Qüestiió - V. 8, n.º 1 (març 1984) pp. 1-8

# MULTIVARIABLE ARMA SYSTEMS AND PRACTICABLE CALCULATIONS

E. J. HANNAN

THE AUSTRALIAN NATIONAL UNIVERSITY

*The representation of ARMA systems in canonical state space form is described and the idea of order and the manifold structure of spaces of systems are introduced. Some neighbourhood systems are described and related to other systems of structures. Algorithms for maximum likelihood estimation are discussed and related to stability of the system. An algorithm for order estimation is briefly discussed and also some asymptotic theory.*

Keywords: ARMA, HANKEL MATRIX, ORDER, MANIFOLD, KALMAN FILTER, MAXIMUM LIKELIHOOD, AUTOREGRESSION, AIC, LAW OF ITERATED LOGARITHM, SIMULATIONS.

## 1. THE STRUCTURE OF ARMA SYSTEMS.

Consider a stationary vector process, $y(t)$, of the form

$$(1)\quad y(t) = \sum_0^\infty K(j)\varepsilon(t-j), \quad E\{\varepsilon(t)\varepsilon(u)'\} = \delta_{tu}\,\Sigma, \quad \Sigma > 0_s$$

where $k(z) = \Sigma K(j)z^{-j}$ is rational in $z$ and $\varepsilon(t)$ is the linear innovation sequence for $y(t)$. Then

$$(2)\quad \det\{k(z)\} \neq 0, \quad |z| > 1; \quad k(z) \text{ is analytic for } |z| \geq 1.$$

If in (2) we require, more strongly, that $\det(k) \neq 0$, $|z| \geq 1$ we shall speak of (2)'. Such rational transfer function models have been widely used, especially for $s=1$, partly because of the influence of /6/ .

Since $k(z)$ is rational we have $k = a^{-1}b$ where $a,b$ are matrices of polynomials. In this matrix fraction description (m.f.d) we may require that $a,b$ be left coprime, for which a necessary and sufficient conditions is

$$(3)\quad [a(z),b(z)] \text{ is of rank } s \text{ for all } z \text{ in the complex plane.}$$

There are many such factorisations as there are many polynomial matrices $u(z)$ with $\det(u) =$ constant $\neq 0$ so that $ua$, $ub$ will provide another prime m.f.d. To attain a unique decomposition consider the (infinite) Hankel matrix, $H$.

$$H = \begin{matrix} K(1) & K(2) & K(3) & \cdots \\ K(2) & K(3) & K(4) & \cdots \\ K(3) & K(4) & K(5) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{matrix}$$

This is of finite rank, $n$ let us say, iff $k$ is rational (see /2/ ). Call $r(u,j)$ the $j$th row, $j=1,\ldots,s$, in the $u$th row of blocks. Then we may always choose a basis for the rows of $H$ of the form

$$(4)\quad r(u,j); \quad u=1,2,\ldots n_j; \quad j=1,\ldots,s; \quad \Sigma n_j = n.$$

Thus the basis is fully specified by the partition $n = \Sigma n_j$. Let $M(n)$ be the set of all $k(z)$ for which (2) holds and for which $H$ is of rank $n$ and call $n$ the order for $M(n)$. Let $U(\{n_j\})$ be the subset of $M(n)$ for which (4) is a basis for $H$. It will be convenient to represent the set $\{n_j\}$ by the

- Edward J. Hannan -The Australian National University - Dep. of Statistics IAS - P.O.Box 4 Canberra ACT 2600 (Australia).
- Article rebut el novembre del 1983.
- Aquest treball es una ponencia invitada al "First Catalan International Symposium on Statistics" a Barcelona del 23 al 24 de setembre de 1983.

symbol $\nu$ , to write $|\nu| = \Sigma n_j$ and to write $U_\nu = U(\{n_j\})$ if $\nu$ indexes the set $\{n_j\}$. It is clear that $M(n) = \bigcup_{|\nu|=n} U_\nu$ . We may topologise the $k(z)$ with the norm $\|k\|$

$$\|k\|^2 = \sum_1^\infty \mathrm{tr}\{K((j)K(j)'\}$$

then each $U_\nu$ , $|\nu| = n$, is open and dense in $M(n)$. As we shall see these $U_\nu$ constitute coordinate neighbourhoods covering the, essentially algebraic, surface $M(n)$. ($M(n)$ is an analytic manifold but no differential geometry is used in this paper.) The dimension of $M(n)$ is $2ns$. Put

$$y(t+u|t) = \sum_u^\infty K(j)\ \varepsilon\ (t+u-j).$$

Then compose $x(t+1|t)$ from the $y_j(t+u|t)$, $u=1,\ldots,n_j$ ; $j=1,\ldots,s$. If $k(z)\in U_\nu$ then

$$(6)\quad y_j(t+u|t) = y_j(t+u|t-1) + \sum_1^s \kappa_{jk}(u)\varepsilon_k(t),$$

$$u=1,\ldots,n_j-1$$

$$y_j(t+n_j|t) + \sum_{k=1}^s \sum_{v=1}^{n_k} \alpha_{jk}(v-1)y_k(t+v-1|t-1) =$$

$$= \sum_1^s \kappa_{jk}(n_j)\varepsilon_k(t).$$

Here $\kappa_{jk}(u)$ is the typical element of $K(u)$. Thus

$$(7)\quad x(t+1|t) = F\ x\ (t|t-1) + K\varepsilon(t)$$

$$y(t) = H\ x(t|t-1) + \varepsilon(t)$$

where $F,H,K$ are easily constructed using (5) (6). The freely varying elements in $F,H,K$ (i.e. those not 0 or 1, identically) are the $2ns$ quantities, which we call "system parameters",

$$(8)\quad \alpha_{jk}(v),\quad v=0,\ldots,n_k-1;$$

$$\kappa_{jk}(v),\quad v=1,\ldots,n_j\ ;\quad j,k=1,\ldots,s.$$

Putting $a(z) = \sum A(v)z^v$ ,

$$A(v) = [\alpha_{jk}(v)]_{j,k=1,\ldots,s},$$

$$\alpha_{jj}(n_j)=1,$$

and $b(z) = a(z)k(z)$, which is also polynomial, then $k = a^{-1}b$ is a left prime m.f.d. for $k$. (See /7/).

Let $V_\nu$ be the set of all $k(z)$ for which (2)' holds and (4) is the first set of linearly independent rows met as you go down the rows of $H$ . The $n_j$, called dynamical indices, for which this is true are uniquely determined, i.e. $k(z)$ lies in a unique $V_\nu$ (these do not intersect) and for $|\nu|=n$ their union is $M(n)$. Then the $\alpha_{jk}(v)$ in (8) satisfy further restrictions, namely

$$(9)\quad \alpha_{jk}(v) = 0,\quad v > n_j;$$

$$\alpha_{jk}(n_j) = 0,\quad j < k.$$

Then $V_\nu \subset U_\nu$ and equality holds only when, putting $n=ps+r$,

$$n_1 = n_2 = \ldots n_r = p+1;\ n_{r+1} = \ldots = n_s = p.$$

Call this neighbourhood just $U(n)$.

For $k\in U_\nu$ put $k(z^{-1}) =$

$$=\{Z_\nu z(z^{-1})\}^{-1}\{Z_\nu\ b(z^{-1})\} \triangleq \phi(z)^{-1}\psi(z),$$

where $Z_\nu$ is diagonal with $z^{n_j}$ in the jth place in the diagonal. Call $\Phi(j)$, $\Psi(j)$ the coefficient matrices of $z^j$ in $\phi(z)$, $\psi(z)$. Then $\Phi(0) = \Psi(0)$ and we have an ARMA representation

$$(10)\quad \sum\Phi(j)y(t-j) = \sum\Psi(j)\varepsilon(t-j).$$

However unless $k = a^{-1}b$ corresponds to the $n_j$ for which $k\in V$ then not all of the coefficients in the $\Psi(j)$ (that are not 0 or 1) are independently varying. (See /8/).

If we put $\bar{M}(n)$ for the part of the closure, $\overline{M(n)}$ , of $M(n)$ for which (2)' holds then $\bar{M}(n) = \bigcup_{j\le N} M(j)$. Since $U(n)$ is dense in $M(n)$ it is dense in $\bigcup_{j\le n} M(j)$.

A further parameterisation discussed below is $M(p,q)$ which consists of all $k(z)$, satisfying (2)', for which $k(z^{-1})$ has a left prime m.f.d, $\phi(z)^{-1}\psi(z)$ with $\Phi(0) = \Psi(0) = 1$ and $[\Phi(p),\Psi(q)]$ of full rank. Put $r = \max(p.q)$ and $m = p-q$ . Then

(11) <u>$M(p,q)$ is equivalent to the part of $U(rs)$ satisfying</u>

$$\underline{K(1) = K(2) = \ldots = K(m-1) = 0_s,\quad \overline{K(m) = I_s}}$$

We give the proof for $p \ge q$. Consider $z^{-m}k(z)\in U(ps)$. Then $k(z) = a(z)^{-1}\{z^m b(z)\}$ where $a(z)$ is of degree $p$ and $b(z)$ is of

degree q = p-m. If $\phi(z) = z^{p}a(z^{-1})$, $\psi(z) = z^{p-m}{}_{-}(z^{-1})$ then $\phi,\psi$ are left prime because $\Phi(0) = A(p) = I_s$ and certainly $[\phi(z),\psi(z)]$ is of full rank for $z \neq 0$. Moreover $[\Phi(p),\Psi(q)]$ is of full rank since it is $[A(0),B(0)] = [a(0),b(0)]$. The argument may be reserved to show that for $k(z)\epsilon M(p,q)$ then $k(z) = z^{m}\{z^{p}\phi(z^{-1})\}^{-1} z^{q}\psi(z^{-1}) = z^{m} a(z)^{-1}b(z)$ where $a(z)^{-1}b(z)\epsilon U(p\ s)$. For $q \geq p$ we consider $k(z)^{-1}$ and proceed similarly.

In relation to the M(p,q) the following is true. (i) There are k(z) that lie in no M(p,q). However since $M(p.p) = U(ps)$ and $U(ps)$ is open and dense in $\bigcup_{j\leq ps} M(j)$ these k(z) may be regarded as exceptional. (ii) The M(p,q) are not disjoint. (iii) The largest dynamical index $n_{max}$, of k(z) satisfies $n_{max} \leq \min\{\max(p,q)\}$, where the minimum is over all (p,q) for which $k(z)\epsilon M(p,q)$. however inequality may hold. For the same kind of reason as was given under (i) it would not be unreasonalbe to strengthen the requirement "$[\Phi(p)\ \Psi(q)]$ is of full rank" to $\Phi(p)$ is of full rank $p > q$; $\Psi(q)$ is of full rank, $q > p$, $[\Phi(p), \Psi(p)]$ is of full rank, p=q. Then $k(z)\epsilon U(s \max(p,q))$.

The virtue in M(p,q) can be only in an, a priori reasonable, belief that a good "fit" will be obtained with |p-q| different from zero. In case s=1 this seems often to be true and may correspond to the fact that spectra often show sharp peaks, which requires high p but not necessarily high q.

The virtue in the *V* lies in the uniqueness of the parameterisation. If the elements of y(t) are permuted then so are the dynamical indices, but not in the same way. For $|\nu|=n$ these indices appear to define an arbitrary division of M(n) so that any gain in the reduction in the number of parameters fitted, due to (9), seems illusory.

The meaning of n is direct, as the rank of *H* or the dimension of the "predictor space", spanned by the $y_j(t+u|t_j)$, j=1,...,s; u=1,2,... . See /2/).

The most attractive situation is that where (10) (or a state space formulation) is a priori so heavily constrained by valid constraints that the representation is uniquely determined on physical grounds. However we do not in general deal with this situation.

## 2. ALGORITHMS

In /15/ a direct search for an appropriate $U_\nu$ is considered. That is not discussed here, partly because nothing seems so far established about the method and partly because it seems better first to find, by some ad hoc method, an estimate of the order parameters (n, $\nu$ or p,q) and of the system parameters for that order. Maximum likelihood (ML) can then be used commencing from these initial values. The computation of the MLE is not trivial. One way to proceed is via the Kalman Filter (KF) to construct the finite past innovations, e(t), and their covariance matrix $\Sigma_t$. Thus

$$(12)\quad e(t) = y(t)-H\hat{x}(t|t-1),$$

$$\hat{x}(t+1|t)=F\hat{x}(t|t-1)+K_t e(t),$$

$$\hat{x}(0|-1) = 0$$

$$(13)\quad K_t = M_t\Sigma_t^{-1},\ M_t = FP_tH'+K\Sigma,\ \Sigma_t=HP_tH'+\Sigma$$

$$P_{t+1} = FP_tF'+K\Sigma K'-M_t\Sigma_t^{-1}M_t',\ P_o=FP_oF'+K\Sigma K'$$

Then the e(t), $\Sigma_t$, t=1,...,T are used to construct the Gaussian likelihood. The optimisation of that is a not inconsiderable task and requires the calculation of derivatives, again using the KF, but we omit any discission here. In /17/ it is suggested that (13) be modified as in /3, pp.228,9/. Thus let $\gamma_{jk}(v) = E\{y_j(t)y_k(t-v)\}$ and $\Gamma(v) = [\gamma_{jk}(v)]$ where the typical matrix element is indicated. Let *G* be composed from the $\Gamma(v)$ as *H* was from the K(v). Then *G,H* have the same rank and the same row dependence structure. Thus in (8) the $\kappa_{jk}(v)$ may be replaced by the corresponding $\gamma_{jk}(v)$. Also (13) may be reorganised as follows, putting $\Gamma$ for the matrix constructed from the $\gamma_{jk}(v)$ as was K from the $\kappa_{jk}(v)$,

$$(13)'\quad K_t=M_t\Sigma_t^{-1},\ M_t=\Gamma-FT_tH',\ \Sigma_t=\Gamma(0)-HT_tH',$$

$$T_{t+1}=FT_tF'\ M_t\Sigma_t^{-1}M_t'\ ,\ T_o = 0.$$

Put

$$\hat{\Gamma}(t) = T^{-1}\sum_{u=1}^{T} y(u)y(u+t)'\ .$$

Here we have omitted mean corrections for convenience but they will be needed. We may use the $\hat{\Gamma}(t)$ to initially estimate $\Gamma(0)$ and $\Gamma$ . To estimate the $\alpha_{jk}(v)$ we may use an instrumental variable (IV) method, using the $y_j(t-\ell)$, for $\ell$ large, via (10), as the instrumental variables. Thus let $G_\nu^\mu$ be the submatrix of $G$ taking $|\nu|$ rows according to the partition $\nu$ and $|\mu| = |\nu|$ columns according to $\mu$. Let A have $\alpha_{jk}(v)$, $v=0,1,\ldots,$ $v=0,1,\ldots,n_{k-1}$, $j,k=1,\ldots,s$ in row j, column $n_1+n_2 \ldots+n_{k-1} + v+1$, where $\nu = \{n_j\}$ . Let $g_\nu^\mu$ contain rows $r(n_j+1,j)$ $j=1,\ldots,s$ from $G$ and columns according to $\mu$. Then if $k(z)\epsilon U_\nu$

$$(14) \quad AG_\nu^\mu = -g_\nu^\mu$$

If we wish to consider $V_\nu$ then we must take account of the relation (9) that constrains A. If $G_\nu^\mu$ is non - singular we may estimate A consistently by using the $\hat{\Gamma}(t)$ to estimate $G_\nu^\mu$, $g_\nu^\mu$ . We may associate n and $\nu$ with the spectral density $f(\omega)$ if $k\epsilon U_\nu$ since $f(\omega)$ uniquely determines $k(z)$ via $2\pi f(\omega) = k(e^{i\omega})\, \sharp\, k^*(e^{i\omega})$. Then $f(-\omega)$ has the same n and $G_\nu^\mu$ is of full rank iff $f(\omega)\epsilon U_\mu$, $f(-\omega)\epsilon U_\mu$ . (See /5/). However this method of estimation may be very inefficient. For example when s=1 (when $f(\omega) = f(-\omega)$ and we may take $\mu=\nu$ ) and $y(t)+0.9y(t-1)=\varepsilon(t)+0.7\varepsilon(t-1)$ then the asymptotic efficiency of the IV method is 0.08. Another problem arises with the stability of (12), (13)'. Consider s=1, $y(t) = \varepsilon(t) + \psi\varepsilon(t-1)$. Then $\psi/(1+\psi^2) = \gamma(1)/\gamma(0) \overset{d}{=} a$. However if $\hat{a} > 1/2$, which is quite likely, then (12), (13)' will not "run". Indeed, then H=1, F=0, $K=\psi, K_t = b_t^{-1} a$, $b_t = 1-a^2/b_{t-1}$, $b_o=1$. Then $b_t =$

$= \text{const.}\ \cos(\theta t+\phi_1)/\cos(\theta t+\phi_2)$ where,

$|a|^t \cos(\theta t+\phi_1)/\cos\phi_1$ is the solution of

$x_t - x_{t-1} + a^2 = 0$, $x_o = 1$, $x_1 = 1-a^2$ while $\phi_2$ corresponds to $x_o = 1 = x_1$.

Thus evidently $b_t$, and hence $K_t$ oscillate increasingly wildly with occasional negative values. (Here $b_t = \sharp_t/\gamma(0)$). It seems preferable to face the problem that $2\pi f(\omega)$ may not factorise, as in this case where the estimate using the $\hat{\Gamma}(t)$ is, $2\pi\hat{f} = \hat{\gamma}(0) + 2\hat{\gamma}(1)\cos\omega$.

A procedure that estimates $\nu$ and the $\alpha_{jk}(v)$ via canonical correlation methods was presented in /1/. It essentially proceeds by examining sucessive rows down $H$ (or equivalently $G$). We omit discussion here for brevity and discuss another method. This is an extension /12/, /14/, of an idea due to Durbin /9/, who considered s=1, n known and gave heuristic proofs. It has also been used in a real time, recursive, calculation of estimates for s=1 by Astrom and Mayne. The method, in outline, is as follows.

I. Form the autoregression

$$(15) \quad \sum_0^h \hat{L}_h(j)y(t-j) = \hat{\varepsilon}_h(t),\ h=0,1,\ldots,\ H(T),$$

$$\hat{L}_h(0) = I_s.$$

This may be recursively calculated by the method due to Whittle /19/. Using $\hat{\sharp}_h$ for the estimate of $\sharp$ got from this algorithm we choose $h_T$ to minimise

$$(16) \quad \log\det\hat{\sharp}_h + hs^2C(T)/T$$

For $C(T) = 2$ we refer to (16) as AIC(h) and for $C(T) = \log T$ we refer to BIC(h). In practice we would use AIC(h). Put $\hat{\varepsilon}_T(h)$, $\hat{L}_T(j)$, $\hat{\sharp}_T$ for the estimate at $h = h_T$.

II. We next act as if $k\epsilon U(ps)$ and calculate the regression

$$\sum_o^p \tilde{\Phi}_p(j)y(t-j) = \sum_1^p \tilde{\Psi}_p(j)\hat{\varepsilon}_T(t-j) + \tilde{\varepsilon}_p(t),$$

$$p \le P(T).$$

Again the Whittle recursion is used with $y(t-j)'$ in (15) being replaced by $(y(t-j)', \hat{\varepsilon}_T(t-j)')$ and where only the first s rows of the resulting $\hat{L}(j)$ and the first s rows and columns of the resulting $\tilde{\sharp}$ are used directly. (The other parts are used in later calculations). However the direct estimate of $\sharp$ obtained, thus, from the recursion needs adjustment by term of order $\log T/T$ to account for a bias due to the use of the $\hat{\varepsilon}(t)$. Choose $p_T$ by AIC(p), BIC(p) where now these correspond to

$$(17) \quad \log\det\tilde{\sharp}_p + 2ps^2C(T)/T.$$

III. In this step the dynamical indices are

estimated. We cannot give details here (See /14/). As argued below it may be reasonable to consider only the $U(n)$. Then we take $n = (p_T-1)s + r$, $r = 1,2,\ldots,s$. Then $\bar{\Sigma}_n$ and estimates of the system parameters are easily found using the calculations in II and $n_T = (p_T-1)s+r_T$ is chosen to minimise AIC(r) or BIC(r) via

$$\log\det \bar{\Sigma}_n + 2nsC(T)/T. \quad n=(p_T-1)s+r,$$

$$r=1,\ldots,s.$$

III'. An alternative would be to consider the M(p,q). A possible economy in parameters (see section 1) is balanced by greater computations.

IV. A final stage is ML using the results at stage III to determine the order parameters and initial estimates of system parameters. This could be done via (12), (13) or via asymptotically equivalent procedure that lends itself to recursive calculation, if one wishes to re-estimate the order parameters.

The problem in relation to (12), (13)' in relation to IV method may recur here, though II seems more efficient for the estimation of the $\Phi(j)$ (i.e. the $\alpha_{jk}(v)$) than is (14). (It is always more efficient for $s = p = 1$, at least, and in the case discussed above where IV has asymptotic efficiency 0.08 that of II is 0.6.) The problem arises because $\det \tilde{\psi}(z)$ may have zeros in $|z| \le 1$. Techniques for adjusting $\tilde{\psi}(z)$ to avoid the consequences of this in III are discussed in /14/. One has to replace $\psi_T$ by $\tilde{\psi}_T(0) + \lambda\{\tilde{\psi}_T - \tilde{\psi}_T(0)\}$, $0 < \lambda < 1$, choosing $\lambda$ so as to achieve stability.

## 3. ASYMPTOTIC THEORY.

We assume there are T data points. There are many reasons for attaining generality in the asymptotic theory. Consider, for example, a model as in /16/, namely

$$(18)\quad \sum_0^h \{\Phi(j)+B_j(t)\}y(t-j) = u(t),\ \Phi(0)=I_s,\quad B_0(t)\equiv 0_s$$

Assuming the $B_j(t)$, $u(t)$ to be i.i.d. and the $B_j(t)$ independent of $u(s)$, all s,t, then when a stationary solution exists (see /16/).

$$\sum_0^h \Phi(j)y(t-j)=\varepsilon(t),\quad \varepsilon(t)=-\sum_1^h B_j(t)y(t-j)+u(t).$$

If $F_t$ is the $\sigma$-algebra generated by $y(s)$, $s\le t$

$$(19)\quad E\{\varepsilon(t)|F_{t-1}\} = 0.$$

However it is <u>not</u> true in this case that

$$(20)\quad E\{\varepsilon(t)\varepsilon(t)'|F_{t-1}\} = E\{\varepsilon(t)\varepsilon(t)'\} = \Sigma.$$

Under (18) and, say,

$$(21)\quad E\{\varepsilon_j(t)^4\} < \infty,\quad j = 1,\ldots,s,$$

the asymptotic properties of the estimates described in (15), (16) hold and evidently these autoregressive estimates, given the estimated order, $h_T$, could provide a basis for a more efficient estimation procedure based on the structure (18). The condition (19) is natural in (1), in general, but (20) is difficult to justify, except by assuming Gaussianity. However many results hold under (19), or (19) and (21), except that variances, covariances of estimates in their limiting distribution are more complicated unless (20) holds. For example if $y(t) + \phi y(t-1) = \varepsilon(t)$ and (19), (21) hold then for $\hat{\phi} = -\gamma(1)/\hat{\gamma}(0)$ we have $(\hat{\phi} - \phi) = 0\{(\log\log T/T)^{1/2}\}$, a.s. and $T^{1/2}(\hat{\phi} - \phi)$ is asymptotically normal with zero mean and variance estimated by

$$(22)\quad T^{-1}\sum_2^T [\{y(t)+\hat{\phi}y(t-1)\}^2 y(t-1)^2]/\{T^{-1}\sum_1^T y(t)^2\}^2.$$

If the data satisfies (20) also then this expression may be replaced by $1-\hat{\phi}^2$. Since (22) involves fourth moments its use is to be avoided if possible and it may be preferable to use the simpler formula since it may be near enough to appropriate for practical purposes.

Some basic results are as follows. (1) Let $k\in\overline{M(n)}$ (so that now (2) is permitted, rather than (2)'). Assuming only ergodicity for $\varepsilon(t)$, if $\hat{f},\hat{\Sigma}$ are the estimates based on maximising the Gaussian likelihood over M(n), then $\hat{k}, \hat{\Sigma} \to k, \Sigma$, a.s. Thus consistency holds if the assumed order is not too small. Analogous results hold for $U_\nu$, M(p,q). (ii) Let $s=1$ and $r(t) = \hat{\gamma}(t)/\hat{\gamma}(0)$. Under (19), (21) and

(23) $E\{\varepsilon(t)\varepsilon(t)'|F_{-\infty}\} = \Sigma$ , a.s.

then, putting $\rho(t) = \gamma(t)/\gamma(0)$,

$$\max_{1\le t\le H(t)} |r(t)-\rho(t)| = 0\{(\log\log T/T)^{1/2}\},\ \text{a.s.};$$

$$H(T) = 0\{(\log T)^{a}\} \ , \ a < \infty.$$

In case $s > 1$ put

$$\dot{\Sigma} = \frac{1}{T}\sum_{1}^{T} \varepsilon(t)\varepsilon(t)' \ , \quad \dot{\Gamma}(t) = \sum_{0}^{\infty} K(j)\ \dot{\Sigma}\ K(j+t)'.$$

Then under the same conditions

$$\max_{1\le t\le H(T)} \|\hat{\Gamma}(t) - \dot{\Gamma}(t)\| 0\{(\log\log T/T)^{1/2}\},\text{a.s.},$$

$$H(T) = 0\{(\log T)^{a}\} \ , \ a < \infty.$$

Using these results we may establish for I of section 2 that, for BIC, $h_T(2\log\lambda)/\log T \to 1$, a.s., where $\lambda$ is the modulus of a zero of $\det k(z)$ nearest to $|z| = 1$. For AIC the same holds, at least in probebility, under somewhat stronger conditions, as was earlier shown for s=1 in /18/. Moreover if $k(z)^{-1} = \Sigma_0^{\infty} L(j)z^j$ then $\|\hat{L}_T(j)-L(j)\| =$ $= 0\{(\log\log T/T)^{1/2}\}$ uniformly in $1 \le j \le h_T$. (See /4/, /12/, for these results and related ones.) As a result in II, using BIC, then $p_T \to \max_j n_j$, where the $n_j$ are the dynamical indices, at least in probability. (See /14/ for details). At stage III analogous results hold so that in IV we may act as if n or the $n_j$ or p,q are known in developing the asymptotic theory.

The methods based on I, II, III, IV have been applied to real data and simulations and work well. In an example with s=2, $k \epsilon U(2)$, $\lambda = 0.8$ the following results were obtained for 1. Here $\tilde{h}_A(T)$, $\tilde{h}_B(T)$ are obtained by solving $\log\det\Sigma_{h+1} - \log\det\Sigma_h + 2ns\,C(T)/T = 0$. They are much smaller than $\ell(T) = \log T/2\log\lambda$. There are 50 replications.

## TABLE 1

VECTOR ARMA MODELS

| T | AIC Med | AIC Min | AIC Max | $\tilde{h}_A(T)$ | BIC Med | BIC Min | BIC Max | $\tilde{h}_B(T)$ |
|---|---|---|---|---|---|---|---|---|
| 100 | 4 | 2 | 9 | 3 | 2 | 1 | 3 | 2 |
| 500 | 7 | 5 | 14 | 6 | 4 | 3 | 6 | 4 |
| 1000 | 8 | 6 | 12 | 7 | 5 | 4 | 7 | 5 |

The following table shows results at Stage II when s=1, $y(t) - 4y(t-1) + .6y(t-2) = \varepsilon(t) + \varepsilon(t-1) + .6\varepsilon(t-2)$, again with 50 replications.

## TABLE 2

SCALAR ARMA ORDER ESTIMATION

| | Stage II | | | | Repeat Stage II | | | Stage IV | | |
|---|---|---|---|---|---|---|---|---|---|---|
| order | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 1 | 2 | 3 |
| sample size | | | | | | | | | | |
| 200 | - | 40 | 10 | - | - | 48 | 2 | - | 49 | 1 |
| 500 | - | 35 | 15 | - | - | 46 | 4 | - | 49 | 1 |
| 1000 | - | 36 | 14 | - | - | 46 | 4 | - | 49 | 1 |

The adjustment to $\tilde{t}_p$ were not made in this calculation and would reduce the over estimation at Stage II.

The question remains of the use of AIC or BIC. If there is a true model (which is most unlikely) then BIC is to be preferred at stages II or III (or IV). Indeed at stage II AIC will be inconsistent in rather a shocking manner if there is a true model as shown in /10/, /11/. Here we consider only the case s=1. Let $\delta > 0$ be such that $|\Psi(1)| < 1-\delta$. Then if $p_T = n_T$ is chosen by AIC

$$\lim_{\delta \to 0} \lim_{T \to \infty} P\{n_T > n\} = 1.$$

However there will be no true order in practice and the best choice is not apparent. For an autoregression Shibata, /18/, has shown strong arguments for AIC when s=1 and there is no true order.

## 4. REFERENCES.


/1/ AKAIKE, H.: "Canonical correlation analysis of time series and the use of an information criterion." Pages 27-29 in System Identification: Advances and Case Studies; C.R.K. Mehra and D.G. Lainiotis eds., Academic Press, New York (1976).

/2/ AKAIKE, H.: "Markovian representations of stochastic processes and its aplication to the analysis of autoregressive moving average processes". Ann. Inst. Statist. Math. 26, 363-387. (1974).

/3/ ANDERSON, B.D.O. AND MOORE, J.B.: "Optimal Filtering", Addison-Wesley, New York. (1979).

/4/ AN, HONG-ZHI; CHEN, ZHAO-GUO AND HANNAN, E.J.: "Autocorrelation, autoregression and autoregressive approximation. Ann. Statist. 10, 926-36. (1982).

/5/ AN, HONG-ZHI; CHEN, ZHAO-GUO AND HANNAN, E.J.: "A note on ARMA estimation. J.Time Series Anal. 4. (1983).

/6/ BOX, G.E.P. AND JENKINS, G.M.: "Time Series Analysis, Forecasting and Control". Holden-Day, San Francisco. (1979).

/7/ CLARK, J.M.C.: "The consistent selection of parameterization in systems identification". Paper presented at JACC Purdue University. (1976).

/8/ DEISTLER, M. AND HANNAN, E.J.: "Some properties of the parameterization of ARMA systems with unknown order". J. Multivariate Anal. 11, 474-484. (1981).

/9/ DURBIN, J.: "The fitting of time series models". Rev.Int.Inst.Statist. 28, 233-244. (1960).

/10/HANNAN, E.J.: "Testing for autocorrelation and Akaike's criterion". Pp 403-12 in Essays in Statistical Science, eds. J.M. Gani and E.J. Hannan Applied Probability Trust, Sheffield. (1982).

/11/HANNAN, E.J.: "Estimating the dimension of a linear system". J. Multivariate Anal 11, 458-73. (1981).

/12/HANNAN, E.J. AND RISSANEN, J.: "Recursive estimation of ARMA order". Biometrika, 69 81-94. (1982).

/13/HANNAN, E.J. AND KAVALIERIS, L.: "The convergence of autocorrelations and autoregressions". Aust.J.Statist.25, 287-297. (1983).

/14/HANNAN, E.J. AND KAVALIERIS, L.: "Multivariate linear time series models". Adv. Applied Probability. (1984).

/15/VAN OVERBEEK, A.J.M.: "On the Structure Selection for the Identification of Multivariable Systems". Department of Electrical Engineering, Linköpping, Sweden. (1982).

/16/NICHOLS, D.F. AND QUINN, B.G.: "Random coefficient autoregressive models". Springer-Verlag, New York. (1983).

/17/SOLO, V.: "The exact likelihood for a multivariate ARMA model". Research report Harvard. (1984).

/18/SHIBATA, R.: "Asymtotically efficient estimation of the order of the models for estimating parameters of a linear process" Ann. Statist. 8, 147-64. (1980).

/19/ WHITTLE, P.: "On the fitting of multivariate autoregression and the approximate canonical factorisation of a spectral density matrix". Biometrika, 50, 129-34. (1973).