INTRODUCTION TO Time Series Analysis and Forecasting

WITH APPLICATIONS OF SAS AND SPSS

ROBERT YAFFEE
with Monnie McGee

C33
Y12

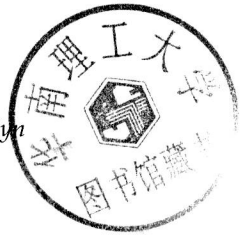
Introduction to Time Series Analysis and Forecasting

with Applications of SAS and SPSS

Robert A. Yaffee
Statistics and Social Science Group
Academic Computing Service of the Information Technology Services
New York University
New York, New York
and
Division of Geriatric Psychiatry
State University of New York Health Science Center at Brooklyn
Brooklyn, NY

with

Monnie McGee
Hunter College
City University of New York
New York, New York



ACADEMIC PRESS, INC.
San Diego London Boston New York Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Copyright © 2000 by ACADEMIC PRESS

All Rights Reserved.
No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Requests for permission to make copies of any part of the work should be mailed to:
Permissions Department, Harcourt, Inc., 6277 Sea Harbor Drive,
Orlando, Florida 32887-6777

Academic Press
A Harcourt Science and Technology Company
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

Academic Press
Harcourt Place, 32 Jamestown Road, London NW1 7BY, UK
http://www.hbuk.co.uk/ap/

Library of Congress Catalog Card Number: 99-62662

International Standard Book Number: 0-12-767870-0

PRINTED IN THE UNITED STATES OF AMERICA
00 01 02 03 04 05 MM 9 8 7 6 5 4 3 2 1

Introduction to

Time Series Analysis and Forecasting

with Applications of SAS and SPSS

For
Liz and Mike

此为试读,需要完整PDF请访问: www.ertongbook.com

Preface

This book is the product of an intellectual odyssey in search of an understanding of historical truth in culture, society, and politics, and the scenarios likely to unfold from them. The quest for this understanding of reality and its potential is not always easy. Those who fail to understand history will not fully understand the current situation. If they do not understand their current situation, they will be unable to take advantage of its latent opportunities or to sidestep the emergent snares hidden within it. George Santayana appreciated the dangers inherent in this ignorance when he said, "Those who fail to learn from history are doomed to repeat it." Kierkegaard lemented that history is replete with examples of men condemned to live life forward while only understanding it backward. Even if, as Nobel laureate Neils Bohr once remarked, "Prediction is difficult, especially of the future," many great pundits and leaders emphasized the real need to understand the past and how to forecast from it. Winston Churchill, with an intuitive understanding of extrapolation, remarked that "the farther back you can look, the farther forward you can see."

Tragic tales abound where vital policies failed because decision makers did not fathom the historical background—with its flow of cultural forces, demographic resources, social forces, economic processes, political processes—of a problem for which they had to make policy. Too often lives were lost or runied for lack of adequate diplomatic, military, political, or economic intelligence and understanding. Obversely, policies succeeded in accomplishing vital objectives because policy makers have understood the likely scenarios of events. After we learned from the past, we needed to study and understand the current situation to appreciate its future possibilities and probabilities. Indeed, the journalistic and scientific quest for "what is" may reveal the outlines of "what can be." The qualitative investigation of "what has been" and "what is" may be the mere beginning of this quest.

The principal objective of this textbook is to introduce the reader to the

fundamental approaches to time series analysis and forecasting. Although the book explores the basic nature of a time series, it presumes that the reader has an understanding of the methodology of measurement and scale construction. In case there are missing data, the book briefly addresses the imputation of missing data. For the most part, the book assumes that there are not significant amounts of missing data in the series and that any missing data have been properly replaced or imputed. Designed for the advanced undergraduate or the beginning graduate student, this text examines the principal approaches to the analysis of time series processes and their forecasting. In simple and clear language, it explains moving average, exponential smoothing, decomposition (Census X-11 plus comments on Census X-12), ARIMA, intervention, transfer function, regression, error correction, and autoregressive error models. These models are generally used for analysis of historical, recent, current, or simulated data with a view toward forecasting. The book also examines evaluation of models, forecasts, and their combinations. Thus, the text attempts to discuss the basic approaches to time series analysis and forecasting.

Another objective of this text is to explain and demonstrate novel theoretical features and their applications. Some of the relatively new features include treatment of Y2K problem circumventions, Census X-12, different transfer function modeling strategies, a scenario analysis, an application of different forecast combination methods, and an analysis of sample size requirements for different models. Although Census X-12 is not yet part of either statistical package, its principal features are discussed because it is being used by governments as the standard method of deseasonalization. In fact, SAS is planning on implementing PROC X12 in a forthcoming version. When dealing with transfer function models, both the conventional Box–Jenkins–Tiao and the linear transfer function approaches are presented. The newer approach, which does not use prewhitening, is more amenable to more complex, multiple input models. In the chapter on event impact or intervention analysis, an approach is taken that compared the impact of an intervention with what would have happened if all things remained the same. A "what if" baseline is posited against which the impact is measured and modeled. The book also briefly addresses cointegration and error correction models, which embed both long-run and short-run changes in the same model. In the penultimate chapter, the evaluation and comparison of models and forecasts are discussed. Attention is paid to the relative advantages and disadvantages of the application of one approach over another under different conditions. This section is especially important in view of the discovery in some of the forecast competitions that the more complex models do not always provide the best forecasts. The methods as well as the relative advantages and disadvantages of combining forecasts

to improve forecast accuracy are also analyzed. Finally, to dispel erroneous conventional wisdom concerning sample size, the final chapter empirically examines the selection of the proper sample size for different types of analysis. In so doing, Monnie McGee makes a scholarly methodological contribution to the study of sample size required for time series tests to attain a power of 0.80, an approach to the subject of power of time series tests that has not received sufficient discussion in the literature until now.

As theory and modeling are explained, the text shows how popular statistical programs, using recent and historical data are prepared to perform the time series analysis and forecasting. The statistical packages used in this book—namely, the Statistical Analysis System (SAS) and the Statistical Package for the Social Sciences (SPSS)—are arguably the most popular general purpose statistical packages among university students in the social or natural sciences today. An understanding of theory is necessary for their proper application under varying circumstances. Therefore, after explaining the statistical theory, along with basic preprocessing commands, I present computer program examples that apply either or both of the SAS Econometric Time Series (SAS/ETS) module or the SPSS Trends module. The programming syntax, instead of the graphic interfaces, of the packages is presented because the use of this syntax tends to remain constant over time while the graphical interfaces of the statistical packages change frequently. In the presentation of data, the real data are first graphed. Because graphical display can be critical to understanding the nature of the series, graphs of the data (especially the SAS Graphs) are elaborately programmed to produce high-resolution graphical output. The data are culled from areas of public opinion research, policy analysis, political science, economics, sociology, and even astronomy and occasionally come from areas of great historical, social, economic, or political importance during the period of time analyzed. The graphs include not only the historical data; after Chapter 7 explains forecasting, they also include forecasts and their profiles. SAS and SPSS computer programs, along with their data, are posted on the Academic Press Web site (http://www.academicpress.com/sbe/authors) to assist instructors in teaching and students in learning this subject matter. Students may run these programs and examine the output for themselves. Through their application of these time series programming techniques, they can enhance their capabilities in the quest for understanding the past, the present, and to a limited extent, *que sera.*

This text is the product of an abiding interest in understanding longitudinal data analysis in general and time series processes in particular. I could not have accomplished this work without the help of many people. Working on three other projects at the time I was writing this book, I asked Professor

Monnie McGee to help me expedite a time consuming analysis of the sample size and statistical power of common time series models. Although Monnie used S plus and I use SAS and SPSS in the rest of the book, researchers and students will find that many of her findings apply to other statistical packages as well. Therefore, a key contributing scholar is Professor Monnie McGee, who contributed an important chapter on a subject that must be a concern to practitioners in the field, sample size and power of times series tests.

There are a number of scholars to whom I owe a great intellectual debt for both inspiration and time series and forecasting education. Although I have quoted them freely, I would like to give special thanks to George E. P. Box, Gwilym Jenkins, Clive W. J. Granger, Robert F. Engle, Paul Newbold, Spyros Makridakis, Steven Wheelwright, Steven McGee, Rob Hyndman, Michelle Hibon, Robert Fildes, Richard McCleary, Richard Hay, Jr., Wayne Fuller, David A. Dickey, T. C. Mills, David F. Hendry, Charles Ostrom, Dan Wood, G. S. Maddala, Jan Kamenta, Robert Pindyck, Daniel Rubenfeld, Walter Labys, George G. Judge, R. Carter Hill, Henri Theil, J. J. Johnston, Frank Diebold, Bill Greene, David Greenberg, Richard Maisel, J. Scott Armstrong, David F. Hendry, Damodir Gujarati, Jeff Siminoff, Cliff Hurvitch, Gary Simon, Donald Rock, and Mark Nicolich.

The research and writing of many other scholars substantially influenced this work. They are numerous and I list the principal ones in alphabetical order. Bovas Abraham, Sam Adams, Isaiah Berlin, Bruce L. Bowerman, Lynne Bresler, Peter J. Brockwell, Courtney Brown, Brent L. Cohen, Jeff B. Cromwell, Russell Davidson, Richard A. Davis, Gul Ege, Dan Ege, Donald Erdman, Robert Fildes, Phillip H. Francis, W. Gilchrist, Jennifer M. Ginn, A. S. Goldberger, Jim Granato, William E. Griffiths, Damodar N. Gujarati, J. D. Hamilton, D. M. Hanssens, Eric A. Hanusheck, Averill Harriman, Andrew C. Harvey, K. Holden, C. C. Holt, R. Robert Huckfeldt, G. B. Hudak, Rob J. Hyndman John Jackson, J. J. Johnston, M. G. Kendall, Paul Kennedy, Minbo Kim, Lyman Kirkpatrick Jr., P. A. Klein, C. W. Kohfeld, Stanley I. Kutler, Walter Labys, Johannes Ledolter, Mike Leonard, R. Lewandowski, Thomas W. Likens, Charles C. Lin, Mark Little, L.-M. Liu, Jeffrey Lopes, Hans Lütkepohol, James MacKinnon, David McDowell, V. E. McGee, G. R. Meek, Errol E. Meidinger, G. C. Montgomery, Meltem A. Narter, C. R. Nelson, M. P. Neimira, Richard T. O'Connell, Keith Ord, Sudhakar Pandit, Alan Pankratz, H. Jin Park, D. A. Peel, C. I. Plosser, James Ramsey, David P. Reilly, T. Terasvirta, Michel Terraza, J. L. Thompson, George C. Tiao, R. S. Tsay, Walter Vandaele, Helen Weeks, William W. S. Wei, John Williams, Terry Woodfield, Donna Woodward, and Shein-Ming Wu.

There are other scholars, writers, statesmen, and consultants whose data, activities research, and teachings directly or indirectly contributed to the

writing of this text. They include D. F. Andrews, Professor Courtney Brown, Dr. Carl Cohen, Professor Jim Granato, Dr. Stanley Greenberg, the Honorable Averill Harriman, Steven A. M. Herzog, R. Robert Huckfeldt, Professor Guillermina Jasso, President Lyndon Baines Johnson, Professor Lyman Kirkpatrick, Jr., Professor Stanley Kutler, Mike Leonard, Dr. Carol Magai, Robert S. McNamara, McGeorge Bundy, Professor Mark Nicolich, David P. Reilly, Professor Donald Rock, Gus Russo, John Stockwell, Professor Pamela Stone, Professor Peter Tuckel, and Donna Woodward.

I also owe a debt of deep gratitude to key people at Academic Press. To Senior Editor Dr. Scott Bentley and his assistants Beth Bloom, Karen Frost, and Nick Panissidi; to production editors and staff members Brenda Johnson, Mark Sherry, and Mike Early; and to Jerry Altman for posting such accompanying teaching materials as the computer program syntax and data sets at *http://www.apnet.com/sbe/authors*, I remain truly grateful. For her invaluable editorial help, I must give thanks to Kristin Landon. Nor can I forget Lana Traver, Kelly Ricci, and the rest of the staff of the PRD Group for their cooperative graphics preparation and composition; I must express my appreciation to them for their professional contribution to this work.

To the very knowledgeable and helpful people at SAS and SPSS, I must acknowledge a debt for their gracious and substantial assistance. SAS consultants, Donna Woodward, Kevin Meyer, and SAS developer Michael Leonard were always gracious, knowledgeable, and very helpful. Other consultants have helped more obliquely. For their very knowledgeable and personal professional assistance, I remain deeply indebted.

I would also like to thank the people at SPSS, to whom I owe a real debt of gratitude for their knowledgeable and professional assistance. Tony Babinec, Director of Advanced Marketing Products; Mary Nelson and David Cody, Managers in charge of Decision Time development; Dave Nichols, Senior Support Statistician; Dongping Fang, Statistician; and David Mathesson, from the Technical Support staff, provided friendly and knowledgeable assistance with various aspects of the SPSS Trends algorithms. Nor can I forget Andy Kodner or Dave Mattingly, who were also very helpful. To David Mattingly and Mary Nelson, I want to express my thanks for the opportunity to beta-test the Trends module. To Mary Nelson and David Cody, I would like to express my gratitude for the opportunity to beta-test the Decision Time Software in the summer of 1999.

The roots of my interest in forecast go back to the mid to late 1960s and 1970s, when my friends from those years saw my concern with forecasting emerge. To Lehigh University professors Joseph A. Dowling, Jerry Fishman, John Cary, and George Kyte, I remain grateful for support and guidance. To Roman Yuszczuk, George Russ, and other dear friends, to

whom I confided those ominous forecasts, there is no need to say that I told you so. In my lectures, I explained what I observed, analyzed, and foresaw, daring to be forthright in hopes of preventing something worse, and risking the consequences. Many people were wont to say that if you remember the 1960s you were not there. However, we sought to understand and do remember. To Professors Stanley Tennenbaum, John Myhill, and Akiko Kino, at the State University of New York at Buffalo Mathematics Department during the late 1960s, a word of thanks should be offered. For other friends from Buffalo, such as Jesse Nash and Laurie McNeil, I also have to be thankful.

To my friends at the University of Waterloo, in Ontario, Canada, where I immersed myself in statistics and its applications, I remain grateful. To the Snyder family, Richard Ernst, Bill Thomas, Professor June Lowe, Professor Ashok Kapur, Professor Carlo Sempi, and Drs. Steve and Betty Gregory, I express my thanks for all their help. Moreover, I confess an indirect obligation to Admiral Hyman Rickover, whose legendary advice inspired me not to waste time.

To Lt. Col. Robert Avon, Ret., Executive Director of the Lake George Opera Festival, who permitted me to forecast student audience development for the Lake George Opera Festival, I still owe a debt of gratitude, as well as to those friends from Skidmore College Professors Daniel Egy, Bill Fox, Bob Smith, and Bob Jones. To librarians Barbara Smith and Mary O'Donnell, I remain grateful. Nor are Jane Marshall and Marsha Levell forgotten.

From the University of Michigan, where I spent several summers, with financial assistance from the Inter-University Consortium for Political and Social Research (ICPSR) and New York University, I am indebted to Hank Heitowit and Gwen Fellenberger for their guidance, assistance, and support. With the inspiration and teachings of Professor Daniel Wood, John Williams, Walter C. Labys, Courtney Brown, and Jim Granato, along with the assistance of Genie Baker, Dieter Burrell, Professor Xavier Martin, and Dr. Maryke Dressing, I developed my knowledge of dynamic regression and time series to include autoregression and ARIMA analysis. As for all of my good friends at and from Ann Arbor, both identified and unnamed, I remain grateful for their contribution to a wonderful intellectual milieu in which our genuine pursuit of knowledge, understanding, and wisdom was really appreciated and supported.

I am deeply grateful for the support of New York University as well. From the Academic Computing Facility, I gleaned much support. To Frank Lopresti, head of the Statistics and Social Science Group, I owe a debt of gratitude for his cooperation and flexibility, as well as to Dr. Yakov Smotritsky, without whose help in data set acquisition much of this would not

此为试读,需要完整PDF请访问: www.ertongbook.com

have been possible. To Burt Holland for his early support of my involvement with ICPSR and to Edi Franceschini and Dr. George Sadowsky, I remain indebted for the opportunity to teach the time series sequence. To Judy Clifford, I also remain grateful for her administrative assistance. For collegial support, instruction, and inspiration, I must specifically thank the New York University Stern School of Business Statistics and Operations Research (including some already acknowledged) faculty, including Professors Jeff Siminoff, Gary Simon, Bill Greene, James Ramsey, Cliff Hurvich, Rohit Deo, Halina Frydman, Frank Diebold, Edward Melnick, Aaron Tenenbein, and Andreas Weigand. Moreover, the support and inspiration of the Sociology Department Faculty including Professors Richard Maisel, David Greenberg, Jo Dixon, Wolf Hydebrand, and Guillermina Jasso, was instrumental in developing my knowledge of longitudinal analysis in areas of event history and time series analysis. These people are fine scholars and good people, who help constitute and maintain a very good intellectual milieu for which I am grateful.

A number of intellectual friends from other fields were the source of inspiration and other kinds of assistance. In research in the field of addictions research. Valerie C. Lorenz, Ph.D., C.A.S., and William Holmes, of the Compulsive Gambling Center, Inc. In Baltimore, Maryland; Robert M. Politzer, Sc.D. C.A.S., Director of Research for the Washington Center for Addictions; and Clark J. Hudak, Jr. Ph.D., Director of the Washington Center for Addictions proved to be wonderful research colleagues studying pathological gambling. Professor John Kindt of the University of Illinois at Champaign/Urbana; Professor Henry Lesieur, formerly of the Department of Sociology at St. Johns University; and Howard Shaffer, Ph.D., C.A.S., Editor-in-chief of the Journal of Gambling Studies, and research assistants Mathew Hall, Walter Bethune, and Emily McNamara at Harvard Medical School, Division of Addictions, have been good, efficient, and supportive colleagues. Nor can I neglect the supportive assistance of Dr. Veronica J. Brodsky at New York University in this area.

In the area of drug addiction research, I thank Steve Titus of the New York University Medical Center Department of Environmental Medicine for the opportunity to assist in structural equation modeling of drug abuse predispositions research. Among former colleagues in sociomedical research, I thank Dr. Karolyn Siegel and Dr. Shelly Kern, formerly of the Department of Social work Research at Memorial Sloan Kettering Cancer Center; Dr. Ann Brunswick, Dr. Peter Messeri, and Dr. Carla Lewis at Columbia University School of Public Health; Dr. Stephanie Auer, Dr. Steven G. Sclan, and Dr. Bary Reisberg of the Aging and Dementia Research Center at New York University Medical School; and more recently Dr. Carl Cohen and Dr. Carol Magai, State University Health Science

Center at Brooklyn. It was a pleasure to work with John Stockwell in researching developing patterns of the AIDS crisis. His findings proved invaluable in analysis of the gathering of data and its evaluation.

JFK assassination study contributed to my analysis of the Watergate scandal. Among those who did prodigious amounts of investigative work in this area were Gus Russo, John Davis, Robert Blakey, Gaeton Fonzi, John Stockwell, Jim Marrs, Dick Russell, and Mary Nichols. Jim Gray and Gilbert Offenhartz should also be mentioned. Special thanks must also go to Michael Bechloss for transcribing the LBJ White House tapes that explained the official basis of the Warren Commission position.

In the fields of political history, political science, and public opinion research, I also thank Professor Marina Mercada, whose courses in international relations permitted me to present some of my former research, out of which my interest in longitudinal analysis and forecast grew. In the field of political science, Professors Adamantia Pollis, Aristide Zolberg, Richard Bensel, and Jacob Landynski of the Graduate Faculty of the New School for Social Research are wonderful people. In the area of International economics, I also thank Professor Giuseppe Ammendola for his recent assistance. As persons who helped in the more quantitative dimension, Professors Donald Rock and Mark Nicolich provided great inspiration and statistical advice, and to both of them I will always remain indebted. To Professors Dan Cohen and Pam Stone, former chairpersons of the Computer Science and Sociology Departments of Hunter College, respectfully, and to Talbot Katz, Ph.D., I am thankful for much assistance during my years of teaching at Hunter College. For inspiration in political history, political polling, and public opinion research, I have to thank Professor Richard Maisel, Professor Kurt Schlicting, Professor Peter Tuckel, and Dr. Stanley Greenberg for their inspiration and cooperation in these areas of research.

Others whose cooperation and support were critical at one time or another included Frank LaFond, Dr. Winthrop Munro, Les Giermanski, Maria Ycasiano, Nancy Frankel, Ralph Duane, Ed DeMoto, Peggy McEvoy, and Professor Yuko Kinoshita. Special thanks must also be given to publishers John Wiley and Sons, Inc., and Prentice Hall, Inc., for permission to publish and post authors' data.

Throughout this wonderful odyssey, I have been blessed with meeting and working with many wonderfully capable, talented, and ethically fine people. This journey has been exhilarating and fascinating. For their understanding of and supportive forbearance with the demands imposed by this project, I must thank the members of my family—Dana, Glenn (and Barb and Steve)—and my parents Elizabeth and Michael Yaffee, to whom I dedicate this book. If this text helps students, researchers, and consultants

learn time series analysis and forecasting, I will have succeeded in substantially contributing to their education, understanding, and wisdom. If the contents of this book inspires them to develop their knowledge of this field, I will have truly succeeded in no small part due to the assistance of those just mentioned. For the contents of the book, I must nonetheless take full responsibility.

Robert A. Yaffee, Ph.D.
Brooklyn, New York
September 1999

About the Authors

Robert A. Yaffee, Ph.D., is a senior research consultant/statistician in the Statistics and Social Science Group of New York University's Academic Computing Facility as well as a Research Scientist/Statistician at the State University of New York Health Science Center in Brooklyn's Division of Geriatric Psychiatry. He received his Ph.D. in political science from Graduate Faculty of Political and Social Research of The New School for Social Research. He serves as a member of the editorial board of the *Journal of Gambling Studies* and was on the Research Faculty of Columbia University's School of Public Health before coming to NYU. He also taught in the Statistical packages in the Computer Science department and the Empirical Research and Advanced Statistics in the Sociology Department of Hunter College. He has published in the fields of statistics, medical research, and psychology.

Monnie McGee, Ph.D., is an assistant professor in the Department of Mathematics and Statistics at Hunter College, City University of New York, New York. She received her Ph.D. in statistics from Rice University, Houston, Texas. She has worked as a bio-statistical consultant for The Rockefeller University and as a computational statistician for Electricité de France. She has published in the areas of time series and medical statistics. Her hobbies include ice-skating and foreign languages.

Contents

此为试读,需要完整PDF请访问: www.ertongbook.com