Multiple Time Series

E. J. HANNAN
The Australian National University
Canberra

John Wiley and Sons, Inc.
New York · London · Sydney · Toronto

This Page Intentionally Left Blank

Multiple Time Series

A WILEY PUBLICATION IN APPLIED STATISTICS

Multiple Time Series

E. J. HANNAN

The Australian National University
Canberra

John Wiley and Sons, Inc.

New York · London · Sydney · Toronto

10 9 8 7 6 5 4 3 2 1

Library of Congress Catalogue Card Number: 77-112847

ISBN 0 471 34805 8

Printed in the United States of America

Preface

The subject of time series analysis has intimate connections with a wide range of topics, among which may be named statistical communication theory, the theory of prediction and control, and the statistical analysis of time series data. The last of these is to some extent subsidiary to the other two, since its purpose, in part at least, must be to provide the information essential to the application of those theories. However, it also has an existence of its own because of its need in fields (e.g., economics) in which at present well-developed, exact theories of control are not possible. It is with the third of these topics that this book is concerned. It extends beyond that in two ways. The first and most important is by the inclusion in the first half of the book of a fairly complete treatment of the underlying probability theory for second-order stationary process. Although this theory is for the most part classical and available elsewhere in book form, its understanding is an essential preliminary to the study of time series analysis and its inclusion is inevitable. I have, however, included a certain amount of material over and above the minimum necessary, in relation, for example, to nonlinear filters and random processes in space as well as time. The statistical development of this last subject is now fragmentary but may soon become important.

The second additional topic is the theory of prediction, interpolation, signal extraction, and smoothing of time series. The inclusion of this material seems justified for two reasons. The first arises from the understanding that the classical "Wiener-Kolmogoroff" theories give of the structure of time series. This understanding is needed, in part at least, for statistical developments (e.g., identification problems and problems associated with the relation between the eigenvalues of the covariance matrix and the spectrum). The second reason is that these developments are becoming important to people who are statisticians concerned with time series (e.g., in missile trajectory estimation and economics). Of course, some of the more practically valuable work here is recent and would require a separate volume for an adequate treatment, but some indications concerning it are needed.

There is one other characteristic that a modern book on time series must have and that is the development of the theory and methods for the case in which multiple measurements are made at each point, for this is usually the case.

Having decided on the scope of the book, one must consider the manner in which the material will be presented and the level of the presentation. This book sets out to give the theory of the methods that appear to be important in time series analysis in a manner that it is hoped will lead finally to an understanding of the methods as they are to be used. On the whole it presents final formulas but often does not discuss computational details and it does not give computer programs. (For the most part the methods discussed are already programmed and these programs are available.) With minor exceptions numerical examples are not given. It is not a book on "practical time series analysis" but on the theory of that subject. There is a need for books of the first kind, of course, but also of this second kind, as any time series analyst knows from requests for references to a definitive discussion of the theory of this or that topic. The level of presentation causes problems, for the theory is both deep and mathematically unfamiliar to statisticians. It would probably be possible to cover the underlying probability theory more simply than has been done by making more special assumptions (or by making the treatment less precise). To make the book more accessible a different device has been used and that is by placing the more difficult or technical proofs in chapter appendices and starring a few sections that can be omitted. It is assumed that the reader knows probability and statistics up to a level that can be described as familiarity with the classic treatise *Mathematical Methods of Statistics* by Harald Cramér. A mathematical appendix which surveys some needed elementary functional analysis and Fourier methods has been added.

Some topics have not been fully discussed, partly because of the range of my interests and partly because of the need to keep the length of the book within reasonable bounds. I have said only a small amount about the spectra of higher moments. This is mainly because the usefulness of this spectral theory has not yet been demonstrated. (See the discussion in Chapter II, Section 8.) Little also has been said about nonstationary processes, and particularly about their statistical treatment. This part of the subject is fragmented at the moment. Perhaps, of necessity, it always will be. A third omission is of anything other than a small discussion of "digitized" data (e.g., "clipped signals" in which all that is recorded is whether the phenomenon surpassed a certain intensity). There is virtually no discussion of the sample path behavior of Gaussian processes, for this subject has recently been expertly surveyed by Cramér and Leadbetter (1967) and its inclusion here is not called for. I have also not discussed those inference procedures for point processes based on the times of occurrence of the events in the process (as distinct from the intervals between these times). This has recently been surveyed by Cox and Lewis (1966). Finally, the second half of the book (on inference problems) discusses only the discrete time case. This is justified by the dominance of digital computer techniques.

I have not attempted to give anything approaching a complete bibliography of writing on time series. For the period to 1959 a very complete listing is available in Wold (1965). It is hoped that the references provided herein will allow the main lines of development of the subject to the present time to be followed by the reader.

I have many people to thank for help. The book developed from a course given at The Johns Hopkins University, Baltimore, Maryland, and an appreciable part of the work on it was supported by funds from the United States Air Force. The book's existence is due in part to encouragement from Dr. G. S. Watson. Dr. C. Rhode at Johns Hopkins and R. D. Terrell, P. Thomson, and D. Nicholls at the Australian National University have all read parts of the work and have corrected a number of errors in its preliminary stages. The typing was entirely done, many times over, by Mrs. J. Radley, to whom I am greatly indebted.

E. J. HANNAN

Canberra, Australia
April, 1970

This Page Intentionally Left Blank

Contents

† These topics are special and may be omitted.

This Page Intentionally Left Blank

WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

Probability and Mathematical Statistics

ANDERSON · An Introduction to Multivariate Statistical Analysis
BLACKWELL and GIRSHICK · Theory of Games and Statistical Decisions
CRAMÉR · The Elements of Probability Theory and Some of Its Applications
DOOB · Stochastic Processes
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume II
FISZ · Probability Theory and Mathematical Statistics, *Third Edition*
FRASER · Statistics—An Introduction
FRASER · The Structure of Inference
GRENANDER and ROSENBLATT · Statistical Analysis of Stationary Time Series
HANNAN · Multiple Times Series
HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volumes I and II
HOEL · Introduction to Mathematical Statistics, *Third Edition*
KEMPTHORNE · The Design and Analysis of Experiments
LARSON · Introduction to Probability Theory and Statistical Inference
LEHMANN · Testing Statistical Hypotheses
PARZEN · Modern Probability Theory and Its Applications
RAO · Linear Statistical Inference and Its Applications
RIORDAN · An Introduction to Combinatorial Analysis
SCHEFFÉ · The Analysis of Variance
WALD · Sequential Analysis
WILKS · Collected Papers: Contributions to Mathematical Statistics
WILKS · Mathematical Statistics

Applied Probability and Statistics

BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BARTHOLOMEW · Stochastic Models for Social Processes
BENNETT and FRANKLIN · Statistical Analysis in Chemistry and the Chemical Industry
BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
BROWNLEE · Statistical Theory and Methodology in Science and Engineering, *Second Edition*

Applied Probability and Statistics (Continued)

CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. I
CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. II
CHERNOFF and MOSES · Elementary Decision Theory
CHIANG · Introduction to Stochastic Processes in Biostatistics
CLELLAND, deCANI, BROWN, BURSK, and MURRAY · Basic Statistics with Business Applications
COCHRAN · Sampling Techniques, *Second Edition*
COCHRAN and COX · Experimental Designs, *Second Edition*
COX · Planning of Experiments
COX and MILLER · The Theory of Stochastic Processes
DAVID · Order Statistics
DEMING · Sample Design in Business Research
DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis
GOLDBERGER · Econometric Theory
GUTTMAN and WILKS · Introductory Engineering Statistics
HALD · Statistical Tables and Formulas
HALD · Statistical Theory with Engineering Applications
HOEL · Elementary Statistics, *Second Edition*
HUANG · Regression and Econometric Methods
JOHNSON and LEONE · Statistics and Experimental Design: In Engineering and the Physical Sciences, Volumes I and II
LANCASTER · The Chi Squared Distribution
MILTON · Rank Order Probabilities: Two-Sample Normal Shift Alternatives
PRABHU · Queues and Inventories: A Study of Their Basic Stochastic Processes
SARHAN and GREENBERG · Contributions to Order Statistics
SEAL · Stochastic Theory of a Risk Business
WILLIAMS · Regression Analysis
WOLD and JURÉEN · Demand Analysis
WONNACOTT and WONNACOTT · Econometrics
YOUDEN · Statistical Methods for Chemists

Tracts on Probability and Statistics

BILLINGSLEY · Ergodic Theory and Information
BILLINGSLEY · Convergence of Probability Measures
CRAMÉR and LEADBETTER · Stationary and Related Stochastic Processes
RIORDAN · Combinatorial Identities
TAKÁCS · Combinatorial Methods in the Theory of Stochastic Processes

PART ONE

Basic Theory

This Page Intentionally Left Blank

CHAPTER I

Introductory Theory

1. INTRODUCTION

In a wide range of statistical applications the data to hand consist of a time series, that is, a series of observations, $x_j(n)$; $j = 1, \ldots, p$; $n = 1, \ldots, N$, made sequentially through time. Here j indexes the different measurements made at each time point n. Sets of observations of this kind dominate in some of the earth sciences (e.g., meteorology, seismology, oceanography, and geomorphology) and the social sciences (particularly economics). They are of great importance in other fields also; for example, in medical research (electrocardiograms and electroencephalograms) and in connection with problems of estimating missile trajectories. Although we have so far thought of n as measuring the passage of time, there are situations in which it might be a space variable. The $x_j(n)$ might be measurements made at equidistant points along a strip of material (e.g., a strip of newsprint, a textile fiber or a coal face). Again n might be the distance, in units of 100 m, downstream from some permanent feature of a river and $x_j(n)$, $j = 1, 2, 3$, might be measures of the speed and direction of flow of the river (or the discharge and direction of discharge).

The feature of the situations we have in mind that distinguishes them from those found in the classical part of statistics is the fact that $x_j(m)$ and $x_j(n)$ will not be independent, even for $m \neq n$. Indeed, it is almost always true that observations made at equidistant intervals of time form a sample of observations from what was, in principle, a continuously observable phenomenon.† Now it is inconceivable that $x_j(s)$ and $x_j(t)$ should remain independent as s approaches t, so that the lack of independence is forced on us by this most fundamental fact, namely, the (seeming) continuity of natural phenomena. Although, as we have said, most discrete time series may be regarded as samples from a continuous time function, we shall at times

† We use the symbol $x_j(t)$ whenever a continuous time phenomenon is being discussed, but also, sometimes, for a discrete time phenomenon. When the distinction between the two cases has to be emphasized, we use $x_j(n)$ for the discrete time phenomenon.

treat such series as probabilistic phenomena in their own right without imbedding them in a continuous time background. We use the symbol $x(t)$ without subscripts for the vector with components $x_j(t)$. Of course, $x(t)$ may, in particular, be a scalar.

The observable $x(t)$ is to be thought of as an observation on a vector-valued random variable. Thus we imply that for every finite set $t_1, t_2, \ldots, t_r$ of values of t the joint probability distribution of all elements of the vectors $x(t_j)$, $j = 1, \ldots, r$, is prescribed. We might obtain that distribution in another way. We might begin from a set of time points $t_1, t_2, \ldots, t_r, t_{r+1}$ and the corresponding probability distribution of the elements of the $x(t_j)$, $j = 1, \ldots, r + 1$ and obtain the marginal distribution of the elements of the $x(t_j)$, $j = 1, \ldots, r$ from it. If this gives a different distribution from that originally prescribed, then clearly the various prescriptions were inconsistent. We thus naturally impose the requirement that such inconsistencies do *not* occur. When this is done, it has been shown that a probability measure space (sample space) Ω may be constructed on which all the random variables $x_j(t)$ may be defined. Thus on a Borel field, $\mathcal{A}$, of sets in Ω there is defined a probability measure P, in terms of which all probability statements concerning the $x_j(t)$ are made. This space Ω may be thought of as the space of all "histories" or "realizations" of the vector function $x(t)$ and, to emphasize the fact that these random variables are definable, for each fixed j, t, as measurable functions on Ω, we sometimes replace the symbol $x_j(t)$ with $x_j(t, \omega)$. (For a further discussion see Billingsley, 1968, p. 228.) When ω is fixed and t varies, we obtain a particular history. As ω varies we get different histories (of which part of only one may be observed). The family of all such histories together with the associated structure of probabilities we call a stochastic (or random) process, and we use the words "time function" (or "time series" in the discrete case) for a particular history (or "realization") of the stochastic process. Often we do not exhibit the variable ω, replaçing, for example, an explicit integration symbolism with respect to the probability measure P on Ω with a symbolism involving the expectation operator $\mathcal{E}$.

We are mainly concerned with situations in which all of the second moments

$$\mathcal{E}\big(x_j(s)\, x_k(t)\big) = \gamma_{j,k}(s, t)$$

are finite. When $j = k$, we write $\gamma_j(s, t)$ for simplicity. We arrange these p^2-quantities, $\gamma_{j,k}(s, t)$, in a symmetric matrix which we call $\Gamma(s, t)$ and we refer to it as a covariance matrix, even though we have not made mean corrections. Moreover, in the continuous time case we almost always assume that $\gamma_{j,k}(s, t)$ is a continuous function of each of its arguments.

In this connection we have the following theorem:

Theorem 1. *In order that the* $\gamma_{j,k}(s, t)$ *may be continuous,* $j, k = 1, \ldots, p$, *it is necessary and sufficient that* $\gamma_j(s, t)$ *be continuous at* $s = t$ *for* $j = 1, \ldots, p$.

Proof. The necessity is obvious. For the sufficiency we have, using Schwartz's and Minkowski's inequalities,†

$$\begin{aligned} |\gamma_{j,k}(s+u, t+v) - \gamma_{j,k}(s, t)| &= |\mathcal{E}\{x_j(s+u)x_k(t+v)\} - \mathcal{E}\{x_j(s)x_k(t)\}| \\ &= \left|\mathcal{E}\{(x_j(s+u) - x_j(s))x_k(t+v)\} + \mathcal{E}\{x_j(s)(x_k(t+v) - x_k(t))\}\right| \\ &\le [\mathcal{E}\{(x_j(s+u) - x_j(s))^2\}\gamma_k(t+v, t+v)]^{1/2} \\ &\quad + [\mathcal{E}\{(x_k(t+v) - x_k(t))^2\}\gamma_j(s, s)]^{1/2}. \end{aligned}$$

If the condition of Theorem 1 is satisfied, then $\gamma_k(t+v, t+v)$ is uniformly bounded in v for v in any finite interval and thus the continuity of the $\gamma_{j,k}(s, t)$ is implied by the condition that, for all s,

$$\lim_{u\to 0} \mathcal{E}\{(x_j(s+u) - x_j(s))^2\} = 0, \qquad j = 1, \ldots, p, \tag{1.1}$$

and, since the left-hand side is

$$\lim_{u\to 0} \{\gamma_j(s+u, s+u) + \gamma_j(s, s) - 2\gamma_j(s+u, s)\},$$

this is implied by the continuity of $\gamma_j(s, t)$ at $s = t$.

We have also shown that the continuity of the $\gamma_{j,k}(s, t)$ is equivalent to (1.1), which is the condition for what is called mean-square continuity. More erratic behavior than this implies is, perhaps, not unthinkable, but presumably it could not be recorded because of the inability of a recording device to respond instantly to a change in the level of the phenomenon being recorded. We put

$$\|x(t)\| = [\mathcal{E}\{x'(t)x(t)\}]^{1/2},$$

where the prime indicates transposition of the vector. Later we shall have occasion to introduce into the theory linear combinations of the $x(t_j)$ with complex coefficients; z being such a combination, we shall then put

$$\|z\| = [\mathcal{E}\{z^*z\}]^{1/2},$$

the star indicating transposition combined with conjugation. This is called the norm of z.

2. DIFFERENTIATION AND INTEGRATION OF STOCHASTIC PROCESSES

Let x_n, $n = 1, 2, \ldots$, be a sequence of random variables for which $\mathcal{E}(|x_n|^2) < \infty$. Then x_n is said to converge in mean square‡ to a random

† See the Mathematical Appendix at the end of this volume.

‡ The notion of mean-square convergence is discussed further in the Mathematical Appendix at the end of this volume.

variable x if

$$\lim_{n} \mathcal{E}(|x - x_n|^2) = \|x - x_n\|^2 = 0. \tag{2.1}$$

We then write $x_n \to x$. We use the modulus sign here to cover the case in which the random variables x_n are complex-valued, so that each x_n can be considered as a pair of real-valued random variables. The necessary and sufficient condition that an x exist such that (2.1) will hold true is the "Cauchy" condition

$$\lim_{m,n\to\infty} \|x_n - x_m\| = 0. \tag{2.2}$$

The limit x is then uniquely defined in the sense that any two random variables x satisfying (2.1) differ only on a set of measure zero. Moreover, $\mathcal{E}(x\bar{y})$ is a continuous function of x and y, so that if $\|x_n\|$ and $\|y_n\|$ are finite and $x_n \to x$, $y_n \to y$ then $\mathcal{E}(x_n\bar{y}_n) \to \mathcal{E}(x\bar{y})$. (See the Mathematical Appendix.) On the other hand, if $\mathcal{E}(x_n\bar{x}_m)$ converges to a limit independently of the way m and n tend to infinity, the Cauchy condition is satisfied, since $\|x_n - x_m\|^2 = \|x_n\|^2 + \|x_m\|^2 - \mathcal{E}(x_n\bar{x}_m) - \mathcal{E}(x_m\bar{x}_n)$.

We now need to introduce the concepts of differentiation and integration of continuous time processes.

(i) We say that the scalar process $x(t)$ is mean-square differentiable at t if $\delta^{-1}\{x(t + \delta) - x(t)\}$ has a unique limit in mean square as δ approaches zero.

We are now considering a variable tending continuously to zero. However, taking an arbitrary sequence δ_n so that $\delta_n \to 0$ as $n \to \infty$, we see that the following Cauchy condition for mean-square differentiability is necessary:

$$\lim_{\delta_1,\delta_2\to 0} \|\delta_1^{-1}\{x(t + \delta_1) - x(t)\} - \delta_2^{-1}\{x(t + \delta_2) - x(t)\}\| = 0. \tag{2.3}$$

However, this condition is also sufficient, for if δ_n and δ_n' are two sequences converging to zero, we may form from them the combined sequence δ_n'', and (2.3) assures us that $\{x(t + \delta_n'') - x(t)\}/\delta_n''$ also converges in mean square and that the limit is independent of the sequence chosen.

From what was said above concerning convergence in mean square we know that the necessary and sufficient condition for (2.3) to hold is the existence of the limit

$$\begin{aligned}&\lim_{\delta_1,\delta_2\to 0} \mathcal{E}[\delta_1^{-1}\{x(t + \delta_1) - x(t)\}\, \delta_2^{-1}\{x(t + \delta_2) - x(t)\}] \\ &\qquad = \lim_{\delta_1,\delta_2\to 0} \frac{\gamma(t + \delta_1, t + \delta_2) - \gamma(t + \delta_1, t) - \gamma(t, t + \delta_2) + \gamma(t, t)}{\delta_1\delta_2}.\end{aligned}$$

For this, in turn, it is sufficient that $\{\partial^2\gamma(s, t)/\partial s\, \partial t\}$ exist and be continuous (e.g., see Goursat and Hedrick, 1904, p. 13). We know that the covariance function of $\dot{x}(t)$, the mean-square derivative, is

$$\lim_{\delta_1,\delta_2\to 0} \mathcal{E}\{\delta_1^{-1}(x(s+\delta_1)-x(s))\,\delta_2^{-1}(x(t+\delta_2)-x(t))\} = \frac{\partial^2\gamma(s,t)}{\partial s\,\partial t}.$$

In the same way

$$\mathcal{E}(x(s)\,\dot{x}(t)) = \frac{\partial\gamma(s,t)}{\partial t}.$$

Theorem 2. *The necessary and sufficient condition that the scalar process $x(t)$ be mean-square differentiable is the existence of the limit*

$$\lim_{\delta_1,\delta_2\to 0} \frac{\gamma(t+\delta_1, t+\delta_2) - \gamma(t+\delta_1, t) - \gamma(t, t+\delta_2) + \gamma(t,t)}{\delta_1\,\delta_2}.$$

If $\dot{x}(t)$ is the mean-square derivative, this has covariance function $\partial^2\gamma(s,t)/\partial s\,\partial t$ and $\mathcal{E}(x(s)\,\dot{x}(t)) = \partial\gamma(s,t)/\partial t$.

(ii) With regard to integration we first consider the definition of integrals to be represented by symbols of the form†

$$\int_{-\infty}^{\infty} x(t)\, m(dt), \tag{2.4}$$

where m is a σ-finite measure adjusted, let us say, so that the corresponding distribution function is continuous from the right. We consider only $x(t)$ which are mean-square continuous and functions for which

$$\iint_{-\infty}^{\infty} \gamma(s,t)\, m(ds)\, m(dt) < \infty. \tag{2.5}$$

We could accomplish the definition of (2.4) by defining first

$$\int_a^b x(t)\, m(dt) \tag{2.6}$$

by considering approximating sums

$$\sum_1^n x(t_j)\{m((s_{j-1}, s_j])\},$$

wherein the points s_j divide the interval $[a, b]$ into subintervals of length less than ϵ and $t_j \in (s_{j-1}, s_j]$. If this sum converges to a unique limit in mean square, as n increases so that ϵ converges to zero, we call this limit the

† We prefer the notation $m(dt)$ to the common one which replaces that by $dm(t)$ for the obvious reason that it is t that is changing, differentially, not necessarily m.

integral of $x(t)$ with respect to $m(t)$ over $[a, b]$ and indicate it by the symbol (2.6). When $x(t)$ is mean-square continuous, the integral (2.6) will exist. Proof of this is obtained by a completely straightforward modification of the well known proof of the existence of the Riemann-Stieltjes integral of a continuous function over a finite interval. Indeed, to obtain the modification we need only to replace the absolute value in the proof with the norm [regarding $x(t)$ as the continuous function]. The integral (2.4) can now be defined as the limit in mean square of a sequence of integrals of the form (2.6) as $a \to -\infty$, $b \to \infty$. A sufficient condition for the existence (and uniqueness) of this limit is the condition (2.5). Indeed, from the definition of (2.6) it follows that

$$\left\| \int_a^b x(t)\, m(dt) \right\|^2 = \iint_a^b \gamma(s, t)\, m(ds)\, m(dt),$$

so that the Cauchy condition becomes, for example, as $b \to \infty$,

$$\lim_{b_1, b_2 \to \infty} \left\| \int_{b_1}^{b_2} x(t)\, m(dt) \right\|^2 = \lim_{b_1, b_2 \to \infty} \iint_{b_1}^{b_2} \gamma(s, t)\, m(ds)\, m(dt) = 0$$

and (2.5) implies that this converges to zero.

We may state these results as follows:

Theorem 3. *If $x(t)$ is mean-square continuous and $m(t)$ is a σ-finite measure, then* (2.6) *is uniquely defined as a limit in mean square of approximating Riemann-Darboux sums. The necessary and sufficient condition that* (2.4) *exist as a limit in mean square of* (2.6) *as $a \to -\infty$, $b \to \infty$, is given by* (2.5).

Of course, the definition (2.4) can be extended to a much wider class of functions than those that are mean-square continuous in the same way as the definition of the Lebesgue integral can be extended from the Riemann integral definition for continuous functions of compact support to the wide class of Lebesgue integrable functions. For a discussion of this problem we refer the reader to Hille and Phillips (1957, Chapter III). We do not need the extensions in this book.

We also wish to define what we mean by

$$\int_{-\infty}^{\infty} g(t)\, \xi(dt), \tag{2.7}$$

where $g(t)$ is now a complex-valued function of t and $\xi(t)$ is the stochastic process. We have in mind, particularly, the situation in which $\xi(t)$ is a process of orthogonal increments, that is, one for which $\xi(t_1) - \xi(t_2)$ and

$\xi(s_1) - \xi(s_2)$ have zero covariance if the intervals $[t_1, t_2]$ and $[s_1, s_2]$ have no common point. This is one situation, however, in which we do not wish to assume that $\xi(t)$ is mean-square continuous. However, it is evidently true that

$$\mathcal{E}\left([\xi(t) - \xi(t_0)]^2\right) \tag{2.8}$$

increases monotonically with t for $t \geq t_0$. This quantity (2.8) defines a Lebesgue-Stieltjes measure on $(-\infty, \infty)$, the increment over the interval $(t_0, t]$ being just the expression (2.8). We write $F(t) - F(t_0)$ for that expression, keeping especially in mind the situation, often occurring later, in which $\mathcal{E}\{\xi^2(t)\} \leq a < \infty$, so that $F(t)$ is a distribution function. Again we may define (2.7) for the case in which the integral is over the interval $[a, b]$ as the limit in mean square of a sequence of approximating sums

$$\sum g(t_j)\{\xi(s_j) - \xi(s_{j-1})\},$$

if this limit exists and is uniquely defined. This will certainly be so if $g(t)$ is continuous. Again the proof is straightforward. We may then extend the definition to infinite intervals by a further limiting process. However, a more inclusive and direct procedure can be used (see the appendix to this chapter) which enables (2.8) to be defined for any function $g(t)$ that is measurable with respect to the Lebesgue-Stieltjes measure induced by $F(t)$ and for which

$$\int_{-\infty}^{\infty} |g(t)|^2 \, F(dt) < \infty.$$

3. SOME SPECIAL MODELS

(i) An important class of scalar time series is that generated by a linear mechanism of the form

$$\sum_0^q \beta(j)\, x(n-j) = \sum_0^s \alpha(k)\, \epsilon(n-k), \qquad \beta_0 = 1, \tag{3.1}$$

wherein the $\beta(j)$ and $\alpha(k)$ are real, of course, and the $\epsilon(n)$ satisfy

$$\mathcal{E}\left(\epsilon(m)\, \epsilon(n)\right) = \sigma^2 \delta_m{}^n.\dagger$$

Such a mechanism should often be a reasonable approximation to reality, since the idea that $x(n)$ is determined by immediate past values of itself, together with past disturbances, is an appealing one. The linearity is a convenient mathematical fiction only. Experience suggests that a model of the form (3.1) will fit a wide range of data. When $s = 0$, (3.1) is said to be

† We shall always assume that $\epsilon(n)$ has these covariance properties unless we indicate otherwise. Of course $\delta_m{}^n$ is "Kronecker's delta."

an autoregressive relation. When $s > 0$, the terminology "mixed autoregression and moving average" is sometimes used. When $q = 0$, we speak of a (finite) moving average.

Evidently, if $x(n)$ is given q initial values $x(-q), \ldots, x(-1)$, (3.1) serves to define it thereafter in terms of the $\epsilon(n)$. We wish to examine the nature of the formula expressing $x(n)$ in terms of $\epsilon(n), \epsilon(n-1), \ldots$. We consider the homogeneous equation, obtained by replacing the right-hand term of (3.1) with zero, and seek for solutions of the form z^n, which procedure leads us to the characteristic equation

$$\sum_0^q \beta(j)z^{-j} = 0. \tag{3.2}$$

A solution $b_u z_u{}^n$ corresponds to each nonrepeated root z_u. If a root z_u is repeated p_u times, it is easily verified that $b_{u,j}n^j z_u{}^n$ for $j = 0, 1, \ldots, p_u - 1$ are all solutions. If z_u is complex, evidently a conjugate corresponds to each of these solutions. Thus we obtain the general solution of the homogeneous equation as

$$\sum_{u=1}^{r}\sum_{j=0}^{p_u-1} b_{u,j}n^j z_u{}^n, \tag{3.3}$$

where the q constants $b_{u,j}$ are sufficient to satisfy q initial conditions. Of course, we could also write (3.3) as a sum of terms of the form

$$n^j\rho_u{}^n(b'_{u,j}\cos\theta_u n + b''_{u,j}\sin\theta_u n), \tag{3.4}$$

where $z_u = \rho_u \exp i\theta_u$ and $b_{u,j} = \frac{1}{2}(b'_{u,j} - ib''_{u,j})$ if $\theta_u \neq 0, \pi$. (For $\theta_u = 0, \pi$, of course, $b_{u,j} = b'_{u,j}$ and $b''_{u,j} = 0$.)

The $b_{u,j}$ in (3.3) may be chosen so that (3.3) takes any prescribed values at q time points; for example, at the points $n = -q + k$, $k = 0, \ldots, q - 1$. Indeed, if that were not so, the q sequences $n^j z_u{}^n$ would be linearly dependent when evaluated at these q values, which implies that a nonnull polynomial of degree not greater than $q - 1$ has q zeros. This is impossible. Provided no z_u has unit modulus, we may find a sequence $f(n)$ such that $|f(n)|$ converges to zero exponentially as $|n| \to \infty$ and which satisfies the equation

$$\sum_0^q \beta(j)f(n-j) = \delta_0{}^n.$$

Indeed, under this condition $(\sum \beta(j)z^j)^{-1}$ has a valid Laurent expansion in an annulus containing the unit circle in its interior. If

$$\sum_{-\infty}^{\infty} f(n)z^n$$

is this expansion, then, since $\sum \beta(j)z^j \sum f(n)z^n \equiv 1$ for z in the annulus, $f(n)$ satisfies the required relation. It evidently converges to zero exponentially as we have said.

Theorem 4. *If $\epsilon(n)$ is a sequence of random variables with covariance $\mathcal{E}(\epsilon(m)\,\epsilon(n)) = \delta_m{}^n\sigma^2$ and $x(-q), \ldots, x(-1)$ are prescribed initial values, there exists a unique random sequence, $x(n)$, which satisfies* (3.1) *and takes these initial values. If no z_u has unit modulus, this solution is of the form*

$$x(n) = \sum_{u=1}^{r} \sum_{j=0}^{p_u-1} b_{u,j} n^j z_u{}^n + \sum_{k=0}^{s} \alpha(k) \sum_{v=-\infty}^{\infty} f(n-v)\,\epsilon(v-k).$$

Proof. It is evident that $x(n)$ is uniquely defined once the initial values are given in terms of these initial values and the $\epsilon(n)$ sequence; for example, if $\beta(q) \neq 0$, then

$$x(-q-1) = -\beta(q)^{-1} \sum_{1}^{q} \beta(j)x(-j) + \sum_{0}^{s} \alpha(k)\epsilon(-q-1-k)$$

and the $x(n)$ for all $n < -q$ may be defined by an iteration of this process. The sum

$$\sum_{v=-M}^{N} f(n-v)\,\epsilon(v-k)$$

certainly converges in mean square as M, N increase, since the $|f(n)|$ decrease exponentially with $|n|$. Thus the expression for $x(n)$ is well defined. Moreover,

$$\sum_{k=0}^{s} \alpha(k) \sum_{v=-\infty}^{\infty} f(n-v)\,\epsilon(v-k)$$

may be verified to satisfy (3.1). Indeed, we have

$$\sum_{0}^{q} \beta(j) \left\{ \sum_{v=-\infty}^{\infty} f(n-j-v)\,\epsilon(v-k) \right\} = \sum_{v=-\infty}^{\infty} \left\{ \sum_{0}^{q} \beta(j) f(n-v-j) \right\} \epsilon(v-k)$$
$$= \epsilon(n-k),$$

since the term in curly brackets is null for $n - v \neq 0$, whereas for $v = n$ it is unity. Since we know that the $b_{u,j}$ may be chosen to satisfy any q initial conditions, the proof is complete.

Putting

$$\lambda(j) = \sum_{k=0}^{s} \alpha(k) f(j-k),$$

we may rewrite the solution as

$$(3.5) \quad x(n) = \sum_{u} \sum_{j} n^j \rho_u{}^n (b'_{u,j} \cos\theta_u n + b''_{u,j} \sin\theta_u n) + \sum_{-\infty}^{\infty} \lambda(j)\,\epsilon(n-j).$$

For a solution, $x(n)$, of the form

$$\sum_{-\infty}^{\infty} \lambda(j)\,\epsilon(n-j) \tag{3.6}$$

we have, let us say,

$$\mathcal{E}\big(x(n)\,x(m)\big) = \sigma^2 \sum_{-\infty}^{\infty} \lambda(j)\,\lambda(j+|m-n|) = \gamma(m-n).$$

A series $x(n)$ whose covariances depend only on $(m-n)$ is said to be second-order (or wide sense) stationary.†

The most important case of Theorem 4 is that in which, for all u, $\rho_u < 1$, for then $f(n) = 0$, $n < 0$; that is, $\lambda(j) = 0, j < 0$. In this case the stationary solution (3.6) of (3.1) is an infinite moving average of past $\epsilon(n)$. This is the case likely to be of interest, for when we "model" some real world system by means of (3.1) we usually think of the $\epsilon(n)$ as disturbances affecting the system for the first time at time n. In this case also the solution of the homogeneous equation approaches zero as $n \to \infty$, so that all solutions approach the form (3.6).

The case in which, for certain u, $\rho_u = 1$ has also attracted attention in both economics and systems engineering (see Orcutt, 1948; Box and Jenkins, 1962). Let us factor the polynomial $\sum \beta(j) z^{q-j}$ into two factors, the first of which accounts for all the zeros of unit modulus and the second accounts for all remaining zeros. Let the degrees of the polynomials be q_1 and q_2, respectively, so that $q = q_1 + q_2$. Let $\beta'(j)$ and $\beta''(j)$ be the coefficients of the polynomials, with $\beta'(0) = \beta''(0) = 1$. We may find a stationary solution to

$$\sum_{0}^{q_2} \beta''(j)\, y(n-j) = \epsilon(n) \tag{3.7}$$

of the form in (3.6).

The first polynomial may be written in the form

$$\sum_{0}^{q_1} \beta'(j) z^{q-j} = \prod_{u=1}^{q_1} (z - e^{i\theta_u}),$$

where, if $\theta_u \neq 0, \pi$ occurs, then so must $-\theta_u$, since the $\beta'(j)$ are real. Call S_1 the operator that replaces a sequence such as $y(n)$ by

$$y_1(n) = S_1\, y(n) = \sum_{j=-q_1}^{n} y(j) e^{i\theta_1(n-j)}, \qquad n \geq -q_1.$$

† We see that $\gamma(m, n) = \gamma(0, n-m)$ in this stationary case and we have, somewhat confusingly, put $\gamma(n-m) = \gamma(0, n-m)$. A similar notation is used elsewhere below. Of course, in this scalar case we also have $\gamma(n-m) = \gamma(m-n)$.

Then

$$y_1(n) - e^{i\theta_1} y_1(n-1) = y(n), \qquad n \geq -q_1 + 1.$$

Now defining S_u in terms of $\exp i\theta_u$, as for S_1, we may form $y_2(n) = S_2\, y_1(n)$ and see that it satisfies

$$y_2(n) - (e^{i\theta_1} + e^{i\theta_2})\, y_2(n-1) + e^{i(\theta_1+\theta_2)}\, y_2(n-2) = y(n), \quad n \geq -q_1 + 2.$$

Repeating the operation q_1 times, we see that

$$\left(\prod_{u=1}^{q_1} S_u\right) y(n)$$

is a solution of (3.1) for $n \geq 0$. Thus we see that

$$\sum_u \sum_j n^j \rho_u^{\,n}(b'_{u,j} \cos \theta_u n + b''_{u,j} \sin \theta_u n) + \left(\prod_{u=1}^{q_1} S_u\right) y(n), \qquad n \geq 0 \tag{3.8}$$

may, by suitable choice of the $b_{u,j}$, be made to satisfy (3.1) for all $n \geq 0$ and to satisfy, at $n = -q, -q+1, \ldots, -1$, any q initial conditions. There will now be no stationary solution to (3.1). The general situation may be understood from a simple case. We put

$$\Delta x(n-1) = x(n) - x(n-1) \tag{3.9}$$

and, considering

$$\Delta x(n-1) = \epsilon(n),$$

we obtain the solution

$$x(n) = x(-1) + \sum_0^n \epsilon(j), \qquad n \geq 0.$$

For this case we have

$$\mathcal{E}\{(x(n) - x(-1))(x(m) - x(-1))\} = \sigma^2 \min(m, n).$$

Thus the correlation between $x(n)$ and $x(m)$ tends to unity if m and n both increase in such a fashion, for example, that $(n - m)$ remains constant. Thus a realization of this process will show a relatively smooth appearance. For a discussion of the nature of the behavior of these realizations we refer the reader to the second chapter of Cox and Miller (1965).

(ii) We now consider the case in which $x(n)$ is a vector. We write the model as

$$\sum_0^q B(j)\, x(n-j) = \sum_0^s A(k)\, \epsilon(n-k), \qquad B(0) = I_p, \tag{3.10}$$

wherein the $B(j)$ are square matrices,† but we do not restrict the $A(k)$. We assume that the $\epsilon(n)$ satisfy

$$\mathcal{E}(\epsilon(n)\,\epsilon'(m)) = \delta_m{}^n G. \tag{3.11}$$

We use the same terminology (autoregression, moving average, etc.) as in the scalar case.

Now we seek for solutions of the homogeneous system of the form $b(u)\,z_u{}^n$, where $b(u)$ is a vector that satisfies

$$\left[\sum_0^q B(j)\,z_u^{q-j}\right] b(u) = 0$$

and z_u satisfies

$$\det\left[\sum_0^q B(j)\,z_u^{q-j}\right] = 0. \tag{3.12}$$

If z_u is a multiple root, we may be able to find more than one solution $b(u)$ corresponding to the same z_u. We call these $b(i, u)$. However, in general we shall also have to adjoin additional terms of the form $n^j z_u{}^n b(i, u)$. The theory now proceeds along "readymade" lines so that we may state Theorem 4′, the proof of which we have inserted as part of the appendix to this chapter.

Theorem 4′. *If $\epsilon(n)$, $n = 0, \pm 1, \ldots$, is a sequence of random vectors satisfying* (3.11), *then, given q initial values, $x(-q), \ldots, x(-1)$, the solution to* (3.10), *is uniquely defined. If no zero, z_u, of* (3.12) *lies on the unit circle, the solution to* (3.10) *is of the form*

$$x(n) = \sum_u \sum_i \sum_j c(i, j, u)\,z_u{}^n n^j b(i, u) + \sum_{k=-\infty}^{\infty} \Lambda(k)\,\epsilon(n-k),$$

where the elements $\lambda_{ij}(k)$ of the $\Lambda(k)$ converge to zero exponentially as $|k| \to \infty$.

Once again, if, for all μ, $|z_u| < 1$, the first term in the expression for $x(n)$ eventually decays to zero and we obtain a solution of the form

$$\sum_0^\infty \Lambda(j)\,\epsilon(n-j),$$

which every solution of (3.10) then eventually approaches. For this solution we have, let us say,

$$\mathcal{E}(x(m)\,x'(n)) = \sum_0^\infty \Lambda(j)\,G\Lambda'(n-m+j) = \Gamma(n-m).$$

† More general formulations are possible (and important, particularly in economics), but if the relation uniquely determines $x(n)$ in terms of its past and the $\epsilon(n-k)$, the system must be reducible to the form used here.