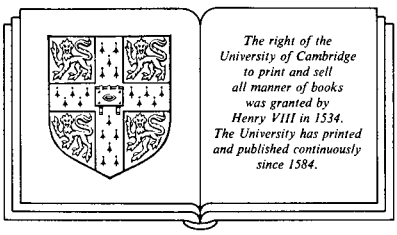
Time-series analysis

A comprehensive introduction for social scientists

Time-series analysis

A comprehensive introduction for social scientists

JOHN M. GOTTMAN
University of Illinois



CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York *New Rochelle*
Melbourne *Sydney*

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521103367

First published 1981
Reprinted 1984, 1988
This digitally printed version 2009

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Gottman, John Mordechai

Time-series analysis: a comprehensive introduction
for social scientists

Bibliography: p.
Includes index
1. Time-series analysis I. Title
HA 30.3.G67 519.5′5′024301 80-25644

ISBN 978-0-521-23597-6 hardback
ISBN 978-0-521-10336-7 paperback

To Robert E. Bohrer

Oh as I was young and easy in the mercy of his means,
Time held me green and dying
Though I sang in my chains like the sea.

Fern Hill, Dylan Thomas

Contents

Preface

Consider the wind on the waters, the panorama of growth, change, development, evolution, or sudden alteration. We are immersed in a universe of natural events that vary with time. The study of all these things is the subject matter of this book. Despite the ubiquitousness and familiarity of these events, we have to learn how to think of processes that occur in time, because the traditional statistics we have learned ignores the dimension of time.

Time-series analysis is the study of variation across some discrete or continuous dimension, which is usually called "time." There are two major approaches to the study of time-series processes, *time-domain* and *frequency-domain analyses*. The first section of this book consists of six chapters that have a minimum of equations. They are designed to introduce the language and the concepts of *both* domains of time-series analysis. Although time and frequency domains are mathematically equivalent, they are, nonetheless, different metaphors for thinking about data. You will need experience with both of these metaphors for them to become useful. Furthermore, it is valuable to be able to switch from one domain to the other.

The book begins with seven chapters that are a nonmathematical encouragement to learn time-series analysis. Concepts are introduced without the mathematics so some readers who ought to be attracted by the subject matter are not frightened away. The first three chapters discuss cycles, because they are part of the most intuitively natural and yet most difficult metaphor of time-series analysis. Chapter 4 is a discussion of the bivariate time-series case. Chapter 5 is a discussion of autoregressive models and forecasting; this is an introduction to the time domain. Chapter 6 is a discussion of experimentation and levels of causal inference, moving from the time to the frequency domains in the discussion.

Just as in classical statistics, in time-series analysis we are interested in making inferences from a sample of data points to something else; in the case of time-series analysis the inference is to the process that may have generated the sample. This leads to the second section of the book, time-series models. The book thus turns to a discussion of the need for models and the

problem of correlated data in Chapter 7; Chapter 8 is a discussion of the basic assumptions of time-series analysis–what is known as stationarity. Chapter 9 takes up the problem of what to do if the data are not stationary, and suggests the use of general models with deterministic and stochastic components.

Part III focuses on stationary time-domain models. Chapter 10 introduces the moving-average models. Chapter 11 begins the discussion of autoregressive models. This discussion continues in Chapter 12, which introduces the second-order autoregressive model, a model that is capable of simulating probabilistic periodicity. In the next two chapters the duality of autoregressive and moving-average models is revealed.

Part IV focuses on stationary frequency-domain models. Chapter 15 begins by discussing the spectral density function and its colorful history; the periodogram is introduced in Chapter 16; the spectral windows are discussed in Chapter 17; and in Chapter 18 an attempt is made to integrate time and frequency domains by a discussion of the spectral densities of moving-average and autoregressive processes, thus explaining the Slutzky effect.

Part V is concerned with estimating the parameters of time-domain models. Chapter 19 is a discussion of the Mann–Wald least-squares autoregressive model-fitting procedure; this procedure is compared to Box–Jenkins models and Wu–Pandit models. Chapter 20 is a more complete discussion of the Box–Jenkins ARIMA models. Most of Part V argues the premise that most users of time-domain methods would do well to use least-squares linear autoregressive models, even if they are not as parsimonious as Box–Jenkins or Wu–Pandit models. Autoregressive models are more familiar models and easier computationally. However, for the sake of thoroughness, all three methods are reviewed. The next two chapters are more applied. Chapter 21 discusses least-squares forecasting, and Chapter 22 presents a worked example of model fitting. This chapter also discusses methods of time-domain and frequency-domain model fitting without the use of a computer.

Throughout the book, primers are included in chapter appendixes that are designed to assist the reader who is unfamiliar with areas such as regression, matrix algebra, the full general linear model, and Fourier series.

Part VI is called "Bivariate Time-Series Analysis." It includes Chapter 23 on bivariate frequency-domain analysis, with a worked example on mother–infant play in Chapter 24. Chapter 25 is about bivariate time-domain analysis; it also includes a worked example, once again on mother–infant play. Chapter 26 is about the interrupted time-series experiment, and Chapter 27 is about multivariate extensions, including time-series regression analysis.

This work is a practical handbook that tries to convey a healthy respect for empirical common sense. An additional goal, however, is to influence the

way social scientists think about problems, stressing the facts that much can be gained by searching for patterns across time and the way these patterns themselves change, and that stochastic rather than purely deterministic models may be better at representing data.

Designed to be everyone's first time-series book and to invite use of these techniques, the work will suit many fine scholars who could benefit from a familiarity with time-series analysis but who would be kept away by all the equations that are necessary to discuss the subject. If I could, I would materialize like a genie from the pages and relax the fear away from these scholars whenever it arises, and coax them back to the text. Instead, what I have chosen to do is to make it possible to read the book at several levels and at several speeds. The chapters are usually short, so that only a few concepts are introduced at a time. Although the effect may be a choppy book, completing chapters always gives me a sense of accomplishment as a reader. I have tried to avoid making the reader wade through long presentations that lose the point. There is a lot of prose, and the concepts are presented for both the mathematical and nonmathematical reader. The book should thus serve as an introduction for both kinds of readers.

This work would not have been possible without the release time provided by a National Institute of Mental Health Research Scientist Development Award 1K02MH00257 and, previous to that, by a University of Illinois Faculty Study in a Second Discipline fellowship. The Second Discipline fellowship made it possible for me to spend a year focusing on time-series analysis in the mathematics department at the University of Illinois.

I am indebted to various colleagues and teachers over the last ten years: Robert Bohrer, Richard Devor, Gene Glass, Robert Kuhn, James Ringland, and Takamitsu Sawa. In particular, I wish to acknowledge the superb tutoring I have received from James Ringland over the last two years. James Ringland is a gifted teacher, a talented statistician, and a valuable colleague. I am indebted to Duane Steidinger, who worked in my laboratory from 1977 to 1979 as a computer programmer and research associate. I have also benefited from conversations with my colleagues Robert Bohrer, Steve Porges, and Jim Sackett. I would like to express my appreciation to those who have shared their data with me for use in this book, in particular Professors Gene Glass, Robert Levenson, Steve Porges, and Edward Tronick. In some cases data from Glass, Willson, and Gottman (1975) were used. I wish to acknowledge the feedback I received from the members of a class I taught at the University of Illinois, particularly the editorial comments of Ms. Janet Kruse. Merle Thorne was an extremely helpful, intelligent typist of the last several drafts of the manuscript. Most important, this work has thrived from the support I have received from my department at Illinois. For the magnificent intellectual climate that my colleagues at Illinois provide, I am grateful.

April 1981

J. M. G.