\{U_n\}$ is a stationary AR model with characteristic roots $\rho_1, \dots, \rho_k$. As before let

$$\tilde{Y}_n = \left(Y_n, Y_{n-1}, \dots, Y_{n-p+1}\right)',$$

$$\tilde{U}_n = (U_n, U_{n-1}, \dots, U_{n-k+1})'$$

and

$$\tilde{V}_n = \left(V_n, V_{n-1}, \dots, V_{n-p+k+1}\right)',$$

Then there is a matrix M such that

$$M \begin{pmatrix} \tilde{U}_n \\ \tilde{V}_n \end{pmatrix} = \tilde{Y}_n. \tag{47}$$

Let $\hat{\beta}_n$ be the L.S. estimators of $\beta = (\beta_1, \dots, \beta_p)'$. Then, using (47),

$$(48)\qquad \begin{aligned}(\hat{\beta}_n - \beta) &= \left(\sum_{j=1}^{n} \tilde{Y}_{j-1}\tilde{Y}_{j-1}'\right)^{-1} \sum_{j=1}^{n} \tilde{Y}_{j-1}\varepsilon_j \\ &= M'^{-1}\left(\sum_{j=1}^{n} \begin{pmatrix}\tilde{U}_{j-1}\\ \tilde{V}_{j-1}\end{pmatrix}\begin{pmatrix}\tilde{U}_{j-1}\\ \tilde{V}_{j-1}\end{pmatrix}'\right)^{-1} \sum_{j=1}^{n} \begin{pmatrix}\tilde{U}_{j-1}\\ \tilde{V}_{j-1}\end{pmatrix}\varepsilon_j.\end{aligned}$$

Now if we let

$$\eta = \begin{pmatrix} \gamma_1 & \gamma_2 & \cdots & \cdots & \gamma_{p-k} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \cdots & \cdots & \\ 0 & & \cdots & 1 & 0 \end{pmatrix}_{(p-k)\times(p-k)}$$

then Theorem 12 holds with δ_n and $\tilde{Y}_n$ there replaced, respectively, by $\eta_n = \eta^{-(n-2)}$ and $\tilde{V}_n$. To proceed further, we need Lemma 15.

LEMMA 15. *The differences*

$$\operatorname{diag}\left\{\frac{I_k}{\sqrt{n}}, \eta_n\right\}\left(\sum_{j=1}^{n} \begin{pmatrix}\tilde{U}_{j-1}\\ \tilde{V}_{j-1}\end{pmatrix}\begin{pmatrix}\tilde{U}_{j-1}\\ \tilde{V}_{j-1}\end{pmatrix}'\right)\operatorname{diag}\left\{\frac{I_k}{\sqrt{n}}, \eta_n'\right\}$$
$$-\operatorname{diag}\left\{\frac{1}{n}\sum_{j=1}^{n} \tilde{U}_{j-1}\tilde{U}_{j-1}', \eta_n\left(\sum_{j=1}^{n} \tilde{V}_{j-1}\tilde{V}_{j-1}'\right)\eta_n'\right\}$$

converge to 0 *in* $P_{\beta,n}$*-probability, where* I_k *denotes the unit matrix of order* k.

PROOF. It only remains to show that the off-diagonal matrix

$$\sum_{j=1}^{n} \frac{\tilde{U}_{j-1}}{\sqrt{n}} \tilde{V}_{j-1}'\eta_n'$$

converges to 0 in probability. Let Z_n be as defined in (41) with $\tilde{Y}_n$ and δ replaced, respectively, by $\tilde{V}_n$ and η. Then

$$\sum_{j=1}^{n} \frac{\tilde{U}_{j-1}}{\sqrt{n}} \tilde{V}_{j-1}\eta_n' = \sum_{j=1}^{n} \frac{\overline{U}_{j-1}}{\sqrt{n}} Z_{j-1}'\eta'^{-(n-j)}.$$

We have

$$\sup_{1\le j\le n} |\tilde{U}_j|/\sqrt{n} \to_p 0$$

and that $\{Z_n\}$, $n \ge 1$, is bounded in probability. Hence it follows easily that there is an increasing sequence $k_n \uparrow \infty$ such that

$$\sum_{j=n-k_n}^{n} \frac{\tilde{U}_{j-1}}{\sqrt{n}} Z'_{j-1}\eta'^{-(n-j)} \to_p 0. \tag{49}$$

Now, first assume that $\tilde{U}_j$ and Z_j are random variables so that

$$\begin{aligned} E\left[\left|\sum_{j=1}^{n-k_n} \frac{U_{j-1}}{\sqrt{n}} Z_{j-1}\eta^{-(n-j)}\right|\right] &\le \sum_{j=1}^{n-k_n} |\eta|^{-(n-j)} E\left[\left|\frac{U_{j-1}}{\sqrt{n}} Z_{j-1}\right|\right] \\ &\le \frac{K}{\sqrt{n}} \sum_{j=1}^{n-k_n} |\eta|^{-(n-j)} \\ &= \frac{K}{\sqrt{n}} \sum_{j=k_n}^{n-1} |\eta|^{-j} \to 0 \quad \text{as } n \to \infty, \end{aligned}$$

since $|\eta| > 1$, where

$$\begin{aligned} K &= \sup_j E\left[|U_{j-1}Z_{j-1}|\right] \\ &\le \sup_j \left\{E\left(U_{j-1}^2\right)E\left(Z_{j-1}^2\right)\right\}^{1/2} < \infty. \end{aligned}$$

With suitable modifications [see Anderson (1959)], it can be shown that the preceding arguments extend to the general vectors $\tilde{U}_j$ and $\tilde{Z}_j$, also. Thus

$$\sum_{j=1}^{n-k_n} \frac{\tilde{U}_{j-1}}{\sqrt{n}} Z'_{j-1}\eta'^{-(n-j)} \to_p 0. \tag{50}$$

Hence the lemma follows from (49) and (50). □

Lemma 15 leads to Theorem 16.

THEOREM 16. *Let* $\delta_n = \operatorname{diag}\{I_k/\sqrt{n}, \eta_n\}$ *and let* M *be as in* (47). *Then*:

(i) *The difference*

$$\delta_n'^{-1}M'\left(\hat{\beta}_n - \beta\right) - \begin{pmatrix} \sqrt{n}\left(\sum_{j=1}^{n} \tilde{U}_{j-1}\tilde{U}'_{j-1}\right)^{-1} \sum_{j=1}^{n} \tilde{U}_{j-1}\varepsilon_j \\ \eta_n'^{-1}\left(\sum_{j=1}^{n} \tilde{V}_{j-1}\tilde{V}'_{j-1}\right)^{-1} \sum_{j=1}^{n} \tilde{V}_{j-1}\varepsilon_j \end{pmatrix}$$

converges to 0 *in* $P_{\beta, n}$*-probability.*

(ii)

$$\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{U}_{j-1}\varepsilon_j, \eta_n \sum_{j=1}^{n} \tilde{V}_{j-1}\tilde{V}'_{j-1}\eta'_n, \eta_n \sum_{j=1}^{n} \tilde{V}_{j-1}\varepsilon_j\right)$$

converges jointly to $(W_1, A, A^{1/2}W_2)$, *where the vectors* W_1 *and* W_2 *and the random matrix* A *are mutually independent.*

PROOF. The first statement follows easily from Lemma 15. Further, since the eigenvalues of η are greater than 1 in absolute value, it follows easily by applying Theorem 12 to the variables V_n that for every $k_n \uparrow \infty$, the differences

$$\eta_n \sum_{j=1}^{n} \tilde{V}_{j-1}\tilde{V}_{j-1}'\eta_n' - \sum_{j=0}^{n} \eta^{-i} Z_n Z_n' \eta'^{-i} \tag{51}$$

and

$$\eta_n \sum_{j=1}^{n} \tilde{V}_{j-1}\varepsilon_j - \sum_{j=1}^{k_n} \eta^{-j} Z_n \varepsilon_{n-i} \tag{52}$$

converge to 0 in probability, where we now let, with $\tilde{\varepsilon}_i' = (\varepsilon_i, 0, \dots, 0)_{1\times(p-k)}$, $i \geq 1$,

$$Z_n = \tilde{\varepsilon}_1 + \eta^{-1}\tilde{\varepsilon}_2 + \cdots + \eta^{-k_n+1}\tilde{\varepsilon}_{k_n} + \eta\tilde{V}_0.$$

Note that the right-hand sides of (51) and (52) depend, for each n, on only the variables ε_i, $1 \leq i \leq k_n$ and $n - k_n \leq i \leq n$. Thus the statement will follow if we can show that the vectors $\sum_{j=1}^{h} \tilde{U}_{j-1}\varepsilon_j/\sqrt{n}$ are asymptotically equivalent to the vectors which depend only on the variables ε_i, $k_n < i < n - k_n$. For simplicity we present the proof of this fact only for the case $k = 1$. Now one can find an increasing sequence k_n such that $k_n \uparrow \infty$ and such that

$$\left[\sum_{j=1}^{k_n} E\left(U_{j-1}^2\right) + \sum_{j=n-k_n}^{n} E\left(U_{j-1}^2\right)\right]\Big/ n \to 0$$

and, since $|\rho_1| < 1$,

$$\sum_{j=1}^{n} E\left[|U_{j-1} - U_{j-1}^*|^2\right]/n \to 0,$$

where

$$U_j^* = \varepsilon_j + \rho_1\varepsilon_{j-1} + \cdots + \rho_1^{j-k_n-1}\varepsilon_{k_n+1}.$$

Hence it follows that, since $U_{j-1}\varepsilon_j$ and $U_{j-1}^*\varepsilon_j$ are martingale differences,

$$E\left[\left(\sum_{j=1}^{k_n} U_{j-1}\varepsilon_j + \sum_{j=n-k_n}^{n} U_{j-1}\varepsilon_j\right)^2\right]\Big/ n \to 0$$

and

$$E\left[\left(\sum_{j=1}^{n} \left(U_{j-1} - U_{j-1}^*\right)\varepsilon_j\right)^2\right]\Big/ n \to 0.$$

Thus the differences

$$\sum_{j=1}^{n} U_{j-1}\varepsilon_j/\sqrt{n} - \sum_{j=k_n+1}^{n-k_n-1} U_{j-1}^*\varepsilon_j/\sqrt{n} \to_p 0.$$

This completes the proof of the theorem since the sum $\sum_{j=k_n+1}^{n-k_n-1} U_{j-1}^*\varepsilon_j$ depends only on the variables ε_i, $k_n < i < n - k_n$. □

Having established the weak convergence, we now establish the strong convergence under the assumption that ε_i's are Gaussian. We present the details only when $\beta = (\beta_1, \dots, \beta_p)'$ is the true parameter; the general case can be obtained by replacing β throughout by $\beta + \delta_n h_n$, where h_n is bounded and δ_n is as in Theorem 16. To proceed further, note that β can be expressed as a function of the parameters $\alpha = (\alpha_1, \dots, \alpha_k)'$ and $\gamma = (\gamma_1, \dots, \gamma_{p-k})'$ which were introduced at the beginning of this section. Let

$$A_n(\alpha) = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\alpha_1 & 1 & 0 & \cdots & \cdots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \cdots & \vdots \\ 0 & & & \cdots & & 1 \end{bmatrix}_{n \times n}$$

and

$$B_n(\gamma) = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\gamma_1 & 1 & 0 & \cdots & \cdots & 0 \\ -\gamma_2 & -\gamma_1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \cdots & \vdots \\ 0 & & & & & 1 \end{bmatrix}_{n \times n}$$

Note that the matrices $A_n(\alpha)$ and $B_n(\gamma)$ are commutative, that is, $A_n(\alpha)B_n(\gamma) = B_n(\gamma)A_n(\alpha)$ for all n and for all α and γ. Further let, as before, $\mathbf{Y}_n = (Y_1, \dots, Y_n)'$, $\mathbf{U}_n = (U_1, \dots, U_n)'$ and $\mathbf{V}_n = (V_1, \dots, V_n)'$. To simplify the further notations, we take $k = 1$ and $p = 2$ so that α and γ are real values such that $|\alpha| < 1$ and $|\gamma| > 1$. We then have the following relations for all α and γ:

$$(\varepsilon_1 + \alpha U_0, \varepsilon_2, \dots, \varepsilon_n)' = A_n(\alpha)\mathbf{U}_n,$$

$$(\varepsilon_1 + \gamma V_0, \varepsilon_2, \dots, \varepsilon_n)' = B_n(\gamma)\mathbf{V}_n,$$

$$(U_1 + \gamma Y_0, U_2, \dots, U_n)' = B_n(\gamma)\mathbf{Y}_n,$$

and

(53) $$(V_1 + \alpha Y_0, V_2, \dots, V_n)' = A_n(\alpha)\mathbf{Y}_n.$$

[Relation (53) refers to all previous four identities.] If we take $(\alpha, \gamma)'$ as the parameter and $(\alpha + h_1/\sqrt{n}, \gamma + \gamma_n h_2)'$ as the alternative with $\gamma_n = \gamma^{-(n-2)}$, it follows from (53) that Condition (B) of Theorem (4) is satisfied with the transformation

(54) $$(A_n(\alpha)B_n(\gamma))^{-1} A_n\left(\alpha + \frac{h_1}{\sqrt{n}}\right) B_n(\gamma + \gamma_n h_2)\mathbf{Y}_n.$$

In what follows, assume for simplicity that $U_0 = 0$ and $V_0 = 0$. Using the fact that the matrices $A_n(\alpha)$ and $B_n(\gamma)$ are commutative, it then follows from (53) and (54) that the vector $\mathbf{U}_n$ transforms into the vector

$$A_n^{-1}(\alpha) A_n\left(\alpha + \frac{h_1}{\sqrt{n}}\right) B_n(\gamma + \gamma_n h_2)\mathbf{Y}_n = \mathbf{U}_n^*, \quad \text{say},$$

where the kth component of $\mathbf{U}_n^*$ is given by

$$U_k^* = U_k + \frac{h_1}{\sqrt{n}} \sum_{i=1}^{k-1} \left(U_{k-i} + h_2 \gamma_n \sum_{j=1}^{k-i-1} \alpha^{k-i-1-j} V_j \right) \alpha^{i-1}. \tag{55}$$

Similarly $\mathbf{V}_n$ transforms into the vector $\mathbf{V}_n^*$ whose kth component is given by

$$V_k^* = V_k + \gamma_n h_2 \sum_{i=1}^{k-1} \left(V_{k-i} + \frac{h_1}{\sqrt{n}} \sum_{j=1}^{k-i-1} \alpha^{k-i-1-j} V_j \right) \gamma^{i-1}. \tag{56}$$

Using (55) and (56) and using the ideas of Lemmas 8, 13 and 15, it can be shown that setting $h = (h_1, h_2)'$, the random vectors

$$\tilde{\theta}_n(h) = \left(\sum_{j=1}^{n} \begin{pmatrix} U_{j-1}^* \\ V_{j-1}^* \end{pmatrix} \begin{pmatrix} U_{j-1}^* \\ V_{j-1}^* \end{pmatrix}' \right)^{-1} \sum_{j=1}^{n} \begin{pmatrix} U_{j-1}^*(U_j^* - \alpha U_{j-1}^*) \\ V_{j-1}^*(V_j^* - \gamma V_{j-1}^*) \end{pmatrix}$$

satisfy Condition (C) of Theorem (4) with $\delta_n = \operatorname{diag}(n^{-1/2}, \gamma_n\}$. For this δ_n it can further be shown using similar arguments that the differences, setting $\theta = (\theta, \gamma)'$ and $\tilde{Z}_j = (U_j, V_j)'$,

$$\log\left(dP_{\theta+\delta_n h, n} / dP_{\theta, n} \right) - \left[h'\delta_n \sum_{j=1}^{n} \tilde{Z}_{j-1}\varepsilon_j - \tfrac{1}{2} h'\delta_n \left(\sum_{j=1}^{n} \tilde{Z}_j \tilde{Z}_j' \right) \delta_n h \right]$$

converge to 0 in $P_{\theta, n}$-probability so that the LAMN condition described in Section 7 is also satisfied. Hence, in view of the remarks made in Section 7 and in view of Theorem 16, we have

THEOREM 17. *Let $\hat{\beta}_n$ be the L.S. estimator of the partially explosive AR model defined previously. Assume that the error variables are Gaussian. Let δ_n and M be as in Theorem* 16 *and let $A_n = \sum_{j=1}^{n} \tilde{V}_{j-1} \tilde{V}_{j-1}'$. Then $\delta_n'^{-1} M'(\hat{\beta}_n - \beta)$ and* $\operatorname{diag}(\sqrt{n} I_k, A_n\} M'(\hat{\beta}_n - \beta)$ *converge strongly to* $(W_1, A^{-1/2} W_2)$ *and* (W_1, W_2), *respectively, where W_1 is a k-variate Gaussian with mean vector* 0 *and covariance matrix the inverse of the limit of $n^{-1} \sum_{j=1}^{n} \tilde{U}_{j-1} \tilde{U}_{j-1}'$, W_2 is a standard $(p - k)$-variate Gaussian and A is as in statement* (ii) *of Theorem* (16). *Furthermore, W_1, W_2 and A are mutually independent.*

Acknowledgments. It is a pleasure to express my thanks to Professor Bruce M. Hill for easing my teaching responsibilities during the progress of Part I of this work and to the referee for having kindly pointed out the fact (mentioned in Section 6) that the Gaussian restriction in Theorems 14 and 17 can be removed.

REFERENCES

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

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

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *On Normal Approximations and Asymptotic Expansions*. Wiley, New York.

BOOS, D. D. (1985). A converse to Scheffé's theorem. *Ann. Statist.* **13** 423–427.

BOSE, A. (1985). Higher order approximations for autocovariances from linear processes with applications. Technical Report, Indian Statistical Institute, Calcutta.

BUSTOS, O. H. (1982). General M-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality. *Z. Wahrsch. verw. Gebiete* **59** 491–504.

DAVIES, R. B. (1985). Asymptotic inference when the amount of information is random. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 841–864. Wadsworth, Monterey, Calif.

FULLER, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

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

JEGANATHAN, P. (1986). On the strong Gaussian approximation of the distributions of estimators in linear stochastic models, I: Autoregressive model. Technical Report 144, Dept. Statistics, Univ. Michigan.

JEGANATHAN, P. (1987a). On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle, I and II. Technical Report 150, Dept. Statistics, Univ. Michigan.

JEGANATHAN, P. (1987b). Strong convergence of distributions of estimators. *Ann. Statist.* **15** 1699–1708.

LAI, T. L. and WEI, C. Z. (1983). A note on martingale difference sequences satisfying the local Marcinkiewicz–Zygmund condition. *Bull. Inst. Math. Acad. Sinica* **11** 1–13.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LEVIT, B. YA (1974). On the behavior of generalized Bayes estimators for Markov observations. *Theory Probab. Appl.* **19** 342–346.

MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.

MARTIN, R. D. and YOHAI, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5** 119–155. North-Holland, Amsterdam.

RUDIN, W. (1974). *Real and Complex Analysis*. McGraw-Hill, New York.

STIGUM, B. P. (1974). Asymptotic properties of dynamic stochastic parameter estimates, III. *J. Multivariate Anal.* **4** 351–381.

SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70.

VENKATARAMAN, K. N. (1967). A note on the least squares estimators of the parameters of a second order linear stochastic difference equation. *Calcutta Statist. Assoc. Bull.* **16** 15–28.

VENKATARAMAN, K. N. (1968). Some limit theorems on a linear stochastic difference equation with a constant term and their statistical applications. *Sankhyā Ser. A* **30** 51–74.

WHITE, J. S. (1958). The limiting distribution of the serial correlation in the explosive case. *Ann. Math. Statist.* **29** 1188–1197.

DEPARTMENT OF STATISTICS
1444 MASON HALL
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1027