COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO: 291

THE ESTIMATION OF MIXED REGRESSION, AUTOREGRESSION, MOVING AVERAGE AND DISTRIBUTED LAG MODELS

E.J. Hannan and D.F. Nicholls

April 9, 1970

THE ESTIMATION OF MIXED REGRESSION, AUTOREGRESSION, MOVING AVERAGE AND DISTRIBUTED LAG MODELS*

by

E.J. Hannan and D.F. Nicholls**

1. Introduction

In this paper, which is closely related to Hannan (1969), we shall deal with the estimation of the parameters in the model

$$(1) \qquad \sum_{0}^{q}\beta(j)y(n-j) + \sum_{1}^{r}\delta(j)x_j(n) = \sum_{0}^{p}\alpha(j)\epsilon(n-j) , \quad \alpha(0) = \beta(0) = 1 ,$$

assuming that we have observations for $n = 1, \ldots, N$ on $x_j(n)$ and $y(n)$. We also make the following assumptions:

(i) The $\epsilon(n)$ are i.i.d. $(0, \sigma^2)$.

(ii) All zeros of the z transforms

$$g(z) = \sum_{0}^{p}\alpha(j)z^j , \qquad h(z) = \sum_{0}^{q}\beta(j)z^j$$

*The research described in this paper was carried out under grants from the National Science Foundation and from the Ford Foundation.

**Of the Australian National University, Canberra, Australia.

lie outside of the unit circle.

(iii) The $x_j(n)$ come from infinite sequences which satisfy, almost surely,

$$\lim_{N\to\infty} N^{-1} \sum_{m=1}^{N} x_j(m)x_k(m+n) = \gamma_{jk}(n) = \gamma_{kj}(-n) , \quad j, k = 1, \ldots, r , \quad n = 0, 1, \ldots,$$

independently of the $\epsilon(n)$ sequence.

As a consequence of (iii) we have

$$\gamma_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF_{jk}(\lambda) .$$

The matrix $F(\lambda)$ with $F_{jk}(\lambda)$ as entry in row j column k is, of course, a spectral distribution matrix.

The condition (ii) is reasonable in the following sense. If the condition on $g(z)$ is not imposed then it is possible to re-express the right hand side of (1) in terms of a new <u>orthogonal</u> sequence with zero mean and constant variance, which we call $\epsilon_1(n)$, so that the right hand side of (1) may be written

$$\sum_{0}^{p_1} \alpha_1(j)\epsilon_1(n-j) , \quad p_1 \leqq p$$

and for which $g_1(z) = \Sigma\alpha_1(j)z^j$ has all of its zeros on or outside of the unit circle. Moreover the $\epsilon_1(n)$ will be the linear, one step, prediction errors for the right hand side of (1) so that they will be the linear prediction errors for $y(n)$ also when the $x_j(n)$ sequences are known. If the data is Gaussian then the $\varepsilon_1(n)$ sequence will be independent also but in any case it seems as reasonable to hope that they are independent as to hope that the $\varepsilon(n)$ are so. The exclusion of zeros on the unit circle from $g(z)$ is required because the other case causes difficulty with the method which we present. It is felt that if there is a zero on the unit circle the location of this zero will be known beforehand, in which case the method we introduce below can be modified to allow for it. The second requirement in (ii) is necessary if there is to exist a <u>stationary</u> solution, $y(n)$, to (1), which is expressible in terms of the $\varepsilon(n-k)$, $x_j(n-k)$, $k \geq 0$. (Of course the possibility of stationarity will depend on the $x_j(n)$ sequences also.)

It seems preferable to treat the $x_j(n)$ as fixed sequences as in (iii), rather than to prescribe them stochastically as this makes the treatment less restrictive. One could begin from a more general specification such that the $y(n)$ and $x_j(n)$ in (1) are residuals from a regression. For example we might have

$$y(n) = \tilde{y}(n) + \mu' z(n)$$

where μ is a vector of regression coefficients and $z(n)$ is a vector of exogenous variables, so that it is $\tilde{y}(n)$ which is observed. Under reasonable conditions on the $x_j(n)$ in (1) it will be possible to remove the effects of $z(n)$ first by regression and then to work with the estimates of $y(n)$ obtained from the computed regression and the results of this paper will continue to hold. In this way non-stationary (trend) components in $\tilde{y}(n)$ can be allowed for. We shall not deal with this explicitly below except for the case where $z(n) \equiv 1$ (i.e. except for a mean correction). Of course such a $z(n)$ could also be incorporated among the $x_j(n)$ but it will be equally as efficient and computationally simpler to account for it by a mean correction.

The equation (1) may be regarded as an alternative form of the distributed lag model

$$(2) \qquad y(n) = \bar{\mu}(n) + \sum_{j=1}^{r} \delta_j \sum_{k=0}^{\infty} a_k x_j(n-k) + u(n)$$

where

$$\sum_0^{\infty} a_k z^k = h(z)^{-1}, \quad \sum_0^{q} \beta(j) u(n-j) = \sum_0^{p} \alpha(j) \epsilon(n-j)$$

and $\mu(n)$ is a solution of the homogeneous equation

$$\sum_0^{q} \beta(j) \mu(n-j) = 0 .$$

This solution decays to zero at an exponential rate as n increases. In particular we might have $x_j(n) = x(n-j)$ in which case we are dealing with the general rational distributed lag model (we put $\mu(n) \equiv 0$),

$$y(n) = \delta \sum_0^{\infty} \lambda_j x(n-j) + u(n)$$

$$(3) \qquad \sum_0^{\infty} \lambda_j z^j = \{\delta(1)^{-1} \sum_1^{r} \delta(j) z^{j-1}\} / \{\sum_0^{q} \beta(j) z^j\} ,$$

with $u(n)$ as before. Of course the model (1) (or its equivalent form (2)) covers also the case of more than one exogenous variable but the generating functions (3) for these different variables will all have the same denominator. (More general cases can be dealt with but lead to very complex calculations.) A special case is that where $p = q$ and $\alpha(j) \equiv \beta(j)$ so that $u(n) = \epsilon(n)$. We shall discuss this special case further in section 3. In that section we shall also discuss problems associated with another way in which there might be unrecognized restrictions on the $\alpha(j)$ and $\beta(j)$, namely that where $\alpha(p)$ or $\beta(q)$ is zero.

Finally in this section we shall describe the (restricted) sense in which we shall speak of (asymptotically) efficient estimates. If the $\epsilon(n)$ and $x_j(n)$ are Gaussian and the $x_j(n)$ processes are suitably restricted we may set up the likelihood function and obtain the maximum likelihood estimators. With small modifications these estimators may be shown to have a limiting normal distribution, after appropriate normalization. (This is discussed in Walker (1964). He deals basically with the univariate case but it is clear that the results extent to the multivariate case.) For example the covariance matrix for the estimators $\tilde{\delta}$, $\tilde{\beta}$ of δ and β in the model (2) becomes

$$(4) \qquad \left\{ \int_{-\pi}^{\pi} \frac{\sigma^2}{|1 - \alpha e^{i\lambda}|^2} \begin{bmatrix} 1 & \dfrac{\beta e^{i\lambda}}{1 - \alpha e^{i\lambda}} \\ \dfrac{\beta e^{-i\lambda}}{1 - \alpha e^{i\lambda}} & \dfrac{\beta^2}{|1 - \alpha e^{i\lambda}|^2} \end{bmatrix} dF(\lambda) \right\}^{-1} .$$

The existing maximum likelihood results would cover the case of only rather special $F(\lambda)$. We shall establish central limit theorems for our estimates, under the conditions (i), (ii), (iii) and shall speak of them as asymptotically efficient if in their limiting distribution they have (4) as covariance matrix (or have the analogous matrix as covariance matrix if, for example, the more general model (1) is being considered).

The likelihood equations are non-linear and our estimates define a sequence of iterations which, no doubt, converge in some sense to these maximum likelihood estimators, but we do not establish that. By a direct examination of our estimates we shall be able to establish more than can at present be said to be true, in general, of maximum likelihood procedures. In particular we prove almost sure convergence and are able to deal with the rather general specification of the $x_j(n)$ in (iii) above. Of course we also prescribe an algorithm for computing our estimates.

2. The Estimation Procedure and the Central Limit Theorem.

The basic statistics on which our computations are based are the finite Fourier transforms

$$w_y(\omega_t) = \frac{1}{\sqrt{2\pi N}} \sum_{n=1}^{N} y(n)e^{in\omega_t}, \quad w_j(\omega_t) = \frac{1}{\sqrt{2\pi N}} \sum_{n=1}^{N} x_j(n)e^{in\omega_t},$$

$$\omega_t = 2\pi t/N, \quad t = 1, \ldots, [1/2\, N], \quad j = 1, \ldots, r.$$

We also put $w(\omega_{-t}) = \overline{w(\omega_t)}$ and

$$\sqrt{2\pi N}\, w_y(\omega_t) = u_y(\omega_t) + iv_y(\omega_t), \quad \sqrt{2\pi N}\, w_j(\omega_t) = u_j(\omega_t) + iv_j(\omega_t)$$

so that, for example,

$$u_j(\omega_t) = \sum_{n=1}^{N} x_j(n)\cos n\omega_t \; .$$

We shall often omit the ω_t argument, even in summations, for simplicity of printing. It is the u_y, u_j, v_y, v_j which are the basic building blocks and the w_y, w_j are introduced so that formulae can be written more perspicuously. We have omitted the terms for $t = 0$ since these are annihilated by mean correction while the others are not affected by this. We introduce the symbol κ_t, $t = 1, \ldots, N$, which is to be 1 unless $t = \frac{1}{2}N$ (i.e. N even) when it is to be $\frac{1}{2}$. We also put

$$I_y = |w_y|^2 = (2\pi N)^{-1}(u_y^2 + v_y^2) \; ,$$

$$I_j = w_y\bar{w}_j = (2\pi N)^{-1}\{c_j - iq_j\} \; , \qquad c_j = u_y u_j + v_y v_j \; , \qquad q_j = u_y v_j - v_y u_j$$

$$I_{jk} = w_j\bar{w}_k = (2\pi N)^{-1}\{c_{jk} - iq_{jk}\} \; , \quad c_{jk} = u_j u_k + v_j v_k \; , \quad q_{jk} = u_j v_k - v_j u_k \; .$$

We arrange the r quantities I_j in a column vector I and the r^2 quantities I_{jk} in the matrix I_x. Then

$$I = (2\pi N)^{-1}\{C - iQ\} \; , \quad I_x = (2\pi N)^{-1}\{C_x - iQ_x\}$$

where C, C_x, Q, Q_x are defined in terms of c_j, c_{jk}, q_j, q_{jk} as were I and I_x in terms of I_j, I_{jk}.

We also need the serial covariances

$$\hat{c}_y(n) = \hat{c}_y(-n) = (N-n)^{-1} \sum_{m=1}^{N-n} (y(m)-\bar{y})(y(m+n)-\bar{y}) , \quad n = 0, 1, \ldots, p+q ,$$

$$\hat{c}_{jk}(n) = \hat{c}_{kj}(-n) = (N-n)^{-1} \sum_{m=1}^{N-n} (x_j(m)-\bar{x}_j)(x_k(m+n)-\bar{x}_k) , \quad j, k = 1, \ldots, r ; \quad n = 0, 1, \ldots, p ,$$

$$\hat{c}_{yj}(n) = \hat{c}_{yj}(-n) = (N-n)^{-1} \sum_{m=1}^{N-n} (y(m)-\bar{y})(x_j(m+n)-\bar{x}_j) , \quad j = 1, \ldots, r ; \quad n = 0, 1, \ldots, p+q .$$

Our estimation procedure consists of two steps, of which only the second is iterated.

<u>Step 1</u>. We may estimate β , δ by means of

$$\sum_{k=1}^{q} \hat{c}_y(p+j-k)\hat{\beta}(k) + \sum_{k=1}^{r} \hat{c}_{yk}(p+j)\hat{\delta}(k) = -\hat{c}_y(j+p) , \quad j = 1, \ldots, q$$

$$\sum_{k=1}^{q} \hat{c}_{yj}(k)\hat{\beta}(k) + \sum_{k=1}^{r} \hat{c}_{jk}(0)\hat{\delta}(k) = -\hat{c}_{yj}(0) , \quad j = 1, \ldots, r , \quad \hat{\beta}(0) = 1 .$$

It is convenient to introduce the partitioned (row) vectors

$$\rho' = (\beta' : \delta') , \quad \hat{\rho}' = (\hat{\beta}' : \hat{\delta}')$$

wherein β has $\beta(j)$ in the j^{th} place, $j = 1, \ldots, q$ (and similarly for $\hat{\beta}$) and δ has $\delta(j)$ in the j^{th} place, $j = 1, \ldots, r$ (and similarly for $\hat{\delta}$) . We may thus write the equations for $\hat{\beta}$, $\hat{\delta}$ as

$$\hat{\rho} = -\hat{C}^{-1}\hat{c} .$$

We observe that $\hat{C}$ converges almost surely[1] to the matrix

$$G = \left[\begin{array}{c:c} \int_{-\pi}^{\pi} e^{i(p+j-k)\lambda} \left\{ \frac{\sigma^2}{2\pi} \frac{|g|^2}{|h|^2} d\lambda + \frac{\delta' dF(\lambda)\delta}{|h|^2} \right\} & -\int_{-\pi}^{\pi} \frac{e^{i(p+j)\lambda}}{h} \delta' dF(\lambda) \\ \hdashline -\int_{-\pi}^{\pi} \frac{e^{-ik\lambda}}{\bar{h}} dF(\lambda)\delta & \int_{-\pi}^{\pi} dF(\lambda) \end{array}\right]$$

Here j , k vary from 1 to q so that the top left hand corner part is a $q \times q$ matrix. Since $F(\lambda)$ is $r \times r$ the top right hand corner is $q \times r$ and the bottom right hand corner is $r \times r$. We have omitted the argument $\exp i\lambda$ from g and h , for convenience.

Our final assumption is:

(iv) G is non-singular.

For given F , h , h the vectors δ making G singular will lie in an $(r-1)$ dimensional subset so that a priori it seems most unlikely that G will be singular. In any case there is no shortage of "instrumental variables" to replace some or all of the $y(n-p-j)$, $j = 1, \ldots, q$; $x_j(n)$, $j = 1, \ldots, p$. In particular the $x_j(n-k)$, $k > 0$, could be used to replace the $y(n-p-j)$. The use of these would have one decided virtue, namely that the equations give valid estimates independent of the size of p so that if p is to be in-

[1] In future we shall often omit the qualification "almost surely" when we speak of this mode of convergence.

creased in a later calculation the initial estimate of ρ does not have to be recomputed. Thus if $r = 1$, for example, we might choose to replace the first q equations defining $\hat{\rho}$ by

$$\sum_1^q \hat{c}_{y1}(k-j)\hat{\beta}(k) + \hat{c}_{11}(j)\hat{\delta}(1) = -\hat{c}_{1y}(j) , \quad j = 0, \ldots, q , \quad \hat{\beta}(0) = 1 .$$

For $r > 1$ if one wishes to use the $x_j(n-k)$, only, as instrumentals one will have to choose which set to adopt. Presumably the maximum lag will be kept as low as possible. We emphasize that Theorem 1 below will hold in its entirety no matter what system of equations is used out of those just discussed, provided only that (iv) is replaced by an equivalent condition ensuring that these initial equations for $\hat{\rho}$ have, asymptotically, a unique solution.

Once having obtained our first estimate $\hat{\rho}$ we next form

$$\hat{h}\left(e^{i\omega_t}\right) = \sum_0^q \hat{\beta}(j)e^{ij\omega_t} , \quad t = 1, \ldots, [\tfrac{1}{2} N] , \tag{5}$$

where, let us say, $\hat{h} = \hat{\xi}_h + i\hat{\eta}_h$. We also form the autocovariances

$$\hat{c}_w(n) = \sum_{j,k=0}^{q} \hat{\beta}(j)\hat{\beta}(k)\hat{c}_y(k-j) + \sum_{j,k=1}^{r} \hat{\delta}(j)\hat{\delta}(k)\hat{c}_{jk}(n)$$

$$+ \sum_{j=0}^{q} \sum_{k=1}^{r} \hat{\beta}(j)\hat{\delta}(k)\{c_{yk}(n+j) + c_{ky}(n-j)\} , \quad n = 0, 1, \ldots, p .$$

From these we obtain

$$\hat{f}_w(\omega_t) = \frac{1}{2\pi}\{\hat{c}_w(0) + 2\sum_{n=1}^{p}\hat{c}_w(n)\cos n\omega_t\} = \hat{f}_w(-\omega_t)\ ,\quad t = 1,\ 2,\ \ldots,\ [\tfrac{1}{2}N]\ .$$

The function $\hat{f}_w$ estimates the spectrum of

$$w(n) = \sum_0^p \alpha(j)\epsilon(n-j)\ .$$

It thus estimates a positive function. Nevertheless it may be negative, perhaps for a value other than some ω_t. If and only if it is positive may it be, uniquely, factored in the form

$$\hat{f}_w = \frac{\hat{\sigma}^2}{2\pi}\left|\sum_0^p \hat{\alpha}(j)e^{ij\omega_t}\right|^2 = \frac{\hat{\sigma}^2}{2\pi}|\hat{g}|^2\ ,\quad \hat{\alpha}(0) = 1\ ,$$

where all zeros of $\hat{g}(z)$ lie outside of the unit circle. We shall call $\hat{\alpha}$ the vector with $\hat{\alpha}(j)$ in the j^{th} place, $j = 1, \ldots, p$. Since the factorization may be troublesome (or impossible) we now show how to avoid it. We form

$$(6)\qquad \hat{a}(k) = \frac{2}{N}\sum_1^{[\frac{1}{2}N]}\kappa_t\{\cos k\omega_t|\hat{h}w_y + \sum_{j=1}^{r}\hat{\delta}(j)w_j|^2\}/\hat{f}_w^2$$

$$= (4\pi/N^2)\sum_1^{[\frac{1}{2}N]}\kappa_t\{\cos k\omega_t(|\hat{h}|^2 2\pi N I_y + \hat{\delta}'C_x\hat{\delta} + \hat{\delta}'C\hat{\xi}_h + \hat{\delta}Q\hat{\eta}_h)\}/(2\pi\hat{f}_w)^2$$

The second formula is written in a form eliminating factors whose introduction causes needless computing which are absorbed into the factor $4\pi/N^2$. Now let $\hat{A}$ have $\hat{a}(k-\ell)$ in row k column ℓ, $k, \ell = 1, \ldots, p$, and $\hat{a}$ have $\hat{a}(k)$ in row k, $k = 1, \ldots, p$ and put

$$\hat{\alpha} = -\hat{A}^{-1}\hat{a} .$$

We have here used $\hat{\alpha}$ as a name for two separate estimates of α but since only one of these will be computed and they occupy the same place in later formulae no confusion will result. This completes Step 1, which is not repeated in later iterations.

Step 2. From the $\hat{\alpha}(j)$ we have

$$\hat{g} = \sum_0^p \hat{\alpha}(j) e^{ij\omega_t} = \hat{\xi}_g + i\hat{\eta}_g ,$$

let us say. If $\hat{f}_w$ was not factored we estimate σ^2 by

$$(7) \qquad \hat{\sigma}^2 = \frac{4\pi}{N} \sum_{t=1}^{[\frac{1}{2}N]} \{\hat{f}_w / |\hat{g}|^2\}$$

which is, of course, in agreement with the definition when $\hat{f}_w$ is factored.

We next form

$$\hat{d}_{k\ell} = (2/N^2) \sum_1^{[\frac{1}{2}N]} \kappa_t \{2\pi N I_y \cos(k-\ell)\omega_t\} / |\hat{g}|^2 , \quad k, \ell = 1, \ldots, q$$

$$= (2/N^2) \sum_1^{[\frac{1}{2}N]} \kappa_t \{c_j \cos k\omega_t + q_j \sin k\omega_t\} / |\hat{g}|^2 , \quad k = 1, \ldots, q , \quad \ell = q+j , \quad j = 1, \ldots, r$$

$$= (2/N^2) \sum_1^{[\frac{1}{2}N]} \kappa_t c_{ij} / |\hat{g}|^2 , \quad k = q+i , \quad \ell = q+j , \quad i, j = 1, \ldots, r .$$

Finally $\hat{d}_{\ell k} = \hat{d}_{k\ell}$, which completes the definition (with some redundancy).

We define $\hat{\sigma}^2\hat{D}$ to have $\hat{d}_{k\ell}$ in row k, column ℓ, k, $\ell = 1, \ldots, q + r$.

We define $\hat{\sigma}^2\hat{d}$ to have $\hat{d}_k$ in row k where $\hat{d}_k = \hat{d}_{ok}$, $k = 1, \ldots, q + r$.

Thus,

$$d_k = (2/N^2) \sum_{1}^{[1/2N]} \kappa_t \{2\pi NI_y \cos k\omega_t\}/|\hat{g}|^2 , \quad k = 1, \ldots, q ,$$

$$= (2/N^2) \sum_{1}^{[1/2N]} \kappa_t c_j/|\hat{g}|^2 , \quad k = g + j , \; j = 1, \ldots, r .$$

We now put

$$\hat{\rho}^{(1)} = -\hat{D}^{-1} \hat{d} ,$$

so that the factor $\hat{\sigma}^2$ does not need to be introduced (for the factors $2/N^2$). We have defined $\hat{D}$ and $\hat{d}$ in the above manner because this definition is needed later.

We now use $\hat{\beta}^{(1)}$ to compute $\hat{h}^{(1)} = \hat{\xi}_h^{(1)} + i\hat{\eta}_h^{(1)}$ as we used $\hat{\beta}$ to compute $\hat{h}$ in (5) and we then compute $\hat{A}^{(1)}$, $\hat{a}^{(1)}$ in the same way as we computed $\hat{A}$, $\hat{a}$, in and below (6), but using $\hat{h}^{(1)}$, $\hat{\delta}^{(1)}$ and $\hat{\sigma}^2|\hat{g}|^2/2\pi$ in place of $\hat{f}_w$. In computing the $\hat{a}^{(1)}(k)$ we may omit all constant factors, such as $\hat{\sigma}^2/2\pi$, but it is convenient to retain the definition as given. We now form

$$\hat{\alpha}(1) = -\hat{A}(1)^{-1} \hat{a}(1) .$$

We next compute $\hat{\Omega}$ whose typical entry is $\hat{\omega}(k-\ell)$, in row k, column ℓ, where $k = 1, \ldots, q$; $\ell = 1, \ldots, p$ and .

$$\hat{\omega}(k) = \frac{1}{N}\sum_{t=1}^{N-1} \frac{1}{\hat{g}\,\bar{\hat{h}}} e^{-ik\omega_t}, \quad -p+1 \leq k \leq q-1$$

$$= \frac{2}{N}\sum_{t=1}^{[1/2N]} \kappa_t \left\{(\hat{\xi}_g\hat{\xi}_h+\hat{\eta}_g\hat{\eta}_h)\cos k\omega_t+(\hat{\xi}_g\hat{\eta}_h-\hat{\eta}_g\hat{\xi}_h)\sin k\omega_t\right\} / \left\{|\hat{g}|^2|\hat{h}|^2\right\}.$$

If p and q are not large it may be easier to compute $\hat{\omega}(k)$ from the following formula for it, namely,

$$\hat{\omega}(k) = \frac{1}{2\Pi}\int_{-\pi}^{\pi} \{\hat{g}(e^{i\lambda})\hat{h}(e^{-i\lambda})\}^{-1} e^{-ik\lambda}\, d\lambda .$$

For $p \geq q$ this gives

$$\hat{\omega}(k) = \hat{\alpha}(p)^{-1}\sum_{j=1}^{p} \{\hat{g}(\xi_j^{-1})^{-1} h_j^{-k-1} \prod_{\substack{i\neq j\\ i=1}}^{p}(\xi_i - \xi_j)^{-1}\}, \quad -q+1\leq k\leq p-1 ,$$

where the ξ_j are the zeros of $\hat{g}(z)$. For $q \geq p$ we may use

$$\hat{\omega}(k) = \hat{\beta}(q)^{-1}\sum_{j=1}^{q} \{g(z_j^{-1})^{-1} z_j^{k-1} \prod_{\substack{i\neq q\\ i=1}}^{q}(z_i - z_j)^{-1}\}, \quad -q+1 \leq k \leq p-1 ,$$ where the z_j are

the zeroes of $\hat{h}(z)$. In particular for $p = q = 1$ we get $\hat{\omega}(0) = \{1 - \hat{\alpha}(1)\hat{\beta}(1)\}^{-1}$.

If $p = 2$, $q = 1$ and ξ_1 is complex then $\hat{\omega}(0) = \{\hat{\alpha}(2)|\xi_1+\hat{\beta}(1)|^2\}^{-1}$ while

$\hat{\omega}(1) = -2\hat{\omega}(0)\hat{\beta}(1)^{x_1 y_1}$ where $\xi_1 = x_1 + i\, y_1$. It is not difficult to derive simple explicit formulae in my case where p and q are small, as we have said.

Now

$$\tilde{\alpha}^{(1)} = \left\{ I_p - (2\pi/\hat{\sigma}^2)\hat{A}^{-1}[\ \hat{\Omega}'\ \vdots\ 0]\hat{D}^{-1}\begin{bmatrix}\hat{\Omega} \\ \cdots \\ 0\end{bmatrix} \right\}^{-1} (\hat{\alpha} - \hat{\alpha}^{(1)}) + \hat{\alpha}\ .$$

If f_w is factored then $\hat{A}$ will not have been computed, in which case it may be replaced by $\hat{A}^{(1)}$. For computations we may cancel the factor $\hat{\sigma}^{-2}$ occurring before $\hat{A}$ with the factor $\hat{\sigma}^2$ introduced in the definition of $\hat{D}$.

Finally

$$\tilde{\rho}^{(1)} = -\tilde{D}^{(1)^{-1}}\tilde{d}^{(1)}$$

where the $\tilde{d}_{k\ell}^{(1)}$, $\tilde{d}_k^{(1)}$ are defined as were $\hat{d}_{k\ell}^{(1)}$, $\hat{d}^{(1)}$ but using $\tilde{g}^{(1)} = \Sigma\tilde{\alpha}^{(1)}(k)\exp ik\omega_t = \tilde{\xi}_g^{(1)} + i\tilde{\eta}_g^{(1)}$ in place of $\hat{g}$.

We shall shortly state a theorem which establishes the asymptotic efficiency of $\tilde{\alpha}^{(1)}$, $\tilde{\rho}^{(1)}$ but nevertheless it will often be necessary to iterate Step 2. Thus we commence from $\tilde{g}^{(1)}$, $\tilde{\alpha}^{(1)}$, $\tilde{\rho}^{(1)}$, $\tilde{h}^{(1)}$, this last being got via (5) but using $\tilde{\beta}^{(1)}$ in place of $\hat{\beta}$. Of course $\tilde{h}^{(1)} = \tilde{\xi}_h^{(1)} + i\tilde{\eta}_h^{(1)}$. The estimate $\hat{\sigma}^2$ in (7) is now replaced by

(7)'
$$\tilde{\sigma}^{(1)^2} = \frac{2\pi}{N}\sum_{t=1}^{N}|\tilde{h}^{(1)}w_y + \sum_{j=1}^{r}\tilde{\delta}^{(1)}(j)w_j|^2/|\tilde{g}^{(1)}|^2$$

$$= \frac{2}{N^2}\sum_{t=1}^{[1/2N]}\left\{(\tilde{\xi}_h^{(1)}u_y - \tilde{\eta}_h^{(1)}v_y + \sum_{j=1}^{r}\tilde{\delta}^{(1)}(j)u_j)^2 + (\tilde{\xi}_h^{(1)}v_y + \tilde{\eta}_h^{(1)}u_y + \sum_{j=1}^{r}\tilde{\delta}^{(1)}(j)v_j)^2\right\}/|\tilde{g}^{(1)}|^2 .$$

We may now repeat Step 2 arriving at $\tilde{\alpha}^{(2)}$, $\tilde{\rho}^{(2)}$, which for the basis for a further iteration. This completes the description of the computations.

We introduce the matrices

$$\Phi = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|g|^2} e^{i(k-\ell)\lambda} d\lambda \right] \qquad k, \ell = 1, \ldots, p \ . \quad ,$$

$$\Psi = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|h|^2} e^{i(k-\ell)\lambda} d\lambda \right] \qquad k, \ell = 1, \ldots, q \ .$$

$$\Omega = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g \bar{h}} e^{i(k-\ell)\lambda} d\lambda \right] \qquad \begin{matrix} k = 1, \ldots, q \ , \\ \ell = 1, \ldots, p \ . \end{matrix}$$

$$\Delta = \begin{bmatrix} \Delta_{11} + \Psi & -\Delta_{12} \\ -\Delta_{21} & \Delta_{22} \end{bmatrix}$$

$$\Delta_{11} = \left[\frac{1}{\sigma^2} \int_{-\pi}^{\pi} \frac{e^{i(k-\ell)\lambda}}{|g|^2 |h|^2} d(\delta' F(\lambda) \delta) \right] \qquad k, \ell = 1, \ldots, q \ . \quad ,$$

$$\Delta_{12} = \Delta_{21}^{*} = \left[\frac{1}{\sigma^2} \int_{-\pi}^{\pi} \frac{e^{ik\lambda}}{|g|^2 h} d(\delta' F(\lambda)) \right] \quad k = 1, \ldots, q \ . \quad ,$$

$$\Delta_{22} = \frac{1}{\sigma^2} \int \frac{1}{|g|^2} dF(\lambda) \quad .$$

Since $F(\lambda)$ is an $r \times r$ matrix Δ_{12} is $q \times r$ and Δ_{22} is $r \times r$. Now we have the following theorem:

<u>Theorem 1</u>. <u>Under conditions (i), (ii), (iii), and (iv) the vector $(\tilde{\alpha}^{(1)'} \vdots \tilde{\rho}^{(1)'})$ converges almost surely to $(\alpha' \vdots \rho')$ and $\sqrt{N}(\tilde{\alpha}^{(1)}-\alpha)' \vdots (\tilde{\rho}^{(1)}-\tilde{\rho})'$ has a distribution converging to the multivariate normal distribution with zero mean and covariance matrix.</u>

$$V = \begin{bmatrix} \Phi & \vdots & -\Omega' \quad 0 \\ \cdots & \cdots & \cdots \\ -\Omega & \vdots & \Delta \\ 0 & \vdots & \end{bmatrix} .$$

<u>The estimate is asymptotically efficient. The estimates $\hat{\Phi} = (\hat{\sigma}^2/2\pi)\,\hat{A},\ \hat{D},\ \hat{\Omega}$ converge almost surely to Φ, Δ, Ω respectively.</u>

Before discussing the proof two points may be made. In the first place the procedure will usually be iterated, say $(j+1)$ times, so that $(\tilde{\sigma}^{(j)^2}/2\pi)$ $\tilde{A}^{(j)}$, $\tilde{D}^{(j)}$, $\tilde{\Omega}^{(j)}$ will be available and these will be asymptotically efficient.

The second point relates to the size of N. We have in mind situations where $N < 500$. In this case the calculations are well within the capacity of modern equipment. However as N increases the labour of computing the $w(\omega_t)$ becomes dominant. If N is highly composite, e.g. $N = 2^s3^t5^u$, then much larger values may be easily handled. If really large values of N are to be used then it might be worthwhile replacing I_y, I_j, I_{jk} by smoothed estimates of spectra. Provided the somoothing is not too radical there is no doubt that Theorem 1 will continue to hold.

<u>Proof of Theorem 1</u>. We shall give the proof in outline only. In detail it is rather complicated though it does not differ in principle from that given in Hannan (1969). The essential differences are due to the occurrence of the $x_j(n)$ and the dropping of the assumption of stationarity for the $y(n)$ (which is mainly a consequence of the introduction of the $x_j(n)$).* First it is shown

*Details of the proof will be contained in a thesis submitted by the second author to the Australian National University.

that $\hat{\rho}$ converges to ρ and hence $\hat{f}_w$ converges to f_w . Since the zeros of g like outside of the unit circle then

$$|\hat{g}|^{-2} = \sum_{-\infty}^{\infty} \hat{\xi}(u)e^{iu\lambda}$$

where the $\hat{\xi}(u)$ converge to $\xi(u)$ and also for N sufficiently large the sequence $\hat{\xi}(u)$ converges to zero exponentially with u . As a result we may show that $\hat{D}$ converges to Δ . For example

$$\frac{1}{N}\sum_{t=1}^{N-1} e^{ik\omega_t} \frac{I_y}{|\hat{g}|^2} = \frac{1}{2\pi}\sum_{-N+1}^{N-1} \hat{c}_y(v)\left(1 - \frac{|v|}{N}\right) \sum_{j=-\infty}^{\infty} \hat{\xi}(k-v+jN)$$

which consequently converges to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\lambda}\left\{\frac{\sigma^2}{2\pi|h|^2}\,d\lambda + \frac{\delta' dF(\lambda)\delta}{|g|^2|h|^2}\right\} .$$

Similar results hold for other components of $\hat{D}$ so that $\hat{D}$ converges to Δ .

Now we introduce the matrix and vector D , d which differ from $\hat{D}$, $\hat{d}$ only in replacing $\hat{g}$ by g . (Of course they are not computable.) Then D converges to Δ and putting $\mathring{\rho} = -D^{-1}d$ we have

$$\sqrt{N}(\hat{\rho}^{(1)}-\mathring{\rho}) = -\sqrt{N}(\hat{D}^{-1}\hat{d}-D^{-1}d)$$

which converges in probability to $-\Delta^{-1}\sqrt{N}\{(\hat{D}-D)\rho + (\hat{d}-d)\}$. In much the same way as for $\hat{D}$, but after considerable manipulations, this may be shown to converge in probability to

$$\sqrt{N}(\mathring{\rho}-\rho)+\Delta^{-1}\left(\begin{smallmatrix}\Omega\\ 0\end{smallmatrix}\right)\sqrt{N}(\hat{\alpha}-\alpha) \quad . \tag{8}$$

For example the typical entry in $\hat{D} - D$ is

$$\frac{1}{\sqrt{N}} \sum_t \left\{ \left(\frac{1}{|\hat{g}|^2} - \frac{1}{|g|^2} \right) I_y e^{i(k-\ell)\omega_t} \right\}$$

which converges in probability to

$$\frac{1}{N} \sum_t \frac{\sqrt{N}(\hat{g}-g)}{|g|^2 g} I_y e^{i(k-\ell)\omega_t} \quad .$$

Collecting terms together from $-\sqrt{N}\{(\hat{D}-D)\rho + (\hat{d}-d)\}$ we find that $\sqrt{N}(\hat{\rho}^{(1)}-\rho)$ converges in probability to (8) . In the same way we introduce $\mathring{\alpha} = -A^{-1}a$ where A , a differ from $\hat{A}^{(1)}$, $\hat{a}^{(1)}$ in replacing $\hat{h}^{(1)}$, $\hat{\delta}^{(1)}$ and $\hat{g}$ by h , δ and g . Then $\sqrt{N}(\hat{\alpha}^{(1)}-\mathring{\alpha})$ converges in probability to

$$\text{(9)} \qquad \sqrt{N}(\mathring{\alpha}-\alpha)+2\sqrt{N}(\hat{\alpha}-\alpha) - \Phi^{-1}\Omega'\sqrt{N}(\hat{\beta}^{(1)} - \beta) \; .$$

Now using (8) and (9) we may express $\sqrt{N}(\tilde{\alpha}^{(1)}-\alpha)$ and $\sqrt{N}(\tilde{\rho}^{(1)}-\rho)$ linearly in terms of $\sqrt{N}(\mathring{\alpha}-\alpha)$ and $\sqrt{N}(\mathring{\rho}-\rho)$ and it remains only to discover the asymptotic joint distribution of these two vectors and convert that to a distribution for $\sqrt{N}(\tilde{\alpha}^{(1)}-\alpha)$, $\sqrt{N}(\tilde{\rho}^{(1)}-\rho)$. Thus $\sqrt{N}(\mathring{\alpha}-\alpha) = -A^{-1}\sqrt{N}(a+A\alpha)$ which we may replace by $-\Phi^{-1}u$ where u has typical element given by

$$\frac{\sigma^2}{2\Pi} u_k = N^{-1/2} \sum_{t=o}^{N-1} |g|^{-2} g^{-1} \left| h w_y + \sum_{j=1}^{r} \delta(j) w_j \right|^2 e^{ik\omega_t}$$

which may be replaced by

$$N^{-1/2} \sum_{t=o}^{N-1} |g|^{-2} g^{-1} I_w e^{ik\omega_t}$$

where I_w is the periodogram for the, uncomputable, $w(n)$. Now exactly as in Hannan (1969) we see that u_k may be replaced by

$$\sigma^{-2}N^{-1/2}\sum_{n=1}^{N}\epsilon(n)\eta(n-k)$$

where the $\eta(n)$ are a stationary sequence satisfying

$$\sum_{0}^{p}\alpha(j)\eta(n-j) = \epsilon(n) \quad .$$

Similarly we may replace $\sqrt{N}(\hat{\rho}-\rho)$ by $-\Delta^{-1}v$ where v has typical element given by

$$(\sigma^2/2\pi)v_k = N^{-1/2}\sum_{t=o}^{N-1}|g|^{-2}\left\{I_y\bar{h}+\Sigma\delta(j)I_j\right\}e^{ik\omega_t} \qquad k = 1, \ldots, q$$

$$= N^{-1/2}\sum_{t=o}^{N-1}|g|^{-2}\left\{I_j h+\Sigma\delta(k)I_{jk}\right\} \qquad k = q+j,\ j = 1, \ldots, r.$$

We may replace $I_y\bar{h}+\Sigma\delta(j)I_j$ by I_{yw} where this is the cross periodogram between $y(n)$ and $w(n)$ and $I_j h+\Sigma\delta(k)I_{jk}$ by I_{jw} where this is the periodogram between $x_j(n)$ and $w(n)$. We may then replace I_{yw} by $h^{-1}\{-\Sigma\delta(j)I_{jw}+I_w\}$.

The part $N^{-1/2}\Sigma|g|^{-2}h^{-1}I_w(\exp ik\omega_t)$ may be replaced, again as Hannan (1969) by

$$\sigma^{-2}N^{-1/2}\sum_{n=1}^{N}\epsilon(n)\xi(n-k) \ ,$$

$$\sum_{o}^{q}\beta(j)\xi(n-j) = \epsilon(n)$$

while the remaining parts are of the form

$$\Sigma\delta(j) \quad \frac{1}{\sqrt{N}}\Sigma \frac{I_{j\varepsilon}}{gh} e^{ik\omega_t} \quad , \qquad \frac{1}{\sqrt{N}} \Sigma \frac{I_{j\varepsilon}}{g} \quad .$$

The remainder of the derivation of the joint distribution of $\sqrt{N}(\overset{\circ}{\alpha}-\alpha)$, $\sqrt{N}(\overset{\circ}{\rho}-\rho)$ reduces to the problem of obtaining that distirbution for estimated autoregression coefficients or regression coefficients. (See Hannan (1960, 1963)). The conversion of these results into the required results for $\tilde{\alpha}^{(1)}$, $\tilde{\rho}^{(1)}$ is straightforward. The covariance matrix for the asymptotic distribution of the maximum likelihood estimators, assuming the $x_j(n)$ to be stationary, may be found by an extension of the results of Walker (1964) and is the same as that stated in Theorem 1.

3. Some Properties of the Estimators and Distributed Lag Models

The first point which we wish to make in this section relates to the possibility that in choosing a mixed autoregressive, moving-average model one may also be choosing to accept certain risks. Let us now call V_p the matrix V of Theorem 1, so as to emphasise the assumed order of the moving average. Now if $\alpha(p) = 0$ then some calculation shows that the part of V_p^{-1} which gives the asymptotic variances and covariances of the parameters of the model, other than $\alpha(p)$, is

$$V_{p-1}^{-1} + (1-e'V_{p-1}^{-1}e)^{-1} V_{p-1}^{-1} e e'V_{p-1}^{-1}$$

where the vector e has zeros everywhere save in the p^{th} to $(p+q)^{th}$ places where it has, in the $(p+j)^{th}$ place,

$$e_{p+j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} h^{-1} e^{-i(p-j)\lambda} d\lambda \ , \ j = 1, \ldots, q \ ,$$

which is, of course, zero for $j > p$. The matrix V_{p-1} gives the variances and covariances which would obtain if we had known that $\alpha(p) = 0$ and had fitted only $\alpha(1), \ldots, \alpha(p-1)$. Thus the overfitting has resulted in too large a covariance matrix. The effect can be most clearly seen when $q=1$ for then $\sqrt{N}(\tilde{\beta}(1) - \beta(1)$ has a limiting distribution with variance $v/(1-v)$ where v is the variance which would obtain if it were recognized that $\alpha(p) = 0$. For example when $p=q=1$, $r=0$ then $v = (1-\beta^2(1))$ and the efficiency of $\tilde{\beta}(1)$ is only $\beta(1)^2$ when $\alpha(1) = 0$. It is to be expected that this efficiency will be low when $\beta(1)$ is small, as we are then near to an unidentified situation, but even for $\beta(1) = 0.7$ the efficiency is under 50%. Of course when $\alpha(p) \neq 0$ and we assume otherwise the estimates of the remaining parameters will not be consistent.

When $\beta(q) = 0$ and this is not recognized similar problems emerge though these are more complicated to describe. The problem does not arise with $\delta(j)$ where over fitting essentially involves only a loss of degrees of freedom.

A situation whereas the above mentioned problem cannot arise is that where $p=q$ and $\alpha(j) = \beta(j)$, $j=1,\ldots, p$. It is instructive to consider this case. We must now require that $\delta \neq 0$ for otherwise the model is not identified. In this case, we have already said, (1) is equivalent to

$$y(n) + \sum_{j=1}^{r} \delta(j) \sum_{0}^{\infty} a_k x_j(n-k) = e(n) \ , \tag{10}$$

which, for $r=1$, becomes the, so called, distributed lag model, with the a_k generated by $g^{-1} = h^{-1}$. As we have said, this case $r=1$ is probably most interesting when $x_j(n) = x(n-j+1)$ so that (10) is

$$y(n)+\delta\Sigma\lambda_j x(n-j) = \epsilon(n) , \quad \delta = \delta(1) , \tag{11}$$

where now the λ_j are generated by

$$\sum_{0}^{r-1} \left\{\delta(j+1)/\delta(1)\right\} z^j / \sum_{0}^{q} \beta(j) z^j ,$$

so that we are dealing with the general rational distributed lag model with independent errors. Let us return to the general model (10) and consider the best linear combination of $\tilde{\alpha}^{(j)}$ and $\tilde{\beta}^{(j)}$ (assuming j iterations to have been completed), say $A_1\tilde{\alpha}^{(j)}+A_2\tilde{\beta}^{(j)}$, $A_1+A_2 = I_p$. Then it is easily established from Theorem 1 that we must take $A_1 = 0$ and $A_2 = I_p$ so that $\tilde{\beta}^{(j)}$ is the best estimator. Thus $\tilde{\rho}^{(j)}$ is our best estimator of the unknown parameters, and this has covariance matrix

$$\begin{bmatrix} \Delta_{11} & -\Delta_{12} \\ -\Delta_{21} & \Delta_{22} \end{bmatrix}^{-1} . \tag{12}$$

This is the covariance matrix for an asymptotically efficient estimator, needless to say. This agrees also with the result obtained for the special case $p = q = r = 1$ in Hannan (1965)[2]. In fact an examination of the proof shows that the asymptotically efficient estimator of ρ may be more easily got, namely as

[2]Formula 4.7 in Hannan (1965) is incorrectly stated as is evident from the

$$(13)\qquad \rho^{(1)} = -\left\{\hat{D} - \frac{\hat{\sigma}^2}{2\pi}\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\right\}^{-1}\left\{\hat{d} + \frac{\hat{\sigma}^2}{2\pi}\begin{pmatrix} \hat{a} \\ 0 \end{pmatrix}\right\} .$$

Moreover the matrix (12) is estimated by the first factor in $-\check{\rho}^{(1)}$. Thus the computations are greatly simplifed. Further iterations[3] are then completed beginning from $\check{\rho}^{(1)}$, computing $\check{D}$ and $\check{A}^{(1)}$ using $\check{g}^{(1)} = \check{h}^{(1)}$ and $\check{\rho}^{(1)}$ in place of $\hat{g}$, $\hat{h}$ and $\hat{\rho}^{(1)}$. When this is done the estimate is asymptotically equivalent to that introduced in Hannan (1965). Indeed the likelihood function being $\mathcal{L}$ and α being a function of β the equations of maximum likelihood become

$$\frac{\partial\mathcal{L}}{\partial\beta} + \left(\frac{\partial\alpha}{\partial\beta}\right)\frac{\partial\mathcal{L}}{\partial\alpha} = 0 \quad , \quad \frac{\partial\mathcal{L}}{\partial\delta} = 0 \ .$$

If $\alpha = \beta$ then $(\partial\alpha/\partial\beta) = I_p$ and we are led to equation (13) .

To test the hypothesis $\alpha = \beta$, in case $p = q$, we may form

$$(14)\qquad \left\{2\pi N(\tilde{\alpha}^{(j)}-\tilde{\beta}^{(j)})'\tilde{A}^{(j-1)}(\tilde{\alpha}^{(j)}-\tilde{\beta}^{(j)})\right\} / \left\{\tilde{\sigma}^{(j-1)}\right\}^2$$

which is asymptotically distributed as chi-square with p degrees of freedom.

---

2 (cont'd)
derivation which precedes it. The formula should read

$$\left[\int_{-\pi}^{\pi}\{2\pi f_w(\theta)\}^{-1}c(\theta)c(\theta)^*dF_x(\theta)\right]^{-1} .$$

Note that $F_x(\theta)$ differs from our $F(\theta)$ by a factor $\gamma_x(o)^{-1}$.

[3] The need for further iterations was not emphasised in Hannan (1965) but in fact these are likely to be needed.

We summarise these results in Theorem 2.

Theorem 2. If $p = q$ we may test the hypothesis $\alpha = \beta$, for large N, by using (14) as chi-square with p degrees of freedom. If $\alpha = \beta$ then ρ may be estimated by $\tilde{\rho}^{(j)}$ or by (13). Then $\sqrt{N}(\rho^{(1)}-\rho$ $\sqrt{N}(\tilde{\rho}^{(j)}-\rho)$ are each asymptotically normal with covariance matrix

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}^{-1} .$$

This matrix is consistently estimated by

$$\left\{ D - \frac{\hat{\sigma}^2}{2\pi} \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} \right\}^{-1} .$$

The procedure is asymptotically efficient and $\hat{\rho}^{(i)}$ converges to ρ almost surely.

In case $x_j(n) = x(n-j+1)$, so that (11) is the relation to be estimated the computation would be simplified as follows. We would replace w_j by

$$w(\omega_t)e^{i(j-1)\omega_t} , \quad w(\omega_t) = (2\pi N)^{-1/2} \sum_{t=1}^{N} X(n)e^{in\omega_t} = u(\omega_t) + iv(\omega_t) .$$

Then all later formulae involving the ω_j simplify. For example u_j and c_{jk} are replaced, respectively, by

$$u \cos(j-1)\omega_t - v \sin(j-1)\omega_t , \quad (u^2+v^2)\cos(j-k)\omega_t .$$

We would also replace $\hat{c}_{jk}(n)$ by $\hat{c}(n+j-k)$ where

$$c(n) = (N-n)^{-1} \sum_{m=1}^{N-n} (x(m)-\bar{x})(x(m+n)-\bar{x}) = c(-n), \quad n=0, 1, \ldots, p+r-1 .$$

We would replace $\hat{c}_{yj}(n)$ by $\hat{c}_{yx}(n-j+1)$ where

$$\hat{c}_{yx}(n) = (N-n)^{-1}\sum_{m=1}^{N-n}(y(m)-\bar{y})(x(m+n)-\bar{x}) = \hat{c}_{xy}(-n), \quad n=0,1,\ldots, p+q+r-1 \quad .$$

There are two ways in which the treatment in Hannan (1965) is more general than in the present paper. (Of course in other ways it is much less general). The first of these is through the introduction of a more general stochastic structure for the term on the right side in (1). This is assumed to be of the form

$$\sum_{o}^{q}\beta(j)u(n-j)$$

where $u(n)$ is stationary and of a sufficiently regular kind (see the references cited). If $f_u(\lambda)$ is the spectral density of $u(n)$ then the covariance matrix of the limiting distribution of the estimates is, for $p = q = r = 1$,

$$(15) \qquad \left\{\int_{-\pi}^{\pi}\frac{1}{2\pi f_w(\lambda)}\begin{bmatrix}1 & e^{i\lambda}/h \\ e^{-i\lambda}/\bar{h} & \delta^2/|h|^2\end{bmatrix} dF(\lambda)\right\}^{-1}$$

where $f_w(\lambda) = |h|^2 f_u(\lambda)$. Of course if $\alpha = \beta$ and $u(n) = \epsilon(n)$ this agrees with our earlier result . The estimation procedure leading to (15) does not involve any parameterisation of the stochastic structure of $u(n)$ but, as we have said, only rather general specifications are made. (This is a common device in time series analysis). An alternative procedure is to proceed as we have in (1), which (when we do <u>not</u> assume $\alpha = \beta$) takes $u(n)$ to be of the form

$$(16) \qquad u(n) = \sum_{o}^{\infty} b_j\epsilon(n-j) \ , \qquad \sum_{o}^{\infty} b_j z^j = g/h \quad .$$

Of course this specification restricts the nature of the process generating $u(n)$ (as we have already said) since it must correspond to a rational generating function having h in the denominator. However, if g is restricted only to be a polynomial this specification is still very general. It is therefore of interest to compare the covariance matrix of the limiting distribution of the estimates of ρ, obtained by the methods of section 2, with the generalization of (15) appropriate to the case of a general rational disbributed lag model and r exogenous variables (all with lag structure such that the denominator of the generating functions is h). Of course we assume that $f_u = (\sigma^2/2\pi)|g|^2/|h|^2$ so that $f_w = (\sigma^2/2\pi)|g|^2$ the difference in the two methods lying in the fact that the second does not take account of the special form of f_u. The generalization of (15) is

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}^{-1}$$

whereas $\sqrt{N}(\tilde{\rho}^{(i)} - \rho)$ will have for its limiting distribution the covariance matrix

$$\begin{bmatrix} \Delta_{11} + \Psi - \Omega\Phi^{-1}\Omega' & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}^{-1}$$

However

$$\Psi - \Omega\Phi^{-1}\Omega' \geq 0 \tag{17}$$

so that the use of additional information has reduce the covariance matrix (as it clearly must). When $g = h$ then (17) is null and the two methods are equivalent as we have already seen. If the degree of g is overstated then, as we know, we lose efficiency but they are never worse then the methods of Hannan (1965) as (17) shows. Of course if degree of g is understated the methods of this paper become inconsistent but it is possible to test whether p should be increased. In addition to a more accurate estimate of ρ the methods also lead to an estimate of g and hence (for example) to a specification of the optimal linear predictor for $y(n)$, once the $x_j(n)$ are predicted. (The prediction of the $x_j(n)$ requires knowledge of their stochastic structure). The relation between the three cases is exhibited by the special example where $p = q = r = 1$, $\alpha(1) = 0$ and $x(n)$ consists of uncorrelated random variables with unit variance while $\sigma^2 = 1$. Then we take the estimates obtained using (a) all information, (b) only the knowledge that g if of degree one and (c) by the method of Hannan (1965). The three estimates of δ have the same asymptotic variance properties but the estimate of β has, in its limiting distribution, the variance respectively, $(1-\beta^2)/(1+\delta^2)$, $(1-\beta^2)/(\beta^2+\delta^2)$ and $(1-\beta^2)/\delta^2$.

The other direction in which the approach of Hannan (1965) was more general than that of the present paper was in the fact that $x(n)$ was allowed to be more general and in particular was allowed to contain components of the form $x_j(n) = n^k$ as well as of the kind treated in the present paper. It seems that the methods of the present paper could be extended (and modified) to deal with the more general specifications but the statement of already complex results would be further compensated. Since we can allow for trends (for example) by the procedures described in the paragraph just above (2) it has been thought best not to seek for this additional generality.

REFERENCES

Hannan, E.J., (1960) Time Series Analysis. Methuen, London.

__________, (1963) Regression for time series. Chapter 2 in Time Series Analysis edited by M. Rosenblatt. Wiley, New York.

__________, (1965) The estimation of relationships involving distributed lags. Econometrica, 33, 206-224.

__________, (1969) The estimation of mixed moving-average autoregressive systems. Biometrika, 56, 579.

Walker, A.M. (1964) Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary nondeterministic time series. J.Aust.Math. Soc., 4, 363-384.