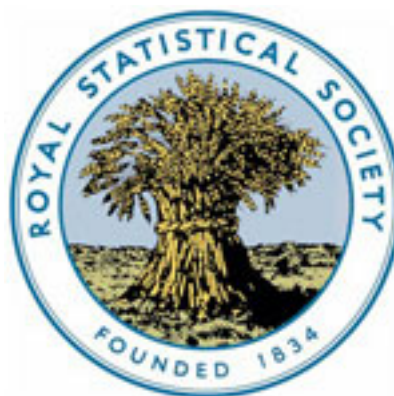




Moving Average Processes

Author(s): O. D. Anderson

Reviewed work(s):

Source: *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 24, No. 4 (Dec., 1975), pp. 283-297

Published by: [Blackwell Publishing](#) for the [Royal Statistical Society](#)

Stable URL: <http://www.jstor.org/stable/2987925>

Accessed: 10/01/2012 07:30

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and Royal Statistical Society are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series D (The Statistician)*.

<http://www.jstor.org>

Moving Average Processes

O. D. ANDERSON,[†] *Department of Mathematics, Nottingham University*

Properties of Moving Average processes are discussed, amongst these being the closure under addition of independent MA models, the sometimes possible "orthogonal" decomposition of an MA process, and stronger restrictions on the autocorrelation function. These ideas would seem to be relevant to practical time series modelling, and some examples of their use are given.

1. Introduction to Box-Jenkins Process

Time Series analysis is concerned with data which is not independent, but serially correlated, and where the relations between consecutive observations are of interest. It is a rapid growth area in statistical practice, and the mass of research and application, currently taking place, has already left its mark on nearly all the numerate disciplines in the sciences, business and technology. The approach to time series developed in the 60s by Professors Box and Jenkins, and summarized in their book (1970), is attracting more and more attention, and is being applied by accountants, sociologists, statisticians, civil, mechanical and telephone engineers, economists, physicists and process chemists, to name but a few.

The Box-Jenkins method only deals with discrete observations $\dots, z_i, \dots$ ordered in some dimension, usually time, and generally observed at intervals which can be formally thought of as fixed. Such a *series* is conveniently written as $\{z_i\}$, and can be considered as arising from some set of random variables $\{Z_i\}$, the associated *time process*.

Box and Jenkins focused interest on a flexible class of stable statistical models, which they believe often give adequate representation for the structure of actually occurring series. A subclass of these is the famous *mixed* or ARMA (p, q) class of models

$$Z_i = \phi_1 Z_{i-1} + \dots + \phi_p Z_{i-p} + A_i + \theta_1 A_{i-1} + \dots + \theta_q A_{i-q} \quad (1)$$

which reduces to the AR(p), or *AutoRegressive* process of order p , when $p > q = 0$; and to the MA(q), the *Moving Average* process of order q , when $0 = p < q$. When both p and q are zero, the simple but important process

$$Z_i = A_i \quad (2)$$

results. In all these models, the A_i are assumed to be independently

[†] Now with the Civil Service College, London.

and normally distributed, with zero mean and constant variance σ_A^2 ; all of which can be summarized by writing

$$A_t \sim \text{IN}(0, \sigma_A^2)$$

Then (2) is termed a *white noise* process, and is a fundamental building block in the theory. In this paper, we shall be concentrating our attention on moving average processes.

If the probabilistic structure of a process is time invariant, the process is termed *stationary*, and one can then define a process mean μ and *auto-covariance function* $\{\gamma_k: k=0, 1, \dots\}$, the set of covariances between Z_t and Z_{t-k} at lag k . Since A_t , the random part is normal, no higher order auto-moments exist for $\{Z_t\}$. Note that $\gamma_0 = \sigma_Z^2$, $\gamma_{-k} = \gamma_k$ and the *autocorrelations* at various lags k are defined by

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (3)$$

The set $\{\rho_k: k=0, 1, \dots\}$ is termed the *autocorrelation function*, and is usually abbreviated to a.c.f.

The *Backshift Operator* B is defined by

$$BZ_t = Z_{t-1}$$

for any process $\{Z_t\}$, and obeys the rules of elementary algebra. Using it, the mixed model (1) can be written

$$(1 - \phi_1 B - \dots - \phi_p B^p)Z_t = (1 + \theta_1 B + \dots + \theta_q B^q)A_t$$

or

$$\phi_p(B)Z_t = \theta_q(B)A_t \quad (4)$$

where the polynomials $\phi_p(B)$ and $\theta_q(B)$ are respectively termed the *Auto-regressive* and *Moving Average Operators* (of order p and q respectively here).

Alternative representations for the mixed model are

$$Z_t = \psi(B)A_t \quad (5)$$

where $\psi(B) = \phi_p^{-1}(B)\theta_q(B)$ is in general an infinite polynomial; and

$$\pi(B)Z_t = A_t \quad (6)$$

where $\pi(B) = \theta_q^{-1}(B)\phi_p(B)$, and is again generally of infinite order. The $\text{MA}(\infty)$ and $\text{AR}(\infty)$ representations (5) and (6) are termed respectively the *random shock* and *inverted* forms. It should be clear then that AR and MA representations are to some extent equivalent, as for instance any $\text{AR}(p)$ can be written as an $\text{MA}(\infty)$.

A necessary and sufficient condition for an ARMA (p, q) process to be stationary is that, for its associated inverted form, the polynomial $\pi(\zeta)$, in the complex variable ζ , has all its zeros outside the unit circle. For instance, the AR(1) model

$$(1 - \phi B)Z_t = A_t$$

is stationary if, and only if, $|\phi| < 1$.

Box and Jenkins have concentrated on building ARMA (p, q) models for observed series, with p and q as small as possible, consistent with adequate representation. In practice, this presents the problem of initially identifying p and q , from statistics such as the sampled a.c.f. This crucial step can be considered difficult to perform by a non-expert. Other workers have investigated simpler, if less powerful, methods of time series modelling. Such approaches are valuable if they can be made fairly automatic and safe to handle by unsophisticated users, provided the penalty associated with failing to achieve optimal results is less than the extra cost of obtaining them. Thus, for instance, *simple exponential smoothing* will still form an *economic* technique for rapidly forecasting large numbers of relatively unimportant series.

Recently considerable attention has been devoted to automatic *step-wise* methods of obtaining an AR representation. For instance Newbold and Granger (1974). The order of such a *fitted* representation will usually be greater than the *parsimonious* $p + q$ obtained by a Box-Jenkins optimal fit. But once programmed, these step-wise algorithms need no further intervention on the part of the user, after inputting the data.

The other possibility of MA modelling seems to have been neglected, but it is hoped that this paper will show that the MA process has some promising properties for identification purposes. The belief is that the results in the third section of this paper are of immediate use in Box-Jenkins identification, and eventually, together with some other ideas being considered, will lead to a more automatic (if less efficient) Box-Jenkins identification procedure, which will be safe for beginners to use.

2. The Moving Average Model

The moving average process of order q , or MA(q), is given by

$$Z_t = \theta_0 A_t + \theta_1 A_{t-1} + \dots + \theta_q A_{t-q} \quad \theta_0 = 1 \quad (7)$$

where, as usual, $A_t \sim \text{IN}(0, \sigma_A^2)$. This can be written

$$Z_t = \theta_q(B) A_t$$

where $\theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ is the MA(q) operator, and the q is

omitted when no confusion is likely. Taking expectations in (7), $\mu_z=0$, while all the autocovariances, which are given by

$$\gamma_k = \sigma_A^2 \sum_{j=0}^{q-|k|} \theta_j \theta_{j+|k|} \quad (8)$$

are evidently fixed. (Note that (8) implies $\gamma_k=0$ for $|k|>q$.) Thus, for finite q , the process is evidently always stationary. This can also be so for some $MA(\infty)$ processes. For instance, writing the $AR(1)$ model in random shock form (5) gives

$$Z_t = \sum_{j=0}^{\infty} \phi^j A_{t-j}$$

which is $MA(\infty)$, and we know that it is stationary provided $|\phi|<1$.

Somewhat analogous to the concept of stationarity is that of *invertibility*. For this the π weights of the inverted form (6) must follow a convergent sequence, and, for the random shock form (5), the zeros of $\psi(\zeta)$ all lie outside the unit circle. Stipulating invertibility avoids certain model multiplicities, which would otherwise occur when identifying a series. Thus, for instance, given an a.c.f.

$$\{\rho_1 \neq 0, \rho_k = 0 \quad k > 1\}$$

two $MA(1)$ models are possible with parameters θ given by the roots of

$$\rho_1 = \frac{\theta}{1+\theta^2} \quad (9)$$

whose product is clearly unity. Thus only one root lies outside the unit circle, and the other will be excluded by the invertibility requirement. If, however, $\rho_1 = \pm \frac{1}{2}$, (9) has coincident solutions, which lie on the unit circle. We then have a case of *borderline non-invertibility*, which is a concept of some importance. This type of non-invertibility never leads to model multiplicity, so there is less reason for avoiding it, though it will give a divergent sequence of π_j weights, which is embarrassing when generating forecasts. A simple example of such a process is the running sum

$$Z_t = A_t + A_{t-1}$$

which is the $MA(1)$ yielding $\rho_1 = \frac{1}{2}$, mentioned above.

Using (8) in (3) we obtain the $MA(q)$ a.c.f., which evidently cuts off after ρ_q . However there are stronger restrictions on the a.c.f. than this.

3. Restrictions on the Autocorrelations of Pure MA Processes

For all integers k , $0 \leq k \leq q$,

$$|\rho_k| \leq \cos \left(\frac{\pi}{[q/k] + 2} \right) \quad (10)$$

where $[x]$ denotes 'the integer part of x '. This was first given in Anderson (1972), and we will call it the *individual inequality*. It is easily proved. We have, for all MA processes, using (8) and (3)

$$\rho_k = \sum_{j=0}^{q-k} \theta_j \theta_{j+k} / \sum_{j=0}^q \theta_j^2 = \theta \Lambda_k \theta^T / \theta \theta^T \quad (11)$$

where Λ_k is a null $q+1$ square matrix, with all elements on the k th super and sub-diagonals replaced by $\frac{1}{2}$. The right of (11) is a *Rayleigh quotient* and so, for instance see Bellman (1970, pp. 112–19), its maximum value is the largest positive eigenvalue of Λ_k , $\bar{\rho}_k$ say, which occurs when θ is a corresponding eigenvector, which is completely fixed by requiring $\theta_0=1$, when we will write it as $\tilde{\theta}$.

It follows that $-\bar{\rho}_k$ is the least attainable value of ρ_k , and this occurs when, defining $J=[j/k]$, $\theta=(\theta_j=(-1)^J \tilde{\theta}_j: j=0, 1, \dots, q)$. Since $-\bar{\rho}_k$ occurs for this choice of θ , it is attainable, but it cannot be exceeded in the negative direction. For, if ρ_k could be $-\rho_k^* < -\bar{\rho}_k$ for $\theta=(\theta_j^*)$, then for $\theta=((1)^J \theta_j^*)$, ρ_k would be $\rho_k^* > \bar{\rho}_k$, which is impossible.

It remains to find $\bar{\rho}_k$. Writing $\Lambda_k = \Lambda_{q,k}$, it can be shown that the non-zero distinct eigenvalues of $\Lambda_{q,k}$ are just those of $\Lambda_{Q,1}$ and possibly $\Lambda_{Q-1,1}$, where $Q=[q/k]$.

Proof: Denote $q-kQ$ by q^* and $|\Lambda_{q,k} - \lambda I|$ by $D_{q,k}$. Consider the permutation of the first $q+1$ positive integers defined by

$$m \rightarrow \begin{cases} (Q+1)(m^*-1) + M + 1 & m^*=1, \dots, q^*+1 \\ (Q+1)(q^*+1) + Q(m^*-q^*-2) + M + 1 & m^*=q^*+2, \dots, k \end{cases}$$

where, for each $m \in (1, 2, \dots, q+1)$, m^* and M are uniquely defined by

$$m = Mk + m^* \quad 1 \leq m^* \leq k$$

Then, if we apply this permutation to first the rows of $D_{q,k}$ and then to the resulting columns, we evidently do not change its value; but we can then expand it directly to give

$$D_{q,k} = (D_{Q,1})^{q^*+1} \times (D_{Q-1,1})^{k-q^*-1}$$

Thus we need only investigate $D_R = D_{R,1}$, for $R=Q$ and $Q-1$. For $R \geq 2$, expanding D_R by its first column, and then by the first row, we obtain

$$D_R = -\lambda D_{R-1} - \frac{1}{2} D_{R-2}$$

This difference equation has the solutions

$$\lambda_r = \cos \left(\frac{r\pi}{R+2} \right) \quad r=1, \dots, R+1$$

which give the $R+1$ distinct eigenvalues, and since $\cos x$ is strictly decreasing over $(0, \pi)$, using an obvious notation,

$$\tilde{\rho}_q, k = \tilde{\rho}_{Q, 1} = \cos \frac{\pi}{Q+2} \quad (12)$$

as required. Then, for $0 \leq j \leq q$

$$\tilde{\theta}_j = \theta_{j^*} \sin \frac{(1+J)\pi}{Q+2} \bigg/ \sin \frac{\pi}{Q+2} \quad \begin{matrix} j^* = 0, 1, \dots, q^* \\ 0 \qquad \qquad \qquad \text{otherwise} \end{matrix} \quad (13)$$

where $j^* = j - kJ$ and the θ_{j^*} are any arbitrary constants, subject to $\theta_0 \equiv 1$, $\theta_{q^*} \neq 0$.

An *overall* inequality on the autocorrelations

$$\sum_{k=1}^q |\rho_k| \leq \frac{q}{2} \quad (14)$$

must also apply, the bounds always being attainable for any order of MA model. Thus the borderline non-invertible running sum

$$Z_t = \sum_{j=0}^q A_{t-j}$$

has

$$\rho_k = \frac{q+1-k}{q+1} \quad k=0, 1, \dots, q \quad (15)$$

which gives equality in (14). (For further "overall" results see Anderson (1975a).)

A relation (12), causes through (13) restrictions on the remaining autocorrelations. These will then satisfy conditional analogues of (10) and (14), which are of exactly the same form, but of lower order, and conditions for conditional bounds, corresponding to (13) and results like (15), can be written down. And so on, until the process is completely "fixed". This will occur when at most q^*+1 (independent) autocorrelations have attained bounds and conditional bounds. (See Anderson, 1975b, for details.) Such a fixed process can be shown to be always borderline non-invertible.

This last point arises from the fact, proved in Anderson (1974a), that the q space of θ maps, through the relation (11), many-one on to the q space of autocorrelation vectors $(\rho_1, \dots, \rho_q)$, which is convex. The boundary between the invertible and non-invertible θ domains goes into the boundary of the convex feasible autocorrelation range.

It is believed that making use of the stronger conditions (10) and (14), rather than just the cut-off point of the a.c.f., will help in the difficult

identification stroke of the Box-Jenkins cycle for modelling linear processes. Of course there will still be the usual problems of the relatively large sampling errors, which occur when estimating autocorrelations from finite histories, and the persistence of bogus patterns in the observed a.c.f. However, some pilot simulations indicate that the bounding a.c.f. profiles can help to discredit an incorrect MA model. And according to a conversation with Professor Granger, the Treasury is finding the individual inequality useful for this purpose.

4. Aggregating Independent Processes

The convexity results mentioned in section 3 can also be used to provide a simple proof of the lemma that, given two independent models $X_i \sim \text{MA}(q_1)$, $Y_i \sim \text{MA}(q_2)$ (where the tilde means "follows a process from the class"), then their sum

$$Z_i = X_i + Y_i \sim \text{MA}(q \leq \max [q_1, q_2])$$

For instance, for second-order MA processes, the feasible region is shown in Figure 1. It is easy to show that, for any two independent processes, with variances σ_X^2 and σ_Y^2 respectively and, using an obvious notation,

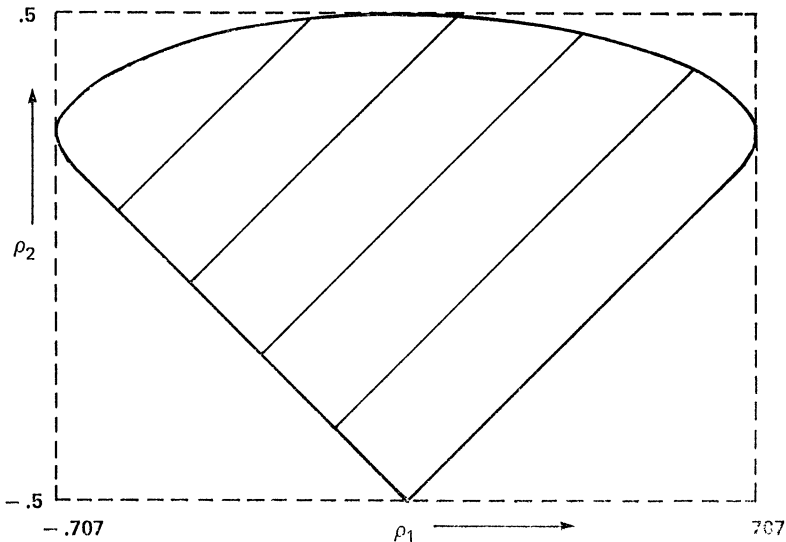


Figure 1

Feasible region for the pair of autocorrelations from an MA(2) process.

autocorrelation vectors $\rho(X)$ and $\rho(Y)$ (necessarily not outside the feasible range), the sum process has

$$\rho(Z) = \frac{\sigma_X^2 \rho(X) + \sigma_Y^2 \rho(Y)}{\sigma_X^2 + \sigma_Y^2} \quad (16)$$

which lies within the join of $\rho(X)$ and $\rho(Y)$, and so inside the feasible range, due to its convexity. Thus $\{Z_i\}$ has an MA(2) representation, which of course could sometimes degenerate to MA(<2). The argument is in fact unaltered by considering general independent MA processes. For, if $q_1 < q_2$ say, then the class MA(q_1) is a subclass of MA(q_2), so $\rho(X)$ and $\rho(Y)$ still fall in the convex range for order q_2 processes, and (16) still applies.

This lemma was discussed by Granger (1972) and Anderson (1975c), and forms the basis for showing how more complicated linear models can often arise from simpler processes, which may more easily lend themselves to a theoretical explanation.

5. The Resolution of a Moving Average Process into "Orthogonal" Components

The lemma of the last section extends immediately by induction to give, for all integers n , and for all independent processes MA(q_1), ..., MA(q_n)

$$\sum_{i=1}^n \text{MA}(q_i) = \text{MA}(q = \max [q_1, \dots, q_n]) \quad (17)$$

where the MA(q) on the right can degenerate into MA(< q). Define the MA $_s(r)$ process to be of the form

$$Z_t = A_t + \theta_1 A_{t-s} + \dots + \theta_r A_{t-rs}$$

where $\{A_t\}$ is a white noise process. (Using this notation, MA(q) would be written MA $_1(q)$.) Then a particular case of (17) is for independent MA $_1(1)$, ..., MA $_q(1)$

$$\sum_{s=1}^q \text{MA}_s(1) = \text{MA}_1(q) \quad (18)$$

Evidently one can always sum the processes on the left of (18) to give the R.H.S., and Anderson (1974b) gives the conditions for the converse to be possible. An MA(q) process can be decomposed into an "orthogonal" set of independent processes $\{\text{MA}_s(1): s=1, \dots, q\}$ if and only if

$$\sum_{k=1}^q |\rho_k| \leq \frac{1}{2} \quad (19)$$

This result can be shown to follow from the convexity property, and

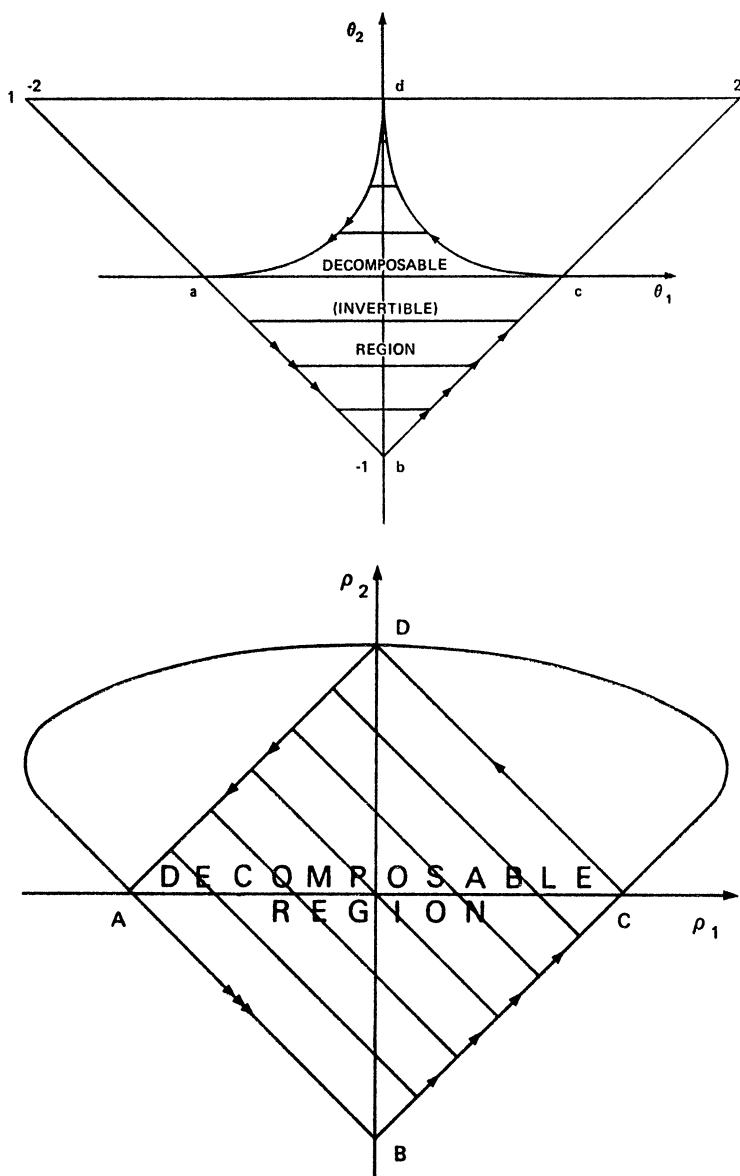


Figure 2

Decomposability regions of the invertible parameter and feasible autocorrelation spaces for MA(2) processes.

Again this idea of orthogonal decomposition, apart from being of theoretical interest, may be of value when attempting to interpret a fitted model in terms of scientific theory.

Associated with the a.c.f. is the p.a.c.f., the *partial autocorrelation function*, defined by

where $\mathbf{P}_k$ is the $k \times k$ autocorrelation matrix

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & & \rho_{k-3} \\ \vdots & & & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{pmatrix}$$

$$\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_k \end{pmatrix}$$

Given an $AR(p)$ process, the p.a.c.f. is a useful concept as it cuts off after ϕ_{pp} . But for a moving average process it experiences no such cut off, though it decays to zero. For instance, for $MA(1)$, writing $\rho_1 = \rho$,

$$|P_k^*| = \begin{vmatrix} 1 & \rho & & & & & \rho \\ \rho & 1 & \rho & & & & 0 \\ & \rho & 1 & \rho & & & . \\ & & . & . & . & & . \\ & & & . & . & . & . \\ & & & & . & . & . \\ & & & & & \rho & 1 & \rho & . \\ & & & & & & \rho & 1 & . \\ & & & & & & & \rho & 0 \end{vmatrix}$$

$$|\mathbf{P}_k^*| = (-1)^{k-1} \rho^k = (-1)^{k-1} \theta^k / (1 + \theta^2)^k \quad (20)$$

While expanding $|P_k|$ by its first column, and then by its first row, gives the recurrence relation

$$|P_k| = |P_{k-1}| - \rho^2 |P_{k-2}|$$

with solution

$$|P_k| = \frac{(1 - \theta^{2k+2})}{(1 - \theta^2)(1 + \theta^2)^k}$$

Together with (20), this yields

$$\phi_{kk} = (-1)^{k-1} \theta^k (1 - \theta^2) / (1 - \theta^{2k+2})$$

which is simply checked for $k=1$ and 2 .

In general, explicit formulae for the p.a.c.f. of higher order MA processes will be involved, and difficult to obtain. Thus the p.a.c.f. is not very helpful for the purpose of identifying an $MA(q)$, though of course it can be used to discredit an AR model of finite order.

7. Some Examples of How the Results can be Applied

We conclude this paper with two demonstrations of how the properties discussed might be useful in analysis.

First consider the stationary ARMA(1, 2) process

$$(1 - \phi B)Z_t = (1 + \theta_1 B + \theta_2 B^2)A_t \quad (21)$$

with $|\phi| < 1$, which can be written in random shock form as the infinite moving average

$$Z_t = A_t + (\phi + \theta_1)A_{t-1} + \{\phi(\phi + \theta_1) + \theta_2\} \sum_{j=2}^{\infty} \phi^{j-2} A_{t-j}$$

Thus the theoretical variance and autocovariances of (21) are given by

$$\begin{aligned} \gamma_0 &= \left[1 + (\phi + \theta_1)^2 + \{\phi(\phi + \theta_1) + \theta_2\}^2 \sum_{r=0}^{\infty} \phi^{2r} \right] \sigma_A^2 \\ \gamma_1 &= \left[(\phi + \theta_1)\{1 + \phi(\phi + \theta_1) + \theta_2\} + \phi\{\phi(\phi + \theta_1) + \theta_2\}^2 \sum_{r=0}^{\infty} \phi^{2r} \right] \sigma_A^2 \\ \gamma_2 &= \left[\{\phi(\phi + \theta_1) + \theta_2\}\{1 + \phi(\phi + \theta_1)\} + \phi^2\{\phi(\phi + \theta_1) + \theta_2\}^2 \sum_{r=0}^{\infty} \phi^{2r} \right] \sigma_A^2 \\ \gamma_k &= \phi \gamma_{k-1} \quad k > 2. \end{aligned}$$

A typical process might have

$$\phi = 0.2, \quad \theta_1 = 0.8, \quad \theta_2 = 1$$

say, giving

$$\gamma_0 = 3.5, \gamma_1 = 2.5, \gamma_2 = 1.5$$

and

$$\gamma_k = 0.2^{k-2} \times 1.5 \quad k > 2.$$

The corresponding autocorrelations are shown in Figure 3.

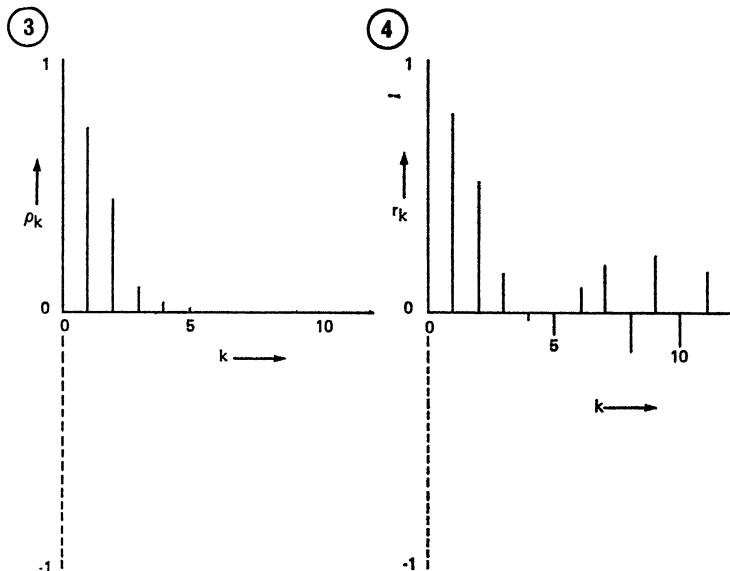


Figure 3

Theoretical autocorrelations for ARMA (1, 2) process.

Figure 4

Sampled autocorrelations for ARMA (1, 2) series.

For this theoretical pattern, though we have two substantial “spikes” at ρ_1 and ρ_2 , the exponential decay for $\{\rho_k: k \geq 1\}$ tells us that we have an ARMA (1, 2) process, not an MA(2). We note, however, that just by looking at ρ_1 and ρ_2 , the MA(2) process is disqualified. This is because

$$\rho_1 = \frac{5}{7} > \frac{1}{\sqrt{2}} = \cos\left(\frac{\pi}{2+2}\right)$$

violates the individual inequality; and because

$$\rho_1 + \rho_2 = \frac{5}{7} + \frac{3}{7} > 1 = \frac{2}{2}$$

the overall inequality is also violated.

Now, except for very long series realizations, the ρ_k , $k > 2$, will be small relative to the sort of sampling errors which are likely to occur, and so it might be expected that the exponential decay will not be apparent in the sampled a.c.f. In this case an MA(2) might well be identified, except that the "inequalities" will still have a good chance of disqualifying it.

To test this, an actual simulation of (21), of length 49, was analysed. Its sample a.c.f., $\{r_k: k=0, 1, \dots\}$, is shown in Figure 4. Though r_1 and r_2 are large, the remaining r_k , $k > 2$, are small and follow no obvious pattern. Under a hypothesis of the series having arisen from an MA(2), all these later r_k can be shown to be insignificantly different from zero. However r_1 and r_2 are both larger than is allowed by the individual inequality, and their sum is considerably larger than that permitted by the overall inequality.

Quite a considerable calculation is required to test whether this pair (r_1, r_2) is significantly inadmissible. However, since we are not committing ourselves to anything at the tentative identification stage, we can afford to work at a level of "hunch", rather than formal significance; and the hunch is that a rather more complicated model is needed. So it would be wise to start with a simple "overfit" to MA(2). Since r_3 is really quite small, an ARMA(1, 2) would seem a wiser choice than an MA(3), though this latter alternative would probably give rise to a very adequate fit. (In fact (21) can be rewritten as

$$Z_t = (1 + B + 1.2B^2 + 0.24B^3 + 0.048B^4 + \dots)A_t$$

which, except for very long series realizations, is effectively equivalent to an MA(3) model.)

We conclude that, for such an example, use of the inequalities would have saved a complete iteration of the Box-Jenkins cycle; as, if an MA(2) had been tried first, we would have needed the resulting residuals analysis before picking up the existence of an autoregressive parameter.

Finally, we give an example, taken from Anderson (1975d), of how a theoretically rather unappealing fit can be more satisfactorily explained by decomposing it into simpler subfits. A series consisting of the daily numbers of traffic wardens off sick in a large city was found to be adequately fitted, statistically, by an ARMA(2, 1) model. The problem was then to explain to ourselves how this fit could have physically come about, and to interpret it for the officials.

Now, it is simple to show, using Granger's lemma, that the sum of two independent stationary first-order autoregressive processes gives rise to an ARMA(2, 1) process; and that the reverse holds, provided a certain "realizability" condition is not violated. (See the appendix for details.)

The obvious guess was then that the observed ARMA(2, 1) had arisen from the superposition of two AR(1) processes, for respectively male and female wardens.

It was easily checked that over the period of observation, the numbers of male and female wardens employed had hardly changed, so stationarity for the two subseries was a not unreasonable assumption. Neither did independence of the two series appear too far-fetched. Also the realizability condition, for such a decomposition to be possible, was obeyed.

So an explanation was offered to the officials that, for each sub-series, there was a model of the form

number sick on a particular day
 = a proportion of those sick on the previous day
 + the newly sick.

This was easy to grasp, and was accepted; and of course, mathematically, the statement can be written in the form

$$S_t = \phi S_{t-1} + C_t \quad 0 < \phi < 1$$

where C_t can be considered as an independent normal random variable, with fixed positive mean and constant variance. That is C_t can be written in the form $C + A_t$, where C is the expected number of "the newly sick" Then

$$E[S_t] = C/(1 - \phi)$$

and, working with $\{Z_t\}$ defined by

$$Z_t = S_t - E[S_t]$$

the process of deviations of the "number of sick on a particular day" from *their* expected number, we get the AR(1) model

$$Z_t = \phi Z_{t-1} + A_t \quad 0 < \phi < 1$$

Unfortunately it was not possible to check out the sub-series for the historic data, as records by sex had not been kept. But kind cooperation from the city, for a further period, showed that in fact AR(1) models fitted very well for this later period.

Appendix

Proposition: If $\{X_t\}$ and $\{Y_t\}$ are independent AR(1) processes, written as, say,

$$(1 - \alpha B)X_t = A_t \quad (22)$$

$$(1 - \beta B)Y_t = B_t \quad (23)$$

where $\{A_t\}$ and $\{B_t\}$ are independent white noise processes, then

$\{Z_t = X_t + Y_t\}$ follows an ARMA(2, 1) process.

Proof:
$$Z_t = (1 - \alpha B)^{-1} A_t + (1 - \beta B)^{-1} B_t$$

so

$$(1 - \alpha B)(1 - \beta B)Z_t = (1 - \beta B)A_t + (1 - \alpha\beta)B_t.$$

The right hand side of this is the sum of two independent MA(1) processes, and so by Granger's lemma can be rewritten as

$$(1 - \gamma B)C_t$$

where $\{C_t\}$ is a white noise process, and it is immediately verified that

$$\gamma \text{ lies between } \alpha \text{ and } \beta \quad (24)$$

Conversely, if (24) holds, the ARMA(2, 1) can be decomposed into (22) and (23). So (24) gives the "realizability" condition.

Acknowledgement

Much of the material of this paper was presented at the 1974 Annual Conference of the Institute of Statisticians, York.

REFERENCES

- ANDERSON, O. D. (1972). An inequality with a time series application. M.Sc. project, Nottingham University.† (Revised version published in *Journal of Econometrics*, 1974, 2, 189–93.)
- ANDERSON, O. D. (1974a). Convex properties of moving average processes. Nottingham Time Series Papers, No. 5.†
- ANDERSON, O. D. (1974b). The orthogonal decomposition of moving average processes. Nottingham Time Series Papers, No. 6.†
- ANDERSON, O. D. (1975a). Bounding sums for the autocorrelations of moving average processes. *Biometrika* 62, 706–7.
- ANDERSON, O. D. (1975b). On a paper by Davies, Pate and Frost concerning maximum autocorrelations for moving average processes. *Australian Journal of Statistics* 17, to appear.
- ANDERSON, O. D. (1975c). On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics*, 3, 331–5.
- ANDERSON, O. D. (1975d). *Time Series Analysis and Forecasting: the Box-Jenkins Approach*. London: Butterworths.
- BELLMAN, R. (1970). *Introduction to Matrix Analysis*, 2nd ed. New York: McGraw-Hill.
- BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- GRANGER, C. W. J. (1972). Time Series Modelling and Interpretation. Paper presented to the European Econometric Congress, Budapest.
- Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts (with Discussion). *Journal of the Royal Statistical Association A*, 138, 131–65.

† Available from the author.