The Annals of Statistics
1975, Vol. 3, No. 4, 975-981

THE ESTIMATION OF ARMA MODELS[1]

BY E. J. HANNAN

The Australian National University

In estimating a vector model, $\sum B(j)x(n-j) = \sum A(j)\varepsilon(n-j)$, $A(0) = B(0) = I_r$, $E(\varepsilon(m)\varepsilon(n)') = \delta_{mn}K$, it is suggested that attention be confined to cases where $g(z) = \sum A(j)z^j$, $h(z) = \sum B(j)z^j$ have determinants with no zeroes inside the unit circle and have I_r as greatest common left divisor and where $[A(p) \vdots B(q)]$ is of rank r, where p, q are the degrees of g, h, respectively. It is shown that these conditions ensure that a certain estimation procedure gives strongly consistent estimates and the last of the conditions is probably necessary for this to be so, when the first two are satisfied. The strongly consistent estimation procedure may serve to initiate an iterative maximisation of a likelihood.

Let $x(n)$ be a strictly stationary, ergodic, vector time series of r components that is generated by the autoregressive-moving average (ARMA) model

$$\sum_0^q B(j)x(n-j) = \sum_0^p A(j)\varepsilon(n-j)\,, \qquad A(0) = B(0) = I_r\,, \qquad E\{\varepsilon(m)\varepsilon(n)'\} = \delta_{mn}K\,. \tag{1}$$

Since mean corrected quantities are used in all statistics introduced below we assume $E\{\varepsilon(n)\} = 0$. Put

$$h(z) = \sum_0^q B(j)z^j\,, \qquad g(z) = \sum_0^p A(j)z^j$$

so that the spectral density matrix is

$$f(\omega) = \frac{1}{2\pi} h(e^{i\omega})^{-1} g(e^{i\omega}) K g^*(e^{i\omega}) h^*(e^{i\omega})^{-1}\,. \tag{2}$$

In future, for brevity, $\exp i\omega$ will often be omitted in such formulae as on the right of (2). In estimating the $A(j)$, $B(j)$ and K we may form the likelihood as if the $\varepsilon(n)$ were Gaussian and maximise this, even though these Gaussian assumptions are not maintained. This will be called "maximising the quasi-likelihood." Structures (1) are equivalent (i.e. give rise to the same quasi-likelihood for all samples) if and only if they correspond to the same $f(\omega)$ since $f(\omega)$ determines and is determined by the covariances. A structure, (1), is identified when a unique member of each equivalence class is chosen. The purpose of this note is to interpret an identification condition given in [1], to point out that if this condition is met then a certain estimation procedure, which is described below, gives (strongly) consistent estimates of the $A(j)$, $B(j)$, K and thus may serve to

Received May 1974; revised October 1974.

[1] This work was done while the author was visiting the Cowles Foundation and the Statistics Department at Yale.

AMS 1970 *subject classifications*. Primary 62M10, 62M15; Secondary 62E20.

Key words and phrases. Autoregressive-moving average process, identification, strongly consistent estimation.

initiate an iterative maximisation of the quasi-likelihood. As a consequence, the (in general) complex problem of parameterising the equivalence classes can be avoided. The condition is probably also necessary for the consistency of the initial procedure but this is not easy to prove. There seems to be some misunderstanding concerning the importance of the identification problem for in [5, page 491] it is said that this is "one of the major outstanding problems of multivariate temporal systems" and in [4, page 138] it said that "a complete and practically convenient solution to the unidentifiability problem is still awaited." Both statements seem unjustified. However, the estimation of (1) may yet prove difficult because the initial estimates that are provided for the $A(j)$, $B(j)$ (or other initial estimates) may be relatively inefficient and may not, in "small" samples, provide a good starting point for iterative solution of the likelihood equations. This problem is, however, related to the number of variables involved and the sample size and is not an identification problem.

A canonical choice of a factorisation of (2) may be made in the following way, following [3]. In future we shall call $g(z)$, $h(z)$, etc. polynomials though they are matrices of polynomials. We explain the meaning of the conditions below. (Write $\det A$ for the determinant of A.)

(3.i) $\det\{g(z)\}$ has no zero inside the unit circle and $\det h(z)$ has no zeros in or on that circle.

(3.ii) $g(z)$, $h(z)$ have I_r as greatest common left divisor (g.c.l.d.).

(3.iii) $h(z)$ is lower triangular with $h_{ij}(z)$ of no higher degree than $h_{jj}(z)$.

(Recall that we have already required $g(0) = h(0) = I_r$.)
As is well known (see [3, Chapter III]) (3.i) is necessary and sufficient in order that $x(n)$ should be (linearly) expressible in terms of $\varepsilon(m)$, $m \leq n$ and that the $\varepsilon(n)$ should be the linear prediction errors. The condition (3.ii) eliminates redundancies in that it eliminates any common left factor, $e(z)$ such that $h = eh_1$, $g = eg_1$, where $e(z)$ has nonconstant determinant. (For more detail see [6, page 35].) One cannot, of course, "divide out" a left factor that has constant determinant (i.e. is unimodular) since such a factor divides any matrix of polynomials. The condition (3.iii) eliminates this indeterminary. It seems that (3.i) and (3.ii) will not conflict with, but rather will agree with, any prior physical meaning given to the structure (1). This may not be true of (3.iii) since one would not necessarily be led to a triangular form for h and if there were prior constraints on the $A(j)$, $B(j)$ they will become unrecognisable when h is transformed to triangular form. The parametrisation of the equivalence classes defined by (3) also is more complex basically because of the large number of integer parameters. This leads to the following theorem.

THEOREM 1. *Let g, h satisfy* (2), (3.i), (3.ii) *and be such that p is as small as* *possible and q as small as possible for that p. The necessary and sufficient condition that g, h be uniquely defined by these requirements is that* $[A(p) \vdots B(q)]$ *be of rank r.*

This is little more than a rephrasing of the theorem in [1] (the latter theorem refers, more generally, to an arbitrary prescribed p, q). Of course the same result holds if the positions of p, q are reversed. The condition is overidentifying (see [3]) in the sense that there are equivalence classes for which it cannot be satisfied. In practice, data will not be generated by any such model as (1). We feel that no advantage in fitting such models will be gained by allowing for the additional equivalence classes in fitting the model since it is a priori very improbable that the best fit will be obtained at a boundary point corresponding to a less than full rank $[A(p) \vdots B(q)]$. Of course the same might be said of the stronger requirement that $A(p)$ and $B(q)$ be nonsingular. For the same kind of reason it seems no essential restriction to require that det $\{g(z)\}$ is never zero on the unit circle. (A model not meeting this condition will not be useful for linear prediction since errors in the initiation of the prediction process will propagate indefinitely.) If the conditions in the theorem and the condition of $g(z)$ are met and if we can find strongly consistent initial estimates of the $A(j)$, $B(j)$ and K a standard iterative solution of the likelihood equations, commencing from this initial estimate, will reach the solution of the likelihood equation, at least for a sufficiently large sample (of a size depending on the realisation of the process). This iteration may proceed through small changes in the elements of the $A(j)$, $B(j)$ and K without concern for the identifying conditions.

The initial estimates of which we speak may be constructed as follows. Put

$$C(n) = \frac{1}{N} \sum_{m=1}^{N-n} \{x(m) - \bar{x}\}\{x(m+n) - \bar{x}\}' = C(-n)' , \qquad n \geqq 0 .$$

Solve

$$\sum_0^q \hat{B}(j)C(j-k) = 0 , \qquad k = p+1, p+2, \cdots, p+q; \ \hat{B}(0) = I_r . \tag{4}$$

Then form

$$\hat{y}(n) = \sum_0^q \hat{B}(j)x(n-j) , \qquad \hat{C}(n) = \frac{1}{N} \sum_{q+1}^{N-n} \hat{y}(m)\hat{y}(m+n)$$

$$\hat{w}(\omega_t) = N^{-\frac{1}{2}} \sum_{q+1}^{N} \hat{y}(n)e^{in\omega_t} , \qquad \omega_t = 2\pi t/N' , \qquad N' \geqq N + q ,$$
$$-\tfrac{1}{2}N' < t \leqq [\tfrac{1}{2}N'] .$$

Here $[x]$ means the integral part of x. Of course N' may be chosen much greater than N to make the computations cheap (see [2, pages 263–273]). Following [2, pages 383–388] put

$$\hat{f}_y(\omega_t) = \frac{1}{2\pi} \sum_{-p}^{p} \hat{C}(n)e^{-in\omega_t}$$

and solve the equations

$$\sum_{j=0}^{p} \left\{\frac{1}{N'} \sum_t \hat{f}_y(\omega_t)^{-1}\hat{w}(\omega_t)\hat{w}(\omega_t)^*\hat{f}_y(\omega_t)^{-1}e^{i(j-k)\omega_t}\hat{A}(j)\right\} = 0 , \tag{5}$$
$$\hat{A}(0) = I_r, \ k = 1, 2, \cdots, p .$$

Finally form

$$\hat{K}^{-1} = \frac{1}{4\pi^2 N'} \sum_t \{\hat{f}_y(\omega_t)^{-1}\hat{w}(\omega_t)\hat{w}(\omega_t)^*\hat{f}_y(\omega_t)^{-1} \sum_{j=0}^{p} \hat{A}(j)e^{ij\omega_t}\} . \tag{6}$$

An alternative to (5) and (6) is to factor $\hat{f}_y$ as $(2\pi)^{-1}\hat{g}\hat{K}\hat{g}^*$ with $\hat{g}$ satisfying (3.i) and (3.ii). Such a factorisation is possible if and only if $\hat{f}_y(\omega) \geqq 0$, $\omega \in [-\pi, \pi]$. See [2, page 66]. Under the conditions of the theorems $\hat{f}_y(\omega) > 0$ will always hold for sufficiently large N. All of the statements about (4)—(6) in the theorem below are true of the estimates obtained from (4) and such a factorisation.

THEOREM 2. *Let $x(n)$ be ergodic with g, h satisfying* (3.i), (3.ii), *and with* $\det\{g(z)\} \neq 0$, $|z| = 1$. *Let the $\hat{B}(j)$, $\hat{A}(j)$, $\hat{K}$ be obtained from* (4), (5), (6). *A sufficient condition that these be strongly consistent is that $[A(p)\vdots B(q)]$ be of rank r.*

PROOF. If it can be shown that the equations

$$\sum_0^q D(j)\Gamma(j-k) = 0 , \qquad k = p+1, p+2, \cdots, p+q,\ D(0) = I_r, \quad \Gamma(n) = \int_{-\pi}^{\pi} e^{in\omega}f(\omega)\,d\omega , \tag{7}$$

have a unique solution (namely $D(j) = B(j)$) then the proof may be completed as follows. First the strong consistency of the $\hat{B}(j)$ follows from the almost sure convergence of the $C(j-k)$ to the $\Gamma(j-k)$, which in turn follows from ergodicity. It then follows that $\hat{C}(n)$ converges almost surely to $\Gamma_y(n) = E\{\sum A(j)\varepsilon(m-j) \sum \varepsilon'(m+n-j)A(j)'\}$. Thus $\hat{f}_y(\omega)$ converges almost surely, and uniformly in ω, to $(2\pi)^{-1}gKg^*$. As will now be shown the coefficient matrix of $\hat{A}(j)$ in (5) then converges to

$$2\pi \int_{-\pi}^{\pi} (gKg^*)^{-1}e^{i(j-k)\omega}\,d\omega . \tag{8}$$

Indeed since $\det\{f(\omega)\} \neq 0$, $\omega \in [-\pi, \pi]$, then $\hat{f}_y(\omega)^{-1}$ converges almost surely, and uniformly in ω, to $f_y(\omega)^{-1}$. Thus the Cesaro sum, $\hat{\Phi}_M(\omega)$, to M terms, of the Fourier series of $\hat{f}_y(\omega)^{-1}$ satisfies, for M large enough, $\|\hat{f}_y(\omega)^{-1} - \hat{\Phi}_M(\omega)\| < \varepsilon$, for any $\varepsilon > 0$ and all sufficiently large N, where $\|\cdot\|$ is any norm in a finite dimensional vector space. Putting $\Phi_M(\omega)$ for the Cesaro sum of the Fourier series for $f_y(\omega)^{-1}$ we have

$$\begin{aligned}\frac{1}{N'} \sum_t \hat{\Phi}_M(\omega_t)\hat{w}(\omega_t)\hat{w}(\omega_t)^*\hat{\Phi}_M(\omega_t)^*e^{i(j-k)\omega_t} &= \sum_{-M}^{M}\sum \hat{C}(m)\{\hat{C}(m-j+k-n) + \hat{C}(m-j+k-n+N') \\ &\quad + \hat{C}(m-j+k-n-N')\}\hat{C}(n)'\left(1-\frac{m}{M}\right)\left(1-\frac{n}{M}\right),\end{aligned}$$

where $\hat{C}(n)$ is null for $|n| \geqq N$. The convergence of the $\hat{B}(j)$ ensures that the second and third terms in the factor in braces converge to the null matrix and the strong consistency of the $\hat{C}(n)$ then gives

$$\lim_{N\to\infty} \frac{1}{N'} \sum_t \hat{\Phi}_M(\omega_t)\hat{w}(\omega_t)\hat{w}(\omega_t)^*\hat{\Phi}_M(\omega_t)^*e^{i(j-k)\omega_t} = \int_{-\pi}^{\pi} \Phi_M(\omega)gKg^*\Phi_M(\omega)^*e^{i(j-k)\omega}\,d\omega , \quad \text{a.s.}$$

The error introduced by replacing $\hat{f}_y$ by $\hat{\Phi}_M$, in forming the expressions under the limit sign on the left, has norm dominated by

$$a\varepsilon \frac{1}{N'} \sum_t \operatorname{tr} \{\hat{w}(\omega_t)\hat{w}(\omega_t)^*\} = a\varepsilon\hat{C}(0) ,$$

where a is a constant. A similar result holds on the right and the assertion concerning (8) now follows. This shows that the $\hat{A}(j)$ converges almost surely to the $A(j)$ since these are the unique solutions of

$$\sum_{j=0}^{p} \{\int_{-\pi}^{\pi} (gKg^*)^{-1} e^{i(j-k)\omega}\, d\omega\} A(j) = 0 , \qquad A(0) = I_r , \qquad k = 1, 2, \cdots, p .$$

That $\hat{K}$ converges almost surely to K now results. Thus the theorem follows from the following lemma.

LEMMA. *If $x(n)$ satisfies* (1), (3.i) *and* (3.ii) *then the necessary and sufficient condition that* (7) *have the unique solution $D(j) = B(j)$ is the condition that* $[A(p) \vdots B(q)]$ *be of rank r.*

Sufficiency. If the result does not hold then the matrix

$$[\Gamma(j - k - p)]_{j,k=1,\cdots,q}$$

is singular, where we have exhibited the block in row j column k in the $(rq \times rq)$ matrix. Reversing the order of rows and columns and transposing we obtain the matrix with $\Gamma(j - k - p)'$ as the (j, k)th block which is the same as the matrix

$$[\Gamma(k + p - j)]_{j,k=1,\cdots,q} .$$

Thus there are matrices $H(j)$ so that

$$\sum_1^q H(j)\Gamma(k + p - j) = 0 , \qquad k = 1, \cdots, q . \tag{9}$$

However, also

$$\Gamma(k + p - j) = -\sum_1^q \Gamma(k + p - j - u)B(u)' , \qquad k > j ,$$

and thus

$$\sum_1^q H(j)\Gamma(k + p - j) = 0 , \qquad k > 0 . \tag{10}$$

Put

$$\sum_1^q H(j)z^{q-j} = \tilde{h}(z) .$$

Then $\tilde{h}$ is of degree $q - 1$. From (10) we have

$$\int \tilde{h}f(\omega)e^{i(k+p-q)\omega}\, d\omega = 0 , \qquad k > 0 . \tag{11}$$

Thus

$$\int \tilde{h}f(\omega)\tilde{h}^* e^{i(k+p)\omega}\, d\omega = 0 , \qquad k > 0 . \tag{12}$$

It follows from (11) that $\tilde{h}(z)h(z)^{-1}g(z)Kg'(z^{-1})h'(z^{-1})^{-1}z^{p-q}$ is analytic within the unit circle. Put $g'(z^{-1})z^a = g_1(z)$ where a is the degree of $g(z)$. Put $h'(z^{-1})z^b = h_1(z)$ where b is the degree of $h(z)$. Then either $a = p$ or $b = q$ (or both) and now $\tilde{h}h^{-1}gKg_1h_1^{-1}z^c$ is analytic where $c = p - q - a + b$. It is shown in [1] that if $h^{-1}g = \hat{h}^{-1}\hat{g}$ and the pairs (g, h), $(\hat{g}, \hat{h})$ each satisfy (3.i), (3.ii) then $h = u\hat{h}$, $g = u\hat{g}$ where u is unimodular. Choose v, w, both unimodular, so that $h^{-1}g = v^{-1}Q^{-1}Pw$ where P, Q are diagonal with terms in the same place in the diagonal having no common factor ([6, page 41]). We may also ensure that q_j divides q_{j+1} where q_j is the jth diagonal element of Q. It follows, putting $\hat{h} = Qv$, $\hat{g} = Pw$, that $g = uPv$, $h = uQw$. Thus $g_1(z)h_1(z)^{-1} = v'P(z^{-1})z^a\{Q(z^{-1})z^b\}^{-1}w'^{-1} = v'\{Q(z^{-1})z^b\}^{-1}\{P(z^{-1})z^a\}w'^{-1}$. We may also find x, unimodular, so that $h^{-1}gKv' = xt$ where t is lower triangular ([6, page 32]). Thus

$$\tilde{h}h^{-1}gKg_1h_1^{-1}z^c = \tilde{h}xt\{Q(z^{-1})z^b\}^{-1}P(z^{-1})z^a w'^{-1} .$$

Now it must be true that $\tilde{h}xt\{Q(z^{-1})z^b\}^{-1}$ is a polynomial and equals $\tilde{h}x\{Q(z^{-1})z^b\}^{-1}t$ and also $\tilde{h}x\{Q(z^{-1})z^b\}^{-1}$ must be a polynomial. Indeed the zeros of the diagonal of $Q(z^{-1})z^b$ lie inside the unit circle while those of the diagonal of t lie outside of it. Thus q_r divides the last column of $\tilde{h}x$. However, the second last column of $\tilde{h}xt$ is $a_{i,r-1}t_{r-1,r-1} + a_{i,r}t_{r-1,r}$, $i = 1, \cdots, r$, where a_{ij}, t_{tj} are the typical elements of $\tilde{h}x$ and t, respectively. Since q_{r-1} divides q_r which divides $a_{i,r}$ and q_{r-1} has no linear factor dividing $t_{r-1,r-1}$ then q_{r-1} divides $a_{i,r-1}$, $i = 1, \cdots, r$. Continuing in this way we establish what we want. Now we observe that the coefficientm atrix of the highest power of z in $Q(z^{-1})z^b$ is I_r. Since $\tilde{h}$ is of degree $q - 1$ and x is unimodular then $b \leqq q - 1$, since otherwise the determinants of $Q(z^{-1})z^b$ will be of degree qr while that of $\tilde{h}$ can be at most $(q - 1)r$. Thus $a = p$ and, moreover, $A(p)$ is nonsingular (since $B(q)$ is null because $b < q - 1$). Now we may also take $Q(z)^{-1}$ through the factors on its right and cancel it with $\tilde{h}'(z^{-1})$ reducing $\tilde{h}h^{-1}gKg^*h^{*-1}\tilde{h}^*$ to the form $kgKg^*k^*$ where k is a polynomial also. However, from (12) we see that this is a trigonometric polynomial of degree at most $(p - 1)$ and this is impossible unless k and hence $\tilde{h}$ is null, since otherwise kg must be at least of degree p, because A_p is nonsingular. Thus the sufficiency is proved.

Necessity. Let α be a vector such that $\alpha'[A(p)\vdots B(q)] = 0$. Consider

$$\int \alpha' hh^{-1}gKg^*h^{*-1}e^{-ik\omega}\,d\omega = \int \alpha' gKg^*h^{*-1}e^{-ik\omega}\,d\omega = 0\,, \qquad k > p\,,$$

since $\alpha'g$ is of degree $p - 1$, only. Thus $h + 1\alpha'h$, where 1 is the vector of units, gives a second solution to (7) and the necessity also is established in the lemma.

It seems very likely that the condition of the theorem is also necessary. However, that is not easy to prove, since (4) may be expected to have a unique solution even when (7) does not and it is not apparent that the deviation of $C(j - k)$ from $\Gamma(j - k)$ might not be such as to ensure that the unique solution of (4) converges to the particular solution, $B(j)$, of (7).

It may be mentioned, that in distinction to the case $p = 0$, it is no longer true that for all N,

$$\hat{h}(z) = \sum_0^q \hat{B}(j)z^j$$

has determinant with all zeros outside of the unit circle (as the simplest examples show). However, the $\hat{A}(j)$ do have this property. One could carry out a further step that obtains obtains from the $B(j)$, $A(j)$ a new estimate of h having the property, but we will not go into that here.

REFERENCES

[1] HANNAN, E. J. (1969). The identification of vector mixed autoregressive-moving average systems. *Biometrika* **56** 223-225.

[2] HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

[3] HANNAN, E. J. (1971). The identification problem for multiple equation systems with moving average errors. *Econometrica* **39** 751-766.

[4] KENDALL, M. G. (1973). *Time Series Analysis*. Griffin, London.

[5] KENDALL, M. G. and STUART, A. S. (1966). *The Advanced Theory of Statistics*, **3**. Griffin, London.

[6] MACDUFFEE, C. C. (1956). *The Theory of Matrices*. Chelsea, New York.

DEPARTMENT OF STATISTICS
INSTITUTE FOR ADVANCED STUDIES
AUSTRALIAN NATIONAL UNIVERSITY
P.O. BOX 4
CANBERRA, 2600, AUSTRALIA