ON THE STRUCTURE OF MOVING AVERAGE PROCESSES

Craig F. ANSLEY, W. Allen SPIVEY, and William J. WROBLESKI

The University of Michigan, Ann Arbor, MI 48109, U.S.A.

This paper uses Hilbert space methods to develop a rigorous proof that the sum of two uncorrelated moving average processes of order q_1 and q_2 is an MA process of order $q \leq$ max (q_1, q_2). The methods establish the existence of suitable random shocks for the summed process, they illuminate relationships between the coefficients of such processes and their random shocks, and they provide means for proving that the random shocks of the summed processes are normal when the shocks of the underlying processes are normal. The role of the Wold decomposition is examined in terms of multiple representations of an MA process.

1. Introduction

In a recent paper in this journal O.D. Anderson (1975a) discusses the theorem that the sum of two independent moving average processes of order q_1 and q_2 is itself a moving average process of order $q \leq$ max (q_1, q_2). This summation theorem has considerable practical importance. Box and Jenkins (1970, p. 121) used it without proof to investigate the effects of correlated noise and white noise on ARMA processes. O.D. Anderson (1975a, b) shows how the theorem can be used to explain the development of composite models from simpler processes which are amenable to practical interpretations. The result has also been used by Zellner and Palm (1974) and by Zellner (1975). In two important papers they have investigated relationships between structural assumptions in simultaneous equation econometric models and their associated final equations and transfer functions. Their approach, which in turn uses the Box–Jenkins ARMA model formulation procedures, makes implicit use of the theorem above [Zellner and Palm (1974, p. 19) and Zellner (1975, p. 378)].

Although this theorem may appear to be intuitively obvious, the work of a number of authors has shown that its proof is far from trivial. T.W. Anderson (1971, pp. 224–225) outlines the proof of a result which can be used to prove this theorem. This development requires lengthy arguments and is not simple. Granger (1972) attempted a proof of the theorem based on frequency domain arguments, but according to O.D. Anderson (1975a, p. 151), 'it has a number of flaws'. O.D. Anderson's attempt to develop a simple proof rests on a convexity property of vectors of MA processes of fixed order q. However, in his develop-

ment of this convexity property, there is a crucial omission which renders his proof incomplete.

This paper uses Hilbert space methods to develop a rigorous proof of the summation theorem. We begin by discussing the Wold decomposition, a now classical result, using a Hilbert space formulation. Directly from the Hilbert space development of the Wold decomposition we are able to give a simple proof of a result of T.W. Anderson, referred to earlier, which showed that any stationary process for which the autocorrelations vanish after lag q has an MA(q) representation. We then show that the random shocks in the MA representation of the sum of two independent, normal MA processes are themselves normal. This result is of particular importance to the work of Zellner and Palm, because it justifies their use of Box–Jenkins estimation methods for final equation and transfer function models.

The Hilbert space approach has several advantages. First, although based on sophisticated mathematical arguments, it is both elegant and conceptually simple. Second, because Hilbert space methods depend only on covariance properties, the proof of the generalization of the summation theorem to uncorrelated rather than independent processes is immediate. Third, the Hilbert space framework specifies clearly the relationships between the MA process and its random shocks. We point out that O.D. Anderson, in attempting to prove the summation theorem, addressed only the problem of the existence of suitable MA coefficients but ignored the question of whether an appropriate stream of random shocks can be found. Our Hilbert space development resolves the problem of the existence of suitable random shocks by giving explicit representations of them as elements in the Hilbert space.

2. Examination of O.D. Anderson's argument

Anderson's argument (1975a) is based on the convexity of the set of vectors of autocorrelation functions of MA(q) processes. The major omission in his approach results from the way in which he attempts to establish the convexity property. To understand the nature of this omission consider the class of MA(q) processes

$$y_t = \sum_0^q \theta_j a_{t-j}, \qquad \theta_0 = 1, \tag{1}$$

whose autocorrelations are

$$\rho_k = \sum_{j=0}^{q-k} \theta_j \theta_{j+k} \bigg/ \sum_{j=0}^q \theta_j^2, \qquad k = 1, \ldots, q. \tag{2}$$

Let

$$w = 1 \bigg/ \sum_{j=0}^q \theta_j^2, \tag{3}$$

and denote each of two MA(q) processes by one and two primes, respectively.[1]
Then for $0 < \lambda < 1$ we have

$$\lambda\rho'_k + (1-\lambda)\rho''_k = \lambda w' \sum_{j=0}^{q-k} \theta'_j\theta'_{j+k} + (1-\lambda)w'' \sum_{j=0}^{q-k} \theta''_j\theta''_{j+k}. \qquad (4)$$

Eq. (4) can be written in the form (2) if the system of nonlinear equations

$$\sum_{j=0}^{q-k} \theta_j\theta_{j+k} \Big/ \sum_{j=0}^{q} \theta_j^2 = \lambda\rho'_k + (1-\lambda)\rho''_k, \qquad k = 1, \ldots, q, \qquad (5)$$

has real solutions $\theta_1, \ldots, \theta_q$. Anderson asserts without proof that such real solutions exist. ['We obtain q equations for the q unknowns $\theta_1, \ldots, \theta_q, \ldots$, and so the resulting equations are soluble.' See O.D. Anderson (1975a, p. 155).] The remainder of his argument rests on this assertion.

However, it is not obvious that (5) has real solutions and a proof of this point is required. We develop such a proof and comment in the final section of this paper on the existence of multiple solution sets corresponding to multiple representations of MA processes.

Besides showing the existence of suitable coefficients, it must also be proved that there exist suitable random shocks $\{a_t\}$ such that

$$y_t = \sum_0^q \theta_j a_{t-j}, \qquad \theta_0 = 1, \qquad (6)$$

with probability 1, where $y_t = y'_t + y''_t$ and where y'_t and y''_t are the MA processes whose autocorrelations appear in (4). Establishing this is especially important if one wants to make inferences concerning distributions of statistics associated with time series data. As mentioned earlier, O.D. Anderson's convexity argument is incomplete because he has not considered this problem. The existence of a stream of such random shocks a_t emerges naturally from the Hilbert space approach we take below. In particular, we show that the random shocks associated with the sum of independent normal processes are themselves normally distributed.

3. Some preliminaries and the Wold decomposition

In order to establish the results indicated earlier, some concepts and properties concerning wide sense stationary processes and their representations in a Hilbert space must be introduced. We first define a wide sense stationary process $\{y_t\}$ as a process with the following properties:

---

[1] We note that the class of MA(q^*) processes, where $q^* < q$, is embedded in the class of MA(q) processes.

(i) $E(y_t)$ is a constant for all t, i.e., $E(y_t) = \mu$ for all t; and

(ii) cov (y_s, y_t) is a function of $s - t$ alone, i.e., cov $(y_s, y_t) = \gamma_{s-t}$ for all s, t.

This definition is sometimes referred to as covariance stationarity. Note that the definition implies that the variance of $\{y_t\}$ (assumed to be finite) is a constant for all t. Without loss of generality we assume that $\mu = 0$ for the remainder of this paper.

Following Rozanov (1967, p. 3) a Hilbert space H can be generated from the closure with respect to mean square convergence of the linear manifold generated by the random variables $\{y_t; -\infty < t < \infty\}$. The elements of the Hilbert space H are random variables and the inner product of any two random variables $x_1, x_2 \in H$ is

$$(x_1, x_2) = \text{cov}\,(x_1, x_2). \tag{7}$$

We write $H(t)$ to represent the (Hilbert) subspace of H formed by the closure in mean square of the linear manifold generated by $\{y_s; s \leq t\}$. Further, let $D(t)$ denote the orthogonal complement of $H(t-1)$ in $H(t)$. Then every $x_1 \in D(t)$ is orthogonal to every $x_2 \in H(t-1)$. $D(t)$ is a closed subspace of $H(t)$, and $H(t-1)$ is a closed subspace of $H(t)$.

If S is a closed subspace of H and $x \in H$, we write $E^*(x|S)$ to represent the projection of x on S.[2]

As Parzen (1961, p. 961) points out, the best linear predictor of y_{t+s}, in terms of mean square error, given all the values of the time series up to and including y_t, is $E^*(y_{t+s}|H(t))$, the projection of y_{t+s} on $H(t)$.

We are interested in processes with no deterministic component, i.e., in processes for which the best linear predictor $E^*(y_{t+s}|H(t)) \to 0$, the mean value, as the prediction interval $s \to \infty$ for all t. Such processes are called *regular* processes.

Rozanov (1967, p. 156) has shown that for regular processes there is a unique sequence of constants $\{c_j; j \geq 0\}$ such that

$$y_t = \sum_0^\infty c_j a_{t-j}, \qquad c_0 = 1, \tag{8}$$

for all t, and where

$$a_t = E^*(y_t|D(t)). \tag{9}$$

---

[2]Parzen (1959) uses this notation because the projection in a Hilbert space of random variables has many of the properties of conditional expectation. For the same reason, Doob (1953, p. 155) refers to projection as 'wide sense conditional expectation'. Parzen discusses a necessary and sufficient condition for projection to be conditional expectation.

The $\{a_t\}$ form a sequence of uncorrelated random variables with constant variance σ_a^2.

This representation, having been first investigated by Wold (1954), is known as the *Wold decomposition*. If $\{y_t\}$ is a real process, the coefficients c_j and the random variables a_t are necessarily real.

An important application of this decomposition is in providing an explicit representation for $E^*(y_{t+s}|H(t))$. From the Wold decomposition (8) we note

$$y_{t+s} = \sum_0^{s-1} c_j a_{t+s-j} + \sum_s^\infty c_j a_{t+s-j}.$$

Now

$$\sum_0^{s-1} c_j a_{t+s-j}$$

is orthogonal to $H(t)$, because each a_{t+s-j} in this sum is an element of $D(t+s-j)$, which is orthogonal to $H(t+s-j-1) \supset H(t)$. Also

$$\sum_s^\infty c_j a_{t+s-j}$$

is an element of $H(t)$ because each a_{t+s-j} in this sum is an element of $H(t+s-j)$ $\subset H(t)$. By the standard uniqueness properties of projections in Hilbert space [see e.g. Parzen (1959, p. 306)] we have

$$E^*(y_{t+s}|H(t)) = \sum_s^\infty c_j a_{t+s-j}. \tag{10}$$

A special case exists where the covariance function is identically zero after a given lag q, i.e., for $s > q$,

$$(y_{t+s}, y_t) = \gamma_s = 0. \tag{11}$$

We show that such a process has a MA(q) representation. It is clear from (11) that y_{t+s} is orthogonal to $H(t)$ for $s > q$, and thus

$$E^*(y_{t+s}|H(t)) = 0.$$

This establishes regularity and for all $s > q$ we have from (10)

$$\left\| E^*(y_{t+s}|H(t)) \right\|^2 = \sigma_a^2 \sum_s^\infty |c_j|^2 = 0,$$

J

giving $c_j = 0$ for all $j \geq s > q$. Thus for such a process the Wold decomposition can be written as

$$y_t = \sum_0^q c_j a_{t-j}, \qquad c_0 = 1, \tag{12}$$

and $\{y_t\}$ has a MA(q) representation as asserted.[3]

Consider now an MA(q) process

$$y_t = \sum_0^q \theta_j u_{t-j}, \qquad \theta_0 = 1, \tag{13}$$

where $\{u_t\}$ is an uncorrelated sequence of random shocks with mean zero and constant variance σ_u^2. This process has autocovariance function $\gamma_s = 0$ for $s > q$, so that from (12) an MA(q) process has the unique Wold decomposition

$$y_t = \sum_0^q c_j a_{t-j}, \tag{14}$$

where

$$a_t = E^*(y_t | D(t)), \tag{15}$$

as in (9). The representations (14) and (15) are not as obvious as they might appear at first sight, because an MA(q) process may have multiple representations. The representation (14) is a canonical representation in which the c_j may not coincide with the θ_j used in the defining eq. (13), nor the a_t coincide with the u_t. We will discuss some of the implications of multiple representations as they affect the summation theorem in the final section.

4. Normal processes

In this section we show that if $\{y_t\}$ is a normal stationary process then every non-zero element in the Hilbert space H generated by $\{y_t; -\infty < t < \infty\}$ has a normal distribution. By a *normal process* we mean a process $\{y_t\}$ where any finite set of random variables $\{y_{t_1}, \ldots, y_{t_n}\}$ from the process has a multivariate normal distribution. A stationary normal process is therefore strictly stationary in the sense that all the moments of the joint distribution of y_s and y_t depend on $s - t$ alone. An example of a normal stationary process would be an MA process generated by independent normal random shocks. We can extend the primary results by showing that any two (non-zero) elements x_1 and x_2 in H, and indeed

---

[3]This is a simple proof of the result of T.W. Anderson (1971, pp. 224–225) and referred to earlier.

any finite set of (non-zero) random variables from H, have a multivariate normal distribution.[4]

To prove these results, we note that for any non-zero $x \in H$ there exists a sequence $\{x_n\}$ of finite linear combinations of $\{y_t; -\infty < t < \infty\}$ such that $x_n \to x$. Suppose that x_n and x have respectively variances $\sigma_n^2 > 0$ and $\sigma^2 > 0$, distribution functions F_n and F, and characteristic functions ϕ_n and ϕ. We have by mean square convergence

$$\sigma_n^2 \to \sigma^2, \tag{16}$$

and

$$F_n \to F, \tag{17}$$

at every continuity point of F [see Fisz (1963, p. 238)], and therefore for all h such that $-\infty < h < \infty$ we have

$$\phi_n(h) \to \phi(h) \tag{18}$$

[see Fisz (1963, p. 188)]. Each x_n is a finite linear combination of random variables with a multivariate normal distribution and, therefore, is itself normally distributed. Now from (16),

$$\phi_n(h) = \exp\left[-\sigma_n^2 h^2 / 2\right] \to \exp\left[-\sigma^2 h^2 / 2\right].$$

Therefore from (18) we have

$$\phi(h) = \exp\left[-\sigma^2 h^2 / 2\right],$$

and x has a normal distribution as asserted.

We extend this result to more than one variable by noting that (17) and (18) extend to multivariate distributions and that if $x_{1n} \to x_1$ and $x_{2m} \to x_2$, then

$$\text{cov}\,(x_{1n}, x_{2m}) = (x_{1n}, x_{2m}) \to (x_1, x_2) = \text{cov}\,(x_1, x_2)$$

as $n, m \to \infty$. Noting that x_{1n} and x_{2m} have a bivariate normal distribution, we have

$$\phi_{nm}(h_1, h_2) = \exp\left[-\tfrac{1}{2}(\sigma_n^2 h_1^2 + 2h_1 h_2 \,\text{cov}\,(x_{1n}, x_{1m}) + \sigma_{2m}^2 h_2^2)\right]$$

$$\to \exp\left[-\tfrac{1}{2}(\sigma_1^2 h_1^2 + 2h_1 h_2 \,\text{cov}\,(x_1, x_2) + \sigma_2^2 h_2^2)\right]$$

$$= \phi(h_1, h_2),$$

[4]In this case uncorrelated random variables are also independent, and we see that projection becomes conditional expectation.

so that x_1 and x_2 have a bivariate normal distribution as asserted. Clearly, using an identical argument this result extends to a finite family $x_1, \ldots, x_n$ of random variables in H.

In particular, if $\{y_t\}$ is a normal stationary process, then the random shocks $\{a_t\}$ in the Wold decomposition (8) are uncorrelated multivariate normal and, therefore, are independent normal, random variables.

5. The summation theorem

The theorem can be formally stated as follows. Suppose that $\{y_{1t}\}$ and $\{y_{2t}\}$ are uncorrelated stochastic processes with $MA(q_1)$ and $MA(q_2)$ representations

$$y_{it} = \sum_{j=0}^{q_i} \theta_{ij} a_{it-j}, \qquad \theta_{i0} = 1, \quad i = 1, 2, \tag{19}$$

where for $i = 1, 2$ and for all s and t,

$$\operatorname{var}(a_{it}) = \sigma_i^2,$$

$$\operatorname{cov}(a_{it}, a_{is}) = 0,$$

$$E(a_{it}) = 0.$$

Then $\{y_t\}$, where $y_t = y_{1t} + y_{2t}$, has a $MA(q^*)$ representation, where

$$q^* \leq q = \max(q_1, q_2).$$

Moreover, if $\{y_{1t}\}$ and $\{y_{2t}\}$ are independent normal processes, then $\{y_t\}$ is normal and its associated random shocks have independent normal distributions.

We begin the proof by observing that $\{y_t\}$ is wide sense stationary with covariance function $\{\gamma_k\}$ such that

$$\gamma_k = 0, \qquad k > q = \max(q_1, q_2).$$

It follows immediately from (12) that $\{y_t\}$ has the (Wold) $MA(q)$ representation

$$y_t = \sum_0^q c_j a_{t-j}, \qquad c_0 = 1,$$

where

$$a_t = E^*(y_t | D(t)),$$

and where $\{c_j\}$ and $\{a_t\}$ are real. Note that we do not exclude the possibility

that $c_j = 0$ for $1 \leq j \leq q$, and therefore we see that $\{y_t\}$ has an MA(q^*) representation for some q^* where $0 \leq q^* \leq q$ and where $c_{q^*} \neq 0$.

If $\{y_{1t}\}$ and $\{y_{2t}\}$ are independent normal processes, we see immediately that $\{y_t\}$ is also normal, and, by the results developed earlier, that its random shocks are independent and normally distributed.   Q.E.D.

6. Comments on the summation theorem

We conclude this paper with some observations on how our results relate to the coefficient structure of MA processes.

First, we have pointed out that in the proof of the summation theorem it is possible that q^* is less than $\max(q_1, q_2)$. It is not difficult to see that this will be the case whenever $q_1 = q_2 = q$ and $\gamma_q = \sigma_1^2 \theta_{1q} + \sigma_2^2 \theta_{2q} = 0$. Moreover, it is possible that the sum $\{y_t\}$ may be an uncorrelated sequence of random variables, i.e., an MA(0) process. An example is

$$y_{1t} = a_{1t} - \theta a_{1t-1}, \qquad \sigma_1^2 = \sigma^2,$$

and

$$y_{2t} = a_{2t} + \theta a_{2t-1}, \qquad \sigma_2^2 = \sigma^2.$$

Then $\{y_t\}$, where $y_t = y_{1t} + y_{2t}$, is a sequence of uncorrelated random variables with variance $2(1 + \theta^2)\sigma^2$. Because an MA process may have multiple representations (see below), there is no simple way to state the conditions which are necessary and sufficient for this situation to arise.

Although it may seem improbable that either of these situations could arise in practice, it is easy to imagine a situation where some of the γ_k were very small after a given lag, and not regarded as significant after identification or estimation of the MA(q) model using standard Box–Jenkins methods. This is an important point to bear in mind when using Zellner's and Palm's technique for examining structural assumptions in econometric models.

We return now to our result (12). If a process $\{y_t\}$ has autocovariances $\{\gamma_k\}$ that vanish after a given lag q, we have in effect shown that the set of $q+1$ equations

$$\sigma_a^2 \sum_0^{q-k} \theta_j \theta_{j+k} = \gamma_k, \qquad \theta_0 = 1, \quad k = 0, \ldots, q, \tag{20}$$

in the $q+1$ unknowns $\theta_1, \ldots, \theta_q, \sigma_a^2$ has at least one real solution. T.W. Anderson (1971, pp. 224–225) reaches the result (12) by showing directly that a real solution exists.

We now discuss how multiple solutions can arise. Box and Jenkins (1970,

pp. 196–197) show that for the MA(q) process,

$$y_t = \sum_0^q \theta_j a_{t-j}, \qquad \theta_0 = 1,$$

there are multiple choices of the θ_j and a_t which lead to multiple representations of $\{y_t\}$. Essentially, they consider the characteristic function

$$C(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$$

$$= \prod_1^q (1 - \alpha_i B),$$

and show that one can construct another function

$$C^*(B) = \prod_1^q (1 - \alpha_i^* B) = 1 + \theta_1^* B + \ldots + \theta_q^* B^q,$$

by arbitrarily setting α_i^* equal to either α_i or $1/\alpha_i$. Then by choosing suitable random shocks $\{a_t^*\}$ we have a new representation

$$y_t = C^*(B) a_t^* = \sum_0^q \theta_j^* a_{t-j}^*, \tag{21}$$

where B is the backward shift operator. There are 2^q different ways of choosing the α_i to be replaced by their inverses in $C^*(B)$; this implies that there are up to 2^q different solutions to eqs. (20). The actual number of solutions may be less than 2^q for two reasons. First, some of the 2^q possible choices of $C^*(B)$ may be identical, because two or more of the factors $(1 - \alpha_i^* B)$ may be identical. Second, some of the α_i may be complex. Such α_i appear in complex conjugate pairs because the coefficients of $C(B)$ are real, and, to ensure that the θ_i^* are real, the α_i^* must also be chosen in complex conjugate pairs and not independently.

Box and Jenkins do not consider the existence of suitable random shocks for (21). However, it can be shown that suitable random shocks can be found for the multiple representations provided $C(B)$, and thus $C^*(B)$ have no zeros on the unit circle. If all the zeros of $C^*(B)$ lie outside the unit circle, then the random shocks a_t^* can be written in the form

$$a_t^* = \sum_0^\infty r_j y_{t-j}. \tag{22}$$

The representation in this case is said to be *invertible*. If all the zeros of $C^*(B)$ lie inside the unit circle, we can write

$$a_t^* = \sum_0^\infty r_j y_{t+q+j},$$

and if some zeros lie outside and some inside the unit circle, we can write

$$a_t^* = \sum_{-\infty}^{\infty} r_j y_{t-j}.$$

In each case

$$\sum |r_j|^2 < \infty,$$

and the limits can be shown to exist in mean square. In the case where all the zeros lie outside the unit circle, we see that a_t^* in (22) is an element of $H(t)$, and the corresponding representation is therefore the Wold representation (14).

Our proof of the result (12) establishes the existence of one set of coefficients (those in the Wold decomposition) which satisfies eqs. (20). By using the procedures of Box and Jenkins to construct multiple representations of the resulting MA process, we see there are corresponding multiple solutions to (20).

The existence of multiple representations and therefore multiple solutions to (20) can be illustrated by the following example. Consider the MA(2) process,

$$y_t = u_t - u_{t-1} + 0.25\, u_{t-2}, \tag{23}$$

where

$$C(B) = 1 - B + 0.25\, B^2 = (1 - 0.5\, B)^2. \tag{24}$$

Both zeros lie outside the unit circle, and (23) is therefore the Wold representation (14). We note that

$$u_t = \sum_{0}^{\infty} (1+j)\,(0.5)^j\, y_{t-j},$$

and

$$\mathrm{var}\,(u_t) = (16/33)\,\gamma_0.$$

We can now construct

$$\tilde{C}(B) = (1 - 2B)^2 = 1 - 4B + 4B^2,$$

and write

$$y_t = \tilde{a}_t - 4\tilde{a}_{t-1} + 4\tilde{a}_{t-2},$$

where

$$\tilde{a}_t = 0.25 \sum_{0}^{\infty} (1+j)\,(0.5)^j\, y_{t+2+j},$$

and

$$\text{var}\,(\tilde{a}_t) = (1/33)\gamma_0.$$

The remaining possibility is to choose

$$C^*(B) = (1-2B)(1-0.5B) = 1-2.5B+B^2,$$

and write

$$y_t = u_t^* - 2.5\,u_{t-1}^* + u_{t-2}^*,$$

where

$$u_t^* = -\tfrac{2}{3}\sum_{-\infty}^{\infty}(0.5)^{|j|}\,y_{t+1-j},$$

and

$$\text{var}\,(u_t^*) = (4/33)\,\gamma_0.$$

Note that the number of possibilities is reduced from $2^2 = 4$ to 3 because there is a double factor of $(1-0.5\,B)^2$ in (24).

Finally, we comment on the case where $C(B)$ has a zero on the unit circle. It is not easy in this case to show directly that suitable random shocks exist corresponding to solutions of (20). T.W. Anderson (1971, p. 226) specifically excludes this case from his discussion. The problem is that we cannot write the random shocks in the Wold representation (14) in the form (22).

This special case is now illustrated by a numerical example. Consider the MA(1) process.

$$y_t = a_t - a_{t-1}; \tag{25}$$

the characteristic function $C(B) = 1-B$ has a zero $B = 1$ on the unit circle. Solving for a_t, we see that

$$a_t = y_t + y_{t-1} + y_{t-2} + \cdots \tag{26}$$

Clearly, this does not converge. To gain some insight into the relationship between the process and its random shocks, we now construct a sequence of finite linear combinations of $\{y_s;\, s \leq t\}$ which does converge to a_t.

Consider first the sequence $\{\tilde{a}_t^{(n)},\, n = 1, 2, \ldots\}$ defined by

$$\tilde{a}_t^{(n)} = \sum_{k=0}^{\infty} 2^{-k/n}\,y_{t-k}.$$

We note that $\tilde{a}_t^{(n)}$ is a well defined element of $H(t)$ because

$$\|\tilde{a}_t^{(n)}\| \leq \gamma_0^{1/2} \sum_{k=0}^{\infty} 2^{-k/n} = \gamma_0^{1/2}/(1-2^{-1/n}) < \infty,$$

where $\gamma_0 = \mathrm{var}\,(y_t)$. The sequence $\{\tilde{a}_t^{(n)}\}$ is a Cauchy sequence because

$$\|\tilde{a}_t^{(n)} - \tilde{a}_t^{(m)}\|^2 = \gamma_0 \left| \frac{1-2^{-1/n}}{1-2^{-2/n}} - \frac{2-2^{-1/n}-2^{-1/m}}{1-2^{-(1/n+1/m)}} + \frac{1-2^{-1/m}}{1-2^{-2/m}} \right|$$

$$\to 0,$$

as $n, m \to \infty$. Thus $\{\tilde{a}_t^{(n)}\}$ approaches a unique limit point in $H(t)$ which, by considering the inner products $(\tilde{a}_t^{(n)}, a_s)$ for $s \leq t$, we can show to be a_t.

If we choose N_n such that

$$\left\| \tilde{a}_t^{(n)} - \sum_{k=0}^{N_n} 2^{-k/n} y_{t-k} \right\| < 1/2^n,$$

and define

$$a_t^{(n)} = \sum_{k=0}^{N_n} 2^{-k/n} y_{t-k}, \tag{27}$$

it is clear that

$$a_t = \lim_{n\to\infty} a_t^{(n)} = \lim_{n\to\infty} \sum_{k=0}^{N_n} 2^{-k/n} y_{t-k}.$$

We have constructed a sequence of finite linear combinations of $\{y_s; s \leq t\}$ which converges to a_t. It has already been shown from (27) that a_t cannot be expressed in the form (22) and it can now be observed that this situation arises because limiting operations in (27) cannot be interchanged. This is analogous to the situation where uniform convergence of a power series no longer holds on its radius of convergence. In fact, this property can be directly related to the divergence of the inverse of $C(B)$ on the unit circle.

Now, by a similar argument we can show that

$$a_t = -\lim_{n\to\infty} \sum_{k=0}^{N_n} 2^{-k/n} y_{t+k+1}, \tag{28}$$

corresponding to (27). Thus we can recover the random shocks a_t by a limiting process on either past or future values of the process. O.D Anderson (1975b) calls this the *semi-invertible* case.

к

As we said earlier, the problem of showing the existence of a sequence $\{a_t^{(n)}\}$ converging to $\{a_t\}$ in the general case where $C(B)$ may have a zero on the unit circle is not a simple matter if approached directly. However, the problem is almost deceptively easy using the Hilbert space approach: a_t is simply the projection of y_t on the subspace $D(t)$.

References

Anderson, O.D., 1975a, On a lemma associated with Box, Jenkins and Granger, Journal of Econometrics 3, 151–156.

Anderson, O.D., 1975b, Moving average processes, Statistician 24, no. 4, 283–297.

Anderson, T.W., 1971, The statistical analysis of time series (Wiley, New York).

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis, forecasting and control (Holden Day, San Francisco).

Doob, J.L., 1953, Stochastic processes (Wiley, New York).

Fisz, M., 1963, Probability theory and mathematical statistics, 3rd ed. (Wiley, New York).

Granger, C.W.J., 1972, Time series modelling and interpretation, Paper presented to the European Econometric Congress (Budapest).

Parzen, E., 1959, Statistical inference on time series by Hilbert space methods, Technical Report no. 23, Jan., O.N.R. Contract 225(21) (Statistics Department, Stanford University, CA). Reprinted in: E. Parzen, ed., 1967, Time series analysis papers (Holden Day, San Francisco) 251–382.

Parzen, E., 1961, An approach to time series analysis, Annals of Mathematical Statistics 32, 951–989. Reprinted in: E. Parzen, ed., 1967, Time series analysis papers (Holden Day, San Francisco) 383–421.

Rosanov, Y.A., 1967, Stationary random processes (Holden Day, San Francisco).

Wold, H.O., 1954, A study in the analysis of stationary time series (Almquist and Wiksell, Stockholm). Originally published in 1937.

Zellner, A. and F. Palm, 1974, Time series analysis and simultaneous equation econometric models, Journal of Econometrics 2, 17–54.

Zellner, A., 1975, Time series analysis and econometric model construction, in: R.P. Gupta, ed., Applied statistics (North-Holland, Amsterdam).