DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT No. 129

August 1967

AN INTRODUCTION TO LINEAR MODELING OF STATIONARY AND NON-STATIONARY TIME SERIES

By

Donald G. Watts

This research was supported in part by the United States Navy through the Office of Naval Research under contract NONR-1202(17), Project NR-042-222.

AN INTRODUCTION TO LINEAR MODELING OF STATIONARY AND NON-STATIONARY TIME SERIES

Donald G. Watts
Associate Professor
Department of Statistics
University of Wisconsin

Summary

Recent work by G. E. P. Box, of the University of Wisconsin, and G. M. Jenkins, of the University of Lancaster has led to simple models for stationary and non-stationary time series. The approach used is essentially that proposed by Zadeh and Ragazzini in 1950 [1] and 1952 [2], but Box and Jenkins have looked at the problem from the point of view of the statistician. Thus, while Zadeh and Ragazzini were concerned with analytic reduction of random processes to white noise, Box and Jenkins are concerned with empirical reduction of time series to uncorrelated residuals [3,4,5,6]. Another important aspect of Box and Jenkins' work is that of developing useful models for seasonal time series. This paper presents an introduction to parametric modeling of time series, using the Box-Jenkins approach [3,4,5,6].

1. Introduction

1.1) Approaches to time series analysis

Figures 1.1, 1.2, and 1.3 show three time series which have arisen in various fields. How would you characterize these series?

The answers given by people confronted with data like that shown in the figures fit into three main categories. First, how much "memory" exists in the waveform — that is, does the value of the series at time t depend very strongly on previous values, say at time $t-\tau$? This is basically the correlation analysis approach. The second mode of investigation is to analyze the "frequency content" of the waveform, the power spectrum approach. Is the time series composed predominantly of high or low frequencies, or is their a tendency to oscillate? The third approach is to fit a random process model to the series. The model, if properly chosen and fitted, can then be used to forecast future values, so that this third method may be somewhat more powerful than the previous ones for some uses. All the approaches are complementary, however, because they all contribute information about the time series, and because they can be used as mutual corroborators.

1.2) Mathematics of time series analysis

A time series analyst is faced with a formidable problem, and as a consequence, must invoke equally formidable mathematics. In particular, he must have considerable facility in likelihood and least squares estimation theory, and in operational and transform calculus. These techniques will be introduced and developed at appropriate points in the text.

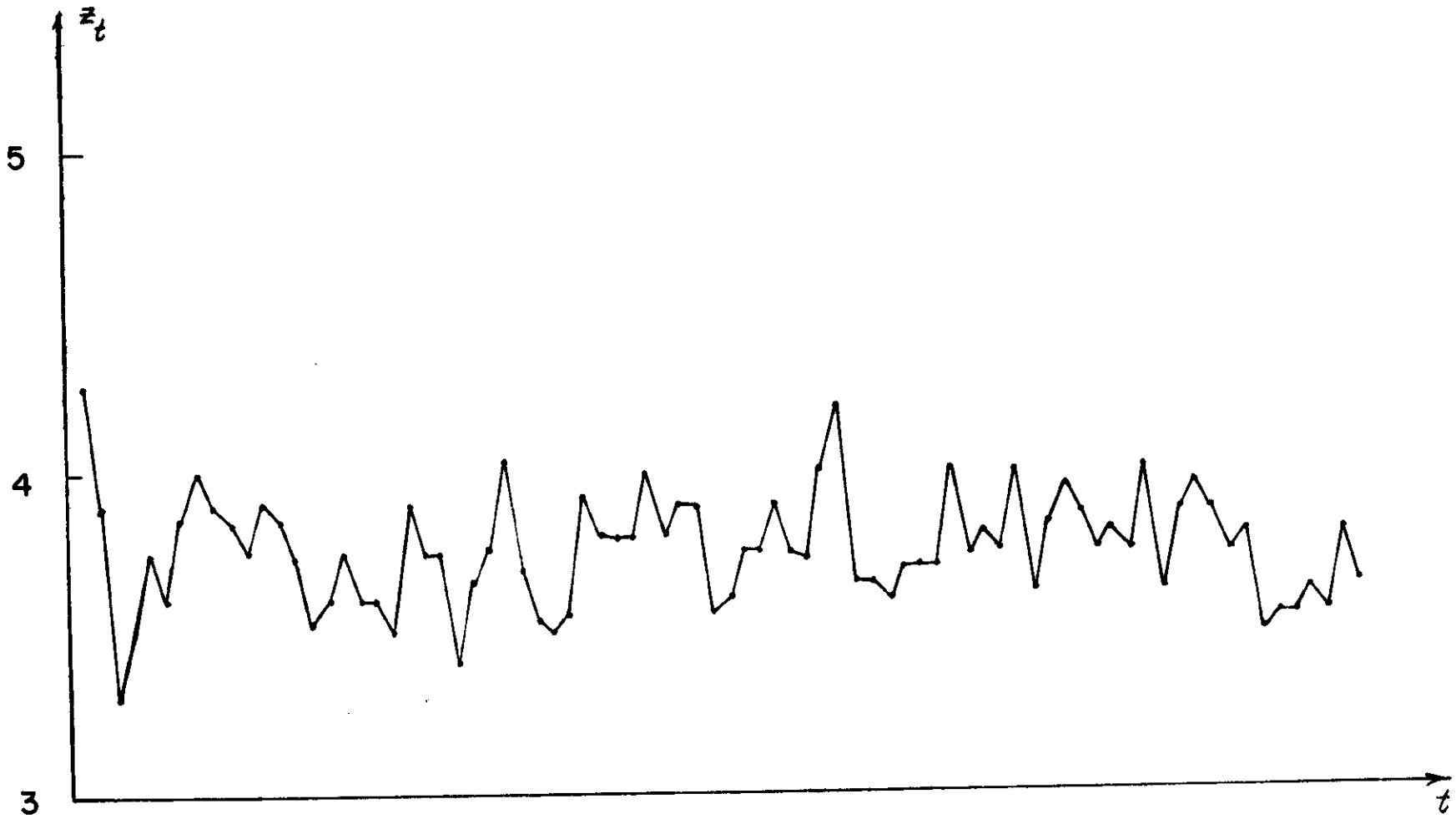

FIGURE 1.1 HOURLY READINGS OF COMPOSITION OF MELT FROM A HEARTH FURNACE.

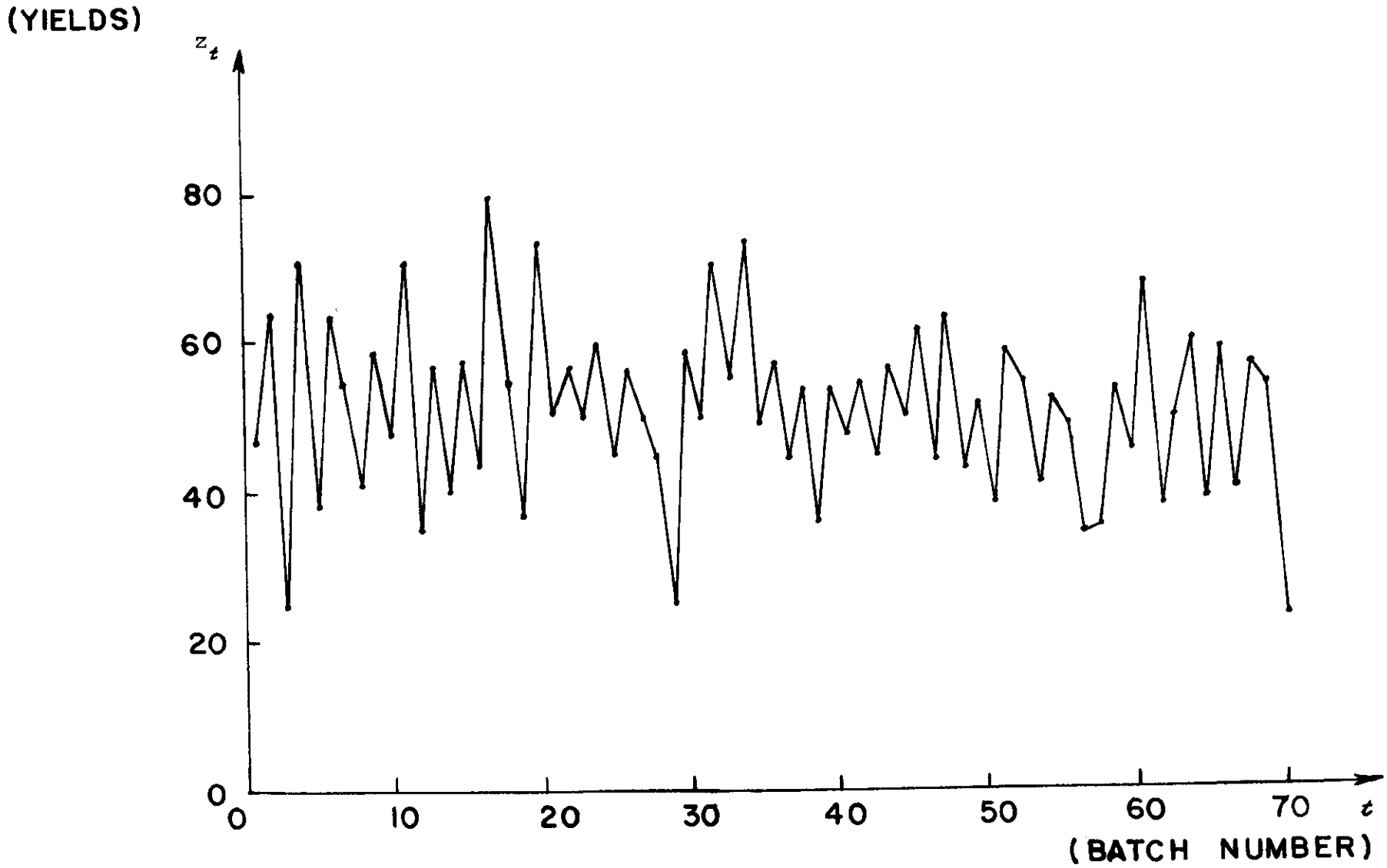

FIGURE 1·2 YIELDS FROM 70 CONSECUTIVE BATCHES OF AN INDUSTRIAL PROCESS

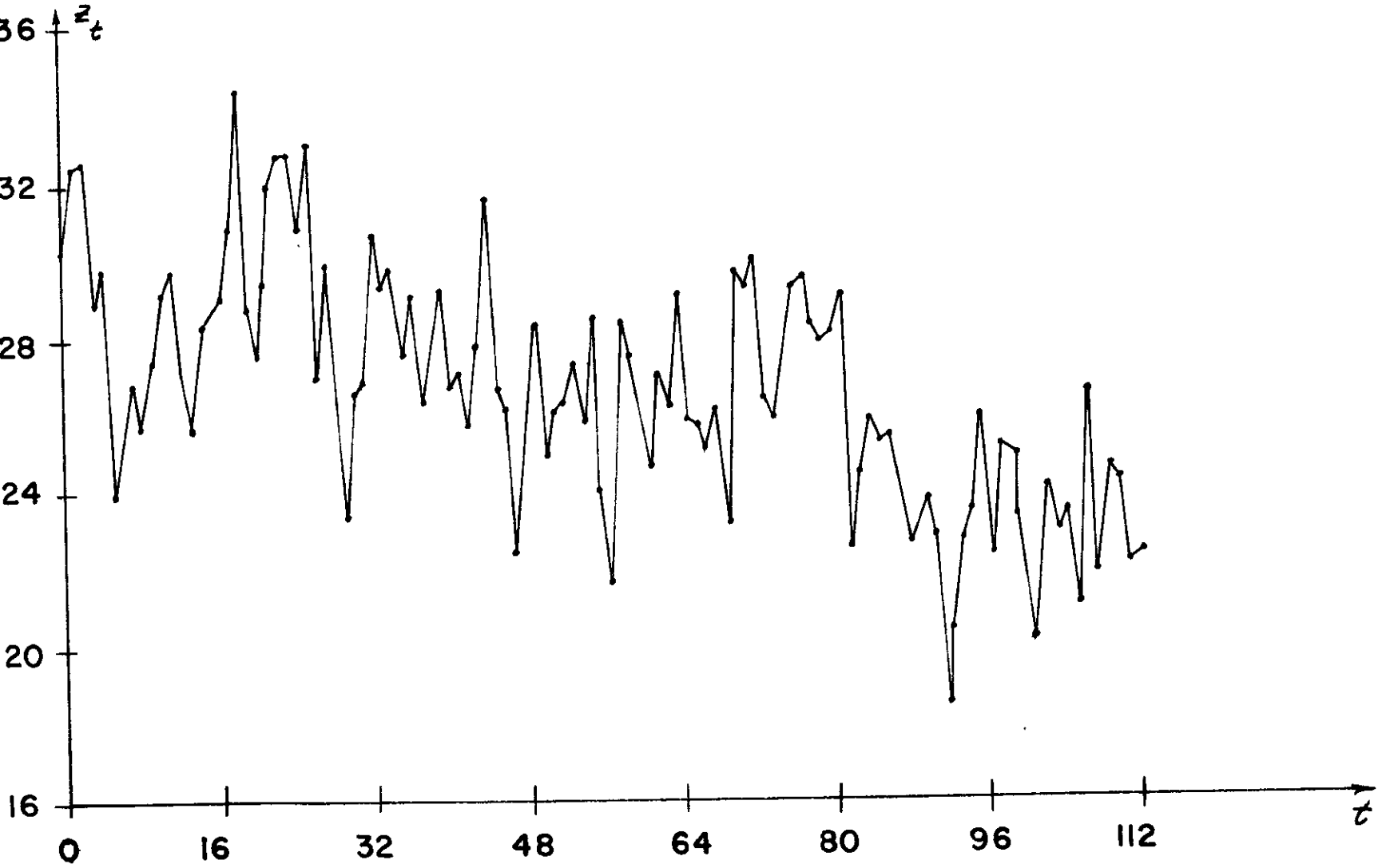

FIGURE 1.3 DAILY TONNAGE FROM MOLDING UNITS AT A FOUNDRY

2. Probability models for time series

The statistician's approach in describing time series such as those shown in Figures 1.1 to 1.3 is the same one he uses to describe any non-deterministic phenomena — he invokes a probability model. The big difference in the time series case, however, is that independence of observations is not assumed — in fact, it is the dependence between observations which is of interest! The particular probability model used depends, of course, on the time series being studied, but there are some general principles which apply.

Continuous or discrete

First, it is useful to distinguish between continuous and discrete time series. An example of the former is the fluctuating thickness of paper as a function of length; an example of the latter is the weight per unit length of paper. Note that these sets of ordered observations are not direct functions of time, and yet they are called time series.

Discrete time series may be obtained by sampling continuous time series. Since much of the computation and analysis is done using digital methods, the remainder of the paper will assume that the time series are discrete, and for simplicity that the observations are at unit intervals. The time series is then denoted by $\{z_t\}$, where the index t runs from 1 to n for a series of n observations.

Random processes

The time series may be described by a set of n random variables (rv) Z_t, $t = 1, 2, \ldots, n$, and and n-dimensional probability density function (pdf) $f_{1\ 2\ \ldots\ n}(z_1, z_2, \ldots, z_n)$ defined on the n dimensional sample space

$-\infty < z_t < \infty$, $t=1,\ldots,n$. The set of rv's may be written as $\{Z_t\}$, which denotes a random process. The pdf $f_{1\,2\,\ldots\,n}(z_1, z_2, \ldots, z_n)$ may then be written $f_{\{Z_t\}}(\underline{z})$, which denotes the temporal nature of the random process $\{Z_t\}$. That is, the pdf $f_{\{Z_t\}}(\underline{z})$ may depend on the origin of time t. In such a case, the random process $\{Z_t\}$ is called non-stationary, since the probabilistic characteristics of the process change with time. If the pdf

$$f_{\{Z_t\}}(\underline{z}) = f_{\{Z_{t+k}\}}(\underline{z}) \tag{2.1}$$

for all k, then $\{Z_t\}$ is stationary. It will be seen that stationary processes may be used to generate and describe non-stationary processes, and so we shall be concerned mainly with stationary ones. For stationary processes, the pdf will be written simply as $f_Z(\underline{z})$.

A particularly useful random process

A particularly useful probability model for the random process $\{Z_t\}$ is the multivariate Normal pdf. This pdf may be written concisely in matrix notation as

$$f_{\{Z_t\}}(\underline{z}) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \exp -\frac{1}{2}(\underline{z}-\underline{\mu})' \underline{V}^{-1} (\underline{z}-\underline{\mu}) \tag{2.2}$$

where $\underline{V}^{-1}$ is the inverse of the covariance matrix $\underline{V}$,

$$\underline{V} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} & \cdots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \sigma_2\sigma_3\rho_{23} & \cdots & \sigma_2\sigma_n\rho_{2n} \\ \sigma_1\sigma_3\rho_{13} & \sigma_2\sigma_3\rho_{23} & \sigma_3^2 & \cdots & \sigma_3\sigma_n\rho_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \sigma_3\sigma_n\rho_{3n} & \cdots & \sigma_n^2 \end{pmatrix} \tag{2.3}$$

Hence the multivariate Normal pdf is completely specified by the n parameters μ_i, the n parameters σ_i, and the $(n-1)(n)/2$ correlation coefficients ρ_{ij}, $i=1,\ldots,n$, $j=i+1,\ldots,n$. If $\{Z_t\}$ is stationary, the covariance matrix reduces to

$$\underline{V} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} \tag{2.4}$$

so that knowledge of the parameters μ, σ^2, and the correlation function ρ_k, $k=1, 2, \ldots$ implies complete knowledge of the pdf.

3. Linear random processes

A particularly useful class of probability models for time series are linear random processes. These are important because they are mathematically tractable, and because they are flexible enough to describe many time series. The role they play is equivalent to that of the linear differential equation in the study of deterministic systems such as in mechanics or control theory or physics.

3.1) Linear discrete random processes

Discrete linear random processes are described by linear difference equations relating an "output" random process Z_t to an "input" random process A_t. That is

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + A_t - \theta_1 A_{t-1} - \cdots - \theta_q A_{t-q} \qquad (3.1)$$

This process is an autoregressive-moving average (ar-ma) process of order (p,q). Special cases of these mixed models are now considered.

3.1.1) White noise

As a particularly simple example, consider a random process $Z_t = A_t$, $t = 0, \pm 1, \pm 2, \ldots$ where A_t is independent of all other values $A_{t-1}, A_{t-2}, \ldots, A_{t+1}, A_{t+2}, \ldots$. The random process A_t is called "a purely random process," or "white noise," and may have any pdf $f_{A_t}(a)$. If the process A_t is stationary, then $f_{A_t}(a)$ may be written $f_A(a)$, and we shall assume that this is the case. As an example of a stationary white noise process, we could imagine random numbers A_t being generated by a computer at a constant rate. An example is given in Figure 3.1.

This concept of a white noise signal is of central importance in time series analysis, just as an independent observational error is in ordinary statistical analysis. The reason is that the next value of the white noise process A_t is unpredictable even if we know all previous and subsequent values $a_{t\pm1}, a_{t\pm2}, \ldots$. We shall reserve the symbols A_t and a_t for white noise processes only.

3.1.2) Moving average processes

The next simplest linear random process is the moving average (ma) process. For example, the first order ma process is

$$Z_t = A_t - \theta_1 A_{t-1} \quad . \qquad (3.2)$$

An example of a first order ma process with $\theta_1 = .25$ is given in Figure 3.2. In general, a q^{th} order ma process is

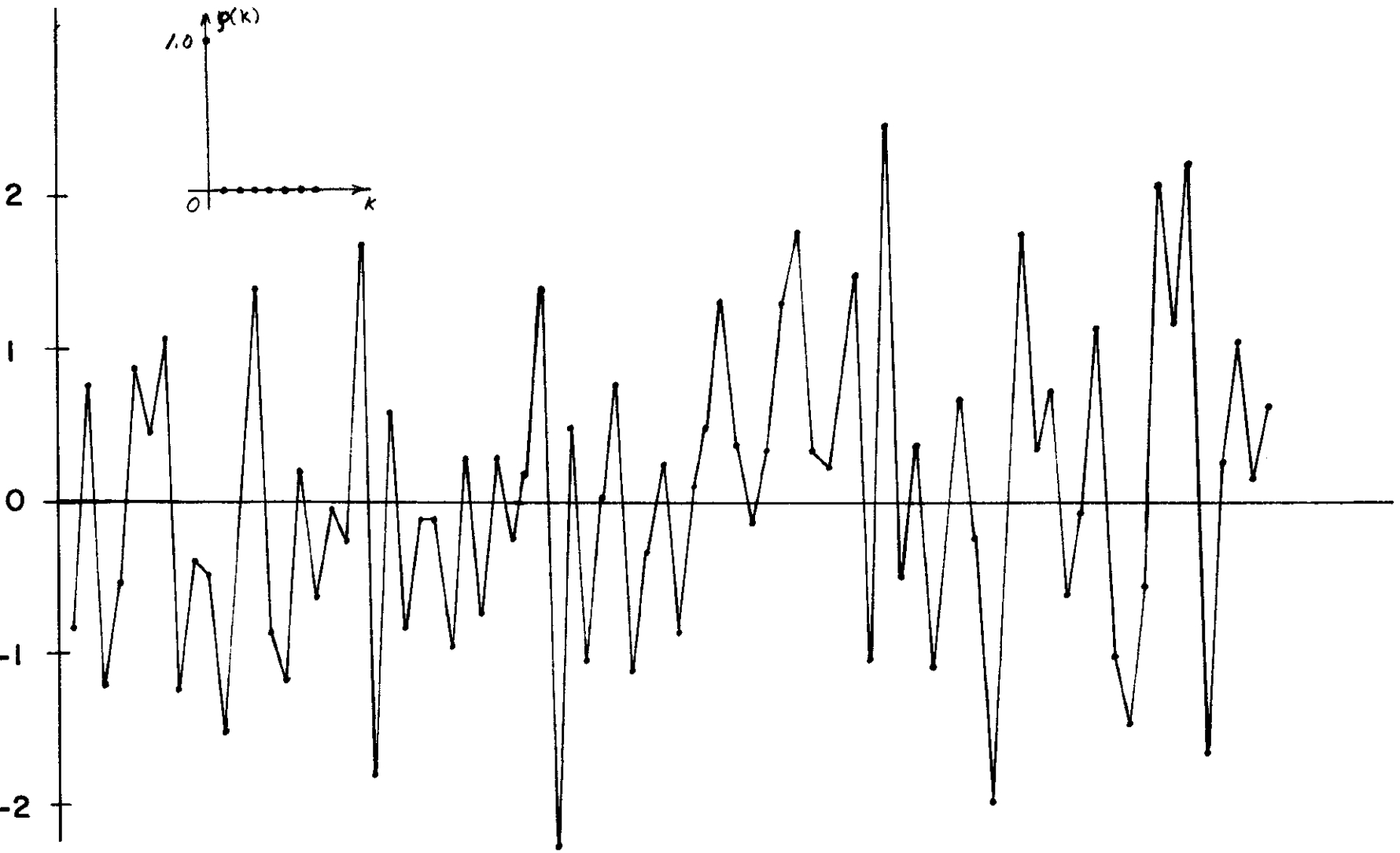

FIGURE 3.1 A SERIES OF 80 RANDOM NUMBERS

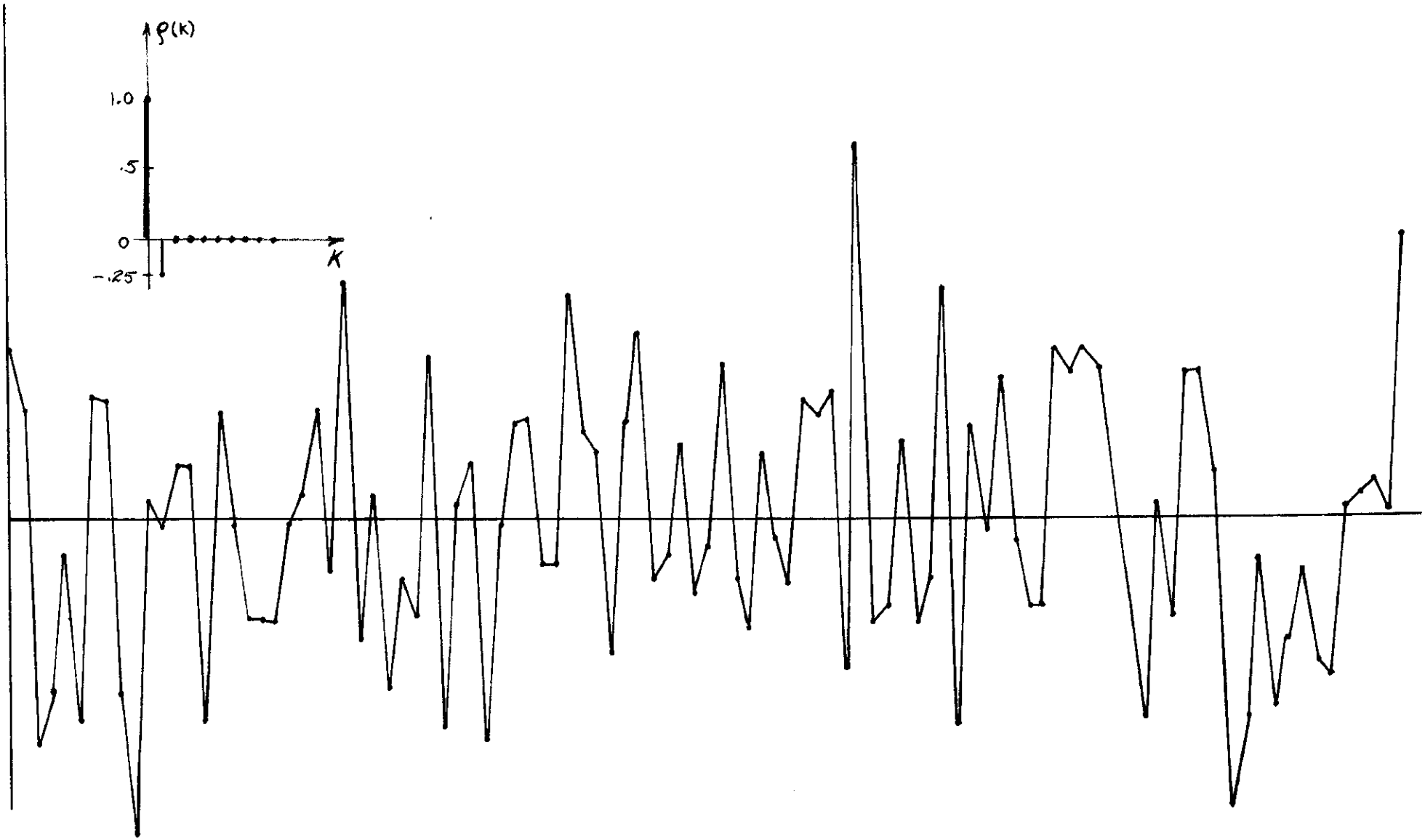

FIGURE 3.2 A REALIZATION OF THE MOVING AVERAGE PROCESS $Z_t = A_t - .25 A_{t-1}$

$$Z_t = A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \ldots - \theta_q A_{t-q} \quad . \tag{3.3}$$

3.1.3) Autoregressive processes

Another simple form of linear process is the autoregressive (ar) process. For example, the first order ar process is

$$Z_t = \phi_1 Z_{t-1} + A_t \quad . \tag{3.4}$$

This states that the next value, Z_t, of the process is given by ϕ_1 times the previous value plus an unpredictable quantity or innovation A_t. Realizations of first order ar processes with $\phi_1 = +.9$ and $-.9$ are shown in Figures 3.3 and 3.4 respectively. The general p^{th} order ar process is

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + A_t \quad . \tag{3.5}$$

A realization of a second order ar process, with $\phi_1 = +1.0$, $\phi_2 = -.5$, is shown in Figure 3.5.

3.2) Properties of stationary random processes

3.2.1) White noise

A white noise process A_t is a random process in which the joint pdf factors,

$$f_{A_{t_1} A_{t_2} \ldots A_{t_n}}(a_1, a_2, \ldots, a_n) = f_{A_{t_1}}(a_1) f_{A_{t_2}}(a_2) \ldots f_{A_{t_n}}(a_n) \tag{3.6}$$

that is, the rv's $A_{t_1}, A_{t_2}, \ldots, A_{t_n}$ are completely independent. In this case

$$\mathrm{Cov}\left[A_{t_1}, A_{t_2}\right] = \begin{cases} 0 & t_1 \neq t_2 \\ \sigma^2_{t_1} & t_1 = t_2 \end{cases} \tag{3.7}$$

and hence the rv's A_{t_1} and A_{t_2} are uncorrelated for all t_1 and t_2 not equal.

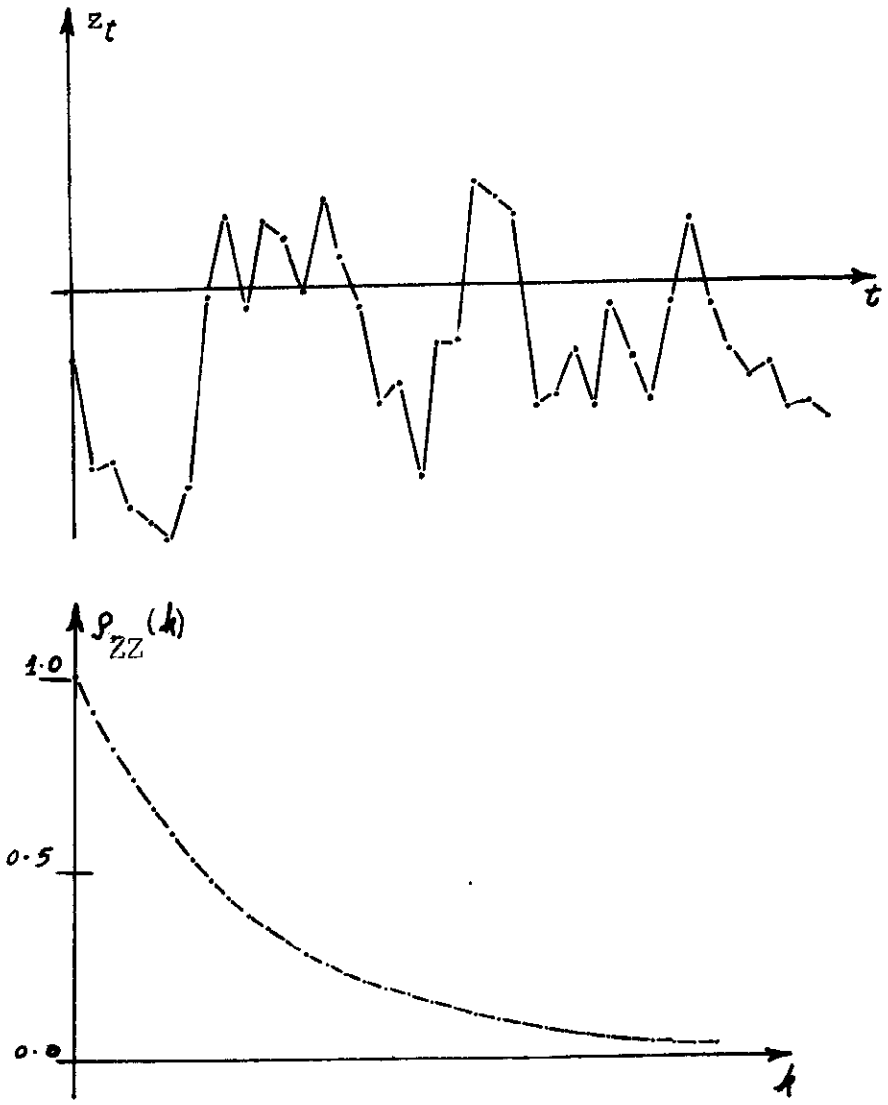

FIGURE 3.3 A REALISATION AND THE AUTOCORRELATIO
FUNCTION OF A DISCRETE FIRST ORDER
AUTOREGRESSIVE PROCESS ($\phi_1 = +.9$)

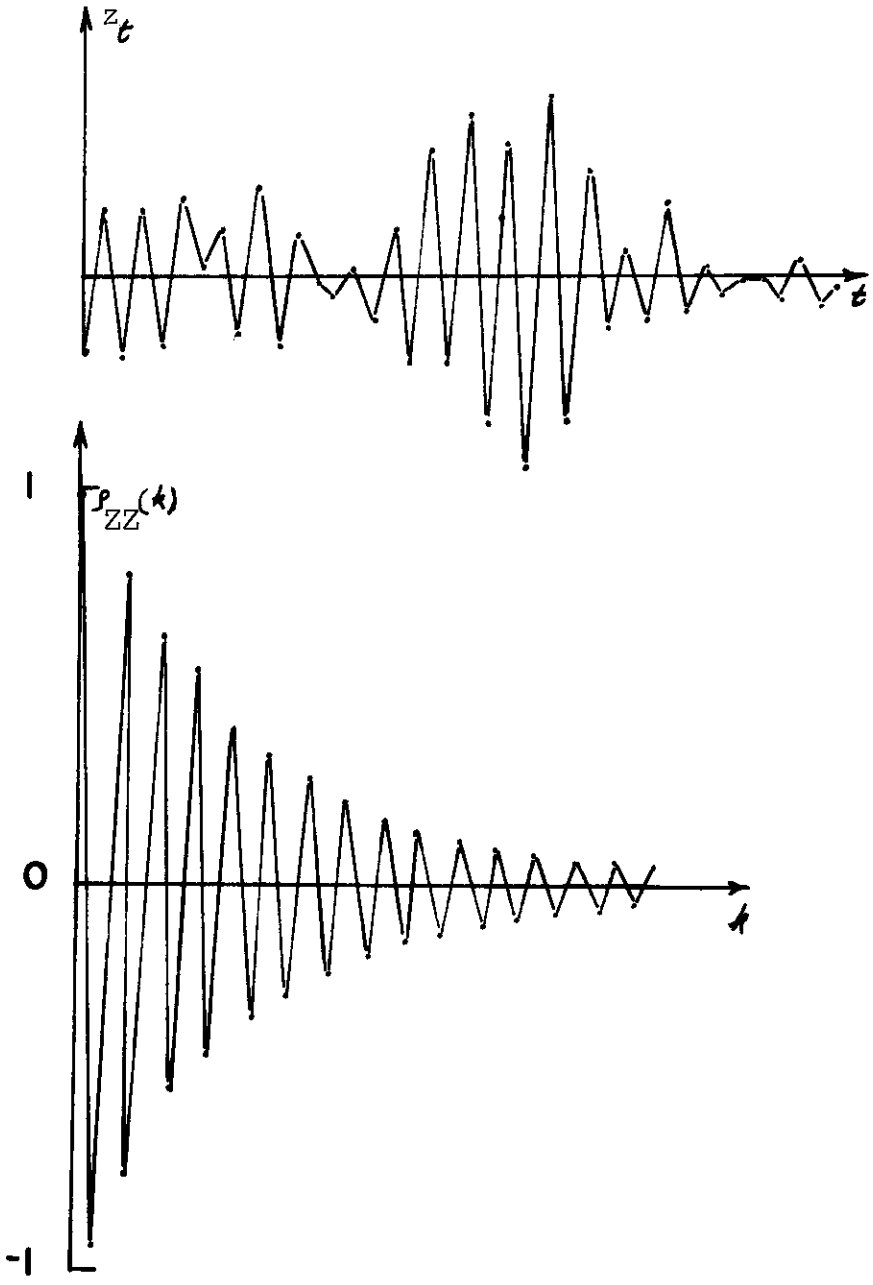

FIGURE 3.4 A REALISATION AND THE AUTOCORRELATION FUNCTION OF A DISCRETE FIRST ORDER AUTOREGRESSIVE PROCESS ($\phi_1 = -.9$)

3.2.2) Autocovariance functions

One useful way of characterizing a time series and a random process is by its autocovariance function (acvf) or its autocorrelation function (acf). This is particularly true if the random process is Normal since, as stated in Section 2, knowledge of the acvf implies complete knowledge of the joint pdf.

For a general random process Z_t, the acvf $\mathrm{Cov}\left[Z_{t_1}, Z_{t_2}\right]$ will be a function of the times t_1 and t_2. For a stationary random process, however, the acvf will be a function only of the time difference or lag $\tau = t_2 - t_1$. For example, for stationary white noise, $\mathrm{Var}[A_t] = \sigma_A^2$ is independent of t, and the acf $\rho(t_1 - t_2) = \rho(u)$ consists of a value unity at $u = 0$, and 0 for $u \neq 0$. As a second example, the first order ma process (3.2) has an acvf and acf of

$$\gamma_{ZZ}(k) = \begin{cases} (1+\theta_1^2)\sigma^2 & k = 0 \\ -\theta_1\sigma^2 & k = \pm 1 \\ 0 & |k| \geq 2 \end{cases} \qquad \rho_{ZZ}(k) = \begin{cases} 1 & k = 0 \\ \dfrac{-\theta_1}{1+\theta_1^2} & k = \pm 1 \\ 0 & |k| \geq 2 \end{cases} \tag{3.8}$$

respectively.

These may be derived as

$$\begin{aligned} \mathrm{Cov}\left[Z_t Z_{t+k}\right] &= \gamma_{ZZ}(k) = \mathrm{Cov}\left[(A_t - \theta_1 A_{t-1}), (A_{t+k} - \theta_1 A_{t+k-1})\right] \\ &= \mathrm{Cov}\left[A_t, A_{t+k}\right] - \theta_1 \mathrm{Cov}\left[A_t, A_{t+k-1}\right] \\ &\quad - \theta_1 \mathrm{Cov}\left[A_{t-1}, A_{t+k}\right] + \theta_1^2 \mathrm{Cov}\left[A_{t-1}, A_{t+k-1}\right] \end{aligned} \tag{3.9}$$

But from (3.7)

$$\text{Cov}\left[A_t, A_{t+k}\right] = \begin{cases} 0 & k \neq 0 \\ \sigma^2 & k = 0 \end{cases}$$

Hence, for example, for $k = 0$, (3.9) becomes

$$\gamma_{ZZ}(0) = \text{Cov}\left[Z_t, Z_t\right] = \sigma^2 + \theta_1(0) + \theta_1(0) + \theta_1^2\sigma^2 = (1+\theta_1^2)\sigma^2 \quad .$$

The correlation function $\rho_{ZZ}(k)$ for the first order ma process is 1 for $k = 0$, is between $-.5$ and $+.5$ for $k = 1$, and is 0 for $|k| > 1$. Note that any value of θ_1 is permissible and yields a stationary process.

For a general q^{th} order ma process,

$$\rho_{ZZ}(k) = \begin{cases} 1 & , \quad k = 0 \\ \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1 + \theta_2^2 + \ldots + \theta_q^2} & , \quad k = \pm 1, \pm 2, \ldots, \pm q \\ 0 & , \quad |k| > q \end{cases} \qquad (3.10)$$

One important point is that $\rho_{ZZ}(k)$ is zero for $|k| > q$.

Correlation functions may also be derived for ar processes. For example, for the first order ar process (3.4)

$$\gamma_{ZZ}(k) = \frac{\phi_1^{|k|}\,\sigma^2}{1-\phi_1^2} \qquad k = 0, \pm 1, \pm 2, \ldots \qquad (3.11)$$

This may also be derived from basic principles as follows. For $k \geq 1$,

$$\text{Cov}\left[Z_{t-k}, Z_t\right] = \gamma_{ZZ}(k) = \text{Cov}\left[Z_{t-k}, \phi_1 Z_{t-1} + A_t\right]$$

$$= \phi_1 \text{Cov}\left[Z_{t-k}, Z_{t-1}\right] + 0 = \phi_1 \gamma_{ZZ}(k-1) \,,$$

since A_t is independent of previous Z_t. Thus, $\gamma_{ZZ}(k)$ satisfies a recursive

equation. It follows, therefore, that

$$\gamma_{ZZ}(k) = \phi_1^k \gamma_{ZZ}(0).$$

But $$\gamma_{ZZ}(0) = \text{Cov}\left[Z_t, Z_t\right]$$

$$= \text{Cov}\left[(\phi_1 Z_{t-1} + A_t), (\phi_1 Z_{t-1} + A_t)\right] = \phi_1^2 \gamma_{ZZ}(0) + \sigma^2.$$

Hence $$\gamma_{ZZ}(0)(1-\phi_1^2) = \sigma^2.$$

Finally, since
$$\gamma_{ZZ}(k) = \text{Cov}\left[Z_t, Z_{t+k}\right] = \text{Cov}\left[Z_{t-k}, Z_t\right]$$
$$= \text{Cov}\left[Z_t, Z_{t-k}\right] = \gamma_{ZZ}(-k) \quad , \tag{3.12}$$

the result (3.11) follows.

The theoretical acf's for the ar processes shown in Figures 3.3, 3.4, and 3.5 are shown in the respective diagrams.

The correlation function for an ar process shows how, on average, the value of Z_t depends on previous values Z_{t-k}. For ϕ_1 near +1 (but < 1), Z_t is strongly dependent on Z_{t-1}, and hence the process "remembers" well, or has much "inertia," or is "smooth." This is reflected in the acf, since $\rho_{ZZ}(k) = \phi_1 \rho_{ZZ}(k-1)$ and hence the acf is very smooth. (See Figure 3.3.) For ϕ_1 near -1 (but > -1), Z_t is strongly dependent on Z_{t-1}, but with opposite sign. (See Figure 3.4.) Hence this process tends to fluctuate strongly — and so does the acf. For ϕ_1 near zero, Z_t depends little on Z_{t-1} and as expected the process tends to be very noisy, or white. For $|\phi_1| > 1$, the process "blows up" and is said to be "unstable" or non-stationary. For $|\phi_1| = 1$, the process is non-stationary in a special peculiar way, and this

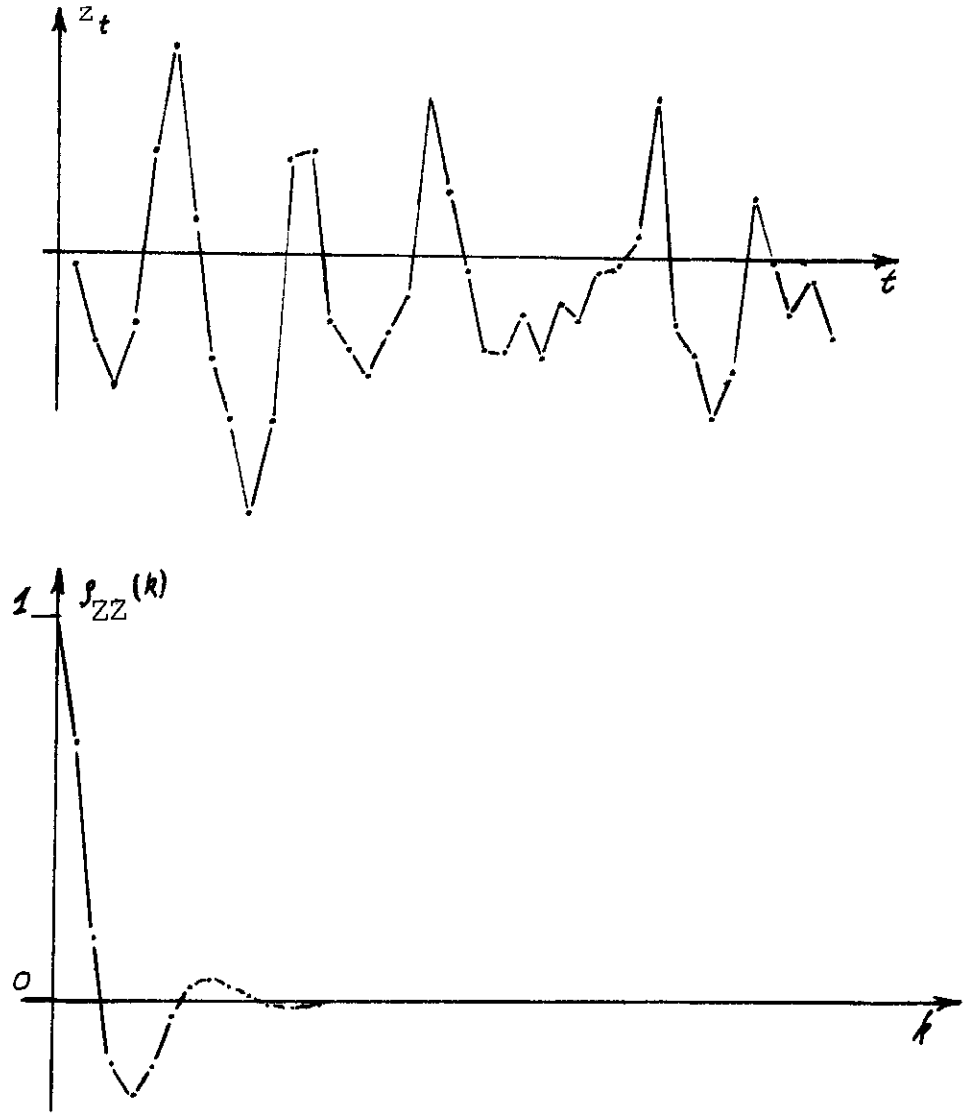

FIGURE 3.5 A REALISATION AND THE AUTOCORRELATION FUNCTION OF A DISCRETE SECOND ORDER AUTOREGRESSIVE PROCESS ($\phi_1 = 1.0$, $\phi_2 = -0.5$)

special case will be discussed more fully later.

For a general p^{th} order ar process, the acvf satisfies the following recursive relationship

$$\gamma_{ZZ}(k) = \phi_1 \gamma_{ZZ}(k-1) + \phi_2 \gamma_{ZZ}(k-2) + \ldots + \phi_p \gamma_{ZZ}(k-p) \qquad k \geq 1 \tag{3.13}$$

with the additional conditions (3.12) and

$$(1 - \phi_1^2 - \phi_2^2 - \ldots - \phi_p^2)\gamma_{ZZ}(0) - 2(\phi_1\phi_2 + \phi_2\phi_3 + \ldots + \phi_{p-1}\phi_p)\gamma_{ZZ}(1)$$
$$- 2(\phi_1\phi_3 + \phi_2\phi_4 + \ldots + \phi_{p-2}\phi_p)\gamma_{ZZ}(2) - \ldots - 2(\phi_1\phi_p)\gamma_{ZZ}(p-1) = \sigma^2 \tag{3.14}$$

Note that the acf of an ar process does not truncate, in general, as the acf of an ma process did.

Of special importance are the first and second order ar processes. For $p = 2$, the acf may be written in closed form as

$$\gamma_{ZZ}(k) = \alpha_1 c_1^{|k|} + \alpha_2 c_2^{|k|} \quad , \qquad k = 0, \pm 1, \pm 2, \ldots \tag{3.15}$$

If the constants c_1 and c_2 are real and distinct, then $\gamma_{ZZ}(k)$ is a sum of two geometrically diminishing terms; if the constants are complex conjegates, then $\gamma_{ZZ}(k)$ is a damped oscillatory function. (See Figure 3.5.) It is seen that this process, involving only two parameters, can produce a wide variety of acf's, and hence the process should be able to model many stationary time series.

Acvf and acf estimates

Estimates $c_{ZZ}(k)$ of the acvf $\gamma_{ZZ}(k)$ of a time series z_t, $t=1,2,\ldots,n$ may be calculated [7] as

$$c_{ZZ}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z}) \tag{3.16}$$

where

$$\bar{z} = \frac{1}{n}\sum_{t=1}^{n} z_t \,. \tag{3.17}$$

For example, for the series $1,6,6,4,3$ with $n=5$,

$$\bar{z} = \frac{1}{5}\{1+6+6+4+3\} = 4$$

so that

$$c_{ZZ}(0) = \frac{1}{5}\{(-3)^2+2^2+2^2+0^2+(-1)^2\} = 3.6$$

$$c_{ZZ}(1) = \frac{1}{5}\{(-3)(2)+2(2)+2(0)+(0)(-1)\} = -.4$$

$$c_{ZZ}(2) = \frac{1}{5}\{(-3)(2)+2(2)+2(-1)\} = -1.6$$

$$c_{ZZ}(3) = \frac{1}{5}\{(-3)(0)+2(-1)\} = -.4$$

$$c_{ZZ}(4) = \frac{1}{5}\{(-3)(-1)\} = .6$$

$$c_{ZZ}(k) = 0\,, \quad k \geqq 5\,.$$

Estimates of the acf $r_{ZZ}(k)$ of $\rho_{ZZ}(k)$ may be obtained from the acvf estimates $c_{ZZ}(k)$ using

$$r_{ZZ}(k) = \frac{c_{ZZ}(k)}{c_{ZZ}(0)} \,. \tag{3.18}$$

For the above series of 5 observations, the sample acf is

$$r_{zz}(1) = \frac{-.4}{3.6} = -.111$$

$$r_{zz}(2) = \frac{-1.6}{3.6} = .444$$

$$r_{zz}(3) = -.111$$

$$r_{zz}(4) = .167$$

$$r_{zz}(k) = 0 \ , \ k \geq 5$$

3.3) Operational methods

3.3.1) Shift operators

A very useful way of handling linear models is to use a shift operator. This is an operator which, when operating on A_t, produces A_{t-1}, that is

$$B A_t = A_{t-1}$$

and

$$B^k A_t = A_{t-k} \quad . \tag{3.19}$$

(Systems engineers may be more at home with the A^{-1} notation.) In the B notation, the first order ma process (3.2) becomes

$$Z_t = (1-\theta_1 B)A_t$$

and similarly the first order ar process (3.4) becomes

$$Z_t = \phi_1 B Z_t + A_t \ ,$$

or

$$(1-\phi_1 B)Z_t = A_t \ . \tag{3.20}$$

Dividing through by $(1-\phi_1 B)$ gives Z_t in terms of A_t as

$$Z_t = \frac{1}{1-\phi_1 B} A_t \quad .$$

Expanding the operator

$$\frac{1}{1-\phi_1 B} = 1 + \phi_1 B + (\phi_1 B)^2 + \ldots$$

gives

$$\begin{aligned} Z_t &= (1 + \phi_1 B + \phi_1^2 B^2 + \ldots) A_t \\ &= A_t + \phi_1 A_{t-1} + \phi_1^2 A_{t-2} + \ldots \end{aligned} \tag{3.21}$$

and hence the ar process can be written equivalently as an infinite ma process. Conversely, for the first order ma process (3.2) we may reverse the roles of Z_t and A_t and regard A_t as an ar process driven by the Z_t. Hence, by writing

$$A_t = \frac{1}{1-\theta_1 B} Z_t \tag{3.22}$$

we find

$$A_t = Z_t + \theta_1 Z_{t-1} - \theta_1^2 Z_{t-2} + \ldots \tag{3.23}$$

Regrouping expresses the ma process A_t as an infinite ar process

$$Z_t = A_t - \theta_1 Z_{t-1} - \theta_1^2 Z_{t-2} - \ldots \tag{3.24}$$

For higher order ma, ar, and mixed ar-ma processes the operational equations become

$$(1-\phi_1 B - \ldots - \phi_p B^p) Z_t = (1-\theta_1 B - \ldots - \theta_q B^q) A_t. \tag{3.25}$$

These may be rewritten to show Z_t as the result of operating on A_t as

$$Z_t = \frac{(1-\theta_1 B-\ldots-\theta_q B^q)}{(1-\phi_1 B-\ldots-\phi_p B^p)} A_t = H(B)A_t \qquad (3.26)$$

or to show A_t as the result of operating on Z_t, as

$$A_t = \frac{(1-\phi_1 B-\ldots-\phi_p B^p)}{(1-\theta_1 B-\ldots-\theta_q B^q)} Z_t = H^{-1}(B)Z_t \qquad (3.27)$$

If A_t is considered the input to a system and Z_t is the response of the system to that input, then H(B) is referred to as the transfer function of the system.

3.3.2) Weighting functions

The transfer function H(B), which for linear systems is a ratio of two polynomials in B, may be expanded as an infinite polynomial in B, that is

$$\begin{aligned} H(B) &= \frac{1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q}{1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p} \\ &= 1+h_1 B+h_2 B^2+\ldots+h_k B^k+\ldots \\ &= \sum_{k=0}^{\infty} h_k B^k \quad . \end{aligned} \qquad (3.28)$$

The set of coefficients h_k, $k = 0, 1, \ldots$, so obtained is called the weighting function or impulse response of the system, and is written simply as $\{h_k\}$.

For example, for the first order ar process (3.20)

$$H(B) = \frac{1}{1-\phi_1 B} \qquad (3.29)$$

If $|\phi_1| < 1$, this may be expanded as

$$H(B) = 1 + \phi_1 B + \phi_1^2 B^2 + \ldots = \sum_{k=0}^{\infty} \phi_1^k B^k \tag{3.30}$$

Thus, $\{h_k\} = \phi_1^k$, $k = 0, 1, 2, \ldots$ (3.31)

For the ma process (3.22)

$$H(B) = 1 - \theta_1 B$$

and

$$\{h_k\} = \begin{cases} 1 & k = 0 \\ -\theta_1 & k = 1 \\ 0 & k = 2, 3, \ldots \end{cases} \tag{3.32}$$

The transfer function H(B) and the weighting function $\{h_k\}$ completely characterize the random process just as the difference equation did. The weighting function, however, provides a pretty interpretation of the process, since it shows how well the process "remembers." For example, for the first order ar process (3.18), the output Z_t at some time t is, from (3.31),

$$Z_t = A_t + \phi_1 A_{t-1} + \phi_1^2 A_{t-2} + \ldots = \sum_{k=0}^{\infty} \phi_1^k A_{t-k} \tag{3.33}$$

Thus, the value of Z at time t is a weighted sum of past values of the input A_t. If ϕ_1 is small, say .1, $h_k = \phi_1^k$ is very small for $k \geq 2$, and hence the output Z_t does not "remember" or depend strongly on previous values of the input. For ϕ_1 large, say .9, h_k is still greater than .5 for $k = 6$, and hence Z_t "remembers" A_{t-6} by about .54.

For a q^{th} order ma process, $h_k = 0$ for $k \geq q+1$, and hence the output

Z_t completely forgets inputs which occurred more than (q+1) intervals ago.

The extent to which the weighting function damps out is a measure of the lack of inertia of the system, and hence of the noisiness of the random process Z_t. Thus, the weighting function conveys the same kind of information as does the acf or acvf. In fact, for white noise input

$$\begin{aligned} \gamma_{ZZ}(k) &= E\left[Z_t Z_{t+k}\right] \\ &= E\left[(h_o A_t + h_1 A_{t-1} + \ldots)(h_o A_{t+k} + h_1 A_{t+k-1} + \ldots \right. \\ &\quad \left. + h_k A_t + h_{k+1} A_{t-1} + \ldots)\right] = \sigma^2 \sum_{u=0}^{\infty} h_u h_{u+k} \end{aligned} \tag{3.34}$$

which shows how the acvf and the weighting function are related.

3.4) Non-stationary processes

It was stated in Section 3.1 that a process is termed stationary if the joint pdf's of the process remain unaltered under a shift of origin. For some time series, such as the one shown in Figure 1.3, it is clearly unrealistic to attempt to describe them by random processes which are stationary. Hence we must invent some ways of describing and characterizing non-stationary time series.

3.4.1) A first order process

One way is already at hand, since our linear models can describe non-stationary processes as well as stationary ones. To illustrate, consider the first order ar process (3.4)

$$Z_t = \phi_1 Z_{t-1} + A_t \tag{3.35}$$

where, as before, A_t is a stationary white noise process.

Writing this entirely in terms of the A_t, as in (3.21), gives

$$Z_t = A_t + \phi_1 A_{t-1} + \phi_1^2 A_{t-1} + \ldots + \phi_1^k A_{t-k} + \ldots \tag{3.36}$$

Hence Z_t "remembers" the input of k units before by an amount ϕ_1^k. For stationary processes $|\phi_1| < 1$ and hence past inputs receive weights which are less than one, and which decrease as the time difference k increases.

If $|\phi_1| \geq 1$, then Z_t "remembers" previous inputs by an amount $|\phi_1^k| \leq 1$, and hence the process is <u>nonstationary</u> or <u>unstable</u>. If $|\phi_1| > 1$, Z_t depends overwhelmingly on remote past values and the process <u>diverges</u> or explodes. Hence this linear process may be used to describe the beginnings of epidemics where the physical system may be safely assumed linear.

If $|\phi_1| = 1$, the process is nonstationary, but in a peculiar, useful way. Thus, if $\phi_1 = -1$, the process tends to act like a noise-driven oscillator whose amplitude of oscillation may vary without bound. A more important case is when $\phi_1 = +1$. In this case, the first <u>difference</u> $Z_t - Z_{t-1}$, written ∇Z_t, is stationary white noise. Operationally,

$$\nabla Z_t = (1-B)Z_t = Z_t - Z_{t-1} \tag{3.37}$$

or

$$Z_t = \nabla^{-1} A_t = A_t + A_{t-1} + A_{t-2} + \ldots = SA_t \tag{3.38}$$

That is, Z_t is the <u>sum</u> of the present plus all past inputs.

This random process has no moments in general, but in contrast to the process for which $|\phi_1| > 1$ its behavior is not explosive, but "homogeneously"[4] nonstationary. In this case, the process $Z_t = SA_t$ tends to fluctuate about a discontinuous level. An example is shown in Figure 3.6. Hence the process Z_t defined by $\nabla Z_t = A_t$ is termed <u>nonstationary in the mean</u>.

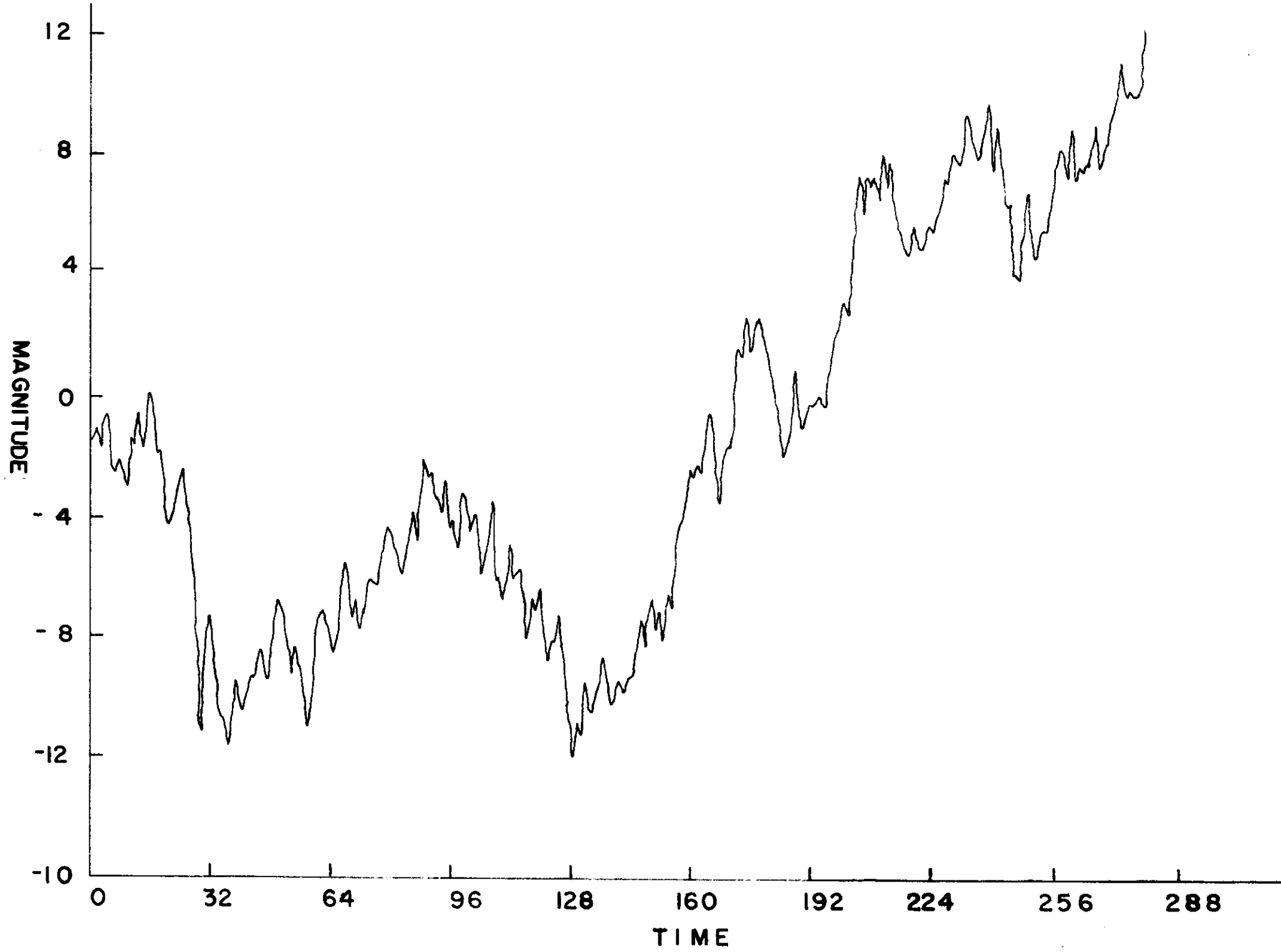

FIGURE 3·6 A REALIZATION OF THE PROCESS $z_t = \sum A_t$

3.4.2) Stability and stationarity regions

The stationarity of an ar process may be very simply determined by investigating the transfer function H(B). Thus, for a general p^{th} order ar process,

$$H(B) = \frac{1}{1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p} \tag{3.39}$$

The denominator may be factored into p simple factors

$$H(B) = \frac{1}{(1-\alpha_1 B)(1-\alpha_2 B)\ldots(1-\alpha_p B)} \tag{3.40}$$

when the α_i will either be real, or will appear in pairs of complex conjugates. For example, for $p=2$,

$$H(B) = \frac{1}{1-\phi_1 B-\phi_2 B^2} = \frac{1}{\left(1-\frac{\left(\phi_1+\sqrt{\phi_1^2+4\phi_2}\right)B}{2}\right)\left(1-\frac{\left(\phi_1-\sqrt{\phi_1^2+4\phi_2}\right)B}{2}\right)} \tag{3.41}$$

and if $4\phi_2 < -\phi_1^2$, α_1 and α_2 are complex conjugates, while if $4\phi_2 > -\phi_1^2$, α_1 and α_2 are real.

When the α_i are all distinct, that is no multiple roots, H(B) may be written as a sum of terms

$$H(B) = \frac{C_1}{(1-\alpha_1 B)} + \frac{C_2}{(1-\alpha_2 B)} + \ldots + \frac{C_p}{(1-\alpha_p B)} \tag{3.42}$$

by using partial fractions. Expanding each term as in (3.30) and writing

$$H(B) = \sum_{k=0}^{\infty} h_k B^k$$

gives

$$h_k = C_1\alpha_1^k + C_2\alpha_2^k + \ldots + C_p\alpha_p^k \tag{3.43}$$

It is easily seen, therefore, that in order for the system to be stationary (or stable) all the terms α_i, $i=1,2,\ldots,p$ must satisfy $|\alpha_i| < 1$. Hence the stationarity condition for a general p^{th} order ar process is that the roots, $B_i = 1/\alpha_i$ of the equation

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) = 0 \tag{3.44}$$

must lie outside the unit circle $|B| = 1$. These relationships are well known to control engineers who have had experience with discrete control systems.

Further, if the roots B_i of the equation lie inside the unit circle, the process is explosive. If the roots lie exactly on the unit circle, the process is termed homogeneously non-stationary.

When the r^{th} root is repeated d times, say, the same stationarity conditions apply, but now the partial fraction expansion (3.42) becomes

$$\begin{aligned} H(B) = {} & \frac{C_1}{(1-\alpha_1 B)} + \ldots + \frac{C_{r_1}}{(1-\alpha_r B)} + \frac{C_{r_2}}{(1-\alpha_r B)^2} + \ldots \\ & + \frac{C_{r_d}}{(1-\alpha_r B)^d} + \ldots + \frac{C_{p-d}}{(1-\alpha_{p-d} B)} \end{aligned} \tag{3.45}$$

A term of the form $1/(1-\alpha_r B)^{d+1}$ corresponds to a weighting function h_k of the form

$$h_k = \binom{k+d}{d} \alpha_r^d \tag{3.46}$$

For example,

$$H(B) = \frac{1}{(1-\phi B)^2}$$

$$= (1+\phi B+\phi^2 B^2+\ldots)^2$$

$$= 1+B(2\phi)+B^2(3\phi^2)+B^3(4\phi^3)+\ldots$$

$$= \sum_{k=0}^{\infty}\binom{k+1}{1}\phi^k = \sum_{k=0}^{\infty}(k+1)\phi^k \quad .$$

In the special case of homogeneous nonstationarity of order d,

$$H(B) = \frac{1}{(1-B)^d} \quad , \tag{3.47}$$

and

$$h_k = \binom{k+d-1}{d-1} \tag{3.48}$$

3.4.3) Integrated ar-ma processes

The above results and the results of Section 3.2 may be combined to give a very powerful model for nonstationary processes. This model is termed [4] an integrated-autoregressive-moving average model (i-ar-ma), of order (p,d,q) and has the operational form

$$(1-B)^d(1-\phi_1 B-\ldots-\phi_p B^p)Z_t = (1-\theta_1 B-\ldots-\theta_q B^q)A_t \tag{3.49}$$

Thus, the d^{th} difference is a stationary mixed ar-ma model. The transfer function is

$$H(B) = \frac{(1-\theta_1 B-\ldots-\theta_q B^q)}{(1-B)^d(1-\phi_1 B-\ldots-\phi_p B^p)} \tag{3.50}$$

which may be written

$$H(B) = \frac{\theta(B)}{(1-B)^d \phi(B)} \tag{3.51}$$

Hence (3.49) may be recast as

$$\phi(B)Z_t = \frac{\theta(B)}{(1-B)^d} A_t \tag{3.52}$$

If now values z_t of Z_t are given for $t = 0, 1, \ldots, d-1$, then the value of Z_t may be written as an ar model in terms of a <u>complementary solution</u> which depends only on initial conditions and which is a polynomial of order (d-1) in t, and a <u>particular solution</u> which involves weighted sums of the A_t. That is, the solution is

$$\phi(B)Z_t = c_o + c_1 t + \ldots + c_{d-1} t^{d-1} + \frac{\theta(B)}{(1-B)^d} A_t \quad , \qquad t = d, d+1, \ldots \tag{3.53}$$

$$A_t = a_t \qquad t = 0, 1, \ldots, d-1$$

For example, suppose $\phi(B) = 1$, $\theta(B) = 1$, $d = 2$ so that

$$\nabla^2 Z_t = A_t$$

or

$$H(B) = \frac{1}{(1-B)^2} \quad .$$

Then if $Z_o = z_o$, $Z_1 = z_1$,

$$Z_t = c_o + c_1 t + \sum_{k=0}^{t-2} (k+1) A_{t-k} \qquad t = 2, 3, \ldots$$

where $c_o = z_o$

$$c_1 = (z_1 - z_o) .$$

Hence Z_t consists of a random component involving the A_t, and a linear deterministic component in t. Hence this random process will fluctuate about a trend line whose initial slope is $c_1 = z_1 - z_o$.

It is apparent, therefore, that by using a general i-ar-ma model of order (p,d,q) we can describe many stationary and nonstationary time series.

4. Fitting Models

Having decided that i-ar-ma processes are sufficiently flexible and physically interpretable models for time series, the remaining problem is to estimate the parameters for a particular case. This involves a three-stage iteration procedure consisting of identification, estimation, and diagnostic checking [5].

Identification consists of using the data and any additional knowledge to suggest whether the series can be described as stationary or nonstationary, and as ma, ar, or mixed ma-ar.

Estimation consists of using the data to estimate, and make inferences about, values of the parameters conditional on the tentatively identified model.

Diagnostic checking involves examination of the residuals from fitted models, which can result in either

(a) no indication of model inadequacy, or

(b) model inadequacy, together with information on how to better describe the series.

Thus the residuals a_t would be examined for any lack of randomness, and if

the residuals are autocorrelated, this information would be used to modify the model. The modified model would then be fitted and subjected to diagnostic checking.

4.1) Identification

The task is to identify an appropriate model of the form

$$(1-B)^d \phi_p(B) z_t = \theta_o + \theta_q(B) a_t \tag{4.1}$$

which may be used to describe a given time series. The approach is to difference to produce stationarity, hopefully reducing (4.1) to the mixed ar-ma model

$$\phi_p(B) y_t = \theta_o + \theta_q(B) a_t \tag{4.2}$$

where $y_t = (1-B)^d z_t$, and then to identify the model. A powerful complementary identification tool, the partial acf [5], can also be used, as illustrated below.

4.1.1) Differencing

Nonstationarity is suggested when the sample acf does not diminish at large lags. When the original series or acf exhibits nonstationarity, successive differencing is carried out until the acf of the differenced series dies out rapidly. It is usually sufficient to look at the sample acf of the original series and of its first and second differences.

The acvf $\gamma_k^{(d)}$ of the differenced series may be computed using the acvf's γ_k of the original series, since

$$\gamma_k^{(1)} = -(\gamma_{k+1} - 2\gamma_k + \gamma_{k-1})$$

and

$$\gamma_k^{(2)} = \gamma_{k+2} - 4\gamma_{k+1} + 6\gamma_k - 4\gamma_{k-1} + \gamma_{k-2}$$

For finite series, these relationships are approximations, due to loss of terms from the ends of the series as a result of differencing. For long series, these end effects are usually small, and hence the sample acvf's of the first and second difference series may be derived from those of the original series by simple differencing, with considerable savings in computer time. These autocorrelations are generally calculated up to a maximum of 20 lags.

4.1.2) Identification via the autocorrelation function

It was shown in Section 3 that the theoretical acf for the pure ma process $Y_t = \theta_q(B)A_t$ truncates, being zero after lag q, while that for the pure ar process $\phi_p(B)Y_t = A_t$ is of infinite extent. Moving average processes are thus characterized by truncation of the acf while autoregressive processes are

5. 1) Maximum likelihood estimation of autoregressive parameters

Suppose that it is required to fit the autoregressive model

$$Z_t = \phi_o + \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + A_t \qquad (5.1)$$

to an observed time series $z_1, z_2, \ldots, z_n$. The fitting procedure involves two stages:

is the sum of squares function

$$S(\phi_o, \ldots, \phi_p | z_1, \ldots, z_p) = \sum_{t=p+1}^{n} \{z_t - \phi_o - \phi_1 z_{t-1} - \ldots - \phi_p z_{t-p}\}^2 \quad (5.4)$$

The maximum likelihood or least squares estimates may then be obtained by differentiating (5.4). A special case is now considered.

First order autoregressive process (p=1)

Differentiation of the sum of squares

$$S(\phi_o, \phi_1) = \sum_{t=2}^{n} \{z_t - \phi_o - \phi_1 z_{t-1}\}^2$$

gives rise to the normal equations

$$\bar{z}_2 = \hat{\phi}_o + \hat{\phi}_1 \bar{z}_1$$

$$\sum_{t=2}^{n} z_{t-1} \{z_t - \hat{\phi}_o - \hat{\phi}_1 z_{t-1}\} = 0$$

where $\bar{z}_1$, $\bar{z}_2$ are the means of the first and last n-1 observations. On substituting the first equation into the second and rearranging, the final form of the estimation equations is

$$\hat{\phi}_o = \bar{z}_2 - \hat{\phi}_1 \bar{z}_1$$

$$\hat{\phi}_1 = \sum_{t=2}^{n} (z_{t-1} - \bar{z}_1)(z_t - \bar{z}_2) \Big/ \sum_{t=2}^{n} (z_{t-1} - \bar{z}_1)^2 \quad (5.5)$$

Since $\bar{z}_1$ and $\bar{z}_2$ will be very close to the overall mean $\bar{z}$, the estimate ϕ_1 may be approximated by r_1.

The residual sum of squares

$$S(\hat{\phi}_o, \hat{\phi}_1) = \sum_{t=2}^{n} \{z_t - \hat{\phi}_o - \hat{\phi}_1 z_{t-1}\}^2$$

may be simplified to give

$$S(\hat{\phi}_o, \hat{\phi}_1) = \sum_{t=2}^{n} (z_t - \bar{z}_2)^2 - \hat{\phi}_1 \sum_{t=2}^{n} (z_t - \bar{z}_2)(z_{t-1} - \bar{z}_1) \qquad (5.6)$$

Since there are effectively n-1 observations in $S(\phi_o, \phi_1)$ and two degrees of freedom are lost in fitting the constants ϕ_o, ϕ_1, the variance of A_t may be estimated from

$$s_a^2 = \frac{1}{n-3} S(\hat{\phi}_o, \hat{\phi}_1)$$

The $100(1-\alpha)\%$ confidence interval for θ_1 is approximately

$$(\phi_1 - \hat{\phi}_1)^2 \leqq \frac{s_a^2 f_{1,n-3}(1-\alpha)}{\sum_{t=1}^{n} (z_t - \bar{z})^2} \qquad (5.7)$$

where $f_{1,n-3}(1-\alpha)$ is the $(1-\alpha)$ probability point for the $F_{1,n-3}$ distribution.

The correct order of the ar process can be determined by looking at the residuals from the fitted model, and by looking at the autocorrelation function of the residuals, as mentioned in Section 4.2.

5.2) An example

The data of Figure 1.2 yields the acf and partial acf shown in Figure 5.1. From the plot of the original data and from the acf's, the original series may be identified as stationary and autoregressive of order p=1 or 2. This identification may be confirmed by using the partial acf test statistic of Section 4.1.3.

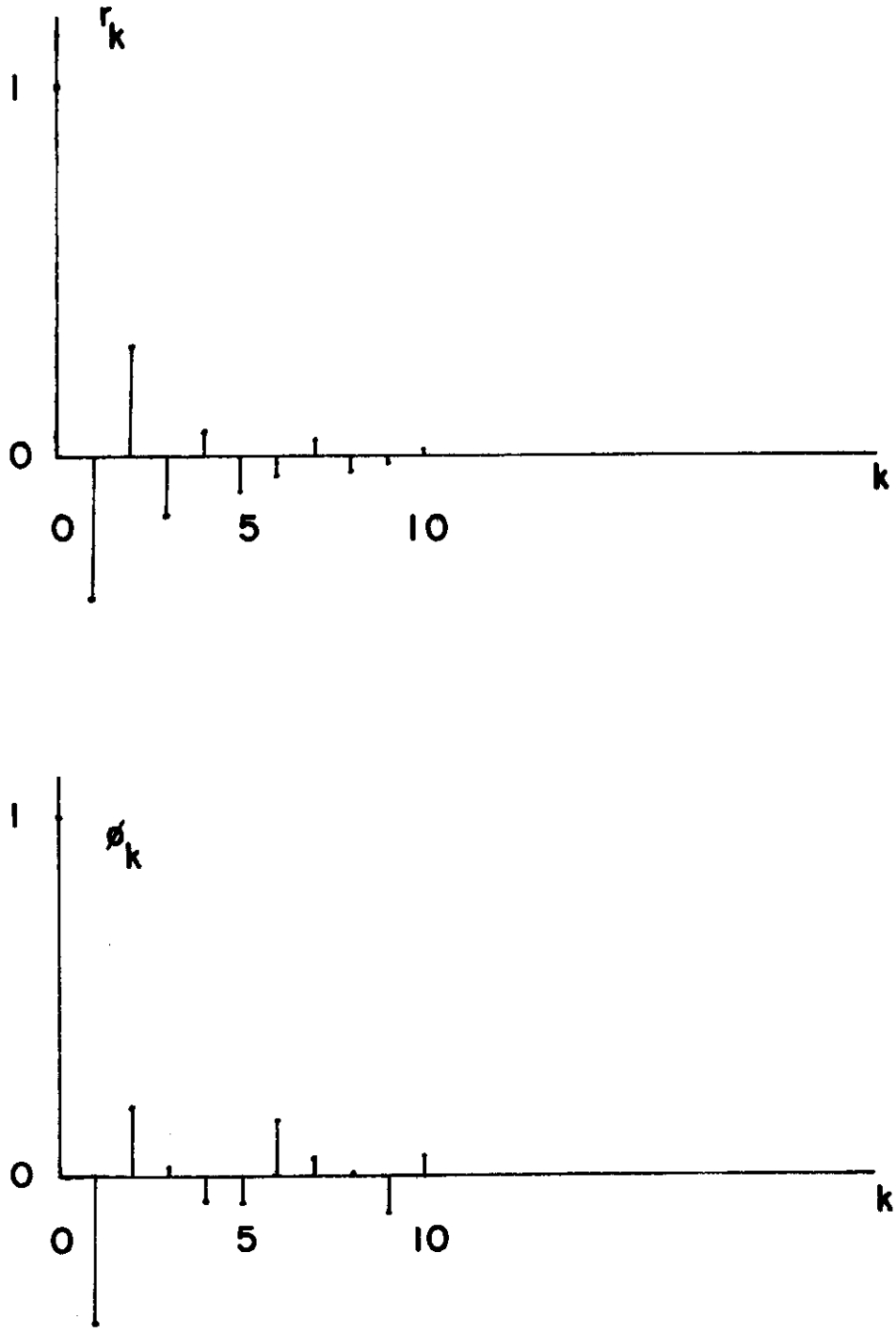

FIGURE 5·1 AUTOCORELATION AND PARTIAL AUTO-CORELATION ESTIMATES FOR THE BATCH DATA

Table 5.1: Test statistic for the order of an ar process

p=k	R_k	$1.96/\sqrt{n}$
1	-.40	.23
2	.19	.23
3	.01	.23
4	-.07	.23
5	-.07	.23
6	.15	.23
7	.05	.23
8	.00	.23

All of the test statistic values lie within the 95% limits of $\pm 1.96/\sqrt{n}$ except for $p = 1$. Hence it may be concluded that the series is described adequately by a model of order $p = 1$. The value for $p = 2$ is quite large, however, and hence it is wise to fit both a first order and second order model.

The approach of Section 5.1 yields the values $\hat{\phi}_1 = -.39$ for the first order model, and $\hat{\phi}_1 = -.32$, $\hat{\phi}_2 = +.18$ for the second order model. The sum of squares contours for the second order fit are shown in Figure 5.2 with the 95% confidence region. It is concluded that a second order model is necessary to adequately fit the data.

5.3) Estimation of the parameters of a moving average process

The first question which has to be decided in order to fit a moving average process

$$Z_t = \theta_o + A_t + \theta_1 A_{t-1} + \ldots + \theta_q A_{t-q}, \tag{5.8}$$

is the appropriate order q for the model. The method of analysis is not as simple as that for the autoregressive process because it is difficult to write down an explicit form for the likelihood function of the process. However,

and contours of constant sums of squares sketched in. If the minimum value of the sum of squares for a fixed value of q is denoted by $S(\hat{\theta}_o, \hat{\theta}_1, \ldots, \hat{\theta}_q)$ the confidence region may be obtained by locating that contour for which

$$S(\theta_o, \theta_1, \ldots, \theta_q) = S(\hat{\theta}_o, \hat{\theta}_1, \ldots, \hat{\theta}_q)(1 + \frac{q+1}{n-(q+1)} f_{q+1, n-q-1}(1-\alpha)) \tag{5.10}$$

5.4) An example

The data of Figure 1.1 yields the acf and partial acf shown in Figure 5.3. From the plot of the original data and from the acf's, the original series may be identified as stationary and first order ma or ar.

order moving average model can be fitted to these differences of the data using the approach of Section 5. 3.

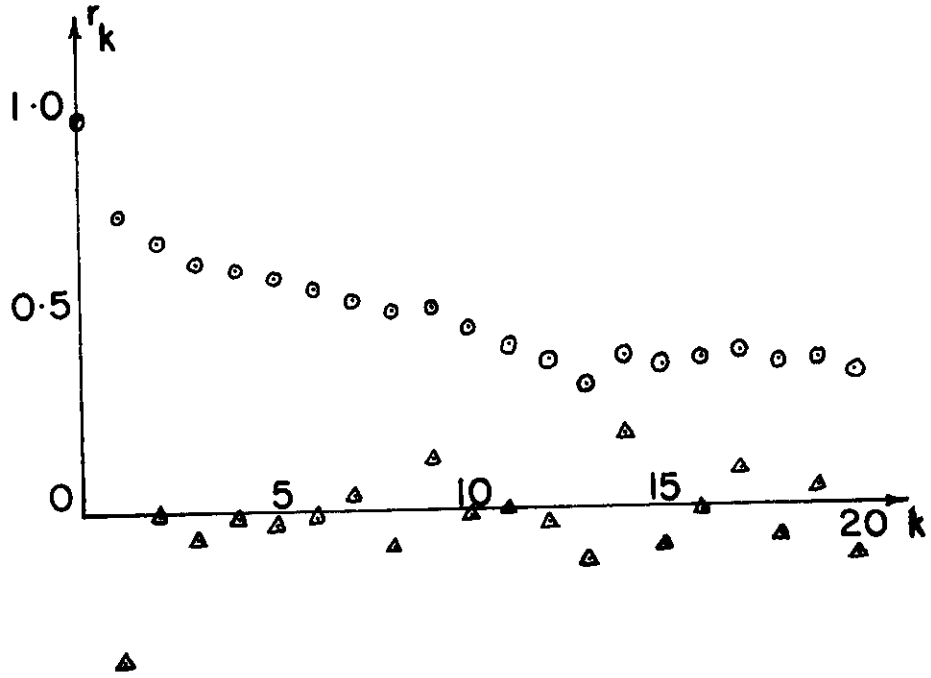

FIGURE 5·4 AUTOCORRELATION FOR THE ORIGINAL AND FIRST DIFFERENCED DATA OF FIGURE 1·3

References

1] Zadeh, L. A., and J. R. Ragazzini. An extension of Wiener's Theory of Prediction, J. Applied Physics, Vol. 21, July, 1950, pp. 645-655.

2] Zadeh, L. A., and J. R. Ragazzini. Optimum filters for the detection of signals in noise, Proc. I.R.E., Vol. 40, October, 1952, pp. 1223-1231.

3] Box, G. E. P., and G. M. Jenkins. Models for prediction and control II Linear stationary models, Univ. of Wis. Dept. of Statistics Technical Report No. 77.

4] Box, G. E. P., and G. M. Jenkins. Models for prediction and control III Linear non-stationary models, Univ. of Wis. Dept. of Statistics Technical Report No. 79.

5] Box, G. E. P., and G. M. Jenkins. Models for prediction and control IV Model identification, Univ. of Wis. Dept. of Statistics Technical Report No. 94.

6] Box, G. E. P., and G. M. Jenkins. Models for prediction and control V Estimation, Univ. of Wis. Dept. of Statistics Technical Report No. 95.

7] Jenkins, G. M., and D. G. Watts. Spectrum Analysis and its Applications. Holden-Day, San Francisco, 1968.

8] Bartlett, M. S. On the theoretical specification and sampling properties of autocorrelated time series, J. Roy. Stat. Soc., Series B, Vol. 8, 1946, p. 27.

9] Jenkins, G. M. Tests of hypotheses in the linear autoregressive model, I, Biometrika, 41, 1954, p. 405; II, Biometrika, 43, 1956, p. 187.

10] Durbin, J. The fitting of time series models, Rev. Int. Inst. Stat., 28, 1960, p. 233.