

On the Prediction of Non-Stationary Processes

N. A. Abdrabbo

M. B. Priestley

University of Manchester

Received August 1966. Revised January 1967

Abstract

The problem of linear least-squares prediction for a class of non-stationary processes which possess “evolutionary spectral representations” is considered. It is shown that, under certain conditions, such processes admit moving-average representations in terms of time-dependent coefficients. This feature enables development of a close analogue of the Wiener-Kolmogorov approach to the corresponding problem for stationary processes.

1 Introduction

The Wiener-Kolmogorov “linear least-squares” approach to the prediction and filtering of stationary processes is now well established, and may, in principle, be applied to any practical problem involving stationary processes whose second-order properties are known. Accounts of this theory have been written at various levels of mathematical rigour (see, for example, [1], [2], [3], [4], [5]). However, the corresponding problem for non-stationary processes has received little attention. There have been a few isolated attempts to deal with this topic, but in the main the approaches have been either too general or too restricted to be useful in practical applications. For example, [6] has solved the non-stationary prediction problem in principle, but the approach is somewhat abstract, and the “solution” for the optimum predictor is expressed as a certain inner-product in a Hilbert space. [7] considered the same problem, and obtained some interesting results in the form of “existence theorems”, but did not present a method for determining the explicit form of a predictor in terms of the observed variables. Similar remarks apply to the work of [9]. On the other hand, several papers written from an engineering standpoint have appeared (see, for example, [1], [1], [1]), but in most cases the “general solution” stops with the construction of the well-known integral equation (cf. (94)) involving the covariance function of the process and the unknown “coefficients” of the optimal predictor, the solution of which is obtainable only when the process obeys some very simple model. On a different level, [1] has studied multivariate non-stationary processes corresponding to dynamical systems which are governed by known linear differential equations, and has established some basic results regarding the analytic structure of optimal predictors for such models. However, from the point of view of practical

application, the most useful results so far obtained are due to [5], who considered non-stationary processes generated by autoregressive models with time-dependent coefficients, and obtained explicit recursive relations for the optimal predictors. In fact, some of the results for these particular models correspond very closely to those obtained by Whittle.

The success of classical prediction theory for stationary processes is due essentially to the fact that such processes admit a spectral representation in terms of an orthogonal process. This feature not only simplifies the solution of the prediction problem, but also enables one to treat a general class of stationary processes by means of a “canonical” representation, so that the discussion need not be restricted to particular models such as the autoregressive, moving-average, etc. Hitherto, the lack of a similar spectral representation for a general class of non-stationary processes has no doubt been one of the major stumbling blocks in attempts to generalize the classical theory. However, it turns out that the recently developed theory of evolutionary spectra representations [1] provides an ideal framework for the formulation and solution of non-stationary prediction problems. In fact, by using evolutionary spectral representations one obtains a prediction theory which is almost an exact parallel of the Wiener-Kolmogorov theory. The basic idea underlying this approach is the introduction of the evolutionary (i.e. time-dependent) spectrum of a non-stationary process, whose form completely determines the values of the coefficients of the optimal linear predictor. This means that even if observations from a process whose structure is completely unknown are presented, this prediction theory may still be usefully applied by first estimating the evolutionary spectrum (methods for estimating evolutionary spectra have been discussed by [1], [1]). Of course, this estimation procedure introduces further complications which will not be discussed here (see also Section 6). Throughout this paper it is assumed that the second-order properties of the process are known a priori. It should be pointed out, further, that in this account the aim is not at complete mathematical rigour; rather the aim is to show how this approach may be used to obtain explicit results for non-stationary processes.

Before discussing the prediction problem, relevant parts of the theory of evolutionary spectra are summarized briefly.

2 Evolutionary Spectral Representations

A class of continuous parameter processes, $\{X(t)\}$, with $E\{X(t)\} = 0$, $E\{X^2(t)\} < \infty$, for all t , is considered, for which there exists a family \mathcal{F} of functions $\{\phi_t(\omega)\}$ (defined on the real line and indexed by the suffix t) and a measure μ on the real line such that $\{X(t)\}$ admits a spectral representation of the form

$$X(t) = \int_{-\infty}^{\infty} \phi_t(\omega) dZ(\omega) \quad (1)$$

(See also [1].) Here $\{Z(\omega)\}$ is an orthogonal process with $E[|dZ(\omega)|^2] = \mu(d\omega)$. If there exists a family \mathcal{F} for which each ϕ_t may be written in the form

$$\phi_t(\omega) = e^{i\omega t} A_t(\omega) \quad (2)$$

where, for each fixed ω , the generalized Fourier transform of $A_t(\omega)$ (considered as a function of t) has an absolute maximum at the origin, $X(t)$ is termed an oscillatory process (cf. [1]),

and the evolutionary spectrum at time t with respect to the family \mathcal{F} is defined by

$$dF_t(\omega) = |A_t(\omega)|^2 \mu(d\omega) \quad (-\infty < \omega < \infty) \quad (3)$$

For discrete parameter oscillatory processes (i.e. processes defined only for $t = 0, \pm 1, \pm 2, \dots$) there is a corresponding evolutionary spectral representation

$$X(t) = \int_{-\pi}^{\pi} e^{i\omega t} A_t(\omega) dZ(\omega) \quad (4)$$

and the evolutionary spectrum at time t with respect to the family \mathcal{F} is defined by

$$dF_t(\omega) = |A_t(\omega)|^2 \mu(d\omega) \quad (-\pi \leq \omega \leq \pi) \quad (5)$$

When the measure μ is absolutely continuous (with respect to Lebesgue measure) the derivative $F'_t(\omega) = f_t(\omega)$ exists for all ω , and is termed the evolutionary spectral density function at time t .

In both the continuous and discrete cases the functions $A_t(\omega)$ may be normalized so that $A_0(\omega) = 1$ for all ω . With this convention, $\mu(d\omega) = dF_0(\omega)$, the evolutionary spectrum at time zero.

3 Moving-Average Representations for Oscillatory Processes

As in the prediction theory of stationary processes, the construction begins with a one-sided moving-average representation for a class of oscillatory processes, the distinction with the stationary case being that the coefficients in the moving-average scheme are now time-dependent.

3.1 Discrete-parameter Processes

Suppose that $X(t)$ has a representation of the form (4), and that μ is absolutely continuous with respect to $d\omega$. Then $f_t(\omega)$ may be written in the form

$$f_t(\omega) = |A_t(\omega)|^2 f(\omega) \quad (6)$$

where $f(\omega) = d\mu/d\omega$. Note that $f(\omega) = f_0(\omega)$ must be integrable. Suppose now that

$$\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty \quad (\text{C}_1)$$

Then it follows that [1, Doob, 1953, p. 160] there exists a function $\psi(\omega)$ such that

$$|\psi(\omega)|^2 = f(\omega) \quad (7)$$

where $\psi(z)$, considered as a function of the complex variable $z = e^{i\omega}$, has no poles or zeros inside the unit circle $|z| < 1$. The function $\psi(\omega)$ may be written as a one-sided Fourier transform, viz.

$$\psi(\omega) = \sum_{u=0}^{\infty} e^{-i\omega u} g^*(u) \quad (8)$$

for some suitable sequence $\{g^*(u)\}$.

Suppose further that:

- (C2) The family \mathcal{F} can be chosen so that, for each t , $A_t(z)$ (considered as a function of z) also has no poles or zeros inside the unit circle, so that, for each t , $A_t(\omega)$ may be written in the form,

$$A_t(\omega) = \sum_{u=0}^{\infty} e^{-i\omega u} g_t(u) \quad (9)$$

(It is remarked that $A_t(\omega)$ is certainly square-integrable, since $|A_t(\omega)|^2 f(\omega) = f_t(\omega)$ is integrable, and by condition (C₁), $f(\omega)$ may vanish on at most a set of zero measure.)

Note that a necessary condition for the validity of (9) is

$$\int_{-\pi}^{\pi} \log |A_t(\omega)|^2 d\omega > -\infty \quad (\text{C}_3)$$

for all t . Now write $X(t)$ in the form

$$X(t) = \int_{-\pi}^{\pi} e^{it\omega} \alpha_t(\omega) dz(\omega) \quad (10)$$

where $\alpha_t(\omega) = A_t(\omega)\psi(\omega)$, so that

$$|\alpha_t(\omega)|^2 = f_t(\omega) \quad (11)$$

and $z(\omega)$ is an orthogonal process on $(-\pi, \pi)$ with $E|dz(\omega)|^2 = d\omega$. Since both $\psi(\omega)$ and $A_t(\omega)$ have one-sided Fourier transforms, it follows that, for each t , $\alpha_t(\omega)$ has a one-sided Fourier transform, i.e. one may write

$$\alpha_t(\omega) = \sum_{u=0}^{\infty} e^{-i\omega u} h_t(u) \quad (12)$$

Note that a necessary condition for the validity of (12) is

$$\int_{-\pi}^{\pi} \log f_t(\omega) d\omega > -\infty \quad (\text{C}_4)$$

for each t . In fact, it is readily seen that (C₁) and (C₃) \leftrightarrow (C₄), but whereas in the stationary case (C₄) is both necessary and sufficient to ensure the validity of (10) and (12), sufficiency has not been proved in the non-stationary case. Condition (C₄) implies that there exist functions $\{\alpha_t^*(\omega)\}$, say, satisfying (12), and such that, for each t , $f_t(\omega) = |\alpha_t^*(\omega)|^2$, so that

$A_t(\omega) = e^{i\lambda(t,\omega)} \alpha_t^*(\omega) / \psi^*(\omega)$, for some function $\lambda(t, \omega)$. In the stationary case, $\exp\{i\lambda(t, \omega)\}$, being independent of t , may be incorporated with $dz(\omega)$, and sufficiency follows. In the non-stationary case it would appear that further conditions are required. (It is interesting to note, however, that Cramer's result [7, Theorem 6] for the prediction of harmonizable processes has the same feature. That is, Cramer has proved only that when the spectrum of a harmonizable process satisfies a condition of the type (C_4) , the process must be "deterministic", i.e. that in such cases a one-sided moving-average representation is not admissible.)

Now define the process $\{\xi(t)\}$ by

$$\xi(t) = \int_{-\pi}^{\pi} e^{it\omega} dz(\omega) \quad (t = 0, \pm 1, \pm 2, \dots) \quad (13)$$

so that $\{\xi(t)\}$ is a stationary uncorrelated process, with

$$\left. \begin{aligned} E\{\xi(t)\} &= 0 && \text{for all } t, \\ E\{|\xi(t)|^2\} &= 2\pi && \text{for all } t, \\ E\{\xi(t)\xi^*(s)\} &= 0 && (t \neq s). \end{aligned} \right\} \quad (14)$$

It now follows from (10) and (12) that $X(t)$ may be written

$$X(t) = \sum_{u=0}^{\infty} h_t(u) \xi(t-u) \quad (15)$$

the above expression existing as a mean-square limit in virtue of the condition

$$2\pi \sum_{u=0}^{\infty} h_t^2(u) = \int_{-\pi}^{\pi} |\alpha_t(\omega)|^2 d\omega = \int_{-\pi}^{\pi} f_t(\omega) d\omega = \text{var}\{X(t)\} < \infty \quad (16)$$

Theorem 3.1. *Let $\{X(t)\}$ be a discrete-parameter oscillatory process. If there exists a family \mathcal{F} satisfying condition ((C2)), and with respect to which $\{X(t)\}$ has an absolutely continuous evolutionary spectrum satisfying condition (C_4) , then $\{X(t)\}$ may be represented as a one-sided moving average process of the form (15). Conversely, if $\{X(t)\}$ has a one-sided moving average representation of the form (15), condition (C_4) must be satisfied.*

Proof. The proof follows from the construction above and the conditions ensuring the validity of the Fourier representations. \square

3.2 Continuous-parameter Processes

As in the case of stationary processes (see [1]), the results for discrete parameter processes can readily be adapted to the continuous case. The measure μ is again assumed to be absolutely continuous (with respect to Lebesgue measure) and, in place of (C_1) , it is assumed that (using the same notation as in (6))

$$\int_{-\infty}^{\infty} \frac{\log f(\omega)}{1 + \omega^2} d\omega > -\infty \quad (C_1^*)$$

Then there exists a function $\psi(\omega)$ such that $|\psi(\omega)|^2 = f(\omega)$, with $\psi(z)$ having no poles or zeros in the lower half-plane. The function $\psi(\omega)$ may now be written as a one-sided Fourier integral. Corresponding to ((C2)) it is assumed now that, for each t , $A_t(\omega)$ may be written in the form

$$A_t(\omega) = \int_0^\infty e^{-i\omega u} g_t(u) du \quad (17)$$

a necessary condition being

$$\int_{-\infty}^\infty \frac{\log |A_t(\omega)|^2}{1 + \omega^2} d\omega > -\infty \quad (C_3^*)$$

for all t . It then follows that $X(t)$ may be written in the form

$$X(t) = \int_{-\infty}^\infty e^{it\omega} \alpha_t(\omega) dz(\omega) \quad (18)$$

where $z(\omega)$ is an orthogonal process on $(-\infty, \infty)$, with $E[|dz(\omega)|^2] = d\omega$, and $\alpha_t(\omega)$ has a one-sided Fourier integral representation of the form

$$\alpha_t(\omega) = \int_0^\infty e^{-i\omega u} h_t(u) du \quad (19)$$

with $|\alpha_t(\omega)|^2 = f_t(\omega)$. Corresponding to (C4), a necessary condition for the validity of (19) is

$$\int_{-\infty}^\infty \frac{\log f_t(\omega)}{1 + \omega^2} d\omega > -\infty \quad (C_4^*)$$

for all t . In the continuous case there are analytic difficulties in constructing an analogue of the uncorrelated process $\{\xi(t)\}$ (see [1], p. 21). However, one may formally define the process $\xi(t)$ by

$$\xi(t) = \int_{-\infty}^\infty e^{i\omega t} dz(\omega) \quad (20)$$

(so that $\xi(t)$ may be interpreted as the “derivative” of a Wiener process), so that

$$\left. \begin{aligned} E\{\xi(t)\} &= 0 && \text{for all } t, \\ E\{|d\xi(t)|^2\} &= 2\pi dt, \\ E\{d\xi(t)d\xi^*(s)\} &= 0 && (s \neq t). \end{aligned} \right\} \quad (21)$$

Then formally one may write $X(t)$ in the form

$$X(t) = \int_0^\infty h_t(u) \xi(t-u) du \quad (22)$$

or, more precisely,

$$X(t) = \int_0^\infty h_t(u) d\xi(t-u) \quad (23)$$

As in case 3.1, $\text{var}\{X(t)\} < \infty$ implies

$$\int_0^\infty h_t^2(u) du < \infty \quad (24)$$

for all t .

4 The Time-Domain Approach

The basic problem of linear least-squares prediction may be stated as follows: the observed values of the process over the semi-infinite interval $(-\infty, t)$ are given and the value of $X(t+m)$ ($m > 0$) is to be predicted. The predictor $\tilde{X}(t+m)$ is to be chosen as that linear combination of $\{X(s), s \leq t\}$ which is such that

$$M(m) = E\{\tilde{X}(t+m) - X(t+m)\}^2 \quad (25)$$

is minimized.

The existence and uniqueness of the optimum predictor $\tilde{X}(t+m)$ can be established by the Hilbert space approach, using exactly the same arguments as employed in the stationary case (see [6], [4]). (Briefly, one represents the process as a curve in a Hilbert space, and shows that $\tilde{X}(t+m)$ is the “projection” of $X(t+m)$ on the smallest subspace containing $\{X(s), s \leq t\}$.) However, assuming that the process has a one-sided moving average form, an explicit expression for $\tilde{X}(t+m)$ may easily be obtained in terms of the process $\{\xi(t)\}$. The continuous and discrete cases are treated by essentially the same argument, but the details differ and the discrete case is considered first.

4.1 Discrete-parameter Processes

It is assumed that $X(t)$ has a one-sided moving average representation of the form (15), so that one may write

$$X(t) = \sum_{u=0}^{\infty} h_t(u) \xi(t-u) = \sum_{u=-\infty}^t h_t(t-u) \xi(u) \quad (26)$$

Then

$$X(t+m) = \sum_{u=-\infty}^t h_{t+m}(t+m-u) \xi(u) + \sum_{u=t+1}^{t+m} h_{t+m}(t+m-u) \xi(u) \quad (27)$$

Now the predictor $\tilde{X}(t+m)$ is to be chosen as a linear combination of $\{X(s), s \leq t\}$ of the form

$$\tilde{X}(t+m) = \sum_{s=-\infty}^t b(s) X(s) \quad (28)$$

However, as each $X(t)$ is a linear combination of the $\{\xi(t-u)\}$ ($u \geq 0$), one may equally well express $\tilde{X}(t+m)$ as a linear combination of $\{\xi(s), s \leq t\}$ of the form

$$\tilde{X}(t+m) = \sum_{u=-\infty}^t a(u) \xi(u) \quad (29)$$

The problem now reduces to finding the values of the coefficients $\{a(u)\}$ which minimize $M(m)$, and, exactly as in the stationary case, the solution follows immediately from (27) (cf. [1], p. 32; [2], p. 199). For, in view of the orthogonality of the ξ 's it follows (either by direct

substitution of (29) and (27) in (25), or by noting that the second term in (27) is orthogonal to the first term) that

$$\tilde{X}(t+m) = \sum_{u=-\infty}^t h_{t+m}(t+m-u) \xi(u) = \sum_{u=m}^{\infty} h_{t+m}(u) \xi(t+m-u) \quad (30)$$

so that, in terms of (29),

$$a(u) = h_{t+m}(t+m-u) \quad (31)$$

for all u . The “prediction variance”, $M(m)$, is given by

$$M(m) = 2\pi \sum_{u=0}^{m-1} h_{t+m}^2(u) \quad (32)$$

In practice, only the values of $X(s)$ ($s \leq t$) are observed, so that in order to compute $\tilde{X}(t+m)$ the expression (30) must be expressed in terms of the X ’s rather than the ξ ’s. Now write

$$\xi(t) = \sum_{v=0}^{\infty} k_t(v) X(t-v) \quad (33)$$

Then, substituting (33) in (26) and equating coefficients of $\xi(t)$, one obtains

$$\sum_{v=0}^p h_{t-v}(p-v) k_t(v) = \delta_{p,0} \quad (p = 0, 1, 2, \dots) \quad (34)$$

for all t . Thus, given the sequence $\{h_t(u)\}$, for all t , one may, in principle, solve (34) for $\{k_t(v)\}$. When $X(t)$ is stationary, $h_t(u)$ and $k_t(v)$ are both independent of t and (34) is easily solved by introducing the generating functions of the two sequences [1?]. In general, there does not appear to be any systematic relationship between corresponding terms of the two sequences, and the most practical approach would seem to be the method of repeated back substitution. That is, if (34) is written in detail one has

$$\begin{aligned} p = 0 : \quad & h_t(0)k_t(0) = 1 \quad \text{for all } t, \\ p = 1 : \quad & h_t(1)k_t(0) + h_{t-1}(0)k_t(1) = 0 \quad \text{for all } t, \\ p = 2 : \quad & h_t(2)k_t(0) + h_{t-1}(1)k_t(1) + h_{t-2}(2)k_t(2) = 0 \quad \text{for all } t, \quad \text{etc.} \end{aligned} \quad (35)$$

This system of equations is easily solved, step by step, as it is triangular. Thus one finds, for all t ,

$$\begin{aligned} k_t(0) &= 1/h_t(0), \\ k_t(1) &= -h_t(1)/h_t(0)h_{t-1}(0), \\ k_t(2) &= \{-h_t(2)/h_t(0)h_{t-2}(2)\} + \{h_t(1)h_{t-1}(1)/h_t(0)h_{t-1}(0)h_{t-2}(0)\}, \\ &\dots \quad \text{etc.} \end{aligned} \quad (36)$$

Alternatively, one may substitute (26) in (33) and equate coefficients, obtaining

$$\sum_{v=0}^p k_{t-v}(p-v) h_t(v) = \delta_{p,0} \quad (p = 0, 1, 2, \dots) \quad (37)$$

for all t . [5] has pointed out that the two systems of equations (34) and (37) are not identical unless the process is stationary (in which case both $\{h_t(u)\}$ and $\{k_t(v)\}$ are independent of t), but that one set of equations still implies the other. However, the triangular form of both (34) and (37) means that these systems of equations are ideally suited for numerical solution.

4.2 Continuous-parameter Processes

The results for continuous-parameter processes are very similar to those obtained in the discrete case. Thus one is given $\{X(s), -\infty < s \leq t\}$ and wishes to predict $X(t+m)$ ($m > 0$) by a linear combination of past values, of the form

$$\tilde{X}(t+m) = \int_{-\infty}^t X(s)b(s) ds \quad (38)$$

or, equivalently,

$$\tilde{X}(t+m) = \int_{-\infty}^t \xi(u)a(u) du \quad (39)$$

Assuming that $X(t)$ has a moving-average representation of the form (23), it follows immediately that

$$\tilde{X}(t+m) = \int_{-\infty}^t h_{t+m}(t+m-u) \xi(u) du = \int_m^\infty h_{t+m}(u) \xi(t+m-u) du \quad (40)$$

In order to express $\tilde{X}(t+m)$ in terms of $\{X(s), s \leq t\}$ one writes

$$\xi(t) = \int_0^\infty k_t(v)X(t-v) dv \quad (41)$$

and substitutes (41) in (23). One then obtains

$$\int_0^v h_t(u)k_{t-u}(v-u) du = \delta(v) \quad (42)$$

for all $v > 0$ and all t , where $\delta(v)$ denotes the Dirac δ -function. Similarly, substituting (23) in (41), one finds

$$\int_0^u k_t(v)h_{t-v}(u-v) dv = \delta(u) \quad (43)$$

for all $u > 0$ and all t . Given the functions $\{h_t(u)\}$ (for all t), the functions $\{k_t(v)\}$ may, in principle, be determined from either of the above integral equations.

In certain cases the time-domain approach may be used to obtain an expression of the form (38) or (39) for $\tilde{X}(t+m)$, i.e. an expression in terms of past values of $X(t)$, without solving either of the systems (42) or (43) directly. Such cases include uniformly modulated processes, and, naturally enough, finite order autoregressive models (with time-dependent coefficients). Finite order moving-average models (with time-dependent coefficients) can be dealt with also by this method, but they are more suited to a frequency-domain approach (see Section 5). Some specific examples are now discussed.

Example 1: Uniformly modulated process. Consider a continuous-parameter process of the form

$$X(t) = C(t) Y(t) \quad (44)$$

where $Y(t)$ is a stationary process with spectral density function $f_y(\omega)$, say, and $C(t)$ is some given function of t . Then, with respect to the family $\mathcal{F} \equiv \{e^{i\omega t} C(t)\}$, the evolutionary spectrum of $\{X(t)\}$ is given by [1]

$$f_t(\omega) = |C(t)|^2 f_y(\omega) \quad (45)$$

Assume that $f_t(\omega)$ satisfies condition (C_4^*) . Then it follows that

$$\int_{-\infty}^{\infty} \frac{\log f_y(\omega)}{1 + \omega^2} d\omega > -\infty \quad (46)$$

so that $Y(t)$ has a moving-average representation of the form

$$Y(t) = \int_0^{\infty} g(u) \xi(t-u) du \quad (47)$$

Hence one may write $X(t)$ in the form (23), i.e.

$$X(t) = \int_0^{\infty} h_t(u) \xi(t-u) du \quad (48)$$

where $h_t(u) = C(t)g(u)$. According to (40), the linear least-squares predictor of $X(t+m)$ is given by

$$\begin{aligned} \tilde{X}(t+m) &= \int_m^{\infty} C(t+m)g(u) \xi(t+m-u) du \\ &= C(t+m) \tilde{Y}(t+m) \end{aligned} \quad (49)$$

where $\tilde{Y}(t+m)$ is the least-square predictor of $Y(t+m)$, given $\{Y(s), s \leq t\}$, and may be determined in the usual way. The referee has pointed out that this result has a more general form, namely: if

$$X(t) = \sum_{u=0}^{\infty} a_t(u) Y(t-u), \quad \text{then} \quad \tilde{X}(t+m) = \sum_{u=0}^{\infty} a_{t+m}(u) Y^*(t+m-u) \quad (50)$$

where

$$Y^*(s) = \begin{cases} Y(s) & (s \leq t), \\ \tilde{Y}(s) & (s > t). \end{cases} \quad (51)$$

For example, using the Wiener predictor (see [2], p. 201),

$$\tilde{Y}(t+m) = \int_0^{\infty} Y(t-u) b(u) du \quad (52)$$

where

$$b(u) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{i\omega u} B(\omega) d\omega \quad (53)$$

and

$$B(\omega) = \frac{e^{i\omega m} \int_m^\infty g(u)e^{-i\omega u} du}{\int_0^\infty g(u)e^{-i\omega u} du} \quad (54)$$

Thus

$$\tilde{X}(t+m) = \int_0^\infty X(t-u)b^*(u) du \quad (55)$$

where

$$b^*(u) = \{C(t+m)/C(t-u)\}\{(2\pi)^{-1} \int_{-\infty}^\infty e^{i\omega u} B(\omega) d\omega\} \quad (56)$$

Example 2: First-order autoregressive process, discrete-parameter. Suppose $\{X(t)\}$ is given by

$$X(t) - \alpha(t)X(t-1) = \xi(t) \quad (t = 0, \pm 1, \pm 2, \dots) \quad (57)$$

where $\{\alpha(t)\}$ is a given sequence with $|\alpha(t)| < 1$ for all t , and $\{\xi(t)\}$ is an uncorrelated stationary process with $E\{\xi(t)\} = 0$, $E\{\xi^2(t)\} = 1$. From (57) one finds (subject to $X(-\infty) = 0$)

$$X(t) = \sum_{u=0}^{\infty} h_t(u) \xi(t-u) \quad (58)$$

where

$$h_t(u) = \begin{cases} \alpha(t)\alpha(t-1)\cdots\alpha(t-u+1) & (u > 0), \\ 1 & (u = 0). \end{cases} \quad (59)$$

Hence, according to (30), the linear least-squares predictor of $X(t+m)$ is given by

$$\begin{aligned} \tilde{X}(t+m) &= \sum_{u=m}^{\infty} \alpha(t+m)\alpha(t+m-1)\cdots\alpha(t+m-u+1) \xi(t+m-u) \\ &= \{\alpha(t+m)\cdots\alpha(t+1)\} \sum_{u=0}^{\infty} \alpha(t)\cdots\alpha(t-u+1) \xi(t-u) \\ &= \{\alpha(t+m)\cdots\alpha(t+1)\} X(t) \end{aligned} \quad (60)$$

(Note that this is the exact expression for $X(t+m)$ when $\xi(s) = 0$ for all $s > t$.)

Example 3: First-order moving-average process, discrete-parameter. Let $\{X(t)\}$ be given by

$$X(t) = \xi(t) - \alpha(t)\xi(t-1) \quad (t = 0, \pm 1, \pm 2, \dots) \quad (61)$$

with $\{\xi(t)\}$ as above. Then clearly

$$\tilde{X}(t+m) = \begin{cases} -\alpha(t+1)\xi(t) & (m = 1), \\ 0 & (m > 1), \end{cases} \quad (62)$$

where

$$\xi(t) = \sum_{v=0}^{\infty} k_t(v) X(t-v) \quad (63)$$

with $k_t(v) = \alpha(t)\alpha(t-1)\cdots\alpha(t-v+1)$.

5 The Frequency-Domain Approach

An alternative approach based on a generalization of the Wiener-Hopf technique for dealing with the prediction of stationary processes is now considered. The continuous-parameter case is treated first, and the predictor of $X(t + m)$ is written in the form

$$\tilde{X}(t + m) = \int_0^\infty b_t(u) X(t - u) du \quad (64)$$

Note that (64) is merely a formal way of writing

$$\tilde{X}(t + m) = \int_0^\infty X(t - u) d\beta_t(u) \quad (65)$$

The function $b_t(u)$ need not exist in the strict sense, as would be the case if, for example, $\tilde{X}(t + m)$ involved derivatives of $X(t - u)$. Note also that the weight-function $b_t(u)$ may depend on both t and m . Now using the evolutionary spectral representation of $X(t)$, (18) and (64) may be written

$$\tilde{X}(t + m) = \int_{-\infty}^\infty e^{i\omega t} B_t(\omega) dz(\omega) \quad (66)$$

where

$$B_t(\omega) = \int_0^\infty b_t(u) \alpha_{t-u}(\omega) e^{-iu\omega} du \quad (67)$$

On the other hand, from (18),

$$X(t + m) = \int_{-\infty}^\infty e^{i\omega(t+m)} \alpha_{t+m}(\omega) dz(\omega) \quad (68)$$

Hence, in virtue of the orthogonality of $z(\omega)$,

$$M(m) = E|\tilde{X}(t + m) - X(t + m)|^2 = \int_{-\infty}^\infty |e^{i\omega m} \alpha_{t+m}(\omega) - B_t(\omega)|^2 d\omega \quad (69)$$

One must now choose $b_t(u)$ (or, equivalently, $B_t(\omega)$) so as to minimize (69). As a preliminary step it is first shown that, for each t , the function $B_t(\omega)$ is a “backward transform”, that is it may be written as a one-sided Fourier transform of the form

$$B_t(\omega) = (2\pi)^{-1} \int_{-\infty}^\infty e^{-iv\omega} K_t(v) dv \quad (70)$$

where $K_t(v) = 0$ for $v < 0$. For

$$\begin{aligned} K_t(v) &= \int_{-\infty}^\infty e^{i\omega v} B_t(\omega) d\omega \\ &= \int_0^\infty b_t(u) \left\{ \int_0^\infty e^{i\omega(v-u)} \alpha_{t-u}(\omega) d\omega \right\} du \end{aligned} \quad (71)$$

Since

$$\alpha_{t-u}(\omega) = \int_0^\infty e^{-i\omega w} h_{t-u}(w) dw \quad (72)$$

one has

$$\int_{-\infty}^{\infty} e^{i\omega(v-u)} \alpha_{t-u}(\omega) d\omega = \begin{cases} h_{t-u}(v-u) & (u \leq v), \\ 0 & (u > v). \end{cases} \quad (73)$$

Hence

$$K_t(v) = \int_0^v b_t(u) h_{t-u}(v-u) du \quad (74)$$

and, in particular, $K_t(v) = 0$ for $v < 0$.

Now write (for each t)

$$e^{i\omega m} \alpha_{t+m}(\omega) = C_t^{(1)}(\omega) + C_t^{(2)}(\omega) \quad (75)$$

where, for each t , $C_t^{(1)}(\omega)$ is a “backward transform”, and $C_t^{(2)}(\omega)$ is a “forward transform” [2?]. That is, write

$$C_t^{(1)}(\omega) = (2\pi)^{-1} \int_0^{\infty} e^{-i\omega u} \left[\int_{-\infty}^{\infty} e^{i\theta u} \{ e^{i\theta m} \alpha_{t+m}(\theta) \} d\theta \right] du \quad (76)$$

and

$$C_t^{(2)}(\omega) = (2\pi)^{-1} \int_{-\infty}^0 e^{-i\omega u} \left[\int_{-\infty}^{\infty} e^{i\theta u} \{ e^{i\theta m} \alpha_{t+m}(\theta) \} d\theta \right] du \quad (77)$$

Then, for each t , $\{C_t^{(1)}(\omega) - B_t(\omega)\}$ and $C_t^{(2)}(\omega)$ are orthogonal with respect to $d\omega$ integration, and from (69) one has

$$M(m) = \int_{-\infty}^{\infty} |C_t^{(1)}(\omega) - B_t(\omega)|^2 d\omega + \int_{-\infty}^{\infty} |C_t^{(2)}(\omega)|^2 d\omega \quad (78)$$

It follows immediately that the minimum of $M(m)$ is attained when $B_t(\omega)$ is given by

$$B_t(\omega) = C_t^{(1)}(\omega) \quad (79)$$

for all t , and the prediction variance is then given by

$$M(m) = \int_{-\infty}^{\infty} |C_t^{(2)}(\omega)|^2 d\omega \quad (80)$$

It now remains to solve (79) to find the functions $\{b_t(u)\}$. Taking Fourier transforms of both sides one finds, for $v > 0$,

$$\begin{aligned} K_t(v) &= \int_{-\infty}^{\infty} e^{i\theta(v+m)} \alpha_{t+m}(\theta) d\theta \quad \text{for all } t, \\ &= h_{t+m}(v+m). \end{aligned} \quad (81)$$

Hence, using (74), one finds

$$\int_0^v b_t(u) h_{t-u}(v-u) du = h_{t+m}(v+m) \quad (82)$$

for all t and all $v > 0$. Given the form of the evolutionary spectral density function $f_t(\omega)$, for all t , one may (in principle) determine the function $h_t(u)$, for all t , and hence find the optimum weight-function $\{b_t(u)\}$ by solving the integral equation (82).

Equation (82), being an integral equation of the first kind, may be solved explicitly for $\{b_t(u)\}$. (Note that in the stationary case $f_t(\omega)$, and consequently $h_t(u)$, are both independent of t , so that (82) may then be solved immediately by noting that the left-hand side is the convolution of $b_t(u)$ and $h_t(u)$.) In the general case one may solve (82) by first expressing it as an integral equation of the second kind. Thus, if it is assumed that

$$h_{t-v}(0) \neq 0 \text{ for all } v \text{ in } (0, \infty) \text{ and } \frac{\partial h_{t-u}(v-u)}{\partial v} = h'_{t-u,v}(v-u) \text{ exists and is continuous in } u \text{ and } v \quad (u < v) \quad (83)$$

then by differentiating both sides of (82) with respect to v one finds

$$h_{t-v}(0)b_t(v) + \int_0^v h'_{t-u,v}(v-u)b_t(u) du = h'_{t+m,v}(v+m) \quad (v \geq 0) \quad (84)$$

or

$$b_t(v) + \int_0^v \frac{h'_{t-u,v}(v-u)}{h_{t-v}(0)} b_t(u) du = \frac{h'_{t+m,v}(v+m)}{h_{t-v}(0)} \quad (85)$$

The solution of (85) is [2?]

$$b_t(v) = \frac{h'_{t+m,v}(v+m)}{h_{t-v}(0)} + \int_0^v H_t^*(v, u, -1) \left\{ \frac{h'_{t+m,v}(u+m)}{h_{t-u}(0)} \right\} du \quad (86)$$

where, for each t , the “resolvent kernel” $H_t^*(v, u, -1)$ is given by

$$H_t^*(v, u, -1) = - \sum_{p=0}^{\infty} (-1)^p K_{t,p+1}^*(u, v) \quad (87)$$

and the “iterated kernels” $K_{t,p}^*(u, v)$ are given by

$$K_{t,p+1}^*(u, v) = \int_u^v K_{t,1}^*(u, z) K_{t,p}^*(z, v) dz \quad (p = 1, 2, \dots) \quad (88)$$

with

$$K_{t,1}^*(u, v) = \frac{h'_{t-u,v}(v-u)}{h_{t-v}(0)} \quad (89)$$

Equation (86) provides an analytic expression for $\{b_t(u)\}$, but in numerical work it may well be more convenient to solve the integral equation (82) directly.

For discrete-parameter processes the predictor $\tilde{X}(t + m)$ will be of the form

$$\tilde{X}(t + m) = \sum_{u=0}^{\infty} b_t(u) X(t - u) \quad (90)$$

so that, using the spectral representation (10),

$$M(m) = \int_{-\pi}^{\pi} |e^{i\omega m} \alpha_{t+m}(\omega) - B_t(\omega)|^2 d\omega \quad (91)$$

where now

$$B_t(\omega) = \sum_{u=0}^{\infty} b_t(u) \alpha_{t-u}(\omega) e^{-iu\omega} \quad (92)$$

Following a similar argument to that used above, one may show that $M(m)$ is minimized when the sequence $\{b_t(u)\}$ is given by

$$\sum_{u=0}^v b_t(u) h_{t-u}(v-u) = h_{t+m}(v+m) \quad (v = 0, 1, 2, \dots; t = 0, \pm 1, \pm 2, \dots) \quad (93)$$

Here, $h_t(u)$ ($u = 0, 1, 2, \dots$) is the sequence defined by (12).

The system of equations (93) is again “triangular” (cf. (35)) so that it is easily solved for $\{b_t(u)\}$ by repeated back-substitution. In fact, this approach has been tried on a numerical example (using an ATLAS computer) and has proved quite successful.

It is interesting to note that had one attempted to obtain $\{b_t(u)\}$ directly from the covariance function $R(s, t) \equiv E\{X(s)X^*(t)\}$, one would have obtained the well-known set of equations (see [1], p. 149),

$$\sum_{u=0}^{\infty} b_t(u) R(t-u, t-v) = R(t-v, t+m) \quad (v \geq 0) \quad (94)$$

(In the continuous case one obtains the corresponding integral equation.)

The above system does not lend itself to either an analytic or a numerical solution and the spectral approach (cf. (82), (93)) may be regarded as a method of reducing (94) to “triangular” form.

The methods of this section are now applied to some examples discussed in Section 4.2.

Example 4. Consider the first-order autoregressive process (discrete-parameter) of (57). Using the form of $\{h_t(u)\}$ given by (58) in (93), one obtains

$$b_t(0) + \sum_{u=1}^v b_t(u) \alpha(t-u) \alpha(t-u-1) \cdots \alpha(t-v+1) = \alpha(t+m) \alpha(t+m-1) \cdots \alpha(t-v+1) \quad (v = 0, 1, 2, \dots) \quad (95)$$

Solving this system step by step, one obtains

$$\begin{cases} b_t(0) = \alpha(t+m) \alpha(t+m-1) \cdots \alpha(t+1), \\ b_t(u) = 0 \quad (u > 0). \end{cases} \quad (96)$$

Hence

$$\tilde{X}(t+m) = \alpha(t+m) \alpha(t+m-1) \cdots \alpha(t+1) X(t) \quad (97)$$

in agreement with Example 2.

Example 5. Consider now the moving-average process (discrete-parameter) of (61). For this process,

$$h_t(u) = \begin{cases} 1 & (u = 0), \\ -\alpha(t) & (u = 1), \\ 0 & (u > 1), \end{cases} \quad (98)$$

for all t . Substituting this expression in (93), one finds

$$b_t(0) = \begin{cases} -\alpha(t+1) & (m = 1), \\ 0 & (m > 1), \end{cases} \quad (99)$$

and

$$b_t(u)h_{t-u}(1) + b_t(u+1)h_{t-u}(0) = 0 \quad (u = 1, 2, \dots) \quad (100)$$

for all m . The solution of this set of equations is, for all u ,

$$b_t(u) = \begin{cases} -\alpha(t+1)\alpha(t)\cdots\alpha(t-u+1) & (m = 1), \\ 0 & (m > 1), \end{cases} \quad (101)$$

so that

$$\tilde{X}(t+m) = \begin{cases} -\sum_{u=0}^{\infty} \alpha(t+1)\alpha(t)\cdots\alpha(t-u+1)X(t) & (m = 1), \\ 0 & (m > 1), \end{cases} \quad (102)$$

in agreement with Example 3.

6 Discussion

As mentioned in Section 1, this paper is concerned primarily with the problem of determining the optimum predictor on the assumption that the second-order properties of the process are fully known. However, it has been shown that the predictor $\tilde{X}(t+m)$ is determined uniquely by the form of the evolutionary spectra $\{f_t(\omega)\}$, so that in cases where there is no precise knowledge of the underlying model the problem may be attacked by first estimating the functions $\{f_t(\omega)\}$. It will be seen from Section 5 that $\tilde{X}(t+m)$ depends on the form of $h_{t+m}(u)$, i.e. on the form of the evolutionary spectrum at the future time point $(t+m)$. In a purely empirical analysis, one would not, of course, be able to estimate $f_{t+m}(\omega)$ directly from observation extending only up to time t , and one would have to assume that the spectra were changing with sufficient “smoothness” over time to enable the form of $f_{t+m}(\omega)$ to be inferred from the spectra which have been estimated up to time t . If the prediction step m is much smaller than the characteristic width B_X of the process (see [1]), it would be a reasonable approximation to replace $f_{t+m}(\omega)$ by $f_t(\omega)$ —the most recently available spectrum. In fact, it seems clear intuitively that the very nature of a non-stationary process precludes long-range prediction, unless one is prepared to make assumptions about the character of the non-stationarity. Consequently, if $m > B_X$ one would have to extrapolate the available spectra to the time point $(t+m)$. This extrapolation could be performed on the spectral ordinates themselves (by fitting suitable functions of t to a range of values of ω), but in

general one would presumably start from a model of the process which determines the form of the evolutionary spectra in terms of a set of time-varying parameters $\{\alpha_i(t)\}$, say, and then extrapolate the values of these parameters. (Thus, for the model of Example 2, the $\{f_t(\omega)\}$ may be expressed in terms of the time-varying parameter $\alpha(t)$.) If the model specifies further the functional time-dependence of the $\{\alpha_i(t)\}$ in terms of another set of constant parameters, $\{\beta_j\}$, say, then the extrapolation is a straightforward problem and involves merely the estimation of the $\{\beta_j\}$. (For example, in some cases one may postulate that the $\{\alpha_i(t)\}$ are periodic functions of t with known periodicities.) If, on the other hand, the functional form of the $\{\alpha_i(t)\}$ is not specified by the model, the extrapolation would have to be performed by regression techniques.

References

- [1] Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.
- [2] Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge: University Press.
- [3] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- [4] Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*. Englewood Cliffs, N.J.: Prentice-Hall.
- [5] Whittle, P. (1965). Recursive relations for predictors of non-stationary processes. *J.R. Statist. Soc. B*, **27**, 523–532.
- [6] Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.*, **32**, 951–989.
- [7] Cramér, H. (1961a). On some classes of non-stationary processes. *Proc. 4th Berkeley Symp. Math. Statist. & Prob.*, **2**, 57–78.
- [8] Cramér, H. (1961b). On the structure of purely non-deterministic stochastic processes. *Ark. Mat.*, **19**, 249–266.
- [9] Davis, R. C. (1952). On the theory of prediction of non-stationary stochastic processes. *J. Appl. Phys.*, **23**, 1047–1053.
- [10] Booton, R. C. (1952). An optimization theory for time-varying linear systems with nonstationary statistical inputs. *Proc. Inst. Radio Engrs*, **40**, 977–981.
- [11] Zadeh, L. A. (1953). Optimum non-linear filters. *J. Appl. Phys.*, **24**.
- [12] Bendat, J. S. (1956). A general theory of linear prediction and filtering. *J. Soc. Industr. Appl. Math.*, **4**, 131–151.
- [13] Bendat, J. S. (1957). Exact integral equation solutions and synthesis for a large class of optimum time variable linear filters. *Trans. Inst. Radio Engrs*, IT-3, 71–80.

- [14] Bendat, J. S. (1959). *Principles and Applications of Random Noise Theory*. New York: Wiley.
- [15] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng. (Trans. A.S.M.E., Series D)*, **82**, 35–45.
- [16] Granger, C. W. J. and Hatanaka, M. (1964). *Spectral Analysis of Economic Time Series*. Princeton: University Press.
- [17] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J.R. Statist. Soc. B*, **27**, 204–237.
- [18] Priestley, M. B. (1966). Design relations for non-stationary processes. *J.R. Statist. Soc. B*, **28**, 228–240.
- [19] Whittle, P. (1963). *Prediction and Regulation*. London: English Universities Press.
- [20] Tricomi, F. G. (1957). *Integral Equations*. New York: Wiley.