BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-83/01

ON THE UNBIASEDNESS PROPERTY OF AIC FOR EXACT OR APPROXIMATING MULTIVARIATE ARMA MODELS

by

David F. Findley
U.S. Bureau of the Census

Recommended by: Myron J. Katzoff

Report completed: February 5, 1983

Report issued: July 8, 1983 (Typographically Corrected: June 15, 1984)

ON THE UNBIASEDNESS PROPERTY OF AIC FOR EXACT OR APPROXIMATING MULTIVARIATE ARMA MODELS

D. F. FINDLEY
U.S. Bureau of the Census, Washington, D.C.

ABSTRACT. A rigorous analysis is given of the asymptotic bias of the log maximum likelihood as an estimate of the expected log likelihood (the negative of the cross-entropy) of the maximum likelihood model, when an invertible, conditional, multivariate gaussian ARMA(p,q) model, with or without coefficient/innovations covariance constraints, is fit to stationary, possibly non-gaussian observations. It is assumed that these data either (i) arise from a model whose spectral density matrix coincides with that of a member of the class of models being fit, or (ii) do not conform to any ARMA model but do come from a process whose spectral density matrix can be well-approximated by invertible ARMA model spectral density matrix functions. For the gaussian sub-case of (i), the innovations covariance matrices of the models need not be parametrized separately from the coefficients, but otherwise a separate parametrization is assumed. The analysis shows that, for the purpose of comparing maximum likelihood models from different model classes, Akaike's AIC is asymptotically unbiased in case (i). In case (ii), its asymptotic bias is of the order of a number less than unity raised to the power max{p,q} and so is negligible if max{p,q} is not too small. These results extend and complete the somewhat heuristic analysis given by Ogata (1980) for exact or approximating univariate autoregressive models.

Keywords: AIC, autoregressive-moving average models (multivariate), bias of log maximum likelihood, model selection.

1. INTRODUCTION AND OVERVIEW.

Let $L_N[\eta]$ denote a log likelihood function of a model from a class, parametrized by the coordinate vector η, a member of which is to be fit to the observations $x(1),\ldots,x(N)$. Using E to denote expectation with respect to the true joint distribution of the observations (assumed continuously differentiable), set $E_N[\eta] = EL_N[\eta]$. If E_N^{true} designates the expected value (assuming it exists) of the true log likelihood function of the data, then $E_N[\eta] < E_N^{true}$ holds unless $L_N[\eta]$ coincides with the true likelihood. Thus, given maximum likelihood estimates, $\hat{\eta}^N$ and $\hat{\zeta}^N$, from two different model classes, if $E_N[\hat{\eta}^N] < E_N[\hat{\zeta}^N]$, it seems appropriate to prefer the model defined by $\hat{\zeta}^N$. (See Akaike (1977) and Findley (1982) for some perspectives on this criterion.) The difficulty with this procedure might appear to be that $E_N[\hat{\eta}^N]$ (and $E_N[\hat{\zeta}^N]$) cannot be calculated unless the true joint density of $x(1),\ldots,x(N)$ is known. Akaike (1973) proposed, however, that if $L_N[\hat{\eta}^N]$ is a reasonable approximation to the true log likelihood, then $L_N[\hat{\eta}^N]$ - $\dim\eta$ would be an essentially unbiased estimate of $E_N[\hat{\eta}^N]$ adequate for such comparisons provided N is large enough. (Here $\dim\eta$ denotes the dimension of η, i.e., the number of independent parameters estimated in the model.) Akaike's $AIC[\hat{\eta}^N]$ is defined to be $-2L_N[\hat{\eta}^N] + 2\dim\eta$, so the model with the smaller AIC value is preferred.

The analysis of AIC's asymptotic bias properties given in Akaike (1973) is not completely rigorous and is restricted to i.i.d. data. Because of this restriction, it does not immediately apply to the situation of fitting ARMA models to time series. The optimality properties of the minimum AIC procedure established in Shibata (1980; 1981a; 1981b) and Taniguchi (1980) are more relevant than the bias properties in certain applications, but the bias question seems basic and in need of resolution, given the influential nature of Akaike's work on AIC (see This Week's Citation Classic (1981)). An additional stimulus for this investigation is the Bayesian modeling procedure proposed by Akaike (1978), in which $L_N[\hat{\eta}^N] - \dim\eta$ is used as an estimate of the log likelihood.

In the present paper, a rigorous analysis is given of the asymptotic bias of $E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N]$, when invertible, and possibly otherwise constrained, stationary, gaussian, multivariate ARMA(p,q) models are fit to stationary r-dimensional observations $x(1),\ldots,x(N)$, under mostly standard data and model assumptions described by (2.I-VI) in section 2. Additional notational conventions and some background results are also presented there.

In section 3, we outline the derivation of the formula (3.1) identifying this bias as the negative of the trace of $F^{-1}[\eta^{p,q}]\{G[\eta^{p,q}] + H[\eta^{p,q}]\}$, an expression involving the matrices defined in (2.16-8) which describe the covariance matrix of the limiting gaussian distribution of $\hat{\eta}^N$. The details required to fill in this outline are given in Appendix 1. Examples 3.1 and 3.2, which present asymptotic properties of a type of univariate sample autocorrelation, help to illustrate the scope of (3.1) and yield information about some proposed procedures for selecting multi-step-ahead forecasting models.

It is shown in section 4 that, for the case (i) described in the Abstract, the trace expression in (3.1) reduces to $\dim\eta$ + (in the non-gaussian situation) a scale invariant term involving only the covariances and fourth cumulants of the components of the innovations series of the observed series x(t). Also for case (i), explicit formulas are obtained for $F[\eta^{p,q}]$, $G[\eta^{p,q}]$ and $H[\eta^{p,q}]$ when p or q is 0 (i.e., moving average or autoregressive models are fit) assuming the model parameters η are the model's coeffficients together with the on- and above-diagonal elements of its innovations covariance matrix. These formulas are used to verify the condition (5.IV) which is imposed in section 5 where the, perhaps more realistic, case (ii) is considered, when ARMA models can only approximately describe the covariance structure of the observations.

The main result of section 5, Proposition 5.3, asserts that in this approximating situation, the asymptotic bias formulas obtained in the exact modeling situation of section 4 are still accurate to order $(\delta-)^{\max\{p,q\}}$ or better. Here $\delta-$ is any positive number less than the number δ used in (2.II) to describe the rate of decay of the coefficients of the infinite autoregressive and innovations representations of the series x(t). The proofs of the preliminary results required for Proposition 5.3 are deferred to Appendix 2.

A less rigorous analysis of the asymptotic bias of AIC for approximating univariate autoregressive models is given in Ogata (1980), along with several interesting examples. After the requisite central limit theorem is established, the derivation of the analogue of (3.1) given there does not go beyond presenting the two Taylor expansions we describe in section 3. Thus a seemingly needed assumption like (2.V) insuring uniform integrability is lacking, and so are proofs of results analogous to those established in Appendix 1 of the present paper. (A similar lapse occurs in the proof of Proposition 3 of Taniguchi (1980)). Also, a condition like (5.IV) is not mentioned or verified, and this too seems essential.

Kozin and Nakajima (1980), who offer a derivation of AIC for time-varying AR models in case (i), do make extra assumptions to deal with the difficulty of approximating means of quadratic forms having random coefficients. However, they neglect to precisely analyze the error in their basic approximation, (2-1-13), and the requisite analysis of this error and the verification of several subsequent assertions seem to require a number of assumptions beyond those that they have made.

The results of the present paper most directly related to the minimum AIC model selection procedure are given in Corollaries 4.1 and 4.2 and in Remark 5.3. Some other research related to AIC is summarized briefly in Remark 4.2.

2. ASSUMPTIONS AND OTHER PRELIMINARIES.

Throughout this paper, $x(1),\ldots,x(N)$ will denote consecutive observations from a r-dimensional stationary time series $x(t)$ admitting a representation of the form

$$x(t) = \sum_{m=0}^{\infty} c(m)e(t-m) \quad , \quad (c(0)=I) \tag{2.1}$$

where I denotes here the rxr identity matrix and where $e(t) = (e_1(t),\ldots,e_r(t))^T$ $(t=0,\pm1,\ldots)$ is a sequence of zero mean random column r-vectors (^{T} denotes transpose) with the following properties:

(2.Ii). *For any integers* t, u *with* $t \neq u$, $e(t)$ *and* $e(u)$ *are independent.*

(2.Iii). *The absolute moments of order* τ *of the coordinate series* $e_j(t)$ $(1 \leq j \leq r;\ t=0,\pm1,\ldots)$ *are uniformly bounded,*

$$\sup_{1\leq j\leq r;\ -\infty<t<\infty} E|e_j(t)|^\tau < \infty \tag{2.2}.$$

Here τ *is a positive number which will be further specified in* (2.VI) *below and will, in any case, be larger than* 8.

(2.Iiii). *The mixed moments of order four or less of the coordinate series* $e_j(t)$ $(1\leq j\leq r;\ -\infty<t<\infty)$ *do not depend on* t.

We denote the fourth-order cumulants $\mathrm{cum}(e_a(t), e_b(t), e_c(t), e_d(t))$ by κ_{abcd} $(1 \le a,b,c,d \le r)$ or, if $r=1$, by κ_4. Our final assumption on the $e(t)$ series is that it has full rank,

(2.Iiv). $\Sigma = Ee(t)e^T(t)$ is nonsingular.

Before we describe the conditions to be imposed on the coefficients $c(m)$ in (2.1), we introduce some notational conventions. We use subscripts to denote coordinate entries of vectors and matrices. For a matrix K, of order v, say, we define $|K|_\infty = \max_{1 \le i,j \le v} |K_{ij}|$. Next, let δ denote a positive number less than 1 whose value is fixed throughout this paper. As m increases, we shall require the magnitudes $|c(m)|_\infty$ of the coefficient matrices in (2.1) to decay exponentially at a rate more rapid than δ^m,

$$\limsup_{m \to \infty} \{|c(m)|_\infty\}^{1/m} < \delta ,$$

a condition which we shall find it convenient to denote by

$$|c(m)|_\infty \sim O[(\delta-)^m] \qquad (2.3),$$

and to describe in words by saying that $|c(m)|_\infty$ is of order $(\delta-)^m$. In general, it will turn out to simplify the exposition to use $\delta-$ to denote a positive number (perhaps a different number at each occurrence!) whose only significant property is that it is less than δ. (Lemma A1.1 in Appendix 1 and its proof provide a simple illustration of the utility of this notation.)

From (2.3), it follows that the series defining the z-transform of the coefficients, $C(z) = \sum_{m=0}^{\infty} c(m)z^m$, converges uniformly in $\{|z| \leq (\delta -)^{-1}\}$. If the determinant $\det C(z)$ has no zeros in this disk, then the coefficients of the power series expansion $\sum_{m=0}^{\infty} d(m)z^m$ of $D(z) = C^{-1}(z)$ will satisfy

$$|d(m)|_{\infty} \sim O[(\delta -)^m] \qquad (2.4).$$

We shall, in fact, assume

(2.II). <u>The condition</u> (2.3) <u>holds</u>, <u>and</u>, <u>also</u>, $\det C(z)$ <u>has no zeros in</u> $\{|z| \leq (\delta -)^{-1}\}$.

As a consequence of (2.I-II), the series $x(t)$ has stationary moments up through order four and its covariance structure will closely resemble that of a series conforming to an invertible ARMA model. With $\|y\|_\gamma$ denoting the γ-norm, $\{E|y|^\gamma\}^{1/\gamma}$, of a random variable y (where $\gamma \geq 1$), we note for later reference that the assumptions (2.I-II) imply that

$$\sup_{1 \leq j \leq r} \|x_j(m) - \sum_{m=0}^{t-u-1} \{c(m)e(t-m)\}_j\|_4 \sim O[(\delta -)^{t-u}] \qquad (2.5)$$

holds for all $t \geq u+1$.

We shall model the observations as though they coincided with N observations $y(1),\ldots,y(N)$ from a stationary time series $y(t)$ which satisfies some invertible ARMA(p,q) model,

$$y(t) + a[\eta](1)y(t-1) +\ldots+ a[\eta](p)y(t-p) =$$

$$\varepsilon(t) + b[\eta](1)\varepsilon(t-1) +\ldots+ b[\eta](q)\varepsilon(t-q) \qquad (2.6),$$

whose r-dimensional innovations series $\varepsilon(t)$ has covariance matrix $\Sigma[\eta]$. Let $A[\eta](z)$ and $B[\eta](z)$ designate the respective matrix polynomials $I + a[\eta](1)z +\ldots+ a[\eta](p)z^p$ and $I + b[\eta](1)z +\ldots+ b[\eta](q)z^q$. The parameterizing vector η is required to belong to a set, $ETA^{p,q}$, about which we shall assume

(2.III). $ETA^{p,q}$ *is a compact, convex set in the* s-*dimensional Euclidean coordinate space* IR^s *having non-empty interior. The coefficients of* $A[\eta](z)$ *and* $B[\eta](z)$ *and the entries of innovations covariance matrix* $\Sigma[\eta]$ *are continuous on* $ETA^{p,q}$ *and three times continuously differentiable in the interior of* $ETA^{p,q}$. Also, *for each* η *in* $ETA^{p,q}$, $\det\Sigma[\eta]$ *is nonzero, and the zeros of* $\det A[\eta](z)$ *and those of* $\det B[\eta](z)$ *belong to* $\{|z|>(\delta-)^{-1}\}$.

This last assumption insures that for each η in $ETA^{p,q}$, $C[\eta](z) = A^{-1}[\eta](z)B[\eta](z)$ and $D[\eta](z) = C^{-1}[\eta](z)$ have power series expansions, $\sum_{m=0}^{\infty} c[\eta](m)z^m$ and $\sum_{m=0}^{\infty} d[\eta](m)z^m$, whose coefficients are or order $(\delta-)^m$, e.g.,

$$\sup_\eta |d[\eta](m)|_\infty \sim O[(\delta-)^m] \qquad (2.7).$$

We now consider the log likelihood function associated with gaussian observations $y(1),\ldots,y(N)$ satisfying (2.6) with the initial conditions

$$y(0) = y(-1) = \ldots = y(-p+1) = 0 \qquad (\text{if } p>0) \qquad (2.8)$$

$$\varepsilon(0) = \varepsilon(-1) = \ldots = \varepsilon(-q+1) = 0 \qquad (\text{if } q>0) \qquad (2.9).$$

When this log likelihood function is evaluated at $x(1),\ldots,X(N)$, we denote the result by $L_N[\eta]$. Thus,

$$L_N[\eta] = -\frac{1}{2} \log \det 2\pi\Sigma[\eta] - \frac{1}{2} \sum_{t=1}^{N} \varepsilon^T[\eta](t)\Sigma^{-1}[\eta]\varepsilon[\eta](t) \qquad (2.10),$$

where

$$\varepsilon[\eta](t) = \sum_{m=0}^{t-1} d[\eta](m)x(t-m) \qquad (1 \leqslant t \leqslant N) \qquad (2.11).$$

To express $L_N[n]$ explicitly as a quadratic form in the components of the observations, we define the Nr-dimensional column vectors $x^N = \mathrm{vec}(x(N),\ldots,x(1))$ $(= [x^T(N)\ x^T(N-1)\ldots x^T(1)]^T)$ and $\varepsilon^N[n] = \mathrm{vec}(\varepsilon[n](N),\ldots,\varepsilon[n](1))$, and we define $\Sigma^{-N}[n]$ to be the NrxNr block diagonal matrix $\mathrm{diag}(\Sigma^{-1}[n],\ldots,\Sigma^{-1}[n])$. If $d^N[n]$ denotes the block upper triangular matrix of order Nr whose (m,n)-block $(m<n)$ is the coefficient matrix $d[n](n-m)$, then, by (2.11), $\varepsilon^N[n] = d^N[n]x^N$, and

$$L_N[n] = -\frac{1}{2}\log\det 2\pi\Sigma[n] - \frac{1}{2}x^{NT}d^{NT}[n]\Sigma^{-N}[n]d^N[n]x^N \qquad (2.12).$$

This is a convenient form for the calculation of $E_N[n]$, the <u>expected value</u> $EL_N[n]$ of $L_N[n]$ with respect to the true joint distribution of x^N. (We use such notation because n will frequently be made random, n , but always <u>after</u> the expectation is calculated. If we wrote $EL_N[n]$, this wouldn't be clear.) If $\Gamma(m) = Ex(t)x^T(t-m)$ $(m = 0,\pm1,\ldots)$ and if we denote by Γ^N the block Toeplitz covariance matrix whose (m,n)-block is $\Gamma^T[m-n]$, then, from (2.12),

$$E_N[n] = -\frac{1}{2}\log\det 2\pi\Sigma[n] - \frac{1}{2}\mathrm{tr}\Gamma^N d^{NT}[n]\Sigma^{-N}[n]d^N[n] \qquad (2.13)$$

Here tr denotes the trace operator, and we follow the convention that <u>matrix products are calculated before traces</u>.

Now we introduce the spectral density matrices $f(\lambda) = (2\pi)^{-1}\,C(e^{i\lambda})\Sigma C^*(e^{i\lambda})$ and $f[n](\lambda) = (2\pi)^{-1}C[n](e^{i\lambda})\Sigma[n]C^*[n](e^{i\lambda})$ $(-\pi<\lambda<\pi)$, where * denotes complex conjugate transpose. We recall from Hannan (1970, p. 162) that

$$\log \det\Sigma[\eta] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \det 2\pi f[\eta](\lambda)d\lambda \qquad (2.14).$$

It will follow from Proposition A1.1 of Appendix 1 that

$$\lim_{N\text{--->}\infty} \quad N^{-1}E_N[\eta] = W[\eta] \qquad (2.15)$$

uniformly on $ETA^{p,q}$, where

$$W[\eta] = -\frac{1}{2}\log \det 2\pi\Sigma[\eta] - \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr} f(\lambda)f^{-1}[\eta](\lambda)d\lambda$$

$$= -r\log 2\pi - \frac{1}{4\pi}\int_{-\pi}^{\pi}\{\mathrm{tr} f(\lambda)f^{-1}[\eta](\lambda)+\log \det f[\eta](\lambda)\}d\lambda \quad (2.16),$$

with the second equality coming from (2.14).

Now, let $\hat{\eta}^N$, η^N, and $\eta^{p,q}$ denote points in $ETA^{p,q}$ at which maximum values are obtained by $L_N[\eta]$, $E_N[\eta]$ and $W[\eta]$, respectively. We shall assume

(2.IVi). *For* N *sufficiently large,* $\hat{\eta}^N$ *is, almost surely, an interior point of* $ETA^{p,q}$.

(2.IVii). η^N *is the only point of* $ETA^{p,q}$ *at which a maximum value of* $E_n[\eta]$ *is obtained*.

(2.IViii). $\eta^{p,q}$ <u>belongs to the interior of</u> $ETA^{p,q}$ <u>and is the only point of</u> $ETA^{p,q}$ <u>at which a maximum value of</u> $W[\eta]$ <u>is obtained</u>.

(2.IViv). <u>The hessian matrix</u> $\{\partial^2 W/\partial\eta\partial\eta^T\}[\eta^{p,q}]$ <u>is non-singular</u>.

For the cases p=0 or q=0 most of interest to us (when the parameters η are the coefficients and the innovations covariances) (2.IViv) will follow from (4.11), (4.13-15) and (5.22) below. For the univariate situation (r=1), a set of ARMA(p,q) models with p and q both non-zero for which (2.IViv) appears to be satisfied is described in Example 1 of Taniguchi (1980, p. 405).

The situation regarding (2.IVii-iii), which are required by the reference we cite for the central limit theorem below, is more difficult, except in the case q=0. An especially important issue is addressed in the following Remark, which focuses on (2.IViii), but which is largely applicable to (2.IViii), as well. However, we mention without giving details that the bias properties of AIC established in this paper can also be obtained in certain circumstances in which $E_N[\eta]$ and $W[\eta]$ are maximized at finitely many points, all in the interior of $ETA^{p,q}$.

<u>Remark 2.1</u>. If $\eta^{p,q}$ is a point $E_{p,q}$ maximizing $W[\eta]$, then (2.IViii) requires, in particular, that there be no $\eta \neq \eta^{p,q}$ in $ETA^{p,q}$ such that $f[\eta](\lambda)$ coincides with $f[\eta^{p,q}](\lambda)$. When p>0 and q>0, and if no special constraints have been imposed, this means, in the univariate case, that $A[\eta^{p,q}](z)$ and $B[\eta^{p,q}](z)$ must not have common roots and that at least one of the final coefficients $a[\eta^{p,q}](p)$, $b[\eta^{p,q}](q)$ should be non-zero, to prevent the possibility that for some η in $ETA^{p,q}$, $A[\eta](z) = (1-\beta z)A[\eta^{p,q}](z)$, and $B[\eta](z) = (1-\beta z)B[\eta^{p,q}](z)$ with $\beta \neq 0$. Generalizations of

these conditions for the case $r>1$ are discussed in Deistler, Dunsmuir and Hannan (1978, pp. 361-366). Hosoya and Taniguchi (1982, p. 149) show that, if the true spectral density matrix $f(\lambda)$ coincides with some $f[\eta^0](\lambda)$, then $W[\eta]$ will be maximized at η^0, so that we must have $f(\lambda) = f[\eta^{p,q}](\lambda)$ if (2.IViii) holds. If $x(t)$ is an unconstrained ARMA(p_0,q_0) process, then (2.IViii) will fail, therefore, if $p>p_0$ and $q>q_0$ (Hannan (1982) describes some pathological behavior of AIC in this case).

Remark 2.2. It follows from (2.IViii) and from the uniform convergence of (2.15) that $\lim_{N\text{--->}\infty} \eta^N = \eta^{p,q}$. Thus, for N sufficiently large, η^N is in the interior of $ETA^{p,q}$, and $\{\partial E_N/\partial\eta\}[\eta^N]$ must be zero, along with $\{\partial W/\partial\eta\}[\eta^{p,q}]$. Also, in the separately parametrized case, i.e., when $\eta = (\xi^T, \theta^T)^T$ with $\Sigma[\eta]$ depending only on ξ and $C[\eta](e^{i\lambda})$ depending only on θ, it follows from (2.I-IV) and (2.5), as in Ljung and Caines (1979), that $\hat{\eta}^N - \eta^N \text{--->}_{a.s.} 0$. The condition (2.IVi) shows that, for sufficiently large N, $\{\partial L_N/\partial\eta\}[\hat{\eta}^N] = 0$ almost surely.

In the separately parametrized case just described, one can, also as in Ljung and Caines (1979), use (2.5) to verify the central limit theorem for $N^{1/2}(\hat{\eta}^N - \eta^N)$ presented as assumption (2.V) below. This assumption is likely to be valid more generally: For example, without assuming such a separate parametrization, Hosoya and Taniguchi (1982) show that $N^{1/2}(\tilde{\eta}^N - \eta^{p,q})$ has the same limiting distribution, when $\tilde{\eta}^N$ maximizes Whittle's likelihood ((3.12) below.) We will not use the gaussian property of the asymptotic distribution. This property is included in (2.V) because it seems always to occur when the rest of (2.V) is satisfied.

(2.V). The asymptotic distribution of $N^{1/2}(\eta^N - \eta^N)$ is gaussian with mean 0 and covariance matrix

$$F^{-1}[\eta^{p,q}](G[\eta^{p,q}] + H[\eta^{p,q}])F^{-1}[\eta^{p,q}] \qquad (2.17),$$

where $\eta^{p,q}$ is as in (2.IVii), where $F[\eta]$ $(= -\{\partial^2 W/\partial\eta\partial\eta^T\}[\eta])$ is given by

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} (\partial^2/\partial\eta\partial\eta^T)\{\mathrm{tr} f(\lambda)f^{-1}[\eta](\lambda) + \log\det f[\eta](\lambda)\}d\lambda \qquad (2.18),$$

where $G[\eta]$ has (j,k)-entry

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr} f(\lambda)\{\partial f^{-1}/\partial\eta_j\}[\eta](\lambda)f(\lambda)\{\partial f^{-1}/\partial\eta_k\}[\eta](\lambda)d\lambda \qquad (2.19),$$

and where $H[\eta]$ has (j,k)-entry

$$\frac{1}{4}\sum_{a,b,c,d=1}^{r} \kappa_{abcd}\left[\frac{1}{4\pi^2}\int_{-\pi}^{\pi} C^*(e^{i\lambda})\{\partial f^{-1}/\partial\eta_j\}[\eta](\lambda)C(e^{i\lambda})d\lambda\right]_{ab}$$

$$\times\left[\frac{1}{4\pi^2}\int_{-\pi}^{\pi} C^*(e^{i\lambda})\{\partial f^{-1}/\partial\eta_k\}[\eta](\lambda)C(e^{i\lambda})d\lambda\right]_{cd} \qquad (2.20),$$

for $1 \leq j,k \leq \dim\eta$ $(=s)$.

Of course, $H[\eta] = 0$ if $x(t)$ is gaussian.

In order to be able to calculate limits of expected values of quadratic forms in $N^{1/2}(\hat{\eta}^N - \eta^N)$, we assume that $\hat{\eta}^N - \eta^N$ converges to zero rapidly enough that, for some $\alpha > 0$

$$(2.\text{VIi}). \qquad \sup_N \; N^{1/2} \sum_{j=1}^{\dim\eta} \|\hat{\eta}_j^N - \eta_j^N\|_{2(1+\alpha)} < \infty \quad .$$

One immediate consequence of (2.VIi) is that for any $\gamma \geq 1$,

$$\sup_N N^{\min\{1/2,(1+\alpha)/\gamma\}} \sum_{j=1}^{\dim\eta} \|\hat{\eta}_j^N - \eta_j^N\|_\gamma < \infty \qquad (2.21).$$

Also, using (2.VIi) and Theorem 4.5.2 of Chung (1968, p. 88), one can establish the following lemma.

<u>Lemma 2.1</u>. <u>If</u> F^N (N=1,2,...,) <u>is a sequence of non-stochastic matrices tending to</u> $F[\eta^{p,q}]$ <u>and if</u> (2.V) <u>and</u> (2.VIi) <u>hold, then</u>

$$\lim_{N\to\infty} \quad NE(\hat{\eta}^N - \eta^N)^T F^N(\hat{\eta}^N - \eta^N) =$$

$$\operatorname{tr} F^{-1}[\eta^{p,q}](G[\eta^{p,q}] + H[\eta^{p,q}]) \qquad (2.22).$$

Without a condition like (2.Vi), it is not possible, in general, to assert even that $NE(\hat{\eta}^N - \eta^N)(\hat{\eta}^N - \eta^N)^T$ tends to the covariance matrix of the asymptotic distribution in (2.V). For integer values of 2α, and for the case of approximating univariate autoregressive models, Bhansali (1981) derives an analogue of (2.VIi) from another condition on $(\hat{\eta}^N - \eta^{p,0})$.

The Taylor expansions of section 3 produce quadratic forms in $N^{1/2}(\hat{\eta}^N - \eta^N)$ with random coefficients. To deal with these, we will use several straightforward consequences of Hölder's inequality, such as the following.

Lemma 2.2. *If $\gamma \geq 1$, and $\mu, \nu > 1$ satisfying $\mu^{-1} + \nu^{-1} = 1$ are given, then for any matrix Δ^N of order $\dim\eta$ with (Borel measurable) entries depending on $x(1),\ldots,x(N)$, we have*

$$\| N(\hat{\eta}^N - \eta^N)^T \Delta^N (\hat{\eta}^N - \eta^N) \|_\gamma \leq$$

$$\sum_{j,k=1}^{\dim\eta} N \|\hat{\eta}_j^N - \eta_j^N\|_{2\gamma\mu} \|\hat{\eta}_k^N - \eta_k^N\|_{2\gamma\mu} \|\Delta_{jk}^N\|_{\gamma\nu} \tag{2.23}$$

Our final assumption augments (2.Iii):

(2.VIii). *For α as in (2.VIi) and for some $\beta > 1$,*

$$\sup_{1 \leq j \leq r; -\infty < t < \infty} \|e_j(t)\|_{2\beta(1+\alpha^{-1})} < \infty.$$

From this condition and (2.1) we get

$$\sup_{1\leq j\leq r;-\infty<t<\infty} \|x_j(t)\|_{2\beta(1+\alpha^{-1})} < \infty \tag{2.24}$$

and, using the Cauchy-Schwarz inequality, also

$$\sup_{1\leq j,k\leq r;-\infty<\tilde{t},t<\infty} \|x_j(\tilde{t})x_k(t)\|_{\beta(1+\alpha^{-1})} < \infty \tag{2.25}$$

3. OUTLINE-DERIVATION OF THE BIAS OF $L_N[\hat{\eta}^N]$ AS AN ESTIMATOR OF $E_N[\hat{\eta}^N]$.

In this section, we outline the proof that, for the maximum likelihood model determined by $\hat{\eta}^N$ and under the assumptions (2.I - VI),

$$\lim_{N\to\infty} E\{E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N]\} =$$

$$- \mathrm{tr}F^{-1}[\eta^{p,q}](G[\eta^{p,q}] + H[\eta^{p,q}]) \tag{3.1}$$

holds, where $F[\eta]$, $G[\eta]$ and $H[\eta]$ are defined by (2.18-20), and η^N (used below) and $\eta^{p,q}$ are characterized by (2.IVi-ii).

We begin with the identity

$$E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N] = \{E_N[\hat{\eta}^N] - E_N[\eta^N]\}$$

$$+ \{E_N[\eta^N] - L_N[\eta^N]\} + \{L_N[\eta^N] - L_N[\hat{\eta}^N]\} \qquad (3.2),$$

and the observation that

$$E\{E_N[\eta^N] - L_N[\eta^N]\} = 0 \qquad (3.3),$$

since, by definition, $E_N[\eta^N] = EL_N[\eta^N]$. Thus (3.1) will follow if we demonstrate that

$$\lim_{N\longrightarrow\infty} \quad E\{E_N[\hat{\eta}^N] - E_N[\eta^N]\} =$$

$$\lim_{N\longrightarrow\infty} \quad E\{L_N[\eta^N] - L_N[\hat{\eta}^N]\} =$$

$$-\frac{1}{2}\,\mathrm{tr}\; F^{-1}[\eta^{p,q}](G[\eta^{p,q}] + H[\eta^{p,q}]) \qquad (3.4).$$

We shall establish (3.4) with the aid of the first degree Taylor polynomial expansions of the first and third expression on the right in (3.2) around η^N and $\hat{\eta}^N$, respectively. We denote $\{\partial^2 E_N/\partial\eta\partial\eta^T\}[\eta]$ by $E_N''[\eta]$, etc..

As Remark 2.1 explains, for sufficiently large N, the first derivative terms are zero, so we have

$$E_N[\hat{\eta}^N] - E_N[\eta^N] = \frac{1}{2}\,(\hat{\eta}^N - \eta^N)^T E_N''[\eta^{N*}](\hat{\eta}^N - \eta^N)$$

$$= \frac{1}{2}\,N(\hat{\eta}^N - \eta^N)^T\{N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N) +$$

$$\frac{1}{2}\,N(\hat{\eta}^N - \eta^N)^T\{N^{-1}E_N''[\eta^{N*}] - N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)$$

for some η^{N*} on the line segment in $\mathbb{R}^s$ between $\hat{\eta}^N$ and η^N. Similarly, for some η^{N**} on this line segment,

$$L_N[\eta^N] - L_N[\eta^N] = \frac{1}{2}\,(\eta^N - \hat{\eta}^N)^T L_N''[\eta^{N**}](\eta^N - \hat{\eta}^N)$$

$$= \frac{1}{2}\,N(\hat{\eta}^N - \eta^N)^T\{N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)$$

$$+ \frac{1}{2}\,N(\hat{\eta}^N - \eta^N)^T\{N^{-1}L_N''[\eta^N] - N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)$$

$$+ \frac{1}{2}\,N(\hat{\eta}^N - \eta^N)^T\{N^{-1}L_N''[\eta^{N**}] - N^{-1}L_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)\;. \tag{3.6}$$

Once we have verified that

$$\lim_{N\longrightarrow\infty} \quad N^{-1}E_N''[\eta^N] = -F[\eta^{p,q}] \tag{3.7}$$

it will follow from Lemma 2.1 that

$$\lim_{N\longrightarrow\infty} \quad E\{N(\hat{\eta}^N - \eta^N)^T\{N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)\} =$$

$$-\frac{1}{2}\,\mathrm{tr}\; F^{-1}[\eta^{p,q}](G[\eta^{p,q}] + H[\eta^{p,q}]) \tag{3.8}$$

Examining (3.5-6), it is clear that (3.8) and (3.9-11) below imply (3.4).

$$\lim_{N\longrightarrow\infty} \quad E\{N(\hat{\eta}^N - \eta^N)^T\{N^{-1}E_N''[\eta^{N*}] - N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)\} = 0 \tag{3.9}$$

$$\lim_{N\longrightarrow\infty} \quad E\{N(\hat{\eta}^N - \eta^N)^T\{N^{-1}L_N''[\eta^{N**}] - N^{-1}L_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)\} = 0 \tag{3.10}$$

$$\lim_{N\longrightarrow\infty} \quad E\{N(\hat{\eta}^N - \eta^N)^T\{N^{-1}L_N''[\eta^N] - N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)\} = 0 \tag{3.11}$$

The assertions (3.7) and (3.9-11) will be established in Appendix 1.

Remark 3.1. It follows from (3.1) that $L_N[\hat{\eta}^N]$ always has an upward bias, asymptotically, as an estimator of $E_N[\hat{\eta}^N]$.

After commenting on some generalizations and variations of (3.1) in Remarks 3.2-5 below, we shall present two examples in which (3.1) is used to obtain properties of an autocorrelation estimate.

Remark 3.2. Using the fact that $\{\partial F/\partial\eta\}[\eta^{p,q}] = 0$, it is easy to verify that $\mathrm{tr}F^{-1}[\eta^{p,q}]G[\eta^{p,q}]$ and $\mathrm{tr}F^{-1}[\eta^{p,q}]H[\eta^{p,q}]$, and therefore also the right hand side of (3.1), are invariant under changes of variable $\eta \longrightarrow \phi$ which are twice continuously differentiable and non-singular in a neighborhood about $\eta^{p,q}$.

Remark 3.3. The central limit theorems supporting Lemma 2.1 apply to certain non-ARMA models as well, as the references given above (2.V) describe. Also, it will be seen in the proofs in Appendix 1 that, for establishing (3.7) and (3.9-11), the only use made of the assumption that the coefficients $d[\eta](m)$ in (2.11) come from an ARMA model is to verify the decay rate condition (A1.7). This condition, in turn, is used only to establish that the (m,n)-entries of the matrices defining the quadratic forms appearing in $L_N[\eta]$ and its first-through-third derivatives are of order $O[(\delta-)^{|m-n|}]$. Thus this property and (2.V - VI) imply (3.1).

Remark 3.4. If η is so chosen that it parametrizes only the coefficients $d[\eta](m)$, and $\Sigma[\eta]$ is a constant matrix, Σ_0, then (3.1) becomes an expression for minus half the asymptotic bias of the positive definite quadratic form $x^{NT}d^{NT}[\eta^N]\ \Sigma_0^{-N} d^N[\eta^N]x^N$ as an estimator of $\mathrm{tr}\Gamma^N d^{NT}[\eta^N]\ \Sigma_0^{-N} d^N[\eta^N]$ (see Example 3.2 below).

Remark 3.5. As was indicated before (2.5), Hosoya and Taniguchi (1982) have established a very general central limit theorem for $N^{1/2}(\tilde{\eta}^N - \eta^{p,q})$, where $\tilde{\eta}^N$ maximizes Whittle's likelihood,

$$\tilde{L}_N[\eta] = -\frac{1}{2}\log\det 2\pi\Sigma[\eta] - \frac{N}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr} I_N(\lambda) f^{-1}[\eta](\lambda)d\lambda \qquad (3.12),$$

where $I_N(\ )$ denotes the periodogram of $x(1),\ldots,x(N)$. Also, the optimality investigation of AIC by Taniguchi (1980) is based on properties of quadratic forms in $N^{1/2}(\tilde{\eta}^N - \eta^{p,q})$. Obvious modifications of our proof of (3.1), with $\hat{\eta}^N - \eta^N$ replaced by $\tilde{\eta}^N - \eta^{p,q}$, yield, in place of (3.1),

$$\lim_{N\to\infty} E\{NW[\tilde{\eta}^N] - \tilde{L}_N[\tilde{\eta}^N]\} =$$

$$-\ \mathrm{tr}\ F^{-1}[\eta^{p,q}]\{G[\eta^{p,q}] + H[\eta^{p,q}]\}$$

$$+\sum_{m=-\infty}^{\infty} \mathrm{tr}\Gamma(|m|)\sum_{n=0}^{\infty} (n+|m|)d^T[\eta^{p,q}](n+|m|)\Sigma^{-1}[\eta^{p,q}]d[\eta^{p,q}](n) \qquad (3.13),$$

the extra term being the asymptotic mean of the expression $NW[\eta^{p,q}] - \tilde{L}_N[\eta^{p,q}]$, in contrast to (3.3). However, it would seem more natural to use $\tilde{L}_N[\tilde{\eta}^N]$ as an estimate of $E_N[\tilde{\eta}^N]$, rather than of $NW[\tilde{\eta}^N]$ (assuming that the analogue of (2.V) can be established for $\tilde{\eta}^N$). That is, (3.1) seems more fundamental than (3.13).

Our main applications of (3.1) come in sections 4 and 5, but the following examples help to illustrate its scope.

Example 3.1. Consider the fitting to univariate observations $x(1),\ldots,x(N)$ of an AR(p) model whose first p-1 coefficients are constrained to be zero, by choosing $\eta = (\xi,\theta)$ to maximize

$$L_N[\eta] = -\frac{N}{2}\log 2\pi\xi - \frac{1}{2\xi}\{\sum_{t=1}^{p} x^2(t) + \sum_{t=p+1}^{N}(x(t) - \theta x(t-p))^2\}.$$

This is a special case of a procedure approximating that used by Gersch and Kitagawa (1982) to obtain a model for making p-step-ahead forecasts. $\hat{\theta}^N$ and $\hat{\xi}^N$ are easily calculated: $\hat{\theta}^N$ coincides with the estimator of the autocorrelation at lag p, $\rho(p)$, given by $r^N(p) = \sum_{t=p+1}^{N} x(t)x(t-p)/\sum_{t=1}^{N-p} x^2(t)$, and

$$\hat{\xi}^N = N^{-1}\{\sum_{t=1}^{p} x^2(t) + \sum_{t=p+1}^{N}(x(t) - \hat{\theta}^N x(t-p))^2\}.$$

We find that $\theta^{p,0} = \rho(p)$, this being the quantity which minimizes $\int_{-\pi}^{\pi} |1-\theta e^{ip\lambda}|^2 f(\lambda)d\lambda$, the variance of the p-step-ahead prediction error when $\theta x(t-p)$ is used to predict $x(t)$. This form of predictor is optimal when $x(t)$ is a gaussian AR(1) process. Even in this AR(1) case, however, it follows from (3.14) below that for the log likelihood, $L_N[n]$, of this example, $L_N[\hat{n}^N] - 2$ is not an asymptotically unbiased estimator of $E_N[\hat{n}^N]$ when $p \geq 2$. Thus the term penalizing for the number of parameters estimated in the criterion of Gersch and Kitagawa for selecting models for p-step-ahead prediction ($p \geq 2$) cannot be viewed as an asymptotic bias correction. (Shibata (1980) and Taniguchi (1980) show that, for series satisfying (2.3), certain classes of asymptotically <u>biased</u> AIC-like criteria share the same optimality properties as AIC.)

Evaluating both sides of minus two times (3.1) for this example, we get

$$\lim_{N \to \infty} NE \left\{ \frac{p\Gamma(0) + (N-p) \int_{-\pi}^{\pi} |1-r^N(p)e^{ip\lambda}|^2 f(\lambda)d\lambda}{\sum_{t=1}^{p} x^2(t) + \sum_{t=p+1}^{N} (x(t) - r^N(p)x(t-p))^2} - 1 \right\}$$

$$= 4\pi(\xi^{p,0})^{-2} \int_{-\pi}^{\pi} |1 - \rho(p)e^{ip\lambda}|^4 f^2(\lambda)d\lambda + \frac{\kappa_4}{\sigma^4}$$

$$+ 8\pi(\Gamma(0)\xi^{p,0})^{-1} \int_{-\pi}^{\pi} (\rho(p) - \cos p\lambda)^2 f^2(\lambda)d\lambda \qquad (3.14),$$

where σ^2 is the innovations variance of $x(t)$, and $\xi^{p,0} = \int_{-\pi}^{\pi} |1 - \rho(p)e^{ip\lambda}|^2 f(\lambda)d\lambda$.
It can be argued that the terms $p\Gamma(0)$ and $\sum_{t=1}^{p} x^2(t)$ can be simultaneously eliminated from the left hand side of (3.14) without affecting the limit.
If $f(\lambda) = (\sigma^2/2\pi)|1-\theta e^{i\lambda}|^{-2}$, the bias formula (3.14) yields the values given in Table 3.1 below for the asymptotic bias of $-2L_N[\hat{\eta}^N]$ as an estimator of $-2E_N[\hat{\eta}^N]$.

Table 3.1. Values of (3.14) When x_t is a Gaussian AR(1) Process with Autoregressive Coefficient θ.

$\lvert\theta\rvert \backslash p$	1	2	3	4	5	6	7	8	9	∞
0.1	4.00	4.69	4.86	4.88	4.89	4.89	4.89	4.89	4.89	4.89
0.2	4.00	5.22	5.76	5.93	5.98	6.00	6.00	6.00	6.00	6.00
0.3	4.00	5.63	6.62	7.10	7.30	7.38	7.41	7.42	7.43	7.43
0.4	4.00	5.96	7.40	8.30	8.80	9.07	9.21	9.28	9.31	9.33
0.5	4.00	6.22	8.08	9.46	10.41	11.04	11.43	11.67	11.81	12.00
0.6	4.00	6.44	8.67	10.53	12.01	13.14	14.00	14.61	15.06	16.00
0.7	4.00	6.62	9.16	11.47	13.49	15.24	16.71	17.94	18.95	22.67
0.8	4.00	6.77	9.58	12.27	14.80	17.15	19.32	21.30	23.11	36.00
0.9	4.00	6.89	9.93	12.95	15.90	18.79	21.60	24.32	26.96	76.00

The fact that the range of values of (3.14) increases with p suggests the possibility that the problem of selecting expressions for prediction with lags $p>1$ increases in difficulty with increasing p.

Example 3.2. Following Remark 3.4, we set $\xi = 1$ in the log likelihood function of Example 3.1 to investigate the quadratic form appearing in $L_N[\eta]$. For the resulting log likelihood, (3.1) yields

$$\lim_{N\to\infty} E\{(N-p)\int_{-\pi}^{\pi}|1 - r^N(p)e^{ip\lambda}|^2 f(\lambda)d\lambda - \sum_{t=p+1}^{N}(x(t) - r^N(p)x(t-p))^2\}$$

$$= 8\pi(\Gamma(0))^{-1}\int_{-\pi}^{\pi}(\rho(p) - \cos p\lambda)^2 f^2(\lambda)d\lambda,$$

which, as (2.V) shows, is $2\Gamma(0)$ times the asymptotic variance of $N^{1/2}(r^N(p) - \rho(p))$, see also Anderson (1971, p. 489).

Example 3.3. By Remark 3.2, the same bias formulas hold if θ in the likelihood functions of the preceding examples is replaced by its p-th power θ^p, provided either that p is odd or that $\rho(p)>0$ if p is even. These likelihoods would arise from the fitting of a predictor of the form $\theta^p x(t)$ for $x(t+p)$ (such as an AR(1) model for $x(t)$ would produce). In this case $\theta^{p,0} = \{\rho(p)\}^{1/p}$. (If p is even and $\rho(p)<0$, then $\theta^{p,0} = 0$ and (2.IViv) fails.).

4. EVALUATION OF THE ASYMPTOTIC BIAS WHEN THE CORRECT MODEL BELONGS TO THE CLASS OF MODELS BEING FIT.

In this section we assume, in addition to (2.I-VI), that

$$f(\lambda) = f[\eta^{p,q}](\lambda) \qquad (-\pi\leq\lambda\leq\pi) \qquad (4.1),$$

so that the observed series $x(t)$ is an ARMA(p_0,q_0) process with $p_0 \leq p$ and $q_0 \leq q$ (see Remark 2.1). When (4.1) holds, we denote the matrices $F[\eta^{p,q}]$, $G[\eta^{p,q}]$ and $H[\eta^{p,q}]$ of (2.17-9) by $\tilde{F}[\eta^{p,q}]$, $\tilde{G}[\eta^{p,q}]$ and $\tilde{H}[\eta^{p,q}]$.

Using the formulas $\{\partial/\partial\eta\}\log \det f[\eta](\lambda) = \operatorname{tr} f^{-1}[\eta](\lambda)\{\partial f/\partial\eta\}[\eta](\lambda)$ and $\{\partial/\partial\eta\} f^{-1}[\eta](\lambda) = -f^{-1}[\eta](\lambda)\{\partial f/\partial\eta\}[\eta](\lambda) f^{-1}[\eta](\lambda)$ (see Dwyer (1967)), one verifies easily that

$$
\begin{aligned}
&\operatorname{tr} f[\eta](\lambda)\{\partial^2 f^{-1}/\partial\eta_j\partial\eta_k\}[\eta](\lambda) + \{\partial^2/\partial\eta_j\partial\eta_k\}\log \det f[\eta](\lambda) \\
&= \operatorname{tr} f[\eta](\lambda)\{\partial f^{-1}/\partial\eta_j\}[\eta](\lambda) f[\eta](\lambda)\{\partial f^{-1}/\partial\eta_k\}[\eta](\lambda) \\
&= \operatorname{tr} f^{-1}[\eta](\lambda)\{\partial f/\partial\eta_j\}[\eta](\lambda) f^{-1}[\eta](\lambda)\{\partial f/\partial\eta_k\}[\eta](\lambda) \\
&= \operatorname{tr} f^{-1}[\eta](\lambda)\{\partial^2 f/\partial\eta_j\partial\eta_k\}[\eta](\lambda) + \{\partial^2/\partial\eta_j\partial\eta_k\}\log \det f^{-1}[\eta](\lambda) \\
&\qquad (1 \leq j,k \leq \dim\eta) \qquad (4.2).
\end{aligned}
$$

Integrating the first equation in (4.2) over $-\pi \leq \lambda \leq \pi$ and setting $\eta = \eta^{p,q}$, we obtain immediately from (2.17-8) that

$$\tilde{F}[\eta^{p,q}] = \tilde{G}[\eta^{p,q}] \qquad (4.3).$$

If the fourth cumulants of $e(t)$ vanish,

$$\kappa_{abcd} = 0 \quad (1 \leq a,b,c,d \leq r) \tag{4.4},$$

as they do when $x(t)$ is gaussian, then $\tilde{H}[\eta^{p,q}]=0$, and (3.1), (3.14) and (4.3) yield

<u>Proposition 4.1</u>. <u>Suppose that</u> (4.1) <u>and</u> (4.4) <u>hold</u>. <u>Then for the maximum likelihood model specified by</u> $\hat{\eta}^N$, <u>we have</u>

$$\lim_{N \to \infty} E\{E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N]\} = -\dim\eta \tag{4.5}.$$

<u>For the model specified by</u> $\tilde{\eta}^N$ <u>maximizing Whittle's likelihood</u> (3.12), <u>we have</u>

$$\lim_{N \to \infty} E\{NW[\tilde{\eta}^N] - \tilde{L}[\tilde{\eta}^N]\} =$$

$$-\dim\eta + \sum_{m=-\infty}^{\infty} \operatorname{tr} \Gamma(|m|) \sum_{n=0}^{\infty} (n+|m|)d^T(n+|m|)\Sigma^{-1}d(n) \tag{4.6}.$$

Suppose that a gaussian ARMA(P,Q) model parametrized by ζ is also fit to $x(1),\ldots,x(N)$ by maximizing $L_N[\zeta]$ or $\tilde{L}_N[\zeta]$, that $\hat{\zeta}^N$ and $\tilde{\zeta}^N$ are the respective maximizing parameter vectors, and that the analogues of the assumptions made for the η-models are satisfied by the ζ-models, including $f(\lambda) = f[\zeta^{P,Q}](\lambda)$, where $\zeta^{P,Q}$ maximizes $W[\zeta]$. Recall from section 1 that, for model selection, the sign of $E_N[\hat{\zeta}^N] - E_N[\hat{\eta}^N]$, or, by analogous reasoning, of $W[\tilde{\zeta}^N] - W[\tilde{\eta}^N]$, can be used to suggest which model to prefer. By subtracting (4.5) and (4.6) from their analogues for the ζ-models, we obtain

Corollary 4.1. *Under the above assumptions, including* (4.4),

$$L_N[\hat{\zeta}^N] - L_N[\hat{\eta}^N] - \{\dim\zeta - \dim\eta\} \tag{4.7}$$

is an asymptotically unbiased estimator of $E_N[\hat{\zeta}^N] - E_N[\hat{\eta}^N]$, *and*

$$\tilde{L}_N[\tilde{\zeta}^N] - \tilde{L}_N[\tilde{\eta}^N] - \{\dim\zeta - \dim\eta\} \tag{4.8}$$

is an asymptotically unbiased estimator of $N\{W[\tilde{\zeta}^N] - W[\tilde{\eta}^N]\}$.

This corollary shows that Akaike's bias correction has the desired asymptotic property under (4.4) even when the parameters of the model are subjected to complicated constraint conditions like those which arise with uniformly sampled data from continuous autoregressions or in ARMA data combined with independent additive observation errors. We included the $\tilde{\eta}^N$-results above because (2.V) has not yet been verified for $N^{1/2}(\hat{\eta}^N - \eta^N)$ in this general-constraint situation, whereas $N^{1/2}(\tilde{\eta}^N - \eta^{p,q})$ has been shown to have the requisite limiting distribution, as we discussed in section 2. For the remainder of the paper, the parametrizations will be one's for which (2.V) has been verified, and we will not present results for $NW[\tilde{\eta}^N] - \tilde{L}_N[\tilde{\eta}^N]$, these being obvious analogues of those we give for $E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N]$.

To generalize Corollary 4.1 to situations in which (4.4) fails, we require the innovations covariance matrix to be parametrized separately from the coefficients: we specify that $\eta = (\xi^T, \theta^T)^T$, that

$$f[\eta](\lambda) = C[\theta](e^{i\lambda})\Sigma^{-1}[\xi]C^*[\theta](e^{i\lambda}) \tag{4.9}$$

and, for simplicity's sake, also that ξ be the column vector of length $r(r+1)/2$ whose entries define consecutively the on- and above-diagonal entries of $\Sigma[\xi]$, according to the lexicographical ordering of the (row, column) indices,

$$\Sigma[\xi] = \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_{r(r-1)/2+1} \\ \cdot & \xi_3 & \cdots & \xi_{r(r-1)/2+2} \\ \vdots & & \cdots & \vdots \\ \cdot & & \cdots & \xi_{r(r+1)/2} \end{bmatrix} .$$

From (4.9) we calculate that

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}[\partial^2/\partial\xi\partial\xi^T\{\mathrm{tr} f[\eta^{p,q}]f^{-1}[\eta](\lambda)+\log\det f[\eta](\lambda)\}]_{\eta=\eta^{p,q}}d\lambda$$

$$= \frac{1}{2}[\partial^2/\partial\xi\partial\xi^T\{\mathrm{tr}\Sigma[\xi^{p,q}]\Sigma^{-1}[\xi]+\log\det\Sigma[\xi]\}]_{\xi=\xi^{p,q}} \tag{4.10}$$

(4.10),

this latter expression being the Fisher information matrix for variances and covariances of an r-variate gaussian process whose covariance matrix is $\xi[\eta^{p,q}]$.

From the left hand side of (4.10), it is clear that this is the matrix of order $r(r+1)/2$ in the upper left corner of $F[\eta^{p,q}]$. If we denote this matrix by $\tilde{F}_{(1)}[\eta^{p,q}]$, we obtain from McCulloch (1982) that

$$\tilde{F}_{(1)}[\xi^{p,q}] = \frac{1}{2}K^T\{\Sigma^{-1}[\xi^{p,q}] \otimes \Sigma^{-1}[\xi^{p,q}]\}K \tag{4.11}$$

where K is a full rank matrix of order $r(r+1)/2 \times r^2$ whose entries do not depend on $\Sigma[\xi^{p,q}]$.

The bottom right corner of order $\dim\eta - r(r+1)/2$ of $\tilde{F}[\eta^{p,q}]$ is

$$\tilde{F}_{(2)}[\eta^{p,q}] = \frac{1}{4\pi}\int_{-\pi}^{\pi} [\partial^2/\partial\theta\partial\theta^T\{\mathrm{tr} f[\eta^{p,q}](\lambda)f^{-1}[\eta](\lambda) + \log\det f[\eta](\lambda)\}]_{\eta=\eta^{p,q}}\, d\lambda$$

$$= \frac{1}{4\pi}\int_{-\pi}^{\pi} [\partial^2/\partial\theta\partial\theta^T\{\mathrm{tr} f^{-1}[\eta^{p,q}]f[\eta](\lambda) + \log\det f^{-1}[\eta](\lambda)\}]_{\eta=\eta^{p,q}}\, d\lambda \tag{4.12}$$

where the second equality comes from integrating the first and last expressions in (4.2). A simpler expression for $\tilde{F}_{(2)}[\eta^{p,q}]$ can be obtained in some important special cases. For example, suppose that $x(t)$ is an $AR(p_0)$ process, that an $AR(p)$ model is being fit with $p \geq p_0$, and that $\theta = \mathrm{vec}(a(1),\dots,a(p))$. Then, with $\Gamma^p[\eta^{p,0}]$ denoting the block Toeplitz matrix whose (m,n)-block is $\int_{-\pi}^{\pi} e^{i(j-k)\lambda} f[\eta^{p,0}](\lambda)d\lambda$ ($= Ex(t-k)x^T(t-j)$ by (4.1), but we use the first definition for later reference) $(1 \leq m,n \leq p)$, the first integral in (4.12) is straightforward to evaluate and leads to

$$\tilde{F}_{(2)}[\eta^{p,0}] = \frac{1}{2}\,\Gamma^p[\eta^{p,0}] \otimes \Sigma^{-1}[\xi^{p,0}] \qquad (4.13).$$

Similarly, suppose $x(t)$ is an $MA(q_0)$ process, that an $MA(q)$ model is fit with $q \geq q_0$ and that $\theta = \mathrm{vec}(b(1),\dots,b(q))$. Then, defining the inverse-autocovariance matrix (Cleveland (1972)) at lag $j-k$ by

$$\Gamma_{inv}[\eta^{0,q}](j-k) = (2\pi)^{-2}\int_{-\pi}^{\pi} e^{i(k-j)\lambda} f^{-1}[\eta^{0,q}](\lambda)d\lambda \ (1 \leq j,k \leq q),$$

we obtain from the second integral in (4.12) the block matrix formula

$$\tilde{F}_{(2)}[\eta^{0,q}] = \frac{1}{2}[\Sigma[\eta^{0,q}] \otimes \Gamma_{inv}[\eta^{0,q}](j-k)]_{1 \leq j,k \leq q} \qquad (4.14).$$

Returning to our general discussion based on (4.9), we now demonstrate that $\tilde{F}[\eta^{p,q}]$ and $\tilde{G}[\eta^{p,q}]$ are block diagonal,

$$\tilde{F}[\eta_{p,q}] = \begin{bmatrix} \tilde{F}_{(1)}[\eta^{p,q}] & 0 \\ 0 & \tilde{F}_{(2)}[\eta^{p,q}] \end{bmatrix} = \tilde{G}[\eta^{p,q}] \qquad (4.15).$$

To verify this, observe that for any coordinate θ_k of θ, $\{\partial C/\partial\theta_k\}[\theta](e^{i\lambda}) = \sum_{m=1}^{\infty}\{\partial c/\partial\theta_k\}[\theta](m)e^{im\lambda}$. Since $C^{-1}[\theta](e^{i\lambda}) = \sum_{m=0}^{\infty} d[\theta](m)e^{im\lambda}$, it is clear that for any constant matrix $K_{(0)}$, $K_{(0)}C^{-1}[\theta](e^{i\lambda})\{\partial C/\partial\theta_j\}[\theta](e^{i\lambda})$ has the form $\sum_{m=1}^{\infty} w[\theta](m)e^{im\lambda}$, and hence that its integral over $-\pi\le\lambda\le\pi$ is 0. Therefore, for any ξ_j and θ_k,

$$\int_{-\pi}^{\pi} \mathrm{tr} f[n](\lambda)\{\partial f^{-1}/\partial\xi_j\}[n](\lambda)f[n](\lambda)\{\partial f^{-1}/\partial\theta_k\}[n](\lambda)d\lambda$$

$$= \int_{-\pi}^{\pi} \mathrm{tr}\{\partial\Sigma/\partial\xi_j\}[\xi][\{\Sigma^{-1}[\xi]C^{-1}[\theta](e^{i\lambda})\{\partial C/\partial\theta_k\}[\theta](e^{i\lambda})\}^* + \{\partial C/\partial\theta_k\}[\theta](e^{i\lambda})\Sigma^{-1}[\xi]C^{-1}[\theta](e^{i\lambda})]d\lambda = 0,$$

from which (4.15) follows.

By a similar calculation, it can be demonstrated that (4.9) implies

$$\tilde{H}[\eta^{p,q}] = \begin{bmatrix} \tilde{H}_{(1)}[\xi^{p,q}] & 0 \\ 0 & 0 \end{bmatrix} \tag{4.16}$$

where $\tilde{H}_{(1)}[\xi^{p,q}]$ is the matrix of order $r(r+1)/2$ whose (j,k)-entry is

$$\frac{1}{4}\ \sum_{a,b,c,d=1}^{r}\ \kappa_{abcd}\{\partial\Sigma^{-1}/\partial\xi_j\}_{ab}[\xi^{p,q}]\{\partial\Sigma^{-1}/\partial\xi_k\}_{cd}[\xi^{p,q}] \tag{4.17}$$

((4.16) is due to Hosoya and Taniguichi (1982, p. 138)). In conjunction with (3.1), the formulas (4.15-16) lead immediately to

Proposition 4.2. *For a parametrization of the form* $\eta = (\xi^T,\theta^T)^T$ *as in* (4.9), *and for the maximum likelihood model specified by* $\hat{\eta}^N$, *we have*

$$\lim_{N\to\infty}\ E\{E_N[\hat{\eta}^N] - L_N[\hat{\eta}^N]\} = -\dim\eta - \mathrm{tr}\tilde{F}^{-1}_{(1)}[\eta^{p,q}]\tilde{H}_{(1)}[\eta^{p,q}] \tag{4.18}$$

where $\tilde{F}_{(1)}[\eta^{p,q}]$ *and* $\tilde{H}_{(1)}[\eta^{p,q}]$ *are given by* (4.11) *and* (4.17), *respectively*.

Suppose an ARMA(p,q) model, parametrized by $\eta = (\xi^T,\theta^T)^T$ as in (4.9), and an ARMA(P,Q) model, parametrized by $\zeta = (\xi^T,\omega^T)^T$ with ξ as before, are fit to the same observations. Obviously, $\dim\zeta - \dim\eta = \dim\omega - \dim\theta$, the difference in the number of ARMA coefficient parameters. The analogue of Corollary 4.1, for separately parametrized models fit to data which need not satisfy (4.5), is, therefore,

Corollary 4.2. *For two separately parametrized models specified by maximum likelihood estimates* $\hat{\zeta}^N$ *and* $\hat{\eta}^N$,

$$L_N[\hat{\zeta}^N] - L_N[\hat{\eta}^N] - \{\dim\omega - \dim\theta\} \tag{4.19}$$

is an asymptotically unbiased estimator of $E_N[\hat{\zeta}^N] - E_N[\hat{\eta}^N]$.

Remark 4.1. The propositions and corollaries of this section hold for certain non-ARMA models as well, assuming (4.1) holds. See Remark 3.2.

Remark 4.2. In the context of using the statistics (4.19) in the manner indicated above to select a univariate AR(p) model from models of orders 0 through p_{max}, when the data come from a gaussian AR(p_0) process with $p_0 < p_{max}$, Shibata (1976) showed that as N increases indefinitely, the model selected will have order at least p_0 with probability 1, and order larger than p_0 with non-zero probability depending on P_{max}. Hannan (1980a) generalized Shibata's results to the case of fitted ARMA(p,q) models when $p \geq p_0$, $q \geq q_0$ and either $p = p_0$ or $q = q_0$. The results of Woodroofe (1982) suggest in this case that the probability of overfitting is never larger than 0.288 and that the expected number of overfitted parameters will always be less than 1. (When $p > p_0$ and $q > q_0$, causing (2.IVii) to fail, see Remark 2.1, Hannan (1980b) showed that the minimum AIC procedure can overestimate p_0 and q_0 with probability arbitrarily close to 1, and also that, even in this situation, the one-step-ahead prediction error performance of the minimum AIC model appears to be good.) In Hannan (1980a), modifications proposed by various authors of the term $\{\dim\omega - \dim\theta\}$ in (4.19) are discussed which yield model selection procedures giving consistent estimates of the model order when (4.1) holds.

However, for some possibly more realistic situations discussed in the next section, where it is assumed that the observations do not conform to any ARMA(p,q) model, Shibata (1980, 1981a) and Taniguchi (1980) show that the minimum AIC procedure selects models for prediction or spectrum estimation in an optimal way according to natural loss functions, whereas the consistent procedures lead to unboundedly large loss for certain series. Thus consistency can be an undesirable property in the context of selecting model orders.

5. APPROXIMATING THE ASYMPTOTIC MEAN OF $E_N[\hat{\eta}^N(p,q)] - L_N[\hat{\eta}^N(p,q)]$ WHEN X(t) IS NOT AN ARMA PROCESS.

In this section, we assume that the series x(t) satisfies (2.II) but is <u>not</u> an ARMA process. We also assume that (3.1) holds for all orders (p,q) in a set S with $\max\{p+q: (p,q)\epsilon S\} = \infty$, when the maximum likelihood estimates $\hat{\eta}^N = \hat{\eta}^N(p,q)$ are obtained from separable parametrizations (4.9) and the compact parameter sets $ETA^{p,q}$ have the cartesian product form $\Xi x \Theta^{p,q}$ (with $\eta=(\xi,\theta)$, $\xi\epsilon\Xi, \theta\epsilon\Theta^{p,q}$). We assume the models parametrized by $\Theta^{p,q}$ have a certain uniformity property described in (5.I) below, and a certain comprehensive property (5.II), as well as the properties implied by (2.III), which is assumed to hold. Two additional assumptions, (5.III) and (5.IV) below, also play a role.

Working from such assumptions, we show that the matrices $F[\eta^{p,q}]$, $G[\eta^{p,q}]$ and $H[\eta^{p,q}]$ can be well approximated, as max{p,q} increases, by the respective matrices $\tilde{F}[\eta^{p,q}]$, $\tilde{G}[\eta^{p,q}]$ and $\tilde{H}[\eta^{p,q}]$ obtained by replacing $f(\lambda)$ by $f[\eta^{p,q}](\lambda)$ in their defining expressions. As a consequence, the asymptotic bias formulas of section 4 will be seen to be correct to within $O[(\delta-)^{\max\{p,q\}}]$ (Proposition 5.3). The most complete results are obtained for autoregressive models and moving average models.

Recall from section 2 that $f(\lambda) = (2\pi)^{-1}C(e^{i\lambda})\Sigma C^*(e^{i\lambda})$ and, for any $\eta = (\xi,\theta)$ in $\Xi x \Theta^{p,q}$, that $f[\eta](\lambda) =$ $(2\pi)^{-1}C[\theta](e^{i\lambda})\Sigma[\xi]C^*[\theta](e^{i\lambda})$. Our basic requirements in this section are (5.I-II):

(5.I). <u>The entries of</u> $C[\theta](e^{i\lambda})$, $D[\theta](e^{i\lambda})(=C^{-1}[\theta](e^{i\lambda}))$, $\{\partial C/\partial\theta_j\}[\theta](e^{i\lambda})$ <u>and</u> $\{\partial^2 C/\partial\theta_j\partial\theta_k\}[\theta](e^{i\lambda})$ $(1<j,k<\dim\theta;\theta\varepsilon\Theta^{p,q})$ <u>are uniformly bounded for</u> $(p,q)\ \varepsilon\ S$ <u>and</u> $-\pi<\lambda<\pi$.

(5.II). <u>Let</u> v <u>be any positive integer such that</u> $(v,q)\varepsilon S$ (<u>resp</u>. $(p,v)\varepsilon S$.) <u>For any given</u> r-th <u>order matrix polynomial</u> $P(z)$ <u>of degree</u> v <u>with</u> $P(0)=I$ <u>and such that the roots of</u> $\det P(z)$ <u>belong to</u> $\{|z|>\delta^{-1}\}$, <u>there exists a</u> $\theta\varepsilon\Theta^{v,q}$ (<u>resp</u>. $\Theta^{p,v}$) <u>such that</u> $C[\theta](e^{i\lambda}) = P(e^{i\lambda})$ (<u>resp</u>. $D[\theta](e^{i\lambda}) = P(e^{i\lambda})$).

The condition (5.I) seems reasonable since $|C(e^{i\lambda})|_\infty$ and $|D(e^{i\lambda})|_\infty$ are bounded. The condition (5.II) insures that for v sufficiently large, the v-th partial sums of $C(e^{i\lambda})$ and $D(e^{i\lambda})$ are associated with candidate models.

For each $\theta\varepsilon\Theta^{p,q}$, define

$$\Sigma_{(\theta)} = \int_{-\pi}^{\pi} D[\theta](e^{i\lambda})f(\lambda)D^{*}[\theta](e^{i\lambda})d\lambda.$$

This is the covariance matrix of the one-step-ahead forecast error when x(t) is forecast as though it conformed to the ARMA(p,q) model with coefficients determined by θ. With $\eta^{p,q} = (\xi^{p,q},\theta^{p,q})$ denoting the asymptotic limit of the maximum likelihood estimates, it is shown in Appendix 2 that the matrix relations

$$\Sigma[\xi^{p,q}] = \Sigma_{(\theta^{p,q})} > \Sigma \tag{5.1}$$

hold, and that $\theta^{p,q}$ is uniquely determined by the property

$$\det \Sigma_{(\theta^{p,q})} = \min_{\Theta^{p,q}} \det \Sigma_{(\Theta)} \tag{5.2}.$$

A least squares analogue $_{ls}\theta^{p,q}$ of $\theta^{p,q}$ can be defined by

$$\mathrm{tr}\Sigma_{(_{ls}\theta^{p,q})} \leqslant \min_{\Theta^{p,q}} \mathrm{tr}\Sigma_{(\theta)} \tag{5.3}.$$

If a parameter vector $_{c}\theta^{p,q}$ in $\Theta^{p,q}$ exists with the property that

$$\Sigma_{(_{c}\theta^{p,q})} \leqslant \Sigma_{(\theta)} \quad \text{for all } \theta\varepsilon\Theta^{p,q} \tag{5.4}$$

holds (as happens when $r=1$, of course, and also, for $r>1$, when $q=0$ and the entries of θ are the AR-coefficients), then we have ${}_{C}\theta^{p,q} = {}_{1S}\theta^{p,q} = \theta^{p,q}$, because (5.4) implies the corresponding inequalities for the determinants and traces, and $\theta^{p,q}$ uniquely satisfies (5.2). In general, however, we will not know whether $\theta^{p,q}$ has the property required of ${}_{1S}\theta^{p,q}$ in (5.3), so we shall make the assumption that a constant M exists, independent of (p,q), such that

(5.III). $$\mathrm{tr}\{\Sigma_{(\theta^{p,q})} - \Sigma\} \leqslant M\, \mathrm{tr}\{\Sigma_{({}_{1S}\theta^{p,q})} - \Sigma\}.$$

The following result is established in Appendix 2.

<u>Proposition</u> 5.1. <u>Under</u> (2.II) <u>and</u> (5.I-III), <u>the</u> <u>assertions</u> (5.5-8) <u>below</u> <u>are</u> <u>valid</u>:

$$\mathrm{tr}\{\Sigma[\xi^{p,q}] - \Sigma\} \sim O[(\delta -)^{\max\{2p,2q\}}] \qquad (5.5),$$

$$\int_{-\pi}^{\pi} \mathrm{tr}\{C[\theta^{p,q}] - C\}(e^{i\lambda})\{C[\theta^{p,q}] - C\}^{*}(e^{i\lambda})d\lambda \sim O[(\delta -)^{\max\{2p,2q\}}] \qquad (5.6),$$

$$\int_{-\pi}^{\pi} \mathrm{tr}\{D[\theta^{p,q}] - D\}(e^{i\lambda})\{D[\theta^{p,q}] - D\}^{*}(e^{i\lambda})d\lambda \sim O[(\delta -)^{\max\{2p,2q\}}] \qquad (5.7),$$

$$\int_{-\pi}^{\pi} |\mathrm{tr}\{f[\eta^{p,q}] - f\}(\lambda)|_{\infty} d\lambda \sim O[(\delta -)^{\max\{p,q\}}] \qquad (5.8).$$

<u>Remark 5.1</u>. The examples constructed by Erohin (1959) show that functions $C(e^{i\lambda})$ satisfying (2.II) exist for which the rates of convergence given in (5.5-7) are the best possible. Presumably the same is true of (5.8).

One immediate consequence of (5.8) is

<u>Corollary 5.2</u>. <u>Under the above assumptions, if</u> $g^{p,q}(e^{i\lambda})$ $((p,q)\varepsilon S)$ <u>is any family of continuous</u> rxr <u>matrix functions which is uniformly bounded</u>, i.e.,

$$\sup_{-\pi\leqslant\lambda\leqslant\pi;(p,q)\varepsilon S} |g^{p,q}(e^{i\lambda})|_\infty < \infty \tag{5.9}$$,

<u>then</u>

$$\int_{-\pi}^{\pi} \mathrm{tr}\{f[\eta^{p,q}] - f\}(\lambda)g^{p,q}(e^{i\lambda})d\lambda \sim O[(\delta-)^{\max\{p,q\}}] \tag{5.10}$$,

<u>and</u>

$$\int_{-\pi}^{\pi} |\{C[\theta^{p,q}]-C\}(e^{i\lambda})g^{p,q}(e^{i\lambda})\{C[\theta^{p,q}]-C\}^*(e^{i\lambda})|_\infty d\lambda$$
$$\sim O[(\delta-)^{\max\{p,q\}}] \tag{5.11}$$.

Now we are ready to analyze $F[\eta^{p,q}] - \tilde{F}[\eta^{p,q}]$, $G[\eta^{p,q}] - \tilde{G}[\eta^{p,q}]$ and $H[\eta^{p,q}] - \tilde{H}[\eta^{p,q}]$. We have $F[\eta^{p,q}] - \tilde{F}[\eta^{p,q}] =$

$$\frac{1}{4\pi}\sum_{1\leq j,k\leq r}\int_{-\pi}^{\pi}\{f_{jk}-f_{jk}[\eta^{p,q}]\}(\lambda)\{\partial^2 f^{-1}/\partial\eta\partial\eta^T\}[\eta^{p,q}](\lambda)d\lambda \tag{5.12}$$

It follows from (5.I), (5.1) and (5.5) that the Hessian matrices $\{\partial^2 f_{jk}^{-1}/\partial\eta\partial\eta^T\}[\eta^{p,q}](\lambda)$ $(1\leq j,k\leq r)$ satisfy (5.9). Hence, from (5.10) and (5.12), we can conclude that

$$|F[\eta^{p,q}]-\tilde{F}[\eta^{p,q}]|_\infty \sim O[(\delta-)^{\max\{p,q\}}] \tag{5.13}$$

An analogous calculation yields

$$|G[\eta^{p,q}]-\tilde{G}[\eta^{p,q}]|_\infty \sim O[(\delta-)^{\max\{p,q\}}] \tag{5.14}$$

and similar arguments based on (5.11) and two obvious applications of the matrix identity $A_1A_2-B_1B_2 = (A_1-B_1)(A_2-B_2) + B_1(A_2-B_2) + (A_1-B_1)B_2$ lead to

$$|H[\eta^{p,q}]-\tilde{H}[\eta^{p,q}]|_\infty \sim O[(\delta-)^{\max\{p,q\}}] \tag{5.15}$$

To go on to an analysis of the approximation of the bias expression $\mathrm{tr}F^{-1}[\eta^{p,q}](G[\eta^{p,q}]+H[\eta^{p,q}])$ by $\mathrm{tr}\tilde{F}^{-1}[\eta^{p,q}](\tilde{G}[\eta^{p,q}]+\tilde{H}[\eta^{p,q}])$ we need a property such as

$$\sup_S |\tilde{F}^{-1}[\eta^{p,q}]|_\infty < \infty \tag{5.16}$$

(see also Remark 5.2 below). Since $\Sigma[\eta^{p,q}] > \Sigma$, by (5.1), the formula (4.11) shows that $\sup_S |\tilde{F}^{-1}_{(1)}[\eta^{p,q}]|_\infty < \infty$. By (4.15) therefore, (5.16) is equivalent to

$$\text{(5.IV).} \qquad \sup_S |\tilde{F}^{-1}_{(2)}[\eta^{p,q}]|_\infty < \infty \quad .$$

For a univariate example of an approximating autoregressive-moving average modeling situation in which (5.IV) appears to be satisfied, see Example 1 of Taniguchi (1980). The next proposition, whose proof is given in Appendix 2, shows that (5.IV) is satisfied by approximating pure autoregressive and pure moving average models.

Proposition 5.2. *The matrices* $\tilde{F}_{(2)}[\eta^{p,0}]$ *and* $\tilde{F}_{(2)}[\eta^{0,q}]$ *given by* (4.13-4) *are such that*

$$\sup_{1 \leqslant p < \infty} \quad |\tilde{F}^{-1}_{(2)}[\eta^{p,0}]|_\infty < \infty \tag{5.17}$$

and

$$\sup_{1 \leqslant q < \infty} \quad |\tilde{F}^{-1}_{(2)}[\eta^{0,q}]|_\infty < \infty \tag{5.18}$$

Now we establish the main result of this section.

Proposition 5.3. Suppose that (5.I-II) hold and that the models being fit to the observed series $x(t)$ are described by either (a), (b) or (c) below:

(a) Autoregressive models in which the θ-parameter coordinates are the coefficients themselves or a one-to-one, twice continuously differentiable transformation thereof;

(b) Moving average models satisfying (5.III) in which the θ-parameter coordinates are the coefficients themselves or a one-to-one, twice continuously differentiable transformation thereof.

(c) Autoregressive-moving average models such that (5.III-IV) are satisfied.

Then, with $\dim\theta(p,q)$ denoting the dimension of the coordinate vectors in $\Theta^{p,q}$, the asymptotic error bounds (5.19-20) below are valid:

$$|\mathrm{tr}F^{-1}[\eta^{p,q}]G[\eta^{p,q}] - \{r(r+1)/2 + \dim\theta(p,q)\}| \sim O[(\delta-)^{\max\{p,q\}}] \tag{5.19}$$,

$$|\mathrm{tr}F^{-1}[\eta^{p,q}]H[\eta^{p,q}] - \mathrm{tr}F_{(1)}^{-1}H_{(1)}| \sim O[(\delta-)^{\max\{p,q\}}]| \tag{5.20}$$,

where $F_{(1)} = \frac{1}{2} K^T\{\Sigma^{-1} \otimes \Sigma^{-1}\}K$ with K as in (4.11), and where the (j,k)-entry of $H_{(1)}$ is given by

$$\sum_{a,b,c,d=1}^{r} \kappa_{abcd}\{\partial\Sigma^{-1}/\partial\xi_j\}_{ab}[\xi^{true}]\{\partial\Sigma^{-1}/\partial\xi_k\}_{cd}[\xi^{true}] \qquad (5.21),$$

with $\Sigma[\xi^{true}]=\Sigma$, so that $\mathrm{tr}F_{(1)}^{-1}H_{(1)}$ is independent of (p,q).

Consequently, $\mathrm{tr}F^{-1}[\eta^{p,q}](G[\eta^{p,q}]+H[\eta^{p,q}]) - \dim\theta(p,q)$ approaches a constant value exponentially rapidly as $\max\{p,q\}$ increases.

Proof. By (4.3) and (4.15-6), the assertions (5.19-20) are equivalent to the two assertions

$$|\mathrm{tr}F^{-1}[\eta^{p,q}]G[\eta^{p,q}] - \mathrm{tr}\tilde{F}^{-1}[\eta^{p,q}]\tilde{G}[\eta^{p,q}]| \sim O[(\delta-)^{\max\{p,q\}}],$$

and

$$|\mathrm{tr}F^{-1}[\eta^{p,q}]H[\eta^{p,q}] - \mathrm{tr}\tilde{F}^{-1}[\eta^{p,q}]\tilde{H}[\eta^{p,q}]| \sim O[(\delta-)^{\max\{p,q\}}].$$

These, in turn, are immediate consequences of (5.13-5) and

$$|F^{-1}[\eta^{p,q}] - \tilde{F}^{-1}[\eta^{p,q}]|_\infty \sim O[(\delta-)^{\max\{p,q\}}] \qquad (5.22),$$

which we shall now verify under the assumption that (5.IV) holds. We start from the equation

$$F^{-1}[\eta p,q] =$$

$$\{I + \tilde{F}^{-1}[\eta p,q](F[\eta p,q] - \tilde{F}[\eta p,q])\}^{-1}\tilde{F}^{-1}[\eta p,q] \qquad (5.23),$$

where I denotes the identity matrix of appropriate order. It follows from (5.13) and (5.16) that $|\tilde{F}^{-1}[\eta p,q](F[\eta p,q] - \tilde{F}[\eta p,q])|_\infty$ is of order $(\delta-)^{\max\{p,q\}}$, which makes it possible to justify the expansion

$$\{I + \tilde{F}^{-1}[\eta p,q](F[\eta p,q] - \tilde{F}p,q])\}^{-1} - I = \sum_{k=1}^{\infty} (-1)^k\{\tilde{F}^{-1}[\eta p,q](F[\eta p,q] - \tilde{F}[\eta p,q])\}^k \qquad (5.24)$$

and to verify that the entries on the left in (5.24) are uniformly of order $(\delta-)^{\max\{p,q\}}$. After right multiplication by $\tilde{F}^{-1}[\eta p,q]$, this left hand side becomes $F^{-1}[\eta p,q] - \tilde{F}^{-1}[\eta p,q]$, by (5.23). Since the entries of $\tilde{F}[\ p,q]$, $(p,q)\varepsilon S$, are bounded, by (5.I) and (5.1), the assertion (5.22) follows. Proposition (5.2) and Remark 3.2 show that (5.IV) holds in the situations (a) and (b), as well as in (c), so the proof is complete.

Remark 5.2. Using the fact that, for any fixed positive interger m, $(p+q)^m(\delta-)^{\max\{p,q\}}$ is of order $(\delta-)^{\max\{p,q\}}$, it is easy to see that (5.19-20) still hold in situation (c) of Proposition 5.3 when the boundedness condition (5.IV) is replaced by one permitting growth of order $(p+q)^m$. In fact, if $|\tilde{F}^{(-1)}[\eta p,q]|_\infty$ grows not faster than $(1/\delta_0)^{\max\{p,q\}}$

with $\delta<\delta_0<1$, then (5.19-20) continue to hold if the asserted decay rate, $\delta-$, is replaced by $(\delta/\delta_0)-$.

Remark 5.3. The preceding remark and Proposition (4.1) together suggest that, in the gaussian case and under the assumptions of this section, the asymptotic bias of $L_N[\hat{\eta}^N] - \{r(r+1)/2 + \dim\theta(p,q)\}$ as an estimate of $E_N[\hat{\eta}^N]$ can quite generally be expected to decay exponentially rapidly as $\max\{p,q\}$ increases and so be negligible, as Akaike proposed, for moderate values of $\max\{p,q\}$ and sufficiently large N. Similarly, based on (5.19-20) and Corollary 4.2, Akaike's use of $L_N(\hat{\xi}^N) - L_N(\hat{\eta}^N) - \{\dim\omega - \dim\theta\}$ as a bias-corrected estimate of $E_N[\hat{\xi}^N] - E_N[\hat{\eta}^N]$ is also justified for certain classes of non-gaussian series, if N is large enough and $\max\{p,q\}$ is not too small.

APPENDIX I. PROOFS OF (2.15), (3.7) and (3.9-11).

We shall have need of two lemmas.

Lemma A1.1. *Suppose* $R^N = [R^N(m,n)]_{1\leq m,n\leq N}$ *and* $S^N = [S^N(m,n)]_{1\leq m,n\leq N}$ *are block-partitioned matrices of order* Nr, for N=1,2,..., *whose blocks (of order* r) *satisfy*

$$\max\{|R^N(m,n)|_\infty, |S^N(m,n)|_\infty\} \leq M_{(\delta-)}(\delta-)^{|m-n|} \qquad (A1.1),$$

for $1 \leq m,n \leq N$ and for some constant $M_{(\delta-)}$ which does not depend on m,n or N. Then a similar constant $\tilde{M}_{(\delta-)}$ exists for which the blocks of order r of the product matrix $[T^N(m,n)]_{1 \leq m,n \leq N} = R^N S^N$ satisfy

$$|T^N(m,n)|_\infty \leq \tilde{M}_{(\delta-)}(\delta-)^{|m-n|} \qquad \text{(A1.2)}$$

for $1 \leq m,n \leq N$ and $N=1,2,\ldots$.

Proof. The meaning of (A1.1) is that, for some $\delta_0 < \delta$ and some $M_{(\delta_0)}$, the magnitudes of the entries of the (m,n)-blocks of R^N and S^N are bounded above by $M_{(\delta_0)}\delta_0{}^{|m-n|}$, for $1 \leq m,n \leq N$ and $N=1,2,\ldots$. A straightforward calculation shows that the quantities $|T^N(m,n)|_\infty$ are therefore bounded above by $M^2_{(\delta_0)}|m-n|\delta_0{}^{|m-n|}$, and thus also by $M_{(\delta_1)}\delta_1{}^{|m-n|}$ for some $M_{(\delta_1)}$, if $\delta_0 < \delta_1 < \delta$. Hence (A1.2) holds.

Lemma A1.2. Suppose that for each η in $ETA^{p,q}$, a matrix function $\Phi[\eta](z) = [\pi_{jk}[\eta](z)/\psi_{jk}[\eta](z)]_{1 \leq j,k \leq r}$ is defined, where the $\pi_{jk}[\eta](z)$ and $\psi_{jk}[\eta](z)$ are polynomials in z of degree not exceeding d_0, whose coefficients, as functions of η, are three-times continuously differentiable. Assume that the zeros of $\psi_{jk}[\eta](z)$ belong to $\{|z|>(\delta-)^{-1}\}$, for all $1 \leq j,k \leq r$, so that the formula (A1.3) is valid,

$$\Phi[\eta](z) = \sum_{m=0}^{\infty} \phi[\eta](m) z^m \qquad (|z| \leq (\delta-)^{-1}) \qquad \text{(A1.3)}.$$

Then, for any differential operator, $\mathcal{D}^J$, *in the coordinates of* η *having order* J, $0 \leq J \leq 3$, *the entries of the matrices* $\mathcal{D}^J \phi[\eta](m)$ $(m=0,1,\ldots)$ *are such that*

$$\sup_\eta |\mathcal{D}^J \phi[\eta](m)|_\infty \sim O[(\delta-)^m] \qquad (A1.4)$$

holds.

Proof. It follows easily from the formula (see Titchmarsh (1939, p. 90)) $\phi[\eta](m) = (2\pi i)^{-1} \int_{|z|=\delta-} z^{-m-1}\, \Phi[\eta](z)dz$ that the entries in the coefficient matrices $\phi[\eta](m)$ are three times continuously differentiable and also that

$$\mathcal{D}^J \Phi[\eta](z) = \sum_{m=0}^{\infty} \mathcal{D}^J \phi[\eta](m) z^m \qquad (A1.5).$$

Clearly, $\mathcal{D}^J \Phi_{jk}[\eta](z) = \mathcal{D}^J \{\pi_{jk}[\eta](z)/\psi_{jk}[\eta](z)\} = \{\psi_{jk}[\eta](z)\}^{J+1} \mu_{jk}[\eta](z)$, where $\mu_{jk}[\eta](z)$ is a polynomial in z of degree not exceeding Jd_0 whose coefficients are continuous functions of η. Since each $\psi_{jk}[\eta](z)$ is bounded away from 0 on $\{|z| \leq (\delta-)^{-1}\}$, it follows from the compactness of $ETA^{p,q}$ that a constant $M_{(\delta-)}$ exists such that

$$\sup_\eta |\mathcal{D}^J \Phi[\eta](z)|_\infty \leq M_{(\delta-)} \qquad (A1.6).$$

The assertion (A1.3) is an immediate consequence of (A1.5-6) and Cauchy's inequality (Titchmarsh (1939, p. 84)).

An important consequence of Lemma A1.2 and (2.III) is that, in addition to (2.7), we also have

$$|\mathcal{D}^J d[\eta](m)|_\infty \sim O[(\delta-)^m] \tag{A1.7}$$

for any differential operator $\mathcal{D}^J$, in the coordinates of η, of order J not exceeding 3. Thus, if we partition the matrix $d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]$ of (2.13-4) into N blocks of order r, denoting the (m,n)-block by $\{d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]\}(m,n)$, it follows from (A1.7) and Lemma A1.1 that a constant $M_{(\delta-)}$ exists such that

$$\sup_\eta|\mathcal{D}^J\{d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]\}(m,n)|_\infty \leq M_{(\delta-)}(\delta-)^{|m-n|} \tag{A1.8}$$

holds for all $1\leq m,n\leq N$, $N=1,2,\ldots$.

Now we establish the following basic result connecting $E_N[\eta]$ of (2.13) with $W[\eta]$ of (2.16).

<u>Proposition A1.1</u>. <u>For</u> $\mathcal{D}^J$ <u>as</u> <u>in</u> (A1.7), <u>the</u> <u>limit</u>

$$\lim_{N \to \infty} \{N \mathcal{E}^J W[\eta] - \mathcal{E}^J E_N[\eta]\} =$$

$$-1/2 \sum_{m=-\infty}^{\infty} \mathrm{tr}\Gamma(|m|) \sum_{n=0}^{\infty} (n+|m|) \mathcal{E}^J \{d^T[\eta](n+|m|)\Sigma^{-1}[\eta]d[\eta](n)\} \tag{A1.9}$$

holds uniformly for η in $\mathrm{ETA}^{p,q}$.

Proof. First, aided by the trace properties $\mathrm{tr}A_1A_2 = \mathrm{tr}A_2A_1$ and $\mathrm{tr}A = \mathrm{tr}A^T$, we directly calculate

$$\mathcal{E}^J \mathrm{tr}\Gamma^N d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta] =$$

$$\sum_{t=1}^{N} \sum_{m,n=0}^{t-1} \mathrm{tr}\Gamma(m-n) \mathcal{E}^J \{d^T[\eta](m)\Sigma^{-1}[\eta]d[\eta](n)\} =$$

$$\sum_{m=-N+1}^{N-1} \mathrm{tr}\Gamma(|m|)[\sum_{n=0}^{N-1-|m|} (N-n-|m|) \mathcal{E}^J \{d[\eta](n+|m|)\Sigma^{-1}[\eta]d[\eta](n)\}] \tag{A1.10}.$$

From Parseval's formula and the same trace properties, we obtain

$$\frac{N}{2\pi} \mathcal{E}^J \int_{-\pi}^{\pi} \mathrm{tr}f(\lambda)f^{-1}[\eta](\lambda)d\lambda =$$

$$\sum_{m=-\infty}^{\infty} \mathrm{tr}\Gamma(|m|) \sum_{n=0}^{\infty} N \mathcal{E}^J \{d^T[\eta](n+|m|)\Sigma^{-1}[\eta]d[\eta](n)\} \tag{A1.11}.$$

Also, as a consequence of (A1.7), a constant $\tilde{M}_{(\delta-)}$ exists such that

$$\sup_\eta |\mathcal{D}^J\{d^T[\eta](n+|m|)\Sigma^{-1}[\eta]d[\eta](n)\}|_\infty \leq \tilde{M}_{(\delta-)}(\delta-)^{2n+|m|} \qquad \text{(A1.12)}$$

holds for $0\leq|m|,n<\infty$. Since minus twice $N\mathcal{D}^J W[\eta] - \mathcal{D}^J E_N[\eta]$ is equal to (A1.11) minus (A1.10), the assertion of the proposition follows from (A1.12) by a straightforward calculation.

The assertion (3.7) is an immediate consequence of (A1.9) with J=2 and the fact (see Remark 2.2) that η^N approaches $\eta^{p,q}$ as N becomes infinite.

PROOF OF (3.9)

Applying (A1.8) with J=3, we observe that a constant M exists which is an upper bound for the magnitudes of the third derivatives of the entries of $N^{-1}E_N[\eta]$ for all η in $ETA^{p,q}$ and all N. Therefore, an application of the mean value theorem leads to

$$|N^{-1}E_N''[\eta^{N*}] - N^{-1}E_N''[\eta^N]|_\infty \leq M \sum_{j=1}^{\dim\eta} |\eta_j^{N*} - \eta_j^N| \qquad \text{(A1.13)},$$

where η^{N*} is as in (3.9). If we set $\Delta^N = N^{-1}E''_N[\eta^{N*}] - N^{-1}E''_N[\eta^N]$, then from (A1.13) and the triangle inequality for the norm $\|\cdot\|_{1+\alpha^{-1}}$, we get

$$\sup_{1 \leq j,k \leq \dim\eta} \|\Delta^N_{jk}\|_{1+\alpha^{-1}} \leq M \sum_{j=1}^{\dim\eta} \|\eta^{N*}_j - \eta^N_j\|_{1+\alpha^{-1}}.$$

Since $\|\eta^{N*}_j - \eta^N_j\|_{1+\alpha^{-1}} \leq \|\hat{\eta}^N_j - \eta^N_j\|_{1+\alpha^{-1}}$ $(1 \leq j \leq \dim\eta)$, it follows from (2.21) that

$$\lim_{N \to \infty} \sup_{1 \leq j,k \leq \dim\eta} \|\Delta^N_{jk}\|_{1+\alpha^{-1}} = 0 \qquad \text{(A1.14).}$$

Now, if we set $\alpha=1$ and $\mu=1+\alpha$ in (2.23), the resulting expression shows that (3.9) is a consequence of (2.V) and (A1.14).

PROOF OF (3.10)

The proof resembles that of (3.9) in outline, but there are differences detail. To start, we demonstrate that, for α,β as in (2.VI), a constant $M_{(\alpha,\beta)}$ exists such that for all partial derivative operators $\mathcal{D}^3$ of order 3, we have

$$\sup_{N,\eta} \|N^{-1}\mathcal{D}^3 L_N[\eta]\|_{\beta(1+\alpha^{-1})} \leq M_{(\alpha,\beta)} \qquad \text{(A1.15).}$$

The uniform boundedness of $\mathcal{D}^3 \log \det\Sigma[\eta]$ follows from (2.III), so that to demonstrate (A1.15) only the terms contributed by the quadratic form of $N^{-1}L_N[\eta]$ need be examined. Using (A1.8), we obtain the existence of a constant $M_{(\delta-)}$ such that

$$|\mathcal{D}^3\{N^{-1}x^{NT}d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]x^N\}| =$$

$$N^{-1}|\sum_{t=1}^{N}\sum_{m,n=0}^{t-1} \operatorname{tr} x(t-n)x^T(t-m)\mathcal{D}^3\{d^T[\eta](m)\Sigma^{-1}[\eta]d[\eta^N](n)\}|$$

$$\leq M_{(\delta-)}\, N^{-1}\sum_{t=1}^{N}\sum_{m,n=0}^{t-1} |x(t-n)x^T(t-m)|_\infty(\delta-)^{m+n} \qquad \text{(A1.16)}.$$

If $\tilde{M}_{(\alpha,\beta)}$ denotes the value of the left hand side of (2.25), then by calculating the norms $\|\cdot\|_{\beta(1+\alpha^{-1})}$ of both sides of (A1.16) and using the triangle inequality, we get

$$\|\mathcal{D}^3\{N^{-1}x^{NT}d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]x^N\}\|_{\beta(1+\alpha^{-1})}$$

$$\leq M_{(\delta-)}\,\tilde{M}_{(\alpha,\beta)}\,[(\delta-)\{1-(\delta-)\}^{-1}]^2 ,$$

from which (A1.15) follows.

Now, for each index pair (j,k) $(1\leq j,k\leq \dim\eta)$ and for each $N = 1,2,\ldots,$ the mean value theorem asserts that

$$\{N^{-1}L_N''[\eta^{N**}] - N^{-1}L_N''[\eta^N]\}_{jk}$$

$$= \sum_{i=1}^{\dim\eta} N^{-1}\{\partial^3 L_N/\partial\eta_i\partial\eta_j\partial\eta_k\}[\overline{\eta}^N](\eta_i^{N**} - \eta_i^N) \qquad (A1.17)$$

holds for some $\overline{\eta}^N = \overline{\eta}^N(j,k)$ between η^{N**} and η^N. Let us set $\Delta^N = N^{-1}L_N''[\eta^{N**}] - N^{-1}L_N''[\eta^N]$. If we take $(1+\alpha^{-1})$-norms in (A1.17) and then apply Hölder's inequality to the norms of products on the right, we find that

$$\|\Delta_{jk}^N\|_{1+\alpha^{-1}} \leq \sum_{i=1}^{\dim\eta} \|N^{-1}\{\partial^3 L_N/\partial\eta_i\partial\eta_j\partial\eta_k\}[\overline{\eta}^N]\|_{\beta(1+\alpha^{-1})} \|\eta_i^{N**} - \eta_i^N\|_{\beta(\beta-1)^{-1}(1+\alpha^{-1})}.$$

Thus, from (A1.15),

$$\|\Delta_{jk}^N\|_{1+\alpha^{-1}} < M_{(\alpha,\beta)} \sum_{i=1}^{\dim\eta} \|\eta_i^N - \eta_i^N\|_{\beta(1-\beta)^{-1}(1+\alpha^{-1})},$$

so that $\lim_{N\to\infty} \|\Delta_{jk}^N\|_{1+\alpha^{-1}} = 0$, by (2.21). This fact, (2.VIi) and (2.23), with $\gamma=1$ and $\mu=1+\alpha$, yield (3.10).

PROOF OF (3.11)

We shall prove a slightly stronger result than (3.11), namely that for some constant M and for α as in (2.VI),

$$|NE(\hat{\eta}^N - \eta^N)^T\{N^{-1}L_N''[\eta^N] - N^{-1}E_N''[\eta^N]\}(\hat{\eta}^N - \eta^N)| \leq MN^{-\min\{1/2,(1+3\alpha)/4\}} \quad \text{(A1.18)}$$

holds. First let us describe how (A1.18) can be obtained once it has been established that, for some constant $\tilde{M}$,

$$\sup_\eta \|\{L_N''[\eta] - E_N''[\eta]\}_{jk}\|_4 \leq \tilde{M}N^{1/2} \quad \text{(A1.19)}$$

holds, for $1\leq j,k\leq \dim\eta$: Using Hölder's inequality, it is easily verified that the left hand side of (A1.18) is bounded above by

$$N\sum_{j,k=1}^{\dim\eta} \|\hat{\eta}_j^N - \eta_j^N\|_{8/3}\|\hat{\eta}_k^N - \eta_k^N\|_{8/3}\|N^{-1}\{L_N''[\eta^N] - N^{-1}E_N''[\eta^N]\}_{j,k}\|_4 \quad \text{(A1.20)}.$$

Next, note that, by (2.21), there is a constant $\overline{M}$ such that

$$\sup_{1\leq j\leq \dim\eta} \|\hat{\eta}_j^N - \eta_j^N\|_{8/3} \leq MN^{-\min\{1/2,3(1+\alpha)/8\}} \quad \text{(A1.21)}.$$

If we set $M=\tilde{M}M$, then (A1.18) follows immediately from (A1.19-21). Thus, it only remains to verify (A1.19).

To do this, it will be convenient to denote the matrix $\partial^2/\partial\eta_j\partial\eta_k\{d^{NT}[\eta]\Sigma^{-N}[\eta]d^N[\eta]\}$ by $Q^N[\eta]$. If $Q^N[\eta]$ is partitioned into blocks of order r, $Q^N[\eta] = [Q^N[\eta](m,n)]_{1\leq m,n\leq N}$, then, by (A1.8), there is a constant $M_{(\delta-)}$ such that

$$\sup_\eta |Q^N[\eta](m,n)|_\infty \leq M_{(\delta-)}(\delta-)^{|m-n|} \qquad \text{(A1.22)}$$

holds, for $1\leq m,n\leq N$, for all N=1,2,... . Since $\{L_N''[\eta] - E_N''[\eta]\}_{jk} = x^{NT}Q^N[\eta]x^N - Ex^{NT}Q^N[\eta]x^N$, the inequality (A1.19) is equivalent to

$$\|x^{NT}Q^N[\eta]x^N - Ex^{NT}Q^N[\eta]x^N\|_4 \leq \tilde{M}N^{1/2} \qquad \text{(A1.23)}.$$

We shall obtain (A1.23) from (A1.22) by using an approximating series $\tilde{x}(t)$ gotten by truncating the innovations representation (2.1) of x(t),

$$\tilde{x}(t) = \sum_{m=0}^{t-1} c(m)e(t-m) \qquad (t=1,2,\dots) .$$

Clearly, from (2.3) and (2.Iii), if $1\leq\nu\leq 8$, then

$$\sup_{1\leq i\leq r} \|x_i(t) - \tilde{x}_i(t)\|_\nu \sim O[(\delta-)^t] \qquad \text{(A1.24)}.$$

Defining $\tilde{x}^N = \mathrm{vec}(\tilde{x}(N), \tilde{x}(N-1)\ldots, \tilde{x}(1))$, we note that $x^{NT}Q^N[n]x^N - \tilde{x}^{NT}Q^N[n]\tilde{x}^N$ is equal to $(x^N + \tilde{x}^N)^T Q^N[n](x^N - \tilde{x}^N)$. Applying (A1.22), (A1.24) and the Cauchy-Schwarz inequality to this last expression, it follows that, for $\gamma=1$ or 4, $\| x^{NT}Q^N[n]x^N - \tilde{x}^{NT}Q^N[n]\tilde{x}^N \|_\gamma$ is bounded by some constant independent of N and n. Consequently, (A1.23) can be obtained by showing that

$$\sup_{N,n} N^{-1/2} \| \tilde{x}^{NT}Q^N[n]\tilde{x}^N - E\tilde{x}^{NT}Q^N[n]\tilde{x}^N \|_4 < \infty \quad .$$

However, from the definition of $\tilde{x}(1),\ldots,\tilde{x}(N)$, we can write $\tilde{x}^{NT}Q^N[n]\tilde{x}^N$ in the form $e^{NT}Q_e^N[n]e^N$, where $e^N = \mathrm{vec}(e(N),\ldots, e(1))$. Further, by (2.3) and Lemma A1.1, if $Q_e^N[n]$ is partitioned into blocks of order r, $Q_e^N[n] = [Q_e^N[n](m,n)]_{1\leq m,n\leq N}$, then the blocks satisfy

$$\sup_n |Q_e^N[n](m,n)|_\infty \leq \tilde{M}_{(\delta-)}(\delta-)^{|m-n|} \qquad \text{(A1.25)},$$

for $1\leq m,n\leq N$ and $N=1,2,\ldots$, where $\tilde{M}_{(\delta-)}$ is a constant. Thus we have finally reduced the demonstration of (A1.19) to the verification that (A1.25) implies

$$\sup_{N,n} N^{-2} E\{e^{NT}Q_e^N[n]e^N - Ee^{NT}Q_e^N[n]e^N\}^4 < \infty \ .$$

This involves a tedious but rather straightforward calculation based on the independence of $e(t)$ $(t=1,2,\dots)$ and the uniform boundedness of the eighth moments of the components $e_j(t)$ $(1\leq j\leq r)$ assumed in (2.Iiii), which we omit for lack of space.

APPENDIX 2. DERIVATIONS OF (5.1-2), (5.5-8) AND (5.17-8).

PROOFS OF (5.1-2)

We note that, from (2.16), it is simple to show, with the aid of the differentiation formulas used to establish (4.2), that the equation $\{\partial W/\partial\xi\}[(\xi,\theta)]=0$ implies

$$\Sigma[\xi] = \Sigma_{(\theta)} \qquad \text{(A2.1).}$$

The assertions of (5.1) follow immediately from this and from (A2.4) below.

One consequence of (A2.1) is that, for maximizing purposes, a concentrated version $W[\theta]$ of $W[n]$ can be obtained by substituting $\Sigma_{(\theta)}$ for $\Sigma[n]$ in (2.16). After simplification, this yields

$$W[\theta] = -\frac{1}{2}\log\det 2\pi e\Sigma_{(\theta)} \qquad \text{(A2.2).}$$

Since $\theta^{p,q}$ uniquely satisfies $W[\theta^{p,q}] = \max_{\Theta^{p,q}} W[\theta]$, it also uniquely satisfies

$$\det\Sigma_{(\theta^{p,q})} = \min_{\Theta^{p,q}} \det\Sigma_{(\theta)} \qquad \text{(A2.3)},$$

which verifies (5.2).

PROOFS OF (5.5-8)

Let L denote the lag operator, $Lx(t) = x(t-1)$. Observe that for any θ in $\Theta^{p,q}$, $\{D[\theta](L) - D(L)\}x(t)$ is a linear function of $x(t-1), x(t-2),\ldots$ and so is uncorrelated with $D(L)x(t) = e(t)$. Hence the covariance matrix, $\Sigma_{(\theta)}$, of $D[\theta](L)x(t)$ is the sum of the covariance matrices of $e(t)$ and $\{D[\theta](L) - D(L)\}x(t)$. Calculating these from the spectral density matrices, we arrive at the following basic formula,

$$\Sigma_{(\theta)} = \int_{-\pi}^{\pi} D[\theta](e^{i\lambda})f(\lambda)D^{*}[\theta](e^{i\lambda})d\lambda = \Sigma +$$

$$\int_{-\pi}^{\pi}\{D[\theta]-D\}(e^{i\lambda})f(\lambda)\{D[\theta]-D\}^{*}(e^{i\lambda})d\lambda \qquad \text{(A2.4)}.$$

The key inequality for establishing rates of convergence is a direct consequence of (A2.4), (5.1), (5.3) and (5.III):

$$\int_{-\pi}^{\pi} \operatorname{tr}\{D[\theta] - D\}(e^{i\lambda})f(\lambda)\{D[\theta] - D\}^*(e^{i\lambda})d\lambda$$

$$\geq M^{-1}\operatorname{tr}\{\Sigma[\xi^{p,q}] - \Sigma\} \quad (\theta\varepsilon\Theta^{p,q}) \qquad \text{(A2.5).}$$

We now choose $(p,q) \varepsilon S$ with p and q large enough that the zeros of the determinants either of the p-th and later partial sums of $D(z)$ or, when $p<q$, of the q-th and later partial sums of $C(z)$ all lie in $\{|z|>\delta^{-1}\}$, and also large enough that these determinants are uniformly bounded away from zero on $\{|z|=1\}$. Then (5.II) insures either that a θ^p in $\Theta^{p,q}$ exists such that $D[\theta^p](z)$ coincides with the p-th partial sum of $D(z)$, or that a θ^q in $\Theta^{p,q}$ exists such that $C[\theta^q](z)$ coincides with the q-th partial sum of $C(z)$. This implies, via (2.3-4), that

$$\sup_{-\pi\leq\lambda\leq\pi} |D[\theta^p](e^{i\lambda}) - D(e^{i\lambda})|_\infty \sim O[(\delta-)^p] \qquad \text{(A2.6),}$$

or

$$\sup_{-\pi\leq\lambda\leq\pi} |C[\theta^q](e^{i\lambda}) - C(e^{i\lambda})|_\infty \sim O[(\delta-)^q] \qquad \text{(A2.7).}$$

It is $D[\theta^q](e^{i\lambda}) - D(e^{i\lambda})$, not $C[\theta^q](e^{i\lambda}) - C(e^{i\lambda})$, about which we need such information. However, observe that for any θ,

$$D[\theta](e^{i\lambda})\{C - C[\theta]\}(e^{i\lambda})D(e^{i\lambda}) = D[\theta](e^{i\lambda}) - D(e^{i\lambda}) \qquad \text{(A2.8),}$$

so that (A2.7) and the uniform boundedness of the $|D[\theta^q](e^{i\lambda})|_\infty$ imply, in fact, that

$$\sup_{-\pi<\lambda<\pi} |D[\theta^q](e^{i\lambda}) - D(e^{i\lambda})|_\infty \sim O[(\delta-)^q] \qquad \text{(A2.9)}.$$

If m_f and M_f are, respectively, the smallest and largest eigenvalues occurring in the family $f(\lambda)$ $(-\pi<\lambda<\pi)$, then

$$0 < m_f I \leq f(\lambda) \leq M_f I \qquad \text{(A2.10)}.$$

Using (A2.6), (A2.9), (A2.10) and (A2.5) in an obvious way, we obtain (5.5). Note that by (A2.4) with $\theta=\theta^{p,q}$, (5.5) is equivalent to

$$\int_{-\pi}^{\pi} \mathrm{tr}\{D[\theta^{p,q}]-D\}(e^{i\lambda})f(\lambda)\{D[\theta^{p,q}]-D\}^*(e^{i\lambda})d\lambda$$

$$\sim O[(\delta-)^{\max\{p,q\}}] \qquad \text{(A2.11)}.$$

This fact, in conjunction with (A2.10), implies (5.7).

Now we turn to (5.6). Using (A2.8), we re-express the left-hand side of (A2.11) as $1/2\pi$ times

$$\int_{-\pi}^{\pi} \mathrm{tr}D[\theta^{p,q}](e^{i\lambda})\{C - C[\theta^{p,q}]\}(e^{i\lambda})\Sigma\{C - C[\theta^{p,q}]\}^*(e^{i\lambda})$$

$$D^*[\theta^{p,q}](e^{i\lambda})d\lambda \qquad \text{(A2.12)}.$$

The assumption (5.I) insures the existence of a positive constant M_C with the property that

$$C[\theta^{p,q}](e^{i\lambda})C^*[\theta^{p,q}](e^{i\lambda}) \leqslant M_C I,$$

holds, for all $-\pi\leqslant\lambda\leqslant\pi$ and $(p,q)\varepsilon S$. Hence we have

$$D^*[\theta^{p,q}](e^{i\lambda})D[\theta^{p,q}](e^{i\lambda}) \geqslant M_C^{-1} I \tag{A2.13}$$

for these λ and (p,q). If m_Σ denote the smallest eigenvalue of Σ, then $\Sigma \geqslant m_\Sigma I$. Together with (A2.13), this shows that (A2.12) is greater than or equal to

$$m_\Sigma M_C^{-1} \int_{-\pi}^{\pi} tr\{C-C[\theta^{p,q}]\}(e^{i\lambda})\{C-C[\theta^{p,q}]\}^*(e^{i\lambda})d\lambda,$$

so it is now apparent that (5.6) is implied by (A2.11).

The assertion (5.8) is a straightforward consequence of (5.5-7).

PROOFS OF (5.17-8)

To establish the assertions (5.17-8) of Proposition 5.2, we begin by verifying

$$\sup_{1 \leq p < \infty} |\{\Gamma^p[\eta^{p,0}]\}^{-1}|_\infty < \infty \qquad \text{(A2.14)}$$

and

$$\sup_{1 \leq q < \infty} |\{\Gamma^q_{inv}[\eta^{0,q}]\}^{-1}|_\infty < \infty \qquad \text{(A2.15)},$$

where $\Gamma^p[\eta^{p,0}]$ and $\Gamma^q_{inv}[\eta^{0,q}]$ are block-Toeplitz autocovariance matrices whose (j,k)-blocks are, respectively,

$$\Gamma[\eta^{p,0}](j-k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(j-k)\lambda} C[\theta^{p,0}](e^{i\lambda}) \Sigma[\xi^{p,0}] C^*[\theta^{p,0}](e^{i\lambda}) d\lambda,$$

for $1 \leq j,k \leq p$, and

$$\Gamma_{inv}[\eta^{0,q}](j-k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(k-j)\lambda} D^*[\theta^{0,q}](e^{i\lambda}) \Sigma^{-1}[\xi^{0,q}] D[\theta^{0,q}](e^{i\lambda}) d\lambda,$$

for $1 \leq j,k \leq q$.

Indeed, from a straightforward multivariate generalization of Theorem 2.2 of Shaman (1976), we obtain

$$\{\Gamma^p[\eta^{p,0}]\}^{-1} \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} D^*[\theta^{p,0}](e^{i\lambda})\Sigma^{-1}[\xi^{p,0}] D[\theta^{p,0}](e^{i\lambda})d\lambda$$

and

$$\{\Gamma^q_{inv}[\eta^{0,q}]\}^{-1} \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} C[\theta^{0,q}](e^{i\lambda})\Sigma[\xi^{0,q}] C^*[\theta^{0,q}](e^{i\lambda})d\lambda.$$

From these matrix inequalities it is clear that the boundedness of $|D[\theta^{p,0}]|_\infty$ and $|C[\theta^{0,q}]|_\infty$ (guaranteed by 5.I) and

$$\lim_{p\to\infty} \Sigma^{-1}[\xi^{p,0}] = \lim_{q\to\infty} \Sigma^{-1}[\xi^{0,q}] = \Sigma^{-1} \tag{A2.16}$$

(see (5.5)) together imply (A2.14-15).

The assertion (5.17) follows immediately from (4.13), (A2.14) and (A2.16).

To complete the derivation of (5.18), we require some properties of the commutator matrix of order r^2, which we denote $I_{(r)}$. If $I_{(r)}$ is partitioned into blocks of order r, the (j,k)-block is a matrix with 1 in the k-th row and j-th column and zeros elsewhere. $I_{(r)}$ has the properties that

$I_{(r)}^{-1} = I_{(r)}$ and that, for any two matrices A_1, A_2 of order r, $A_1 \otimes A_2 = I_{(r)}(A_2 \otimes A_1)I_{(r)}$, see Magnus and Neudecker (1979). If $I_{(r)}^{(q)}$ denotes a block diagonal matrix of order r^2q whose diagonal blocks coincide with $I_{(r)}$, it is readily verified that

$$[\Sigma[\xi^{0,q}] \otimes \Gamma^q{}_{inv}[\theta^{0,q}](j-k)]_{1 \leqslant j,k \leqslant q}$$

$$= I_{(r)}^{(q)}(\Gamma^q_{inv}[\theta^{0,q}] \otimes \Sigma[\xi^{0,q}])I_{(r)}^{(q)}.$$

We conclude, therefore, from (4.14) that

$$F_{(2)}^{-1}[\eta^{0,q}] = I_{(r)}^{(q)} \ (\{\Gamma^q{}_{inv}[\eta^{0,q}]\}^{-1} \otimes \Sigma^{-1}[\xi^{0,q}])I_{(r)}^{(q)} \qquad \text{(A2.17)},$$

from which

$$|F_{(2)}^{-1}[\eta^{0,q}]|_\infty \leqslant |\{\Gamma^q_{inv}[\eta^{0,q}]\}^{-1}|_\infty |\Sigma^{-1}[\xi^{0,q}]|_\infty$$

follows immediately. Thus (5.18) is a consequence of (A2.15-16).

We remark, in passing, that the formula (A2.17) leads to a formula for the large-sample covariance matrix of the maximum likelihood estimates of the coefficients of a multivariate moving average process, expressed in terms of the coefficients and the innovations covariance matrix of the associated backward moving average model, see Findley (1983).

REFERENCES

AKAIKE, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In _2nd International Symposium on Information Theory_, (Eds. B. W. Petrov and F. Csaki) Budapest: Akademaii-Kiado, 267-281.

AKAIKE, H. (1977). On Entropy Maximization Principle. In _Applications of Statistics_ (Ed. P. R. Krishnaiah) Amsterdam: North Holland.

AKAIKE, H. (1978). On the Likelihood of a Time Series Model. _Statistician_ 27, 217-35.

ANDERSON, T. W. (1971). _The Statistical Analysis of Time Series_ New York: Wiley.

BHANSALI, R. J. (1981). Effects of Not Knowing the Order of An Autoregressive Process. _J. Am. Statist. Assoc._ 76, 588-597.

CHUNG, K. L. (1968). _A Course in Probability Theory_ New York: Harcourt, Brace and World.

CLEVELAND, W. C. (1972). The Inverse Autocorrelations of a Time Series and Their Applications. _Technometrics_ 14, 277-293.

DEISTLER, M., W. DUNSMUIR and E. J. HANNAN (1978). Vector Linear Time Series Models: Corrections and Extensions. Advances in Applied Probability 10, 360-72.

DWYER, P. S. (1967). Some Applications of Matrix Derivatives in Multivariate Analysis. J. Amer. Statist. Assoc. 62, 607-625.

EROHIN, V. (1959). On the Best Approximation of Analytic Functions by Rational Fractions with Free Poles. Dokl. Akad. Nauk. SSSR., 128, 29-32 (Russian) MR21#7312.

FINDLEY, D. F. (1982). A Special Property of the Expected Log Likelihood. Communications in Statistics, A11(21), 2379-2387.

FINDLEY, D. F. (1983). A Formula for the Large-Sample Covariance Matrix of the Maximum Likelihood Estimates of the Coefficients of a Vector Moving Average Process. Statistical Research Division Report 83/02, U.S. Census Bureau.

GERSCH, W. and G. KITAGAWA (1982). The Prediction of Time Series with Trends and Seasonalities. Statistical Research Division Report 82/03, U.S. Census Bureau.

HANNAN, E. J. (1970). Multiple Time Series New York: Wiley.

HANNAN, E. J. (1980a). The Estimation of the Order of an ARMA Process. Annals of Statistics 8, 1071-1081.

HANNAN, E. J. (1982). Testing for Autocorrelation and Akaike's Criterion. In _Essays in Statistical Science_ (Eds. J. Gani and E. J. Hannan) Sheffield: Applied Probability Trust, 403-412.

HOSOYA, Y. and M. TANIGUCHI (1982). A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes. _Annals of Statistics_ 10, 132-153.

KOZIN, F. and F. NAKAJIMA (1980). The Order Determination Problem for Linear Time-Varying AR Models. _IEEE Transactions on Automatic Control_ AC-25, 250-7.

LJUNG, L. and P. CAINES (1979). Asymptotic Normality of Prediction Error Estimates for Approximate System Models. _Stochastics_ 3, 29-45.

MAGNUS, R. R. and H. NEUDECKER (1979). The Commutation Matrix: Some Properties and Applications. _Annals of Statistics_ 7, 381-94.

McCULLOCH, C. E. (1982). Symmetric Matrix Derivatives with Applications. _J. Amer. Statist. Assoc._ 77, 679-82.

OGATA, Y. (1980). Maximum Likelihood Estimates of Incorrect Markov Models and the Derivation of AIC. _Journal of Applied Probability_ 17, 59-72.

SHAMAN, P. (1976). Approximations for Stationary Covariance Matrices and their Inverses with Application to ARMA Models. _Annals of Statistics_ 4, 292-301.

SHIBATA, R. (1980). Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. Annals of Statistics 8, 147-164.

SHIBATA, R. (1981a). An Optimal Autoregressive Spectral Estimate. Annals of Statistics 9, 300-306.

SHIBATA, R. (1981b). An Optimal Selection of Regression Variables. Biometrika 68, 45-54.

TANIGUCHI, M. (1980). On the Selection of the Order of the Spectral Density Model for a Stationary Process. Ann. Inst. Stat. Math. 32A, 401-409.

THIS WEEK'S CITATION CLASSIC (1981). Current Contents 51, 22.

TITCHMARSH, E. C. (1937). The Theory of Functions (2nd ed.) London: Oxford University Press.

WOODROOFE, M. (1982). On Model Selection and the Arc Sine Laws. Annals of Statistics 10, 1182-94.