Wiley Classics Library

ANDERSON

The Statistical Analysis of Time Series

The Statistical Analysis of Time Series

T. W. ANDERSON
Professor of Statistics and Economics
Stanford University

Wiley Classics Library Edition Published 1994

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore

This text is printed on acid-free paper.

Wiley Classics Library Edition Published 1994.

This work was supported by the Logistics and Mathematical Statistics Branch of the Office of Naval Research under Contracts Nonr-3279(00), NR 042-219, Nonr-266(33), NR 042-034; Nonr-4195(00), NR 042-233; Nonr-4259(08), NR 042-034; Nonr-225(53), NR 042-002; N 00014-67-A-0012-0030, NR 042-034.

ISBN 0-471-04745-7

Library of Congress Catalog Number 70-126222

Printed in the United States of America

10 9 8 7 6 5 4

Preface

In writing a book on the statistical analysis of time series an author has a choice of points of view. My selection is the mathematical theory of statistical inference concerning probabilistic models that are assumed to generate observed time series. The probability model may involve a deterministic trend and a random part constituting a stationary stochastic process; the statistical problems treated have to do with aspects of such trends and processes. Where possible, the problem is posed as one of finding an optimum procedure and such procedures are derived. The statistical properties of the various methods are studied; in many cases they can be developed only in terms of large samples, that is, on information from series observed a long time. In general these properties are derived on a rigorous mathematical basis.

While the theory is developed under appropriate mathematical assumptions, the methods may be used where these assumptions are not strictly satisfied. It can be expected that in many cases the properties of the procedures will hold approximately. In any event the precisely stated results of the theorems give some guidelines for the use of the procedures. Some examples of the application of the methods are given, and the uses, computational approaches, and interpretations are discussed, but there is no attempt to put the methods in the form of programs for computers.

This book grew out of a graduate course that I gave for many years at Columbia University, usually for one semester and occasionally for two semesters. By now the material included in the book cannot be covered completely in a two-semester course; an instructor using this book as a text will select the material that he feels most interesting and important. Many exercises are given. Some of these are applications of the methods described; some of the problems are to work out special cases of the general theory; some of the exercises fill in details in complicated proofs; and some extend the theory.

Besides serving as a text book I hope this book will furnish a means by which statisticians and other persons can learn about time series analysis without resort to a formal course. Reading this book and doing selected exercises will lead to a considerable knowledge of statistical methodology useful for the analysis of time series. This book may also serve for reference. Much material which has not been assembled together before is presented here in a fairly coherent fashion. Some new theorems and methods are presented. In other cases the assumptions of previously stated propositions have been weakened and conclusions strengthened.

Since the area of time series analysis is so wide, an author must select the topics he will include. I have described in the introduction (Chapter 1) the material included as well as the limitations, and the Table of Contents also gives an indication. It is hoped that the more basic and important topics are treated here, though to some extent the coverage is a matter of taste. New methods are constantly being introduced and points of view are changing; the results here can hardly be definitive. In fact, some material included may at the present time be rather of historical interest.

In view of the length of this book a few words of advice to readers and instructors may be useful in selecting material to study and teach. Chapter 2 is a self-contained summary of the methods of least squares; it may be largely redundant for many statisticians. Chapters 3 and 4 deal with models with independent random terms (known sometimes as "errors in variables"); some ideas and analysis are introduced which are used later, but the reader interested mainly in the later chapters can pass over a good deal (including much of Sections 3.4, 4.3, and 4.4). Autoregressive processes, which are useful in applications and which illustrate stationary stochastic processes, are treated in Chapter 5; Sections 5.5 and 5.6 on large-sample theory contain relevant theorems, but the proofs involve considerable detail and can be omitted. Statistical inference for these models is basic to analysis of stationary processes "in the time domain." Chapter 6 is an extensive study of serial correlation and tests of independence; Sections 6.3 and 6.4 are primarily of theoretical statistical interest; Section 6.5 develops the algebra of quadratic forms and ratios of them; distributions, moments, and approximate distributions are obtained in Sections 6.7 and 6.8, and tables of significance points are given for tests. The first five sections of Chapter 7 constitute an introduction to stationary stochastic processes and their spectral distribution functions and densities. Chapter 8 develops the theory of statistics pertaining to stationary stochastic processes. Estimation of the spectral density is treated in Chapter 9; it forms the basis of analyzing stationary processes "in the frequency domain." Section 10.2 extends regression analysis (Chapter 2) to stationary random terms; Section

10.3 extends Chapters 8 and 9 to this case; and Section 10.4 extends Chapter 6 to the case of residuals from fitted trends. Parts of the book that constitute units which may be read somewhat independently from other parts are (i) Chapter 2, (ii) Chapters 3 and 4, (iii) Chapter 5, (iv) Chapter 6, (v) Chapter 7, and (vi) Chapters 8 and 9.

The statistical analysis of time series in practical applications will also invoke less formal techniques (which are now sometimes called "data analysis"). A graphical presentation of an observed time series contributes to understanding the phenomenon. Transformations of the measurement and relations to other data may be useful. The rather precisely stated procedures studied in this book will not usually be used in isolation and may be adapted for various situations. However, in order to investigate statistical methods rigorously within a mathematical framework some aspects of the analysis are formalized. For instance, the determination of whether an effect is large enough to be important is sometimes formalized as testing the hypothesis that a parameter is 0.

The level of this book is roughly that of my earlier book, *An Introduction to Multivariate Statistical Analysis*. Some knowledge of matrix algebra is needed. (The necessary material is given in the appendix of my earlier book; additional results are developed in the text and exercises of this present book.) A general knowledge of statistical methodology is useful; in particular, the reader is expected to know the standard material of univariate analysis such as t-tests and F-statistics, the multivariate normal distribution, and the elementary ideas of estimation and testing hypotheses. Some more sophisticated theory of testing hypotheses, estimation, and decision theory that is referred to is developed in the exercises. [The reader is referred to Lehmann (1959) for a detailed and rigorous treatment of testing hypotheses.] A moderate knowledge of advanced calculus is assumed. Although real-valued time series are treated, it is sometimes convenient to write expressions in terms of complex variables; actually the theory of complex variables is not used beyond the simple fact that $e^{i\theta} = \cos\theta + i\sin\theta$ (except for one problem). Probability theory is used to the extent of characteristic functions and some basic limit theorems. The theory of stochastic processes is developed to the extent that it is needed.

As noted above, there are many problems posed at the end of each chapter except the first which is the introduction. Solutions to these problems have been prepared by Paul Shaman. Solutions which are referred to in the text or which demonstrate some particularly important point are printed in Appendix B of this book. Solutions to most other problems (except solutions that are straightforward and easy) are contained in a Solutions Manual which is issued as a separate booklet. This booklet is available free of charge by writing to the publisher.

I owe a great debt of gratitude to Paul Shaman for many contributions to this book in matters of exposition, selection of material, suggestions of references and problems, improvements of proofs and exposition, and corrections of errors of every magnitude. He has read my manuscript in many versions and drafts. The conventional statement that an acknowledged reader of a manuscript is not responsible for any errors in the publication I feel is usually superfluous because it is obvious that anyone kind enough to look at a manuscript assumes no such responsibility. Here such a disclaimer may be called for simply because Paul Shaman corrected so many errors that it is hard to believe any remain. However, I admit that in this material it is easy to generate errors and the reader should throw the blame on the author for any he finds (as well as inform him of them).

My appreciation also goes to David Hinkley, Takamitsu Sawa, and George Styan, who read all or substantial parts of the manuscript and proofs and assisted with the preparation of the bibliography and index. There are many other colleagues and students to thank for assistance of various kinds. They include Selwyn Gallot, Joseph Gastwirth,Vernon Johns, Ted Matthes, Emily Stong Myers, Emanuel Parzen, Lloyd Rosenberg, Ester Samuel, and Morris Walker as well as Anupam Basu, Nancy David, Ronald Glaser, Elizabeth Hinkley, Raul Mentz, Fred Nold, Arthur V. Peterson, Jr., Cheryl Schiffman, Kenneth Thompson, Roger Ward, Larry Weldon, and Owen Whitby. No doubt I have forgotten others. I also wish to thank J. M. Craddock, C. W. J. Granger, M. G. Kendall, A. Stuart, and Herman Wold for use of some material.

In preparing the typescript my greatest debt is to Pamela Oline Gerstman, my secretary for four years, who patiently went through innumerable drafts and revisions. (Among other tribulations her office was used as headquarters for the "liberators" of Fayerweather Hall in the spring of 1968.) The manuscript also bore the imprints of Helen Bellows, Shauneen Nelson, Carol Andermann Novak, Katherine Cane, Carol Hallett Robbins, Alexandra Mills, Susan Parry, Noreen Browne Ettl, Sandi Hochler Frost, Judi Campbell,and Polly Bixler.

An important factor in my writing this book was the sustained financial support of the Office of Naval Research over a period of some ten years. The Logistics and Mathematical Statistics Branch has been helpful, encouraging, accommodating and patient in its sponsorship.

Stanford University
Stanford, California
February, 1970

T. W. ANDERSON

Contents

CHAPTER 1

Introduction

A time series is a sequence of observations, usually ordered in time, although in some cases the ordering may be according to another dimension. The feature of time series analysis which distinguishes it from other statistical analyses is the explicit recognition of the importance of the order in which the observations are made. While in many problems the observations are statistically independent, in time series successive observations may be dependent, and the dependence may depend on the positions in the sequence. The nature of a series and the structure of its generating process may also involve in other ways the sequence in which the observations are taken.

In almost every area there are phenomena whose development and variation with the passing of time are of interest and importance. In daily life one is interested in aspects of the weather, in prices that one pays, and in features of one's health; these change in time. There are characteristics of a nation, affecting many individuals, such as economic conditions and population, that evolve and fluctuate over time. The activity of a business, the condition of an industrial process, the level of a person's sleep, and the reception of a television program vary chronologically. The measurement of some particular characteristic over a period of time constitutes a time series. It may be an hourly record of temperature at a given place or the annual rainfall at a meteorological station. It may be a quarterly record of gross national product; an electrocardiogram may compose several time series.

There are various purposes for using time series. The objective may be the prediction of the future based on knowledge of the past; the goal may be the control of the process producing the series; it may be to obtain an understanding of the mechanism generating the series; or simply a succinct description of the salient features of the series may be desired. As statisticians we shall be interested in statistical inference; on the basis of a limited amount of information, a time series of finite length, we wish to make inferences about the probabilistic mechanism that produced the series; we want to analyze the underlying structure.

In principle the measurement of many quantities, such as temperature and voltage, can be made continuously and sometimes is recorded continuously in the form of a graph. In practice, however, the measurements are often made discretely in time; in other cases, such as the annual yield of grain per acre, the measurement must be made at definite intervals of time. Even if the data are recorded continuously in time only the values at discrete intervals can be used for digital computations. In this book we shall confine ourselves to time series that are recorded discretely in time, that is, at regular intervals, such as barometric pressure recorded each hour on the hour. Although the effect of one quantity on another and the interaction of several characteristics over time are often of consequence, in many studies a great deal of knowledge may be gained by the investigation of a single time series; this book (except with respect to autoregressive systems) is concerned with statistical methods for analyzing a univariate time series, that is, one type of measurement made repeatedly on the same object or individual. We shall, furthermore, suppose that the measurement is a real number, such as temperature, which is not limited to a finite (or denumerable) number of values; such a measurement is often called a continuous variable. Some measurements we treat mathematically as if they were continuous in time; for example, annual national income can at best be measured to the nearest penny, but the number of values that this quantity can take on is so large that there is no serious slight to reality in considering the variable as continuous. Moreover, we shall consider series which are rather stable, that is, ones which tend to stay within certain bounds or at least are changing slowly, not explosively or abruptly; we would include many meteorological variables, for instance, but would exclude shock waves.

Let an observed time series be $y_1, y_2, \ldots, y_T$. The notation means that we have T numbers, which are observations on some variable made at T equally distant time points, which for convenience we label $1, 2, \ldots, T$. A fairly general mathematical, statistical, or probabilistic model for the time series can be written as follows:

$$y_t = f(t) + u_t, \qquad t = 1, 2, \ldots, T. \tag{1}$$

The observed series is considered as made up of a completely determined sequence $\{f(t)\}$, which may be called the systematic part, and of a random or stochastic sequence $\{u_t\}$, which obeys a probability law. (Signal and noise are sometimes used for these two components.) These two components of the observed series are not observable; they are theoretical quantities. For example, if the measurements are daily rainfall, the $f(t)$ might be the climatic norm, which is the long-run average over many years, and the u_t would represent those vagaries and irregularities in the weather that describe fluctuations from the climatic norm. Exactly what the decomposition means depends not only on the data, but, in part, depends on what is thought of as repetitions of the experiment giving rise to the data. The interpretation of the random part made here is the so-called "frequency" interpretation. In principle one can repeat the entire situation, obtaining a new set of observations; the $f(t)$ would be the same as before, but the random terms would be different, as a new realization of the stochastic process. The random terms may include errors of observation. [In effect $f(t) = \mathscr{E} y_t$.]

We have some intuitive ideas of what time should mean in such a model or process. One notion is that time proceeds progressively in one direction. Another is that events which are close together in time should be relatively highly related and happenings farther apart in time should not be as strongly related. The effect of time in the mathematical model (1) can be inserted into specifications of the function or sequence $f(t)$; it can be put into the formulation of the probability process that defines the random term u_t; or it can be put into both components. The first part of this book will be devoted to time series represented by "error" models, in which the observations are considered to be independent random deviations from some function representing trend. In the second part we shall be concerned with sequences of dependent random variables, in general stationary stochastic processes with particular emphasis on autoregressive processes. Finally, we shall treat models in which there is a trend and the random terms constitute a stationary stochastic process. Stationary stochastic processes are explained in Chapter 7.

In many cases the model may be completely specified except for a finite number of parameters; in such a case the problems of statistical inference concern these parameters. In other cases the model may be more loosely defined and the corresponding methods are nonparametric. The model is to represent the mechanism generating the relevant series reasonably well, but as a mathematical abstraction the model is only an approximation to reality. How precisely the model can be determined depends on the state of knowledge about the process being studied, and, correspondingly, the information that can be supplied by statistical analysis depends on this state of knowledge. In this book many methods and their properties will be described, but, of course, these are only a selection from the many methods which are useful and available. Here the emphasis is on statistical inference and its mathematical basis.

The early development of time series analysis was based on models in which the effect of time was made in the systematic part, but not in the random part.

For convenience this case might be termed the classical model, because in a way it goes back to the time when Gauss and others developed least squares theory and methods for use in astronomy and physics. In this case we assume that the random part does not show any effect of time. More specifically, we assume that the mathematical expectation (that is, the mean value) is zero, that the variance is constant, and that the u_t are uncorrelated at different points in time. These specifications essentially force any effects of time to be made in the systematic part $f(t)$. The sequence $f(t)$ may depend on unknown coefficients and known quantities which vary over time. Then $f(t)$ is a "regression function." Methods of inference for the coefficients in a regression function are useful in many branches of statistics. The cases which are peculiar to time series analysis are those cases in which the quantities varying over time are known functions of t.

Within the limitations set out we may distinguish two types of sequences in time, $f(t)$. One is a slowly moving function of time, which is often called a trend, particularly by economists, and is exemplified by a polynomial of fairly low degree. Another type of sequence is cyclical; this is exemplified by a finite Fourier series, which is a finite sum of pairs of sine and cosine terms. A pair may be $\alpha \cos \lambda t + \beta \sin \lambda t$ $(0 < \lambda < \pi)$, which can also be written as a cosine function, say $\rho \cos (\lambda t - \theta)$. The period of such a function of time is $2\pi/\lambda$; that is, the function repeats itself after t has gone this amount; the frequency is the reciprocal of the period, namely $\lambda/(2\pi)$. The coefficient $\rho = \sqrt{\alpha^2 + \beta^2}$ is the amplitude and θ is the phase. The observed series is considered to be the sum of such a series $f(t)$ and a random term. Figure 1.1 presents $y_t = 5 + 2 \cos 2\pi t/6 + \sin 2\pi t/6 + u_t$, where u_t is normally distributed with mean 0 and variance 1. [The function $f(t)$ is drawn as a function of a continuous variable t.] The successive values of y_t are scattered randomly above and below $f(t)$. If we know this curve and the error distribution, information about $y_1, \ldots, y_{t-1}$ gives us no help in predicting y_t; the plot of $f(s)$ for $s > t - 1$ does not depend on $y_1, \ldots, y_{t-1}$.

Such a model may be appropriate in certain physical or economic problems. In astronomy, for example, $f(t)$ might be one coordinate in space of a certain planet at time t. Because telescopes are not perfect, and because of atmospheric variations, the observation of this coordinate at time t will have a small error. This error of observation does not affect later positions of the planet nor our observations of them. In the case of a freely swinging pendulum the displacement of the pendulum is a trigonometric function $\rho \cos (\lambda t - \theta)$ when measured from its lowest point.

One general model with the effect of time represented in the random part is a

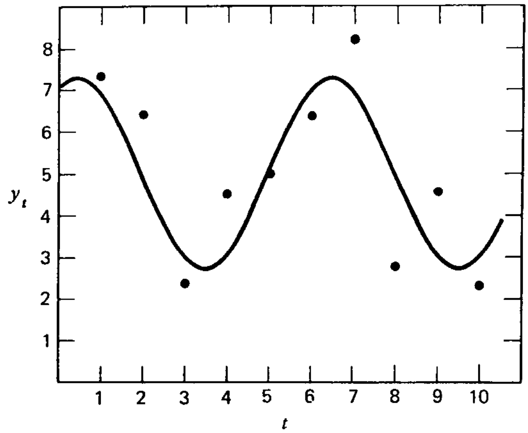

Figure 1.1. Series with a trigonometric trend.

stationary stochastic process; we can illustrate this with an autoregressive process. Suppose y_1 has some distribution with mean 0; let y_1 and y_2 have the joint distribution of y_1 and $\rho y_1 + u_2$, where u_2 is distributed with mean 0, independently of y_1. We write $y_2 = \rho y_1 + u_2$. Define in turn the joint distribution of $y_1, y_2, \ldots, y_{t-1}, y_t$ as the joint distribution of $y_1, y_2, \ldots, y_{t-1}$, $\rho y_{t-1} + u_t$, where u_t is distributed with mean 0, independently of $y_1, \ldots, y_{t-1}$, $t = 3, 4, \ldots$. When the (marginal) distributions of $u_2, u_3, \ldots$ are identical and the distribution of y_1 is chosen suitably, $\{y_t\}$ is a stationary stochastic process, in fact, an autoregressive process, and

$$y_t = \rho y_{t-1} + u_t \tag{2}$$

is a stochastic difference equation of first order. This construction is illustrated in Figure 1.2 for $\rho = \frac{1}{2}$. In this model the "disturbance" u_t has an effect on y_t and all later y_r's. It follows from the construction that the conditional expectation of y_t, given $y_1, \ldots, y_{t-1}$, is

$$\mathscr{E}(y_t \mid y_1, \ldots, y_{t-1}) = \rho y_{t-1}. \tag{3}$$

(In fact, for a first-order process y_t and $y_{t-2}, \ldots, y_1$ are conditionally independent given y_{t-1}.) If we want to predict y_t from $y_1, \ldots, y_{t-1}$ and know the parameter ρ, our best prediction (in the sense of minimizing the mean square error) is ρy_{t-1}; in this model knowledge of earlier observations assists in predicting y_t.

An autoregressive process of second order is obtained by taking the joint

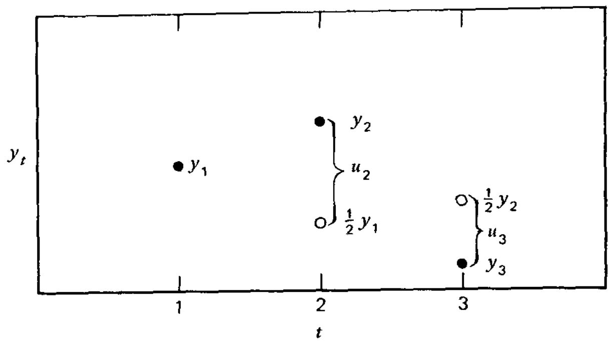

Figure 1.2. Construction of an autoregressive series.

distribution of $y_1, y_2, \ldots, y_{t-1}, y_t$ as the joint distribution of $y_1, y_2, \ldots, y_{t-1}$, $\rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t$, where u_t is independent of $y_1, y_2, \ldots, y_{t-1}$, $t = 3, 4, \ldots$, and the distribution of y_1 and y_2 is suitably chosen; graphs of such series are given in Appendix A.2. Graphs of other randomly generated series are given by Kendall and Stuart (1966), Chapter 45, and Wold (1965), Chapter 1. The variable y_t may represent the displacement of a swinging pendulum when it is subjected to random "shocks" or pushes, u_t. The values of y_t tend to be a trigonometric function $\rho \cos(\lambda t - \theta)$ but with varying amplitude, varying frequency, and varying phase. An autoregressive process of order 4, generated by $y_t = \sum_{s=1}^{4} \rho_s y_{t-s} + u_t$, tends to be like the sum of two trigonometric functions with varying amplitudes, frequencies, and phases.

A general stationary stochastic process can be approximated by an autoregressive process of sufficiently high order, or it can be approximated by a process

$$\sum_{j=1}^{q} (A_j \cos \lambda_j t + B_j \sin \lambda_j t), \tag{4}$$

where $A_1, B_1, \ldots, A_q, B_q$ are independently distributed with $\mathscr{E}A_j = \mathscr{E}B_j = 0$ and $\mathscr{E}A_j^2 = \mathscr{E}B_j^2 = \phi(\lambda_j)$. The process is the sum of q trigonometric functions, whose amplitudes and phases are random variables. On the average the importance of the trigonometric function with frequency $\lambda_j/(2\pi)$ is proportional to the expectation of its squared amplitude, which is $2\phi(\lambda_j)$. In these terms a stationary stochastic process (in a certain class) may be characterized by a *spectral density* $f(\lambda)$ such that $\int_a^b f(\lambda)\, d\lambda$ is approximated by the sum of $\phi(\lambda_j)$ over λ_j such that $a \leq \lambda_j < b$. A feature of stationary stochastic processes is that the covariance $\mathscr{E}(y_t - \mathscr{E}y_t)(y_s - \mathscr{E}y_s)$ only depends on the difference in time $|t - s|$ and hence can be denoted by $\sigma(t - s)$. The covariance sequence and the spectral density (when it exists) are alternative ways of describing the