Simple | Smart | Speed
CHOOLS

Time Series Analysis

Univariate and Multivariate Methods

SECOND EDITION

William W. S. Wei
Department of Statistics
The Fox School of Business and Management
Temple University

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Publisher	**Greg Tobin**
Executive Editor	**Deirdre Lynch**
Assistant Editor	**Sara Oliver**
Managing Editor	**Ron Hampton**
Production Supervisor	**Kathleen A. Manley**
Marketing Manager	**Phyllis Hubbard**
Marketing Assistant	**Celena Carr**
Senior Designer	**Barbara Atkinson**
Prepress Supervisor	**Caroline Fell**
Manufacturing Buyer	**Evelyn Beaton**
Composition, Production Coordination, and Text Design	**Progressive Information Technologies**
Illustrations	**Jim McLaughlin**
Cover Designer	**Leslie Haimes**

The cover image appears as Figure 17.1 within the text and embodies all the concepts of time series analysis. For instance, the thin solid line represents stationary series as discussed in Chapters 2 and 3; the heavy solid line is nonstationary series as discussed in Chapter 4. As you learn more, you may also see the concepts of a second nonstationary series, forecasting, backcast, seasonal series, heterocedasticity and GARCH models, input and output series, noise series, vector time series, nonlinear time series, and cointegration in the cover image.

Library of Congress Cataloging-in-Publication Data

Wei, William W. S.
Time series analysis: univariate and multivariate methods / William W. S. Wei.— 2nd ed.
p. cm.
Includes index.
ISBN 0-321-32216-9
1. Time-series analysis. I. Title.

QA280.W45 2006
519.5'5—dc22 2004058701

Copyright © 2006 Pearson Education, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise; without the prior written permission of the publisher. Printed in the United States of America.

2 3 4 5 6 7 8 9 10—CRW—09 08 07 06 05

To Susanna

About the Author

William W. S. Wei is Professor of Statistics at Temple University in Philadelphia, PA. He earned his B.A. in Economics from the National Taiwan University (1966), B.A. in Mathematics from the University of Oregon (1969), and M.S. (1972) and Ph.D. (1974) in Statistics from the University of Wisconsin–Madison. From 1982–87, he was the Chair of the Department of Statistics at Temple University. His research interest includes time series analysis, forecasting methods, statistical modeling, and applications of statistics in business and economics. He has developed new methodology in seasonal adjustment, aggregation and disaggregation, outlier detection, robust estimation, and vector time series analysis. Some of his most significant contributions include extensive research on the effects of aggregation, methods of measuring information loss due to aggregation, new stochastic procedures of performing data disaggregation, model-free outlier detection techniques, robust methods of estimating autocorrelations, and statistics for analyzing multivariate time series. He is a Fellow of the American Statistical Association, a Fellow of the Royal Statistical Society, and an Elected Member of the International Statistical Institute. He is an Associate Editor of the *Journal of Forecasting* and the *Journal of Applied Statistical Science.*

Contents

	Preface	xix
CHAPTER 1	Overview	1
	1.1 Introduction	1
	1.2 Examples and Scope of This Book	1
CHAPTER 2	Fundamental Concepts	6
	2.1 Stochastic Processes	6
	2.2 The Autocovariance and Autocorrelation Functions	10
	2.3 The Partial Autocorrelation Function	11
	2.4 White Noise Processes	15
	2.5 Estimation of the Mean, Autocovariances, and Autocorrelations	16
	2.5.1 Sample Mean	17
	2.5.2 Sample Autocovariance Function	18
	2.5.3 Sample Autocorrelation Function	20
	2.5.4 Sample Partial Autocorrelation Function	22
	2.6 Moving Average and Autoregressive Representations of Time Series Processes	23
	2.7 Linear Difference Equations	26
	Exercises	30
CHAPTER 3	Stationary Time Series Models	33
	3.1 Autoregressive Processes	33
	3.1.1 The First-Order Autoregressive AR(1) Process	34
	3.1.2 The Second-Order Autoregressive AR(2) Process	39
	3.1.3 The General pth-Order Autoregressive AR(p) Process	45
	3.2 Moving Average Processes	47
	3.2.1 The First-Order Moving Average MA(1) Process	48
	3.2.2 The Second-Order Moving Average MA(2) Process	51
	3.2.3 The General qth-Order Moving Average MA(q) Process	52

3.3 The Dual Relationship Between AR(p) and MA(q) Processes 54
3.4 Autoregressive Moving Average ARMA(p, q) Processes 57
3.4.1 The General Mixed ARMA(p, q) Process 57
3.4.2 The ARMA(1, 1) Process 59
Exercises 66

CHAPTER 4 Nonstationary Time Series Models 68
4.1 Nonstationarity in the Mean 69
4.1.1 Deterministic Trend Models 69
4.1.2 Stochastic Trend Models and Differencing 70
4.2 Autoregressive Integrated Moving Average (ARIMA) Models 71
4.2.1 The General ARIMA Model 72
4.2.2 The Random Walk Model 72
4.2.3 The ARIMA(0, 1, 1) or IMA(1, 1) Model 74
4.3 Nonstationarity in the Variance and the Autocovariance 80
4.3.1 Variance and Autocovariance of the ARIMA Models 82
4.3.2 Variance Stabilizing Transformations 83
Exercises 86

CHAPTER 5 Forecasting 88
5.1 Introduction 88
5.2 Minimum Mean Square Error Forecasts 88
5.2.1 Minimum Mean Square Error Forecasts for ARMA Models 89
5.2.2 Minimum Mean Square Error Forecasts for ARIMA Models 91
5.3 Computation of Forecasts 94
5.4 The ARIMA Forecast as a Weighted Average of Previous Observations 96
5.5 Updating Forecasts 99
5.6 Eventual Forecast Functions 100
5.7 A Numerical Example 103
Exercises 105

CHAPTER 6 Model Identification 108
6.1 Steps for Model Identification 108
6.2 Empirical Examples 111
6.3 The Inverse Autocorrelation Function (IACF) 122

6.4 Extended Sample Autocorrelation Function and Other Identification Procedures 128
6.4.1 The Extended Sample Autocorrelation Function (ESACF) 128
6.4.2 Other Identification Procedures 133
Exercises 134

CHAPTER 7 Parameter Estimation, Diagnostic Checking, and Model Selection 136
7.1 The Method of Moments 136
7.2 Maximum Likelihood Method 138
7.2.1 Conditional Maximum Likelihood Estimation 138
7.2.2 Unconditional Maximum Likelihood Estimation and Backcasting Method 140
7.2.3 Exact Likelihood Functions 143
7.3 Nonlinear Estimation 145
7.4 Ordinary Least Squares (OLS) Estimation in Time Series Analysis 151
7.5 Diagnostic Checking 152
7.6 Empirical Examples for Series W1–W7 153
7.7 Model Selection Criteria 156
Exercises 158

CHAPTER 8 Seasonal Time Series Models 160
8.1 General Concepts 160
8.2 Traditional Methods 162
8.2.1 Regression Method 162
8.2.2 Moving Average Method 163
8.3 Seasonal ARIMA Models 164
8.4 Empirical Examples 170
Exercises 182

CHAPTER 9 Testing for a Unit Root 186
9.1 Introduction 186
9.2 Some Useful Limiting Distributions 186
9.3 Testing for a Unit Root in the AR(1) Model 189
9.3.1 Testing the AR(1) Model without a Constant Term 189
9.3.2 Testing the AR(1) Model with a Constant Term 192
9.3.3 Testing the AR(1) Model with a Linear Time Trend 195
9.4 Testing for a Unit Root in a More General Model 196

9.5 Testing for a Unit Root in Seasonal Time Series Models 206
9.5.1 Testing the Simple Zero Mean Seasonal Model 207
9.5.2 Testing the General Multiplicative Zero Mean Seasonal Model 207
Exercises 211

CHAPTER 10 Intervention Analysis and Outlier Detection 212
10.1 Intervention Models 212
10.2 Examples of Intervention Analysis 215
10.3 Time Series Outliers 223
10.3.1 Additive and Innovational Outliers 223
10.3.2 Estimation of the Outlier Effect When the Timing of the Outlier Is Known 224
10.3.3 Detection of Outliers Using an Iterative Procedure 226
10.4 Examples of Outlier Analysis 228
10.5 Model Identification in the Presence of Outliers 230
Exercises 235

CHAPTER 11 Fourier Analysis 237
11.1 General Concepts 237
11.2 Orthogonal Functions 238
11.3 Fourier Representation of Finite Sequences 241
11.4 Fourier Representation of Periodic Sequences 242
11.5 Fourier Representation of Nonperiodic Sequences: The Discrete-Time Fourier Transform 247
11.6 Fourier Representation of Continuous-Time Functions 254
11.6.1 Fourier Representation of Periodic Functions 254
11.6.2 Fourier Representation of Nonperiodic Functions: The Continuous-Time Fourier Transform 256
11.7 The Fast Fourier Transform 258
Exercises 261

CHAPTER 12 Spectral Theory of Stationary Processes 264
12.1 The Spectrum 264
12.1.1 The Spectrum and Its Properties 264
12.1.2 The Spectral Representation of Autocovariance Functions: The Spectral Distribution Function 267
12.1.3 Wold's Decomposition of a Stationary Process 271
12.1.4 The Spectral Representation of Stationary Processes 272

12.2 The Spectrum of Some Common Processes 274
12.2.1 The Spectrum and the Autocovariance Generating Function 274
12.2.2 The Spectrum of ARMA Models 274
12.2.3 The Spectrum of the Sum of Two Independent Processes 278
12.2.4 The Spectrum of Seasonal Models 279
12.3 The Spectrum of Linear Filters 281
12.3.1 The Filter Function 281
12.3.2 Effect of Moving Average 283
12.3.3 Effect of Differencing 284
12.4 Aliasing 285
Exercises 286

CHAPTER 13 Estimation of the Spectrum 289
13.1 Periodogram Analysis 289
13.1.1 The Periodogram 289
13.1.2 Sampling Properties of the Periodogram 290
13.1.3 Tests for Hidden Periodic Components 292
13.2 The Sample Spectrum 298
13.3 The Smoothed Spectrum 301
13.3.1 Smoothing in the Frequency Domain: The Spectral Window 301
13.3.2 Smoothing in the Time Domain: The Lag Window 304
13.3.3 Some Commonly Used Windows 306
13.3.4 Approximate Confidence Intervals for Spectral Ordinates 313
13.4 ARMA Spectral Estimation 318
Exercises 321

CHAPTER 14 Transfer Function Models 322
14.1 Single-Input Transfer Function Models 322
14.1.1 General Concepts 322
14.1.2 Some Typical Impulse Response Functions 324
14.2 The Cross-Correlation Function and Transfer Function Models 325
14.2.1 The Cross-Correlation Function (CCF) 325
14.2.2 The Relationship between the Cross-Correlation Function and the Transfer Function 328
14.3 Construction of Transfer Function Models 329
14.3.1 Sample Cross-Correlation Function 329
14.3.2 Identification of Transfer Function Models 331
14.3.3 Estimation of Transfer Function Models 332

14.3.4 Diagnostic Checking of Transfer Function Models 334
14.3.5 An Empirical Example 335
14.4 Forecasting Using Transfer Function Models 341
14.4.1 Minimum Mean Square Error Forecasts for Stationary Input and Output Series 342
14.4.2 Minimum Mean Square Error Forecasts for Nonstationary Input and Output Series 343
14.4.3 An Example 346
14.5 Bivariate Frequency-Domain Analysis 349
14.5.1 Cross-Covariance Generating Functions and the Cross-Spectrum 349
14.5.2 Interpretation of the Cross-Spectral Functions 351
14.5.3 Examples 355
14.5.4 Estimation of the Cross-Spectrum 357
14.6 The Cross-Spectrum and Transfer Function Models 359
14.6.1 Construction of Transfer Function Models through Cross-Spectrum Analysis 359
14.6.2 Cross-Spectral Functions of Transfer Function Models 360
14.7 Multiple-Input Transfer Function Models 361
Exercises 363

CHAPTER 15 Time Series Regression and GARCH Models 366
15.1 Regression with Autocorrelated Errors 366
15.2 ARCH and GARCH Models 368
15.3 Estimation of GARCH Models 373
15.3.1 Maximum Likelihood Estimation 373
15.3.2 Iterative Estimation 374
15.4 Computation of Forecast Error Variance 375
15.5 Illustrative Examples 376
Exercises 380

CHAPTER 16 Vector Time Series Models 382
16.1 Covariance and Correlation Matrix Functions 382
16.2 Moving Average and Autoregressive Representations of Vector Processes 384
16.3 The Vector Autoregressive Moving Average Process 386
16.3.1 Covariance Matrix Function for the Vector AR(1) Model 391
16.3.2 Vector AR(p) Models 394
16.3.3 Vector MA(1) Models 396
16.3.4 Vector MA(q) Models 397
16.3.5 Vector ARMA(1, 1) Models 398

16.4 Nonstationary Vector Autoregressive Moving Average Models 400
16.5 Identification of Vector Time Series Models 401
16.5.1 Sample Correlation Matrix Function 401
16.5.2 Partial Autoregression Matrices 402
16.5.3 Partial Lag Correlation Matrix Function 408
16.6 Model Fitting and Forecasting 414
16.7 An Empirical Example 416
16.7.1 Model Identification 417
16.7.2 Parameter Estimation 417
16.7.3 Diagnostic Checking 420
16.7.4 Forecasting 420
16.7.5 Further Remarks 421
16.8 Spectral Properties of Vector Processes 421
Supplement 16.A Multivariate Linear Regression Models 423
Exercises 426

CHAPTER 17 More on Vector Time Series 428
17.1 Unit Roots and Cointegration in Vector Processes 428
17.1.1 Representations of Nonstationary Cointegrated Processes 430
17.1.2 Decomposition of Z_t 434
17.1.3 Testing and Estimating Cointegration 435
17.2 Partial Process and Partial Process Correlation Matrices 442
17.2.1 Covariance Matrix Generating Function 442
17.2.2 Partial Covariance Matrix Generating Function 443
17.2.3 Partial Process Sample Correlation Matrix Functions 447
17.2.4 An Empirical Example: The U.S. Hog Data 448
17.3 Equivalent Representations of a Vector ARMA Model 451
17.3.1 Finite-Order Representations of a Vector Time Series Process 453
17.3.2 Some Implications 457
Exercises 460

CHAPTER 18 State Space Models and the Kalman Filter 463
18.1 State Space Representation 463
18.2 The Relationship between State Space and ARMA Models 464
18.3 State Space Model Fitting and Canonical Correlation Analysis 470
18.4 Empirical Examples 474
18.5 The Kalman Filter and Its Applications 478
Supplement 18.A Canonical Correlations 483
Exercises 487

CHAPTER 19 Long Memory and Nonlinear Processes 489
19.1 Long Memory Processes and Fractional Differencing 489
19.1.1 Fractionally Integrated ARMA Models and Their ACF 489
19.1.2 Practical Implications of the ARFIMA Processes 492
19.1.3 Estimation of the Fractional Difference 492
19.2 Nonlinear Processes 494
19.2.1 Cumulants, Polyspectrum, and Tests for Linearity and Normality 495
19.2.2 Some Nonlinear Time Series Models 498
19.3 Threshold Autoregressive Models 499
19.3.1 Tests for TAR Models 500
19.3.2 Modeling TAR Models 502
Exercises 506

CHAPTER 20 Aggregation and Systematic Sampling in Time Series 507
20.1 Temporal Aggregation of the ARIMA Process 507
20.1.1 The Relationship of Autocovariances between the Nonaggregate and Aggregate Series 508
20.1.2 Temporal Aggregation of the IMA(d, q) Process 511
20.1.3 Temporal Aggregation of the AR(p) Process 512
20.1.4 Temporal Aggregation of the ARIMA(p, d, q) Process 513
20.1.5 The Limiting Behavior of Time Series Aggregates 515
20.2 The Effects of Aggregation on Forecasting and Parameter Estimation 520
20.2.1 Hilbert Space 520
20.2.2 The Application of Hilbert Space in Forecasting 521
20.2.3 The Effect of Temporal Aggregation on Forecasting 522
20.2.4 Information Loss Due to Aggregation in Parameter Estimation 524
20.3 Systematic Sampling of the ARIMA Process 526
20.4 The Effects of Systematic Sampling and Temporal Aggregation on Causality 528
20.4.1 Decomposition of Linear Relationship between Two Time Series 528
20.4.2 An Illustrative Underlying Model 531
20.4.3 The Effects of Systematic Sampling and Temporal Aggregation on Causality 532
20.5 The Effects of Aggregation on Testing for Linearity and Normality 534
20.5.1 Testing for Linearity and Normality 534
20.5.2 The Effects of Temporal Aggregation on Testing for Linearity and Normality 537

20.6 The Effects of Aggregation on Testing for a Unit Root 540
20.6.1 The Model of Aggregate Series 541
20.6.2 The Effects of Aggregation on the Distribution of the Test Statistics 542
20.6.3 The Effects of Aggregation on the Significance Level and the Power of the Test 543
20.6.4 Examples 545
20.6.5 General Cases and Concluding Remarks 547
20.7 Further Comments 549
Exercises 550

References 553

Appendix 565
Time Series Data Used for Illustrations 565
Statistical Tables 565

Author Index 601

Subject Index 605

Preface to the Second Edition

Since the publication of the first edition, this book has been used by many researchers and universities worldwide. I am very grateful for the numerous encouraging letters and comments received from researchers, instructors, and students. Although the original chapters in the book still form the necessary foundation for time series analysis, many new theories and methods have been developed during the past decade, and the time has come to incorporate these new developments into a more comprehensive view of the field. In the process of updating this book, I took the opportunity to clarify certain concepts and correct previous errors.

In time series analysis, we often encounter nonstationary time series, and a formal testing procedure for a unit root has now become the standard routine in time series modeling. To address this procedure, Chapter 9 on unit root tests for both nonseasonal and seasonal models has been added.

Regression analysis is the most commonly used statistical method, and time series data are widely used in regression modeling, particularly in business and economic research. The standard assumptions of uncorrelated errors and constant variance are often violated when time series variables are used in the model. In a separate new Chapter 15, the use of time series variables in regression analysis is discussed. In particular, this chapter introduces models with autocorrelated errors and ARCH/GARCH models for heteroscedasticity that are useful in many economic and financial studies.

Although the basic procedures of model building for univariate time series and vector time series are the same, there are some important phenomena unique to vector time series. After an introduction to various vector time series models in Chapter 16, cointegration, partial processes, and equivalent representations of a vector time series model are covered in the new Chapter 17. They are useful in understanding and analyzing relationships of time series variables.

Many time series exhibit characteristics that cannot be described by linear models. Therefore Chapter 19 on long memory processes and nonlinear time series models that are useful in describing these long memory and nonlinear phenomena was included.

To aid understanding, two supplements have been added: Supplement 16A on multivariate linear regression models, and Supplement 18A on canonical correlations. In the chapter on aggregation, some new results are included on the effects of aggregation on testing for linearity, normality, and unit roots.

In this revision, I follow the fundamental theme of the first edition and continue to balance the emphasis between both theory and applications. Methodologies are introduced with proper theoretical justifications and illustrated with empirical data sets that may be downloaded

from the Web site: http://www.sbm.temple.edu/~wwei/. As with the first edition, exercise problems are included at the end of each chapter to enhance the reader's understanding of the subject. The book should be useful for graduate and advanced undergraduate students who have proper backgrounds and are interested in learning the subject. It should also be helpful as a reference for researchers who encounter time series data in their studies.

As indicated in the first edition, the book was developed from a one-year course given in the Department of Statistics at Temple University. Topics of univariate time series analysis from Chapters 1 through 13 were covered during the first semester, and the remaining chapters related to multivariate time series plus supplemental journal articles were discussed in the second semester. With the proper selection of topics, the book can be used for a variety of one- or two-semester courses in time series analysis, model building, and forecasting.

I wish to thank Dr. Olcay Akman of the College of Charleston, Dr. Mukhtar Ali of the University of Kentucky, Dr. H. K. Hsieh of the University of Massachusetts, Dr. Robert Miller of the University of Wisconsin, Dr. Mohsen Pourahamadi of Northern Illinois University, Dr. David Quigg of Bradley University, and Dr. Tom Short of Indiana University of Pennsylvania for their numerous suggestions and comments that have improved this revision. I am grateful to Ceylan Yozgatligil for her help in preparing some of the updated examples and tables. Finally, I would like to thank Ms. Deirdre Lynch, Executive Editor of Statistics, Addison-Wesley for her continuing interest and assistance with this project as well as Ms. Kathleen Manley, Ms. Barbara Atkinson, Mr. Jim McLaughlin, and the staff at Progressive Publishing Alternatives who provided wonderful assistance in the production of the book.

William W. S. Wei
Department of Statistics
The Fox School of Business and Management
Temple University
Philadelphia, Pennsylvania, USA
April 2005

Preface to the First Edition

Time series analysis has been an area of considerable activity in recent years. Several books on the subject have been published, most of them concentrating either on time-domain analysis, or solely on frequency-domain analysis. Within each of these categories some books provide inadequate theoretical background material, while others offer very few applications. The majority of these books concentrate only on univariate time series, and those few books which discuss multivariate time series focus only on theoretical aspects of the subject. This book, however, attempts to provide a comprehensive introduction to both time-domain and frequency-domain analyses of univariate and multivariate time series.

Since my own student days at the University of Wisconsin–Madison, my attitude toward time series analysis has been influenced greatly by Professor George E. P. Box's motto that statistics will not flourish without a happy marriage between theory and practice. I hope, therefore, that a proper balance between theory and applications has been achieved within the pages of this volume.

Included in the book are many recent advances in univariate and multivariate time series methods such as inverse autocorrelation functions, extended sample autocorrelation functions, Akaike's information criterion (AIC), intervention analysis and outlier detection, vector ARMA models, partial lag correlation matrix functions, partial processes, the state-space model and Kalman filtering, aggregation problems, and many others.

This book was developed from a one-year course given in the Department of Statistics at Temple University. The course was offered to those graduate students with backgrounds in statistical theory and regression analysis. The first nine chapters were covered in the first semester; the remaining chapters were covered in the second semester. The book, however, provides sufficient flexibility regarding the order in which topics can be covered. With an appropriate selection of topics, the book can be used for a variety of one- or two-semester courses in time series analysis and forecasting. For those interested in frequency-domain analysis, Chapters 10, 11, and 12 can be covered earlier. However, this author believes students have a better appreciation of the frequency-domain approach after studying time-domain analysis.

Exercise problems are included at the end of each chapter. The purpose of these problems is to enhance the student's understanding of the subject and not solely to test the student's mathematical skills. Students are encouraged to attempt to solve as many of these exercises as possible.

The book should be useful for graduate and advanced undergraduate students in statistics, economics, business, engineering, meteorology, various areas of social science, and any

other field where analysis and research of time series are prevalent. The book should also be useful as a reference for research workers interested in learning the subject.

I am greatly indebted to Professor George E. P. Box, who introduced me to time series analysis, and Professor George C. Tiao, who supervised my research in the field while I was a graduate student at the University of Wisconsin – Madison. This book could not have been written without their inspiration as great teachers and researchers.

I wish to thank all those who assisted in the preparation of this book, which evolved from a series of lecture notes. I am indebted to Ruth Jackson for typing, and Dror Rom and Hewa Saladasa for proofreading several chapters of the notes. The preparation of the final manuscript began in the summer of 1987, when my administrative responsibility as Department Chairman at Temple was completed.

I am very grateful to my then thirteen-year-old son, Stephen, for typing most of the final manuscript, and to Wai-Sum Chan, Leonard Cupingood, and Jong-Hyup Lee for their careful reading and checking of the entire manuscript, as well as their assistance with numerous computations and the initial draft of figures using SAS (SAS Institute, 1985) and SCA (Scientific Computing Associates, 1986) software. Many of their valuable suggestions were incorporated into the book.

I would also like to thank David Reilly for his excellent time-domain software AUTOBOX and MTS, which were used to analyze many univariate and multivariate time series data sets in the book. Further acknowledgments to John Schlater and Alan Izenman for their valuable comments, and to the various manuscript reviewers for their helpful suggestions and criticisms.

Special thanks goes to my wife, Susanna, for helping to make this project a family effort. She carefully checked the final manuscript, provided many helpful suggestions, developed a sorting system for indices, and supervised our children in a variety of sorting tasks. Of course, any errors and omissions which remain are entirely my responsibility, and I would appreciate having them brought to my attention.

I wish to extend my gratitude to Professor Donald B. Owen and Addison-Wesley Publishing Company for permission to adopt the t and Chi-Square tables from their book, *Handbook of Statistical Tables,* and to the late Professor E. S. Pearson and the Biometrika trustees for permission to adopt the F table from their book, *Biometrika Tables for Statisticians,* Volume I.

Finally, I would like to thank Allan Wylde, Karen Garrison, and the rest of the Addison-Wesley staff, and Grace Sheldrick and Frederick Bartlett, for their interest and assistance with this project.

William Wu-Shyong Wei
March 1989

1 Overview

1.1 Introduction

A time series is an ordered sequence of observations. Although the ordering is usually through time, particularly in terms of some equally spaced time intervals, the ordering may also be taken through other dimensions, such as space. Time series occur in a variety of fields. In agriculture, we observe annual crop production and prices. In business and economics, we observe daily closing stock prices, weekly interest rates, monthly price indices, quarterly sales, and yearly earnings. In engineering, we observe sound, electric signals, and voltage. In geophysics, we record turbulence, such as ocean waves and earth noise in an area. In medical studies, we measure electroencephalogram (EEG) and electrocardiogram (EKG) tracings. In meteorology, we observe hourly wind speeds, daily temperature, and annual rainfall. In quality control, we monitor a process according to a certain target value. In the social sciences, we study annual birth rates, mortality rates, accident rates, and various crime rates. The list of areas in which time series is observed and studied is endless.

A time series, such as electric signals and voltage, that can be recorded continuously in time, is said to be continuous. A time series, such as interest rates, yields, and volume of sales, which are taken only at specific time intervals, is said to be discrete. In this book, we deal exclusively with discrete time series observed at equal intervals. We do so because even continuous time series provide only digitized values at discrete intervals for computations.

There are various objectives for studying time series. They include the understanding and description of the generating mechanism, the forecasting of future values, and optimal control of a system. The intrinsic nature of a time series is that its observations are dependent or correlated, and the order of the observations is therefore important. Hence, statistical procedures and techniques that rely on independence assumption are no longer applicable, and different methods are needed. The body of statistical methodology available for analyzing time series is referred to as time series analysis.

1.2 Examples and Scope of This Book

Figure 1.1 shows four time series that are studied later in the book. Although they each show the salient feature of dependence among observations, they also exhibit other very distinctive characteristics.

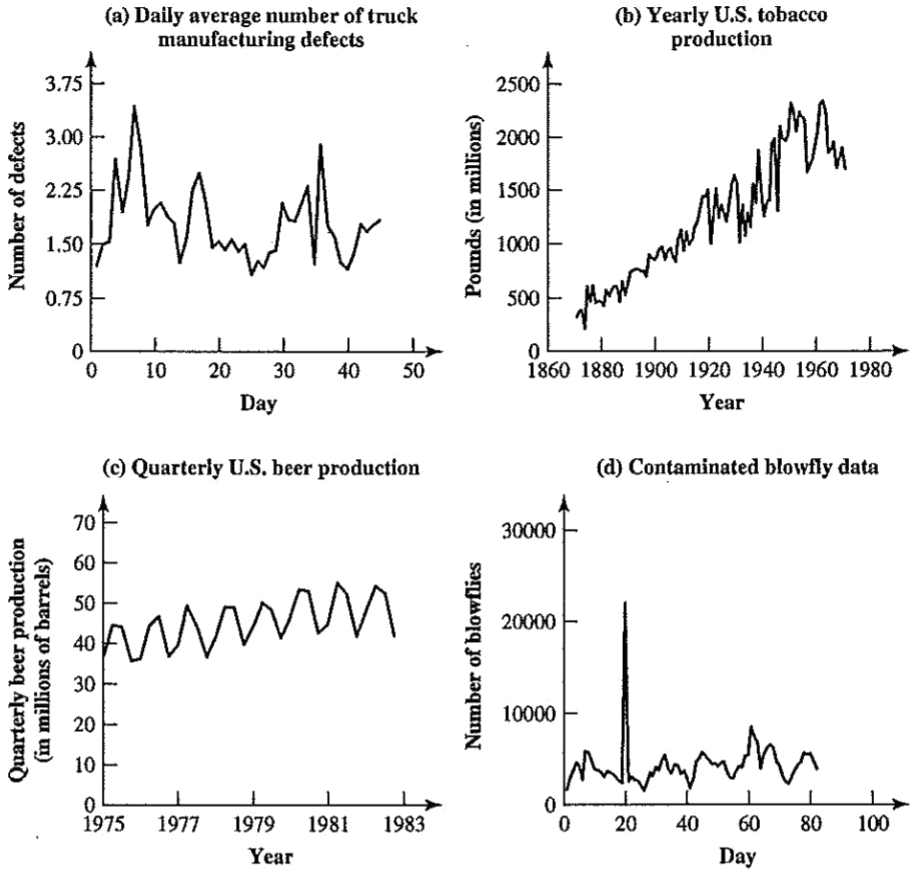


FIGURE 1.1 Examples of some time series.

The daily average number of defects per truck found at the end of the assembly line of a truck manufacturing plant, shown in Figure 1.1(a), appears to vary about a fixed level. Time series that exhibit this phenomenon are said to be stationary in the mean and are special cases of stationary time series. The yearly U.S. tobacco production, shown in Figure 1.1(b), does not vary about a fixed level and exhibits an overall upward trend instead. Moreover, the variance of this tobacco series increases as the level of the series increases. Time series that exhibit these phenomena are said to be nonstationary in mean and variance and are examples of nonstationary time series. The U.S. quarterly beer production in Figure 1.1(c) presents another characteristic pattern that is repetitive in nature due to seasonal variation. Time series containing seasonal variation are called seasonal time series. Nonstationary time series such as those shown in Figures 1.1(b) and (c) can be reduced to stationary series by proper transformations.

After introducing in Chapter 2 fundamental concepts needed to characterize time series, we study in Chapters 3 to 8 a very general class of parametric models: the autoregressive integrated moving average (ARIMA) models that are useful for describing stationary, nonstationary, seasonal, and nonseasonal time series such as those shown in Figures 1.1(a),

(b), and (c). Specifically, in Chapter 2, we introduce the basic ideas of stationarity and nonstationarity, autocorrelation function, partial autocorrelation function, representations of time series processes, and linear difference equations needed for parametric modeling. Chapter 3 introduces stationary autoregressive moving average (ARMA) models, and Chapter 4 discusses nonstationary ARIMA models. These three chapters give the necessary background to identify a tentative time series model. Chapter 5 develops theory and methods of minimum mean square error (MMSE) forecasting, which are needed for later discussion of parameter estimation. Chapters 6 and 7 illustrate the iterative procedure of time series model building. Additional identification tools, such as the inverse autocorrelation function (IACF) and the extended sample autocorrelation function (ESACF), are introduced. Because several different models may adequately represent a given series, we also discuss useful model selection criteria such as Akaike's information criterion (AIC). Chapter 8 extends these ideas to seasonal time series models. Many time series are nonstationary, and a formal testing procedure for nonstationarity has become a standard routine in time series modeling. In Chapter 9, we discuss formal unit root tests for both nonseasonal and seasonal models.

The fourth time series shown in Figure 1.1(d) is a contaminated laboratory series of blowfly data. It reflects another phenomenon of nonstationarity due to a change in structure in the series from some external disturbance. This type of nonstationarity cannot be removed by a standard transformation. Time series are often interrupted by external events. Additional examples include production series affected by strikes and recorded series marred by a machine failure. Such external disturbances are termed *interventions* or *outliers*. To model external interruptions, intervention and outlier models are studied in Chapter 10.

The time series approach presented in Chapters 2 through 10, which uses autocorrelation and partial autocorrelation functions to study the evolution of a time series through parametric models, is known as time domain analysis. An alternative approach, which uses spectral functions to study the nonparametric decomposition of a time series into its different frequency components, is known as frequency domain analysis; it is presented in Chapters 11 to 13. Although these two approaches are mathematically equivalent in the sense that the autocorrelation function and the spectrum function form a Fourier transform pair, on occasion one approach is more advantageous than the other. Basic knowledge of both approaches is a necessary tool for time series data analysis and research.

Chapters 2 through 13 deal with the analysis of a single (univariate) time series. Time series data, however, often involve simultaneous observations on several variables. For example, Figure 1.2 shows the Lydia Pinkham annual data of advertising and sales. In business, one often tries to promote sales by advertising. It seems natural to build a dynamic model to relate current sales to present and past values of advertising to study the effect of advertising and to improve sales forecasting. Thus, in Chapter 14 we present transfer function models useful for relating an output series to one or more input series. Using material from earlier chapters, both time domain and frequency domain approaches to the problem are discussed. In analyzing multivariate time series, the fundamental tool used in the time domain approach is the cross-correlation function (CCF), and the fundamental tool used in the frequency domain approach is the cross-spectral function. They also form a Fourier transform pair in time series analysis. The other useful method to analyze the relationship between variables is regression analysis. The regression model is without a doubt the most commonly used statistical model, but the standard regression assumptions are often violated when time series variables are used in the model. Therefore, in Chapter 15 we discuss the

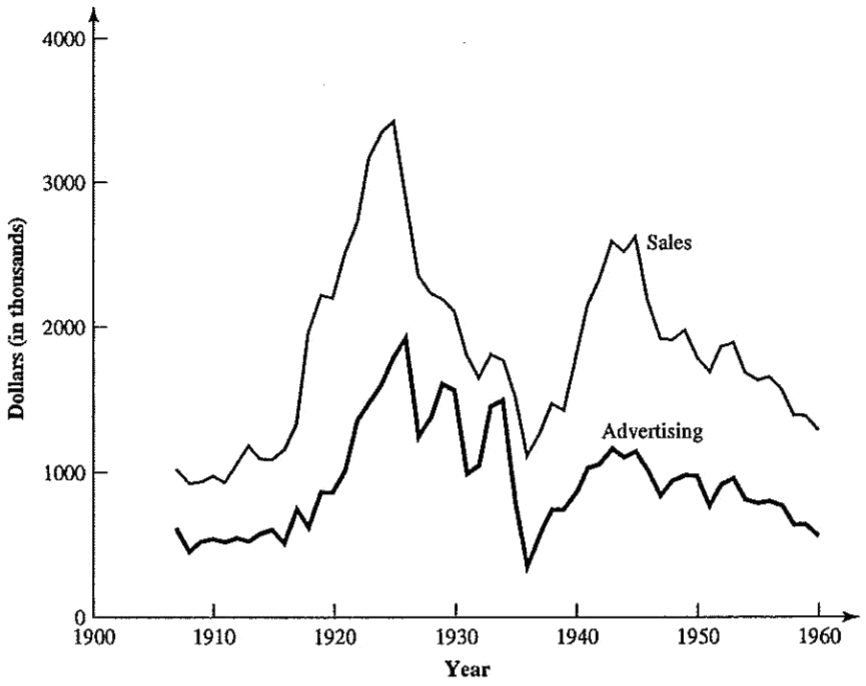


FIGURE 1.2 The Lydia Pinkham advertising and sales data.

use of time series variables in regression analysis including models with autocorrelated errors and various autoregressive conditional heteroscedasticity (ARCH) models for heterogeneous variances.

In many fields of applications, the transfer function and regression models may not be appropriate because complicated feedforward and feedback relationships may exist among variables. For example, in the Lydia Pinkham data, although advertising is expected to promote sales, sales may also affect advertising because advertising budgets are often set as a percentage of sales. In Chapter 16, we extend results to multivariate vector time series and study the joint relationship of several time series variables. Vector autoregressive moving average models are introduced. Detailed discussions on identification tools such as correlation matrix functions, partial autoregression matrices, and partial lag correlation matrix functions are given. Spectral properties of vector processes are also introduced. In Chapter 17, we discuss useful notions such as cointegration, partial processes, and equivalent representations of a vector time series model.

In Chapter 18, we discuss the state space model and Kalman filtering. This alternative approach in time series modeling can be used for both univariate and multivariate time series. Like other methods introduced in the book, the approach is illustrated with empirical data sets. The relationship between the state space and ARIMA models is also established.

After discussing various univariate and multivariate linear time series models and their applications, we present, in Chapter 19, long memory processes and nonlinear time series models. These methods have been found useful in modeling many time series.

Finally, we address a nontrivial, important problem in time series analysis. In working with time series data, one must first decide on the time unit to be used in the analysis. In some cases, such as the annual yield of grain per acre, there exists a natural time unit for the series; there are other cases, however, in which several different time units may be available for adaptation. For example, in analyzing sales, one may consider monthly sales, quarterly sales, or yearly sales. Does the choice of time interval affect the type of conclusions that are drawn? In Chapter 20, we try to supply some answers to this question when we discuss aggregation and sampling in time series.

2

Fundamental Concepts

In this chapter, we introduce some fundamental concepts that are necessary for proper understanding of time series models discussed in this book. We begin with a simple introduction to stochastic processes, the autocorrelation and partial autocorrelation functions, and the notion of white noise processes. We then discuss estimation of the mean, autocovariances, and the autocorrelation and partial autocorrelation functions. Thus, we can illustrate the sampling phenomena of time series models starting in Chapter 3, which, we believe, will enhance the appreciation of model identification discussed in later chapters. In addition to its own theoretical merits, the notion of the moving average and autoregressive representations of processes is useful for understanding the logic underlying parsimonious linear processes used in time series analysis. We also give a simple introduction to linear difference equations, with a special emphasis on the solution of homogeneous difference equations, which play a fundamental role in the parsimonious linear time series processes.

2.1 Stochastic Processes

A stochastic process is a family of time indexed random variables $Z(\omega, t)$, where ω belongs to a sample space and t belongs to an index set. For a fixed t, $Z(\omega, t)$ is a random variable. For a given ω, $Z(\omega, t)$, as a function of t, is called a sample function or realization. The population that consists of all possible realizations is called the ensemble in stochastic processes and time series analysis. Thus, a time series is a realization or sample function from a certain stochastic process. To have a proper appreciation for time series analysis, we introduce some essential concepts of stochastic processes in this section.

In our discussion, we assume that the index set is the set of all integers unless mentioned otherwise. Consider a finite set of random variables $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\}$ from a stochastic process $\{Z(\omega, t): t = 0, \pm 1, \pm 2, \ldots\}$. The n-dimensional distribution function is defined by

$$F_{Z_{t_1},\ldots,Z_{t_n}}(x_1, \ldots, x_n) = P\{\omega: Z_{t_1} \leq x_1, \ldots, Z_{t_n} \leq x_n\}, \tag{2.1.1}$$

where x_i, $i = 1, \ldots, n$ are any real numbers. A process is said to be first-order stationary in distribution if its one-dimensional distribution function is time invariant, i.e., if $F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1)$ for any integers t_1, k, and $t_1 + k$; second-order stationary in distribution if

$F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ for any integers t_1, t_2, k, $t_1 + k$, and $t_2 + k$; and nth-order stationary in distribution if

$$F_{Z_{t_1}, \ldots, Z_{t_n}}(x_1, \ldots, x_n) = F_{Z_{t_1+k}, \ldots, Z_{t_n+k}}(x_1, \ldots, x_n) \tag{2.1.2}$$

for any n-tuple $(t_1, \ldots, t_n)$ and k of integers. A process is said to be strictly stationary if (2.1.2) is true for any n, i.e., $n = 1, 2, \ldots$. The terms *strongly stationary* and *completely stationary* are also used to denote a strictly stationary process. Clearly, if (2.1.2) is true for $n = m$, it is also true for $n \leq m$ because the mth-order distribution function determines all distribution functions of lower order. Hence, a higher order of stationarity always implies a lower order of stationarity.

With proper understanding that a stochastic process, $Z(\omega, t)$, is a set of time indexed random variables defined on a sample space, we usually suppress the variable ω and simply write $Z(\omega, t)$ as $Z(t)$ or Z_t, just as we often denote a random variable by X rather than by $X(\omega)$. The process is called a real-valued process if it assumes only real values. Unless mentioned otherwise, processes discussed in this book are referred to as the real-valued processes. For a given real-valued process $\{Z_t : t = 0, \pm 1, \pm 2, \ldots\}$, we define the mean function of the process

$$\mu_t = E(Z_t), \tag{2.1.3}$$

the variance function of the process

$$\sigma_t^2 = E(Z_t - \mu_t)^2, \tag{2.1.4}$$

the covariance function between Z_{t_1} and Z_{t_2}

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}), \tag{2.1.5}$$

and the correlation function between Z_{t_1} and Z_{t_2}

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}. \tag{2.1.6}$$

For a strictly stationary process, since the distribution function is the same for all t, the mean function $\mu_t = \mu$ is a constant, provided $E(|Z_t|) < \infty$. Likewise, if $E(Z_t^2) < \infty$, then $\sigma_t^2 = \sigma^2$ for all t and hence is also a constant. Moreover, since $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ for any integers t_1, t_2, and k, we have

$$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k)$$

and

$$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k).$$

Letting $t_1 = t - k$ and $t_2 = t$, we get

$$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k \tag{2.1.7}$$

and

$$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k. \tag{2.1.8}$$

Thus, for a strictly stationary process with the first two moments finite, the covariance and the correlation between Z_t and Z_{t+k} depend only on the time difference k.

So far, we have discussed a strong sense of stationarity of a process in terms of its distribution function. A trivial example of a strictly stationary process is a sequence of independent identically distributed (i.i.d.) random variables. This sequence of independent random variables usually does not exist or renders no interest in time series. Other than this simple i.i.d. case, however, it is very difficult or impossible actually to verify a distribution function, particularly a joint distribution function from an observed time series. Thus, in time series analysis, we often use a weaker sense of stationarity in terms of the moments of the process.

A process is said to be nth-order weakly stationary if all its joint moments up to order n exist and are time invariant, i.e., independent of time origin. Therefore, a second-order weakly stationary process will have constant mean and variance, with the covariance and the correlation being functions of the time difference alone. Sometimes, the terms *stationary in the wide sense* or *covariance stationary* are also used to describe a second-order weakly stationary process. It follows from the definitions that a strictly stationary process with the first two moments finite is also a second-order weakly or covariance stationary process. Yet, a strictly stationary process may not have finite moments and hence may not be covariance stationary. A trivial example is the process consisting of a sequence of i.i.d. Cauchy random variables. The process is clearly strictly stationary, but it is not weakly stationary of any order because no joint moments exist.

EXAMPLE 2.1 Let Z_t be twice the value of a true die shown on the tth toss. If the die is tossed three times independently, we have a stochastic process $Z_t = Z(\omega, t)$ where t belongs to the index set $= \{1, 2, 3\}$ and ω belongs to the sample space

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

For a particular ω, say (1, 3, 2), the realization or sample function will be (2, 6, 4). The ensemble contains a total of 216 possible realizations. If a true die is tossed independently and repeatedly, then we have a stochastic process $Z_t = Z(\omega, t)$ with the index set being all positive integers and the sample space being

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \cdots.$$

A particular ω of (1, 4, 2, 3, . . .) would have the realization (2, 8, 4, 6, . . .). In this case, the total number of possible realizations in the ensemble is infinite. Clearly, the process in either case is strictly stationary as it relates to a sequence of i.i.d. random variables.

EXAMPLE 2.2 Consider the time sequence

$$Z_t = A \sin(\omega t + \theta), \tag{2.1.9}$$

where A is a random variable with a zero mean and a unit variance and θ is a random variable with a uniform distribution on the interval $[-\pi, \pi]$ independent of A. Then

$$\begin{aligned}
E(Z_t) &= E(A)E[\sin(\omega t + \theta)] = 0 \\
E(Z_t Z_{t+k}) &= E\{A^2 \sin(\omega t + \theta) \sin[\omega(t + k) + \theta]\} \\
&= E(A^2)E\left\{\frac{1}{2}[\cos(\omega k) - \cos(\omega(2t + k) + 2\theta)]\right\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}E\{\cos(\omega(2t + k) + 2\theta)\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}\int_{-\pi}^{\pi} \cos(\omega(2t + k) + 2\theta) \cdot \frac{1}{2\pi} d\theta \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{8\pi}[\sin(\omega(2t + k) + 2\theta)]_{-\pi}^{\pi} \\
&= \frac{1}{2}\cos(\omega k),
\end{aligned} \tag{2.1.10}$$

which depends only on the time difference k. Hence, the process is covariance stationary.

EXAMPLE 2.3 Let Z_t be a sequence of independent random variables alternately following a standard normal distribution $N(0, 1)$ and a two-valued discrete uniform distribution with equal probability 1/2 of being 1 or -1. Clearly, $E(Z_t) = 0$ and $E(Z_t^2) = 1$ for all t. Now

$$E(Z_t Z_s) = \begin{cases} 0, & \text{if } t \neq s, \\ 1, & \text{if } t = s, \end{cases}$$

and

$$\rho(t, s) = \frac{E(Z_t Z_s)}{\sqrt{E(Z_t^2)}\sqrt{E(Z_s^2)}} = \begin{cases} 0, & \text{if } t \neq s, \\ 1, & \text{if } t = s. \end{cases}$$

Hence, the process is covariance stationary. The process, however, is not strictly stationary. It is, in fact, not stationary in distribution for any order.

From these discussions and examples, it is clear that "covariance stationary" is a much weaker form of stationarity than are "strictly stationary" or "stationary in distribution." We often work, however, with covariance stationary or second-order weakly stationary processes

in time series analysis because it is relatively simple to check the first two moments. Henceforth, unless mentioned otherwise, we use the term *stationary* to refer to all processes that are covariance stationary. With this, the following important remark is in order.

A stochastic process is said to be a normal or Gaussian process if its joint probability distribution is normal. Because a normal distribution is uniquely characterized by its first two moments, strictly stationary and weakly stationary are equivalent for a Gaussian process. Unless mentioned otherwise, processes which we discuss are assumed to be Gaussian. Like other areas in statistics, most time series results are established for Gaussian processes. Thus, the autocorrelation functions and the partial autocorrelation functions discussed in the next two sections become fundamental tools in time series analysis.

2.2 The Autocovariance and Autocorrelation Functions

For a stationary process $\{Z_t\}$, we have the mean $E(Z_t) = \mu$ and variance $\text{Var}(Z_t) = E(Z_t - \mu)^2 = \sigma^2$, which are constant, and the covariances $\text{Cov}(Z_t, Z_s)$, which are functions only of the time difference $|t - s|$. Hence, in this case, we write the covariance between Z_t and Z_{t+k} as

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu), \tag{2.2.1}$$

and the correlation between Z_t and Z_{t+k} as

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \tag{2.2.2}$$

where we note that $\text{Var}(Z_t) = \text{Var}(Z_{t+k}) = \gamma_0$. As functions of k, γ_k is called the autocovariance function and ρ_k is called the autocorrelation function (ACF) in time series analysis because they represent the covariance and correlation between Z_t and Z_{t+k} from the same process, separated only by k time lags.

It is easy to see that for a stationary process, the autocovariance function γ_k and the autocorrelation function ρ_k have the following properties:

1. $\gamma_0 = \text{Var}(Z_t)$; $\rho_0 = 1$.
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$ for all k, i.e., γ_k and ρ_k are even functions and hence symmetric about the lag $k = 0$. This property follows from the time difference between Z_t and Z_{t+k} and Z_t and Z_{t-k} being the same. Therefore, the autocorrelation function is often plotted only for the nonnegative lags such as the one shown in Figure 2.1. This plot is sometimes called a correlogram.
4. Another important property of the autocovariance function γ_k and the autocorrelation function ρ_k is that they are positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j\gamma_{|t_i - t_j|} \geq 0 \tag{2.2.3}$$

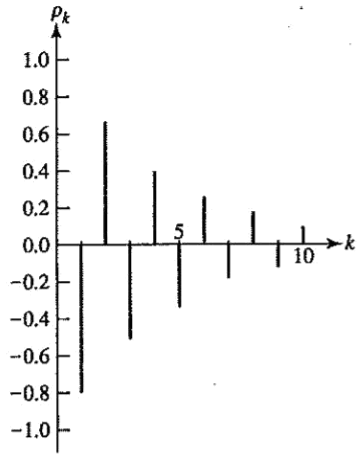


FIGURE 2.1 Example of autocorrelation function (ACF).

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_{|t_i-t_j|} \geq 0 \tag{2.2.4}$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any real numbers $\alpha_1, \alpha_2, \ldots,$ and α_n. By defining the random variable $X = \sum_{i=1}^{n}\alpha_i Z_{t_i}$, the result (2.2.3) follows from

$$0 \leq \mathrm{Var}(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \mathrm{Cov}(Z_{t_i}, Z_{t_j}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|}.$$

The similar result for ρ_k in (2.2.4) follows immediately by dividing the inequality (2.2.3) through by γ_0. Thus, it is important to know that not every arbitrary function satisfying properties (1) to (3) can be an autocovariance or autocorrelation function for a process. A necessary condition for a function to be the autocovariance or autocorrelation function of some process is that it be positive semidefinite.

2.3 The Partial Autocorrelation Function

In addition to the autocorrelation between Z_t and Z_{t+k}, we may want to investigate the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, \ldots,$ and Z_{t+k-1} has been removed. The conditional correlation

$$\mathrm{Corr}(Z_t, Z_{t+k}|Z_{t+1}, \ldots, Z_{t+k-1}) \tag{2.3.1}$$

is usually referred to as the partial autocorrelation in time series analysis.

Consider a stationary process $\{Z_t\}$ and, without loss of generality, we assume that $E(Z_t) = 0$. Let the linear dependence of Z_{t+k} on Z_{t+1}, Z_{t+2}, . . . , and Z_{t+k-1} be defined as the best linear estimate in the mean square sense of Z_{t+k} as a linear function of Z_{t+1}, Z_{t+2}, . . . , and Z_{t+k-1}. That is, if $\hat{Z}_{t+k}$ is the best linear estimate of Z_{t+k}, then

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+k-1} + \alpha_2 Z_{t+k-2} + \cdots + \alpha_{k-1} Z_{t+1}, \tag{2.3.2}$$

where $\alpha_i (1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained from minimizing

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 = E(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2. \tag{2.3.3}$$

The routine minimization method through differentiation gives the following linear system of equations

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \cdots + \alpha_{k-1} \gamma_{i-k+1} \qquad (1 \leq i \leq k-1). \tag{2.3.4}$$

Hence,

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \cdots + \alpha_{k-1} \rho_{i-k+1} \qquad (1 \leq i \leq k-1). \tag{2.3.5}$$

In terms of matrix notation, the above system of (2.3.5) becomes

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}. \tag{2.3.6}$$

Similarly,

$$\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \cdots + \beta_{k-1} Z_{t+k-1}, \tag{2.3.7}$$

where $\beta_i (1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained by minimizing

$$E(Z_t - \hat{Z}_t)^2 = E(Z_t - \beta_1 Z_{t+1} - \cdots - \beta_{k-1} Z_{t+k-1})^2. \tag{2.3.8}$$

Hence,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}, \tag{2.3.9}$$

which implies that $\alpha_i = \beta_i (1 \leq i \leq k-1)$.

It follows that the partial autocorrelation between Z_t and Z_{t+k} will equal the ordinary autocorrelation between $(Z_t - \hat{Z}_t)$ and $(Z_{t+k} - \hat{Z}_{t+k})$. Thus, letting P_k denote the partial autocorrelation between Z_t and Z_{t+k}, we have

$$P_k = \frac{\text{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{\text{Var}(Z_t - \hat{Z}_t)}\sqrt{\text{Var}(Z_{t+k} - \hat{Z}_{t+k})}}. \tag{2.3.10}$$

Now,

$$\begin{aligned}
\text{Var}(Z_{t+k} - \hat{Z}_{t+k}) &= E[(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})^2] \\
&= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&\quad - \alpha_1 E[Z_{t+k-1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&\quad - \cdots - \alpha_{k-1}E[Z_{t+1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})]
\end{aligned}$$

because all other remaining terms reduce to zero by virtue of Equation (2.3.4). Hence,

$$\text{Var}(Z_{t+k} - \hat{Z}_{t+k}) = \text{Var}(Z_t - \hat{Z}_t) = \gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}. \tag{2.3.11}$$

Next, using that $\alpha_i = \beta_i (1 \le i \le k-1)$, we have

$$\begin{aligned}
&\text{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1}Z_{t+k-1})(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1}Z_{t+k-1})Z_{t+k}] \\
&= \gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1.
\end{aligned} \tag{2.3.12}$$

Therefore,

$$P_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}} = \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \cdots - \alpha_{k-1}\rho_{k-1}}. \tag{2.3.13}$$

Solving the system in (2.3.6) for α_i by Cramer's rule gives

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_{i-1} & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_{i-2} & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}} \tag{2.3.14}$$

as the ratio of two determinants. The matrix in the numerator is the same as the symmetric matrix in the denominator except for its ith column being replaced by $(\rho_1, \rho_2, \ldots, \rho_{k-1})$. Substituting α_i in (2.3.14) to (2.3.13) and multiplying both the numerator and denominator of (2.3.13) by the determinant

$$\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{vmatrix},$$

the resulting P_k in (2.3.13) can be easily seen to equal the ratio of the determinants in (2.3.15) each expanded by its last column,

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}. \tag{2.3.15}$$

The partial autocorrelation can also be derived as follows. Consider the regression model, where the dependent variable Z_{t+k} from a zero mean stationary process is regressed on k lagged variables $Z_{t+k-1}, Z_{t+k-2}, \ldots$ and Z_t, i.e.,

$$Z_{t+k} = \phi_{k1}Z_{t+k-1} + \phi_{k2}Z_{t+k-2} + \cdots + \phi_{kk}Z_t + e_{t+k}, \tag{2.3.16}$$

where ϕ_{ki} denotes the ith regression parameter and e_{t+k} is an error term with mean 0 and uncorrelated with Z_{t+k-j} for $j=1, 2, \ldots, k$. Multiplying Z_{t+k-j} on both sides of the above regression equation and taking the expectation, we get

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}; \tag{2.3.17}$$

hence,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}. \tag{2.3.18}$$

For $j = 1, 2, \ldots, k$, we have the following system of equations:

$$\begin{aligned} \rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0. \end{aligned}$$

Using Cramer's rule successively for $k = 1, 2, \ldots$, we have

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$\vdots$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}. \tag{2.3.19}$$

Comparing Equation (2.3.19) with (2.3.15), we see that ϕ_{kk} equals P_k. Thus, the partial autocorrelation between Z_t and Z_{t+k} can also be obtained as the regression coefficient associated with Z_t when regressing Z_{t+k} on its k lagged variables $Z_{t+k-1}, Z_{t+k-2}, \ldots$, and Z_t as in (2.3.16). Because ϕ_{kk} has become a standard notation for the partial autocorrelation between Z_t and Z_{t+k} in time series literature, we use this notation in our book. As a function of k, ϕ_{kk} is usually referred to as the partial autocorrelation function (PACF).

2.4 White Noise Processes

A process $\{a_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $E(a_t) = \mu_a$, usually assumed to be 0, constant variance $\text{Var}(a_t) = \sigma_a^2$ and $\gamma_k = \text{Cov}(a_t, a_{t+k}) = 0$ for all $k \neq 0$. By definition, it immediately follows that a white noise process $\{a_t\}$ is stationary with the autocovariance function

$$\gamma_k = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{2.4.1}$$

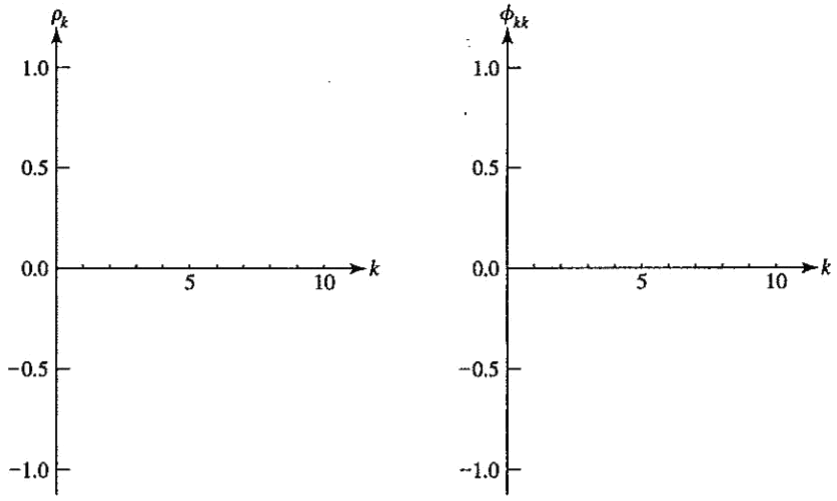


FIGURE 2.2 ACF and PACF of a white noise process: $Z_t = \mu + a_t$.

the autocorrelation function

$$\rho_k = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{2.4.2}$$

and the partial autocorrelation function

$$\phi_{kk} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases} \tag{2.4.3}$$

The ACF and PACF of a white noise process are shown in Figure 2.2.

Because by definition $\rho_0 = \phi_{00} = 1$ for any process, when we talk about the autocorrelation and partial autocorrelations, we refer only to ρ_k and ϕ_{kk} for $k \neq 0$. The basic phenomenon of the white noise process is that its ACF and PACF are identically equal to zero.

Although this process hardly ever occurs in practice, it plays an important role as a basic building block in the construction of time series models, just like the roles played by the sine and the cosine functions $\{\sin(nx), \cos(nx)\}$ in Fourier analysis. More precisely, it plays the role of an orthogonal basis in the general vector and function analysis.

A white noise process is Gaussian if its joint distribution is normal. In the following discussion, unless mentioned otherwise, $\{a_t\}$ is always referred to as a zero mean Gaussian white noise process.

2.5 Estimation of the Mean, Autocovariances, and Autocorrelations

A stationary time series is characterized by its mean μ, variance σ^2, autocorrelations ρ_k, and partial autocorrelations ϕ_{kk}. The exact values of these parameters can be calculated if the ensemble of all possible realizations is known. Otherwise, they can be estimated if

multiple independent realizations are available. In most applications, however, it is difficult or impossible to obtain multiple realizations. Most available time series constitute only a single realization, which makes it impossible to calculate the ensemble average. For a stationary process, however, we have a natural alternative of replacing the ensemble average by the time average. In the following discussion, we examine conditions under which we can estimate with good statistical properties the mean and autocovariances and hence the autocorrelations by using time averages.

2.5.1 Sample Mean

With only a single realization, a natural estimator for the mean $\mu = E(Z_t)$ of a stationary process is the sample mean

$$\overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t, \tag{2.5.1}$$

which is the time average of n observations. The question becomes whether the above estimator is a valid or good estimator. Clearly,

$$E(\overline{Z}) = \frac{1}{n}\sum_{t=1}^{n} E(Z_t) = \frac{1}{n}\cdot n\mu = \mu, \tag{2.5.2}$$

which implies that $\overline{Z}$ is an unbiased estimator of μ. It can also be easily shown that

$$\begin{aligned} \mathrm{Var}(\overline{Z}) &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} \mathrm{Cov}(Z_t, Z_s) = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_{(t-s)} \qquad (2.5.3)\\ &= \frac{\gamma_0}{n^2}\sum_{k=-(n-1)}^{n-1}(n - |k|)\rho_k \\ &= \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1}\left(1 - \frac{|k|}{n}\right)\rho_k, \qquad (2.5.4) \end{aligned}$$

where we let $k = (t - s)$. Thus, if

$$\lim_{n\to\infty}\left[\sum_{k=-(n-1)}^{n-1}\left(1 - \frac{|k|}{n}\right)\rho_k\right]$$

is finite, then $\mathrm{Var}(\overline{Z}) \to 0$ as $n \to \infty$, and $\overline{Z}$ is a consistent estimator for μ. That is,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} Z_t = \mu \tag{2.5.5}$$

in mean square. The process is said to be ergodic for the mean if the result in (2.5.5) holds. A sufficient condition for this result to hold is that $\rho_k \to 0$ as $k \to \infty$, because $\rho_k \to 0$ as $k \to \infty$ implies that for any $\epsilon > 0$, we can choose an N such that $|\rho_k| < \epsilon/4$ for all $k > N$. Hence, for $n > (N + 1)$, we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\rho_k\right| &\leq \frac{2}{n}\sum_{k=0}^{n-1}|\rho_k| \\ &= \frac{2}{n}\sum_{k=0}^{N}|\rho_k| + \frac{2}{n}\sum_{k=N+1}^{n-1}|\rho_k| \\ &\leq \frac{2}{n}\sum_{k=0}^{N}|\rho_k| + \frac{1}{2}\epsilon \\ &\leq \epsilon, \end{aligned} \tag{2.5.6}$$

where we choose an n large enough so that the first term in the next to last inequality above is also less than $\epsilon/2$. Thus, when $\rho_k \to 0$ as $k \to \infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\rho_k = 0,$$

which implies that, in Equation (2.5.4),

$$\lim_{n\to\infty}\mathrm{Var}(\overline{Z}) = 0. \tag{2.5.7}$$

Intuitively, these results simply say that if Z_t and Z_{t+k} sufficiently far apart are almost uncorrelated, then some useful new information can be continually added so that the time average will approach the ensemble average.

2.5.2 Sample Autocovariance Function

Similarly, with only a single realization, we employ the following estimators using the time average to estimate the autocovariance function γ_k, i.e.,

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z}) \tag{2.5.8}$$

or

$$\hat{\gamma}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z}). \tag{2.5.9}$$

Now,

$$\begin{aligned}\sum_{t=1}^{n-k}(Z_t-\bar{Z})(Z_{t+k}-\bar{Z}) &= \sum_{t=1}^{n-k}[(Z_t-\mu)-(\bar{Z}-\mu)][(Z_{t+k}-\mu)-(\bar{Z}-\mu)]\\ &= \sum_{t=1}^{n-k}(Z_t-\mu)(Z_{t+k}-\mu)-(\bar{Z}-\mu)\sum_{t=1}^{n-k}(Z_t-\mu)\\ &\quad -(\bar{Z}-\mu)\sum_{t=1}^{n-k}(Z_{t+k}-\mu)+(n-k)(\bar{Z}-\mu)^2\\ &\approx \sum_{t=1}^{n-k}(Z_t-\mu)(Z_{t+k}-\mu)-(n-k)(\bar{Z}-\mu)^2, \qquad (2.5.10)\end{aligned}$$

where we approximate the terms $\sum_{t=1}^{n-k}(Z_t-\mu)$ and $\sum_{t=1}^{n-k}(Z_{t+k}-\mu)$ by $(n-k)(\bar{Z}-\mu)$. Hence,

$$E(\hat{\gamma}_k) \simeq \gamma_k - \frac{k}{n}\gamma_k - \left(\frac{n-k}{n}\right)\mathrm{Var}(\bar{Z}), \qquad (2.5.11)$$

$$E(\mathring{\hat{\gamma}}_k) \simeq \gamma_k - \mathrm{Var}(\bar{Z}). \qquad (2.5.12)$$

It is clear that both of these estimators are biased. When we ignore the term $\mathrm{Var}(\bar{Z})$ representing the effect of estimating μ, $\mathring{\hat{\gamma}}_k$ becomes unbiased but $\hat{\gamma}_k$ is still biased. In general, $\hat{\gamma}_k$ has a larger bias than $\mathring{\hat{\gamma}}$, especially when k is large with respect to n. Hence, for a given n, often at most $n/4$ estimates are suggested to be calculated in time series analysis. If $\rho_k \to 0$ as $k \to \infty$ and hence the process is ergodic for the mean and $\lim_{n\to\infty}\mathrm{Var}(\bar{Z}) = 0$ as shown in (2.5.7), however, then both estimators $\hat{\gamma}_k$ and $\mathring{\hat{\gamma}}_k$ are asymptotically unbiased. Some may argue that because both $\hat{\gamma}_k$ and $\mathring{\hat{\gamma}}_k$ are biased, it would be more appropriate to compare their mean square errors. It can be shown that for certain types of processes $\hat{\gamma}_k$ has smaller mean square error than $\mathring{\hat{\gamma}}_k$ (see, e.g., Parzen [1961b]). In addition, the estimate $\hat{\gamma}_k$ is always positive semidefinite like γ_k, but $\mathring{\hat{\gamma}}_k$ is not necessarily so. As a result, we use $\hat{\gamma}_k$ in (2.5.8) as the sample autocovariance function to estimate the autocovariance function γ_k.

When the process $\{Z_t\}$ is Gaussian, Bartlett (1946) has shown the following approximate results:

$$\mathrm{Cov}(\hat{\gamma}_k, \hat{\gamma}_{k+j}) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\gamma_i\gamma_{i+j} + \gamma_{i+k+j}\gamma_{i-k}) \qquad (2.5.13)$$

and

$$\mathrm{Var}(\hat{\gamma}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\gamma_i^2 + \gamma_{i+k}\gamma_{i-k}). \qquad (2.5.14)$$

Similarly,

$$\text{Cov}(\hat{\hat{\gamma}}_k, \hat{\hat{\gamma}}_{k+j}) \simeq \frac{1}{n-k}\sum_{i=-\infty}^{\infty}(\gamma_i\gamma_{i+j} + \gamma_{i+k+j}\gamma_{i-k}), \tag{2.5.15}$$

and

$$\text{Var}(\hat{\hat{\gamma}}_k) \simeq \frac{1}{n-k}\sum_{i=-\infty}^{\infty}(\gamma_i^2 + \gamma_{i+k}\gamma_{i-k}). \tag{2.5.16}$$

Thus, the variance of $\hat{\hat{\gamma}}_k$ is larger than the variance of $\hat{\gamma}_k$. In fact, from (2.5.16) we see that the variance of $\hat{\hat{\gamma}}_k$, $\text{Var}(\hat{\hat{\gamma}}_k)$, can be substantial for large k resulting in an unstable and erratic estimate.

Next, we would like to ask when the process is ergodic for the autocovariance function so that in mean square we have

$$\lim_{n\to\infty}\hat{\gamma}_k = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) = \gamma_k. \tag{2.5.17}$$

A rigorous proof of the above statement is complicated. For our purpose, it suffices to note that for any given k, because the sample autocovariance $\hat{\gamma}_k$ is an asymptotically unbiased estimator of γ_k, hence a sufficient condition for $\hat{\gamma}_k$ to be mean square consistent and the process to be ergodic for the autocovariances is that the autocovariance is absolutely summable, i.e., $\sum_{-\infty}^{\infty}|\gamma_i| < \infty$ and hence $\lim_{n\to\infty}\text{Var}(\hat{\gamma}_k) = 0$. For relevant readings, interested readers are referred to Gnedenko (1962), Hannan (1970, p. 201), and Fuller (1996, p. 308), among others. Ergodicity is assumed to hold for the remainder of this book.

2.5.3 Sample Autocorrelation Function

For a given observed time series $Z_1, Z_2, \ldots, Z_n$, the sample ACF is defined as

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}, \qquad k = 0, 1, 2, \ldots \tag{2.5.18}$$

where $\bar{Z} = \sum_{t=1}^{n}Z_t/n$, the sample mean of the series. A plot of $\hat{\rho}_k$ versus k is sometimes called a sample correlogram.

For a stationary Gaussian process, Bartlett (1946) has shown that for $k > 0$ and $k + j > 0$,

$$\text{Cov}(\hat{\rho}_k, \hat{\rho}_{k+j}) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i\rho_{i+j} + \rho_{i+k+j}\rho_{i-k} - 2\rho_k\rho_i\rho_{i-k-j} - 2\rho_{k+j}\rho_i\rho_{i-k} + 2\rho_k\rho_{k+j}\rho_i^2). \tag{2.5.19}$$

For large n, $\hat{\rho}_k$ is approximately normally distributed with mean ρ_k and variance

$$\text{Var}(\hat{\rho}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2). \tag{2.5.20}$$

For processes in which $\rho_k = 0$ for $k > m$, Bartlett's approximation of (2.5.20) becomes

$$\text{Var}(\hat{\rho}_k) \simeq \frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_m^2). \tag{2.5.21}$$

In practice, $\rho_i (i = 1, 2, \ldots, m)$ are unknown and are replaced by their sample estimates $\hat{\rho}_i$, and we have the following large-lag standard error of $\hat{\rho}_k$:

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_m^2)}. \tag{2.5.22}$$

To test a white noise process, we use

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}. \tag{2.5.23}$$

EXAMPLE 2.4 To illustrate the computation of the sample ACF, consider the following ten values of a time series:

t	Z_t	Z_{t+1}	Z_{t+2}	Z_{t+3}	$\cdots;$	Z_{t-1}	Z_{t-2}
1	13	8	15	4			
2	8	15	4	4		13	
3	15	4	4	12		8	13
4	4	4	12	11		15	8
5	4	12	11	7		4	15
6	12	11	7	14		4	4
7	11	7	14	12		12	4
8	7	14	12			11	12
9	14	12				7	11
10	12					14	7

The sample mean of these ten values is $\bar{Z} = 10$. Thus,

$$\hat{\rho}_1 = \frac{(13-10)(8-10) + (8-10)(15-10) + \cdots + (7-10)(14-10) + (14-10)(12-10)}{(13-10)^2 + (8-10)^2 + \cdots + (14-10)^2 + (12-10)^2}$$

$$= \frac{-27}{144} = -.188$$

$$\hat{\rho}_2 = \frac{(13-10)(15-10)+(8-10)(4-10)+\cdots+(11-10)(14-10)+(7-10)(12-10)}{144}$$

$$= \frac{-29}{144} = -.201$$

$$\hat{\rho}_3 = \frac{(13-10)(4-10)+(8-10)(4-10)+\cdots+(12-10)(14-10)+(11-10)(12-10)}{144}$$

$$= \frac{26}{144} = .181$$

$$\vdots$$

and we note that

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Z_t-\bar{Z})(Z_{t+k}-\bar{Z})}{\sum_{t=1}^{n}(Z_t-\bar{Z})^2}$$

$$= \frac{\sum_{t=k+1}^{n}(Z_t-\bar{Z})(Z_{t-k}-\bar{Z})}{\sum_{t=1}^{n}(Z_t-\bar{Z})^2} = \hat{\rho}_{-k}. \tag{2.5.24}$$

In other words, the sample ACF is also symmetric about the origin $k = 0$.

2.5.4 Sample Partial Autocorrelation Function

The sample PACF $\hat{\phi}_{kk}$ is obtained by substituting ρ_i by $\hat{\rho}_i$ in Equation (2.3.19). Instead of calculating the complicated determinants for large k in (2.3.19), a recursive method starting with $\hat{\phi}_{11} = \hat{\rho}_1$ for computing $\hat{\phi}_{kk}$ has been given by Durbin (1960) as follows:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_j} \tag{2.5.25}$$

and

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}, \qquad j = 1,\ldots,k. \tag{2.5.26}$$

The method holds also for calculating the theoretical PACF ϕ_{kk}.

It was shown by Quenouille (1949) that on the hypothesis that the underlying process is a white noise sequence, the variance of $\hat{\phi}_{kk}$ can be approximated by

$$\mathrm{Var}(\hat{\phi}_{kk}) \simeq \frac{1}{n}. \tag{2.5.27}$$

Hence, $\pm 2/\sqrt{n}$ can be used as critical limits on ϕ_{kk} to test the hypothesis of a white noise process.

EXAMPLE 2.5 Using the data in Example 2.4, we have from (2.3.19), (2.5.25), and (2.5.26)

$$\hat{\phi}_{11} = \hat{\rho}_1 = -.188$$
$$\hat{\phi}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{-.201 - (-.188)^2}{1 - (-.188)^2} = -.245$$
$$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22} \cdot \hat{\phi}_{11} = (-.188) - (-.245)(-.188) = -.234.$$

Thus, by (2.5.25) we have

$$\hat{\phi}_{33} = \frac{\hat{\rho}_3 - \hat{\phi}_{21}\hat{\rho}_2 - \hat{\phi}_{22}\hat{\rho}_1}{1 - \hat{\phi}_{21}\hat{\rho}_1 - \hat{\phi}_{22}\hat{\rho}_2}$$
$$= \frac{.181 - (-.234)(-.201) - (-.245)(-.188)}{1 - (-.234)(-.188) - (-.245)(-.201)} = \frac{.088}{.907} = .097.$$

Other $\hat{\phi}_{kk}$ can be calculated similarly.

2.6 Moving Average and Autoregressive Representations of Time Series Processes

In time series analysis, there are two useful representations to express a time series process. One is to write a process Z_t as a linear combination of a sequence of uncorrelated random variables, i.e.,

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \tag{2.6.1}$$

where $\psi_0 = 1$, $\{a_t\}$ is a zero mean white noise process, and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Here and in the following an infinite sum of random variables is defined as the limit in quadratic mean (mean square) of the finite partial sums. Thus, Z_t in (2.6.1) is defined such that

$$E\left[\left(\dot{Z}_t - \sum_{j=0}^{n} \psi_j a_{t-j}\right)^2\right] \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty,$$

where $\dot{Z}_t = Z_t - \mu$. By introducing the backshift operator $B^j x_t = x_{t-j}$, we can write (2.6.1) in the compact form

$$\dot{Z}_t = \psi(B) a_t, \tag{2.6.2}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

It is easy to show that for the process in (2.6.1)

$$E(Z_t) = \mu, \tag{2.6.3}$$

$$\text{Var}(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2, \tag{2.6.4}$$

and

$$E(a_t Z_{t-j}) = \begin{cases} \sigma_a^2, & \text{for } j = 0, \\ 0, & \text{for } j > 0. \end{cases} \tag{2.6.5}$$

Hence,

$$\begin{aligned} \gamma_k &= E(\dot{Z}_t \dot{Z}_{t+k}) \\ &= E\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t+k-j} \right) \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \end{aligned} \tag{2.6.6}$$

and

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}. \tag{2.6.7}$$

Clearly, the autocovariance and autocorrelation functions in (2.6.6) and (2.6.7) are functions of the time difference k only. Because they involve infinite sums, however, to be stationary we have to show that γ_k is finite for each k. Now,

$$|\gamma_k| = |E(\dot{Z}_t \dot{Z}_{t+k})| \leq [\text{Var}(Z_t)\, \text{Var}(Z_{t+k})]^{1/2} = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2.$$

Hence, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is a required condition for the process in (2.6.1) to be stationary.

The form in (2.6.1) is called a moving average (MA) representation of a process. Wold (1938) proved that a stationary process that is purely nondeterministic (i.e., a process contains no deterministic component that can be forecast or predicted exactly from its own past) can always be expressed in the form of (2.6.1). Hence, the representation is also known as Wold's representation in the literature, and any process that can be represented in this form is called a nondeterministic process. Sometimes the term *linear process* is also used to refer to the process in (2.6.1).

For a given sequence of autocovariances γ_k, $k = 0 \pm 1, \pm 2, \ldots$, the autocovariance generating function is defined as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k, \tag{2.6.8}$$

where the variance of the process, γ_0, is the coefficient of B^0 and the autocovariance of lag k, γ_k, is the coefficient of both B^k and B^{-k}. Using (2.6.6) and stationarity, we can write (2.6.8) as

$$\begin{aligned}\gamma(B) &= \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} \\ &= \sigma_a^2 \psi(B)\psi(B^{-1}),\end{aligned} \tag{2.6.9}$$

where we let $j = i + k$ and note that $\psi_j = 0$ for $j < 0$. This method is a convenient way of calculating the autocovariances for some linear processes. The corresponding autocorrelation generating function will be

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}. \tag{2.6.10}$$

Another useful form is to write a process Z_t in an autoregressive (AR) representation, in which we regress the value of Z at time t on its own past values plus a random shock, i.e.,

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \cdots + a_t$$

or, equivalently,

$$\pi(B)\dot{Z}_t = a_t, \tag{2.6.11}$$

where $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, and $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$. The autoregressive representation is useful in understanding the mechanism of forecasting. Box and Jenkins (1976) call a process invertible if it can be written in this form. They argue that in forecasting, a noninvertible process is meaningless. It is easily seen that not every stationary process is invertible. For a linear process $Z_t = \psi(B)a_t$ to be invertible so that it can be written in terms of the AR representation, the roots of $\psi(B) = 0$ as a function of B must lie outside of the unit circle. That is, if β is a root of $\psi(B)$, then $|\beta| > 1$, where $|\ |$ is the standard Euclidian metric. When β is a real number, $|\beta|$ is equal to the absolute value of β and when β is a complex number, $\beta = c + id$, then $|\beta| = \sqrt{c^2 + d^2}$.

It should be noted that an invertible process is not necessarily stationary. By Wold's result, it follows that for the process presented in (2.6.11) to be stationary, the process must be able to be rewritten in a MA representation, i.e.,

$$\dot{Z}_t = \frac{1}{\pi(B)} a_t = \psi(B) a_t, \tag{2.6.12}$$

such that the condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is satisfied. To achieve that, the required condition is that the roots of $\pi(B) = 0$ all lie outside the unit circle. That is, if δ is a root of $\pi(B)$, we have $|\delta| > 1$.

Although the autoregressive and moving average representations are useful, they are not the model forms we use at the beginning stage of model building because they contain an infinite number of parameters that are impossible to estimate from a finite number of available observations. Instead, in modeling a phenomenon, we construct models with only a finite number of parameters.

In the autoregressive representation of a process, if only a finite number of π weights are nonzero, i.e., $\pi_1 = \phi_1, \pi_2 = \phi_2, \ldots, \pi_p = \phi_p$ and $\pi_k = 0$ for $k > p$, then the resulting process is said to be an autoregressive process of order p. This process is written as

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \cdots - \phi_p \dot{Z}_{t-p} = a_t. \tag{2.6.13}$$

Similarly, in the moving average representation, if only a finite number of ψ weights are nonzero, i.e., $\psi_1 = -\theta_1, \psi_2 = -\theta_2, \ldots, \psi_q = -\theta_q$ and $\psi_k = 0$ for $k > q$, then the resulting process is said to be a moving average process of order q, which is given by

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \tag{2.6.14}$$

If we restrict ourselves to the above finite-order autoregressive models in (2.6.13) and moving average models in (2.6.14), however, the number of parameters may still be prohibitively large. A natural alternative is the mixed autoregressive moving average model

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \cdots - \phi_p \dot{Z}_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \tag{2.6.15}$$

For a fixed number of observations, the more parameters in a model, the less efficient is the estimation of the parameters. Other things being equal, in general, we choose a simpler model to describe the phenomenon. This modeling criteria is the principle of parsimony in model building recommended by Tukey (1967) and Box and Jenkins (1976). In the following chapters, we discuss some very useful parsimonious time series models and their properties.

2.7 Linear Difference Equations

Linear difference equations play an important role in the time series models discussed in this book. In fact, all models of finite number of parameters mentioned above relate an output Z_t to an input a_t in terms of linear difference equations. Thus, they are sometimes referred to as

linear difference equation models. The properties of these models often depend on the characteristics of the roots of these difference equations. To better understand these models, we give a brief introduction to linear difference equations, especially to the method of solving these equations. A general nth-order linear difference equation with constant coefficients is given by

$$C_0Z_t + C_1Z_{t-1} + C_2Z_{t-2} + \cdots + C_nZ_{t-n} = e_t, \tag{2.7.1}$$

where C_i, $i = 0, 1, \ldots, n$, are constants. Without loss of generality, we can set $C_0 = 1$. The function e_t in (2.7.1) is called a forcing function. Equation (2.7.1) is said to be nonhomogeneous (or complete) if $e_t \neq 0$, and homogeneous if $e_t = 0$.

Using the backshift operator and letting $C(B) = (1 + C_1B + C_2B^2 + \ldots + C_nB^n)$, we can rewrite (2.7.1) as

$$C(B)Z_t = e_t. \tag{2.7.2}$$

As a function of B, $C(B) = 0$ is called the auxiliary equation associated with the given linear difference equation. The solution of linear difference equations rests primarily on the following lemmas, which can be easily proved using the definition of a solution and the property of linearity.

Lemma 2.7.1. If $Z_t^{(1)}$ and $Z_t^{(2)}$ are solutions of the homogeneous equation, then $b_1Z_t^{(1)} + b_2Z_t^{(2)}$ is also a solution for any arbitrary constants b_1 and b_2. ■

Lemma 2.7.2. If $Z_t^{(H)}$ is a solution to the homogeneous equation and $Z_t^{(P)}$ is a particular solution of the nonhomogeneous equation, then $Z_t^{(H)} + Z_t^{(P)}$ is the general solution of the complete equation. ■

The particular solution of a nonhomogeneous difference equation depends on the form of the forcing function. The solution of a homogeneous difference equation depends on the roots of the associated auxiliary equation. As will be seen, most of the time series models considered in this book are of the difference equation type and the behavior of the time series generated by such models is governed by the nature of the roots of the associated auxiliary equation. In the remainder of this section, we concentrate on the solution of a general homogeneous linear difference equation. When B is used as an operator, it operates on time index t.

Lemma 2.7.3. Let $(1 - B)^m Z_t = 0$. Then a solution is given by $Z_t = bt^j$, where b is any constant and j is a nonnegative integer less than m.

Proof For $m = 1$, $Z_t = bt^0 = b$. Clearly, we have $(1-B)Z_t = (1 - B)b = b - b = 0$. Now assume that $(1 - B)^{m-1} Z_t = 0$, where $Z_t = bt^j, j < (m - 1)$. Then for $Z_t = bt^j, j < m$,

$$\begin{aligned}(1 - B)^m Z_t &= (1 - B)^{m-1}(1 - B)bt^j \\ &= (1 - B)^{m-1}b\{t^j - (t - 1)^j\} \\ &= (1 - B)^{m-1}\left\{-b\sum_{i=0}^{j-1}\binom{j}{i}(-1)^{j-i}t^i\right\}.\end{aligned}$$

Now each term in the last expression involving t contains only integer powers less than $(m - 1)$. Thus, by our induction hypothesis, each term is reduced to zero by the operator $(1 - B)^{m-1}$. The lemma is proved. ∎

It follows from Lemmas 2.7.1 and 2.7.3 that $(1 - B)^m\left[\sum_{j=0}^{m-1} b_j t^j\right] = 0$ for any constants $b_0, b_1, \ldots$ and b_{m-1}.

Lemma 2.7.4 Let $(1-RB)^m Z_t = 0$. Then a solution is given by $Z_t = t^j R^t$, where j is any nonnegative integer less than m, and the general solution is given by $Z_t = \left(\sum_{j=0}^{m-1} b_j t^j\right)R^t$, where b_j is a constant.

Proof First, we note that

$$\begin{aligned}(1 - RB)Z_t &= (1 - RB)t^jR^t = t^jR^t - R(t - 1)^jR^{t-1} \\ &= [(1 - B)t^j]R^t.\end{aligned}$$

Repeated application of the above result gives

$$(1 - RB)^mZ_t = (1 - RB)^mt^jR^t = [(1 - B)^mt^j]R^t,$$

which equals zero because $(1 - B)^m t^j = 0$ by Lemma 2.7.3. The result then follows immediately from Lemma 2.7.1. ∎

Finally, we have the following main result.

Theorem 2.7.1. Let $C(B)Z_t = 0$ be a given homogeneous linear difference equation where $C(B) = 1 + C_1B + C_2B^2 + \cdots + C_nB^n$. If $C(B) = \prod_{i=1}^{N}(1 - R_iB)^{m_i}$, where $\sum_{i=1}^{N} m_i = n$ and $B_i = R_i^{-1}$ $(i = 1, 2, \ldots, N)$ are roots of multiplicity m_i of $C(B) = 0$, then $Z_t = \sum_{i=1}^{N}\sum_{j=0}^{m_i-1} b_{ij}\, t^jR_i^t$. In particular, if $m_i = 1$ for all i and R_i^{-1} $(i = 1, 2, \ldots, n)$ are all distinct, we have $Z_t = \sum_{i=1}^{n} b_iR_i^t$.

Proof The result follows immediately from Lemmas 2.7.1, 2.7.3, and 2.7.4. ∎

Note that for a real-valued linear difference equation, a complex root of $C(B) = 0$ must appear in pairs. That is, if $(c + di)$ is a root, then its complex conjugate $(c + di)^* = (c - di)$ is also a root. A general complex number can always be written in the polar form, i.e.,

$$(c \pm di) = \alpha(\cos\phi \pm i \sin\phi), \tag{2.7.3}$$

where

$$\alpha = (c^2 + d^2)^{1/2} \tag{2.7.4}$$

and

$$\phi = \tan^{-1}(d/c). \tag{2.7.5}$$

Because $(c \pm di)^t = \alpha^t(\cos \phi t \pm i \sin \phi t)$, it follows that for each pair of complex roots of multiplicity m, the solution of the homogeneous difference equation should contain the sequences $\alpha^t \cos \phi t$, $\alpha^t \sin \phi t$; $t\alpha^t \cos \phi t$, $t\alpha^t \sin \phi\, t$; . . . ; and $t^{m-1}\alpha^t \cos \phi t$, $t^{m-1}\alpha^t \sin \phi t$.

For an illustration, consider the second-order difference equation with the following auxiliary equation

$$(1 - C_1B - C_2B^2) = (1 - R_1B)(1 - R_2B) = 0$$

with

$$R_1 = c + di = \alpha(\cos \phi + i \sin \phi) \qquad \text{and} \qquad R_2 = c - di = \alpha(\cos \phi - i \sin \phi).$$

By Theorem 2.7.1, we have

$$Z_t = e_1(c + di)^t + e_2(c - di)^t = e_1[\alpha(\cos \phi + i \sin \phi)]^t + e_2[\alpha(\cos \phi - i \sin \phi)]^t,$$

where e_1 and e_2 are any (complex) constants. For a real-valued process, we claim that we can always write

$$Z_t = b_1\alpha^t \cos \phi t + b_2\alpha^t \sin \phi t,$$

where b_1 and b_2 are real constants. This result follows because we can choose e_2 to be the complex conjugate of e_1. Thus, if $e_1 = x + iy$, then $e_2 = x - iy$, and

$$\begin{aligned} Z_t &= (e_1 + e_2)(\alpha^t \cos \phi t) + (e_1 - e_2)i(\alpha^t \sin \phi t) \\ &= b_1\alpha^t \cos \phi t + b_2\alpha^t \sin \phi t, \end{aligned}$$

where b_1 and b_2 can be easily seen as real.

EXAMPLE 2.6 Let $Z_t - 2Z_{t-1} + Z_{t-2} = 0$. Find the closed form solution for Z_t.

The auxiliary equation is given by

$$C(B) = (1 - 2B + B^2) = (1 - B)^2 = 0.$$

$R^{-1} = 1$ is a root of multiplicity 2. Hence, by Theorem 2.7.1,

$$Z_t = (b_1 + b_2t)1^t = b_1 + b_2t.$$

EXAMPLE 2.7 Find the solution for $Z_t - 2Z_{t-1} + 1.5Z_{t-2} - .5Z_{t-3} = 0$.

The auxiliary equation is given by

$$C(B) = (1 - 2B + 1.5B^2 - .5B^3) = (1 - B + .5B^2)(1 - B) = 0.$$

Hence, $B_1 = R_1^{-1} = 1$ is a root of multiplicity one; and

$$B_2 = R_2^{-1} = \frac{1 + \sqrt{1 - 4(.5)}}{2(.5)} = 1 + i$$

and $B_3 = R_3^{-1} = 1 - i$ are a pair of complex roots of multiplicity one.

Now $R_1 = 1$ and $R_2 = (1 - i)/2$ and $R_3 = (1 + i)/2$. To express R_2 and R_3 in the polar form, we have from (2.7.4) and (2.7.5)

$$\alpha = (c^2 + d^2)^{1/2} = \sqrt{\frac{1}{2}}$$

$$\phi = \tan^{-1}\left(\frac{d}{c}\right) = \tan^{-1}(1) = \frac{\pi}{4}.$$

Hence, by the remarks following Theorem 2.7.1, we have

$$Z_t = b_1 + b_2\left(\sqrt{\frac{1}{2}}\right)^t \cos\left(\frac{\pi}{4}t\right) + b_3\left(\sqrt{\frac{1}{2}}\right)^t \sin\left(\frac{\pi}{4}t\right).$$

Remark The R_i used in the solution of $C(B) = 0$ in Theorem 2.7.1 is the inverse of the root B_i, i.e., $R_i = B_i^{-1}$. Thus, as shown in Example 2.7, to find R_i, we first find the root B_i of $C(B) = 0$ and then calculate its inverse. This cumbersome procedure can be avoided by noting that if we let $R = B^{-1}$ and multiply the equation

$$1 - C_1B - C_2B^2 - \cdots - C_nB^n = 0, \tag{2.7.6}$$

by $R^n = (B^{-1})^n$, we get

$$R^n - C_1R^{n-1} - C_2R^{n-2} - \cdots - C_n = 0. \tag{2.7.7}$$

It can be easily seen that B_i is a root of (2.7.6) if and only if R_i is a root of (2.7.7). Thus, sometimes we can compute $R_i = B_i^{-1}$ needed in Theorem 2.7.1 more easily from solving (2.7.7). The reader can easily verify that each R_i obtained in Example 2.7 is indeed a solution of $R^3 - 2R^2 + 1.5R - .5 = (R^2 - R + .5)(R - 1) = 0$.

EXAMPLE 2.8 Solve $(1 - \phi B)(1 - B)^2Z_t = 0$.

Because $R_1 = \phi$, and $R_2 = 1$ of multiplicity 2, we have from Theorem 2.7.1, $Z_t = b_1\phi^t + b_2 + b_3t$.

EXERCISES

2.1 Let Z_t, where t is even, be a sequence of independent random variables defined as $Z_t = +1$ or -1 with equal probability of 1/2, and $Z_t = Z_{t-1}$ if t is odd, where t is an integer.

(a) Is the process first order stationary in distribution?

(b) Is it second order stationary in distribution?

2.2 Let $Z_t = U \sin(2\pi t) + V \cos(2\pi t)$, where U and V are independent random variables, each with mean 0 and variance 1.

(a) Is Z_t strictly stationary?

(b) Is Z_t covariance stationary?

2.3 Prove or disprove that the following process is covariance stationary:

(a) $Z_t = A \sin(2\pi t + \theta)$, where A is a constant, and θ is a random variable that is uniformly distributed on $[0, 2\pi]$

(b) $Z_t = A \sin(2\pi t + \theta)$, where A is a random variable with zero mean and unit variance, and θ is a constant.

(c) $Z_t = (-1)^t A$, where A is a random variable with zero mean and unit variance.

2.4 Is the following a valid autocorrelation function for a real-valued covariance stationary process? Why?

$$\rho_k = \begin{cases} 1, & \text{if } k = 0, \\ \phi, & \frac{1}{2} < |\phi| < 1, \quad \text{if } |k| = 1, \\ 0, & \text{if } |k| \geq 2. \end{cases}$$

2.5 Verify the following properties for the autocorrelation function of a stationary process:

(a) $\rho_0 = 1$,

(b) $|\rho_k| \leq 1$,

(c) $\rho_k = \rho_{-k}$.

2.6 It is important to note that not every sequence satisfying the properties specified in Exercise 2.5 is an ACF of a process. Show that the following function satisfies the properties stated in Exercise 2.5, but is not an ACF for any stationary process:

$$\rho_k = \begin{cases} 1, & k = 0, \\ .8, & k = \pm 1, \\ .1, & k = \pm 2, \\ 0, & \text{otherwise.} \end{cases}$$

2.7 Given the time series 53, 43, 66, 48, 52, 42, 44, 56, 44, 58, 41, 54, 51, 56, 38, 56, 49, 52, 32, 52, 59, 34, 57, 39, 60, 40, 52, 44, 65, 43:

(a) Plot the series.

(b) Can you guess an approximate value for the first lag autocorrelation coefficient ρ_1 based on the plot of the series?

(c) Plot Z_t against Z_{t+1}, and try again to guess the value of ρ_1.

(d) Calculate and plot the sample ACF, $\hat{\rho}_k$, for $k = 0, 1, 2, 3, 4, 5$.
(e) Calculate and plot the sample PACF, $\hat{\phi}_{kk}$, for $k = 0, 1, 2, 3, 4, 5$.

2.8 Show that the estimate $\hat{\gamma}_k$ is always positive semi-definite, but $\hat{\hat{\gamma}}_k$ is not necessarily so.

2.9 Consider a stationary series with theoretical autocorrelation function

$$\rho_k = \phi^{|k|}, |\phi| < 1, \quad k = 1, 2, 3, \ldots.$$

Find the variance of $\hat{\rho}_k$ using Bartlett's approximation.

2.10 Let $Z_t = \mu + \sum_{j=0}^{\infty}\psi_j a_{t-j}$ or, equivalently, $Z_t = \mu + \psi(B)a_t$, where $\psi(B) = \sum_{j=0}^{\infty}\psi_j B^j$, $\psi_0 = 1$, $\sum_{j=0}^{\infty}|\psi_j| < \infty$, and the a_t are white noise with mean 0 and variance σ_a^2. Prove that $\sum_{j=-\infty}^{\infty}|\gamma_j| < \infty$, where γ_j is the jth autocovariance of the process Z_t.

2.11 Prove that $n\,\mathrm{Var}(\bar{z}) \to \sum_{k=-\infty}^{\infty}\gamma_k$ as $n \to \infty$, if $\sum_{k=-\infty}^{\infty}|\gamma_k| < \infty$.

2.12 Find a closed form solution for the following difference equations:
(a) $Z_t - 1.1Z_{t-1} + .3Z_{t-2} = 0$,
(b) $Z_t - Z_{t-1} + Z_{t-2} - Z_{t-3} = 0$,
(c) $Z_t - 1.8Z_{t-1} + .81Z_{t-2} = 0$.

3

Stationary Time Series Models

Limited by a finite number of available observations, we often construct a finite-order parametric model to describe a time series process. In this chapter, we introduce the autoregressive moving average model, which includes the autoregressive model and the moving average model as special cases. This model contains a very broad class of parsimonious time series processes found useful in describing a wide variety of time series. After giving detailed discussions on the characteristics of each process in terms of the autocorrelation and partial autocorrelation functions, we illustrate the results with examples.

3.1 Autoregressive Processes

As mentioned earlier in Section 2.6, in the autoregressive representation of a process, if only a finite number of π weights are nonzero, i.e., $\pi_1 = \phi_1, \pi_2 = \phi_2, \ldots, \pi_p = \phi_p$, and $\pi_k = 0$ for $k > p$, then the resulting process is said to be an autoregressive process (model) of order p, which is denoted as AR(p). It is given by

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t \tag{3.1.1}$$

or

$$\phi_p(B)\dot{Z}_t = a_t, \tag{3.1.2}$$

where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$, and $\dot{Z}_t = Z_t - \mu$.

Because $\sum_{j=1}^{\infty} |\pi_j| = \sum_{j=1}^{p} |\phi_j| < \infty$, the process is always invertible. To be stationary, the roots of $\phi_p(B) = 0$ must lie outside of the unit circle. The AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock. Yule (1927) used an AR process to describe the phenomena of sunspot numbers and the behavior of a simple pendulum. First, let us consider the following simple models.

3.1.1 The First-Order Autoregressive AR(1) Process

For the first-order autoregressive process AR(1), we write

$$(1 - \phi_1 B)\dot{Z}_t = a_t \tag{3.1.3a}$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t. \tag{3.1.3b}$$

As mentioned above, the process is always invertible. To be stationary, the root of $(1 - \phi_1 B) = 0$ must be outside of the unit circle. That is, for a stationary process, we have $|\phi_1| < 1$. The AR(1) process is sometimes called the Markov process because the distribution of $\dot{Z}_t$ given $\dot{Z}_{t-1}, \dot{Z}_{t-2}, \dot{Z}_{t-3}, \ldots$ is exactly the same as the distribution of $\dot{Z}_t$ given $\dot{Z}_{t-1}$.

ACF of the AR(1) Process The autocovariances are obtained as follows:

$$\begin{aligned} E(\dot{Z}_{t-k}\dot{Z}_t) &= E(\phi_1 \dot{Z}_{t-k}\dot{Z}_{t-1}) + E(\dot{Z}_{t-k}a_t) \\ \gamma_k &= \phi_1 \gamma_{k-1}, \qquad k \geq 1, \end{aligned} \tag{3.1.4}$$

and the autocorrelation function becomes

$$\rho_k = \phi_1 \rho_{k-1} = \phi_1^k, \qquad k \geq 1, \tag{3.1.5}$$

where we use that $\rho_0 = 1$. Hence, when $|\phi_1| < 1$ and the process is stationary, the ACF exponentially decays in one of two forms depending on the sign of ϕ_1. If $0 < \phi_1 < 1$, then all autocorrelations are positive; if $-1 < \phi_1 < 0$, then the sign of the autocorrelations shows an alternating pattern beginning with a negative value. The magnitudes of these autocorrelations decrease exponentially in both cases, as shown in Figure 3.1.

PACF of the AR(1) Process For an AR(1) process, the PACF from (2.3.19) is

$$\phi_{kk} = \begin{cases} \rho_1 = \phi_1, & k = 1, \\ 0, & \text{for } k \geq 2. \end{cases} \tag{3.1.6}$$

Hence, the PACF of the AR(1) process shows a positive or negative spike at lag 1 depending on the sign of ϕ_1 and then cuts off as shown in Figure 3.1.

EXAMPLE 3.1 For illustration, we simulated 250 values from an AR(1) process, $(1 - \phi_1 B)(Z_t - 10) = a_t$, with $\phi_1 = .9$. The white noise series a_t are independent normal $N(0, 1)$ random variables. Figure 3.2 shows the plot of the series. It is relatively smooth.

Table 3.1 and Figure 3.3 show the sample ACF and the sample PACF for the series. Clearly $\hat{\rho}_k$ decreases exponentially and $\hat{\phi}_{kk}$ cuts off after lag 1 because none of the sample

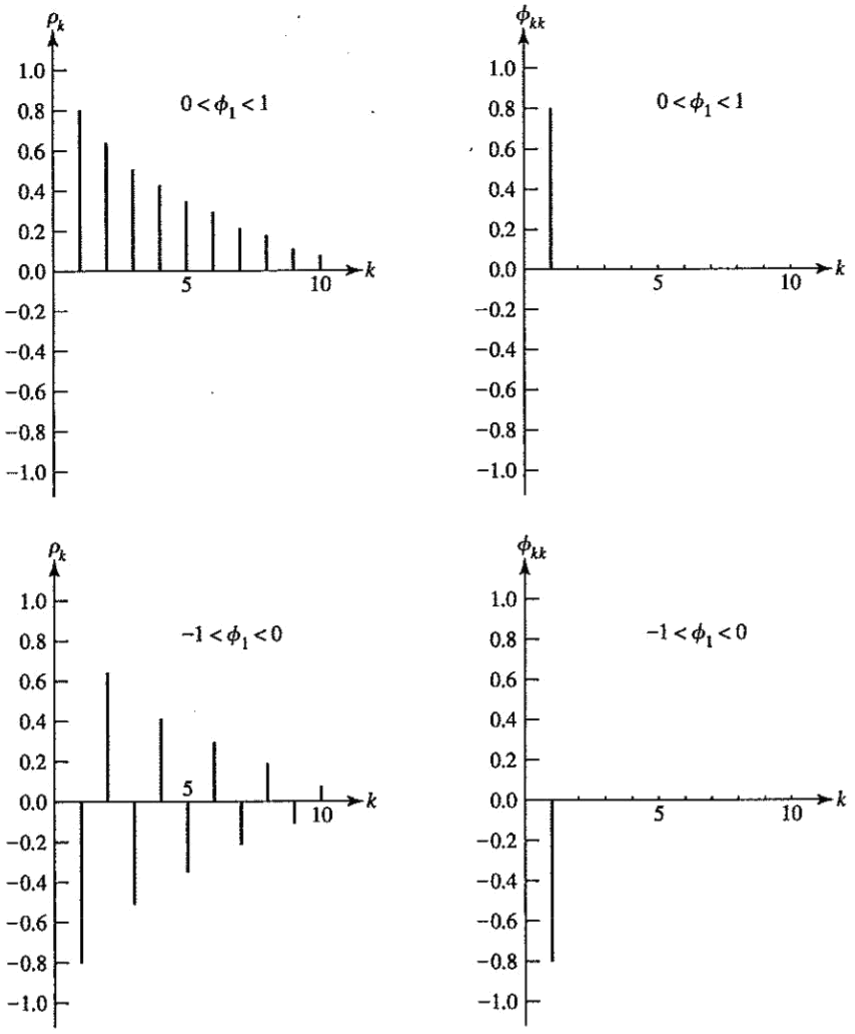


FIGURE 3.1 ACF and PACF of the AR(1) process: $(1 - \phi_1 B)\dot{Z}_t = a_t$.

PACF values is significant beyond that lag and, more important, these insignificant $\hat{\phi}_{kk}$ do not exhibit any pattern. The associated standard error of the sample ACF $\hat{\rho}_k$ is computed by

$$S_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_{k-1}^2)}, \tag{3.1.7}$$

and the standard error of the sample PACF $\hat{\phi}_{kk}$ is set to be

$$S_{\hat{\phi}_{kk}} \simeq \sqrt{\frac{1}{n}}, \tag{3.1.8}$$

which are standard outputs used in most time series software.

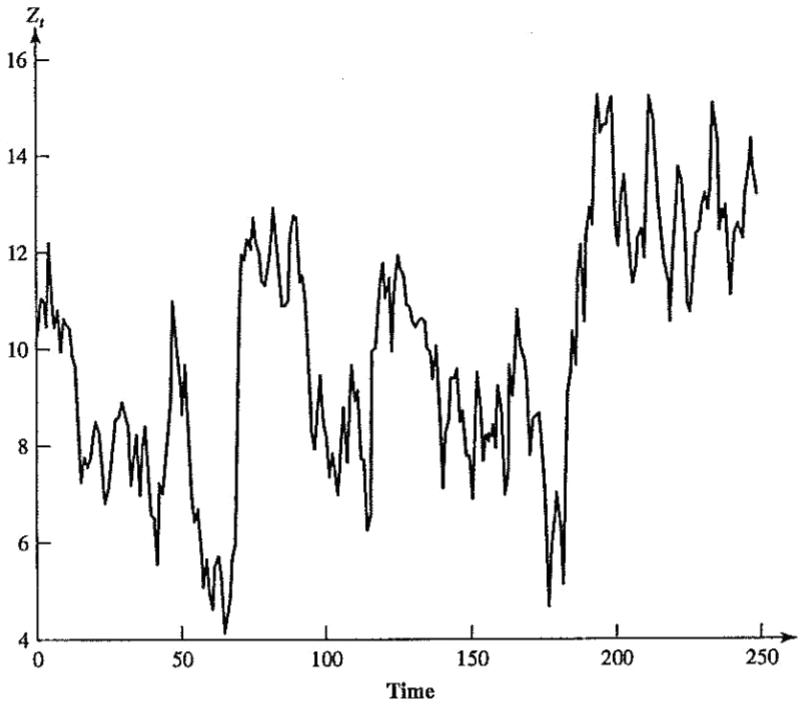


FIGURE 3.2 A simulated AR(1) series, $(1 - .9B)(Z_t - 10) = a_t$.

EXAMPLE 3.2 This example shows a simulation of 250 values from the AR(1) process $(1 - \phi_1 B)(Z_t - 10) = a_t$, with $\phi_1 = -.65$ and a_t being Gaussian $N(0, 1)$ white noise. The series is plotted in Figure 3.4 and is relatively jagged.

The sample ACF and sample PACF of the series are shown in Table 3.2 and Figure 3.5. We see the alternating decreasing pattern beginning with a negative in the sample ACF and the cutoff property of the sample PACF. Because $\hat{\phi}_{11} = \hat{\rho}_1$, $\hat{\phi}_{11}$ is also negative. Also, even though only the first two or three sample autocorrelations are significant, the overall pattern clearly indicates the phenomenon of an AR(1) model with a negative value of ϕ_1.

TABLE 3.1 Sample ACF and sample PACF for a simulated series from $(1 - .9B)(Z_t - 10) = a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.88	.76	.67	.57	.48	.40	.34	.28	.21	.17
St.E.	.06	.10	.12	.14	.14	.15	.16	.16	.16	.16
$\hat{\phi}_{kk}$	.88	.01	−.01	−.11	.02	−.01	.01	−.02	−.06	.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

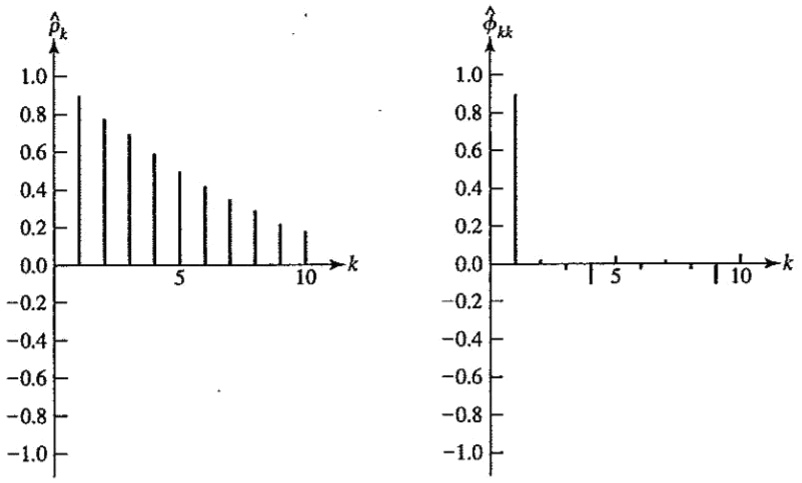


FIGURE 3.3 Sample ACF and sample PACF of a simulated AR(1) series: $(1 - .9B)(Z_t - 10) = a_t$.

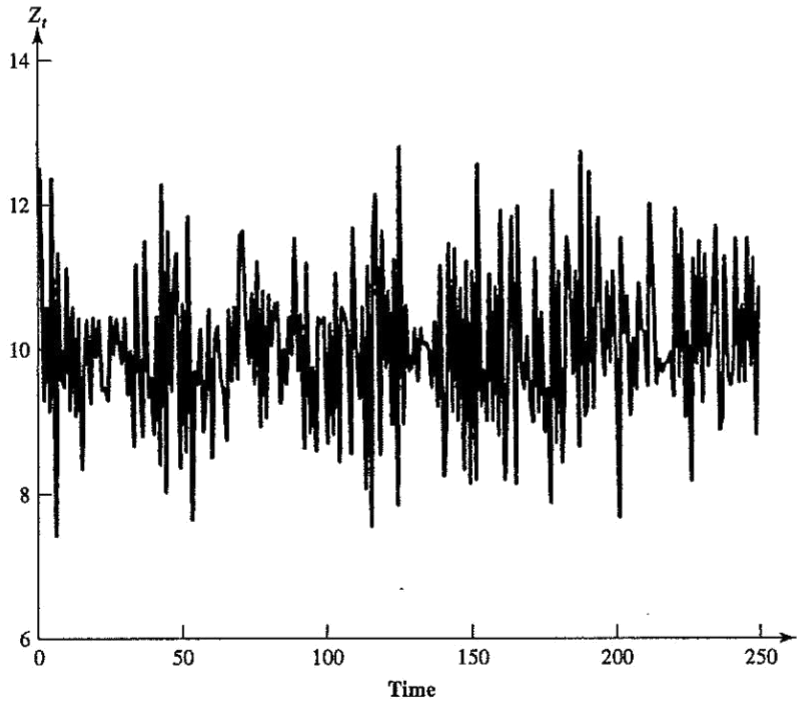


FIGURE 3.4 A simulated AR(1) series $(1 + .65B)(Z_t - 10) = a_t$.

TABLE 3.2 Sample ACF and sample PACF for a simulated series from $(1 + .65B)(Z_t - 10) = a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.63	.36	−.17	.09	−.07	.06	−.08	.10	−.11	.06
St.E.	.06	.08	.09	.09	.09	.09	.09	.09	.09	.09
$\hat{\phi}_{kk}$	−.63	−.06	.05	.02	−.04	−.01	−.06	.04	−.03	−.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

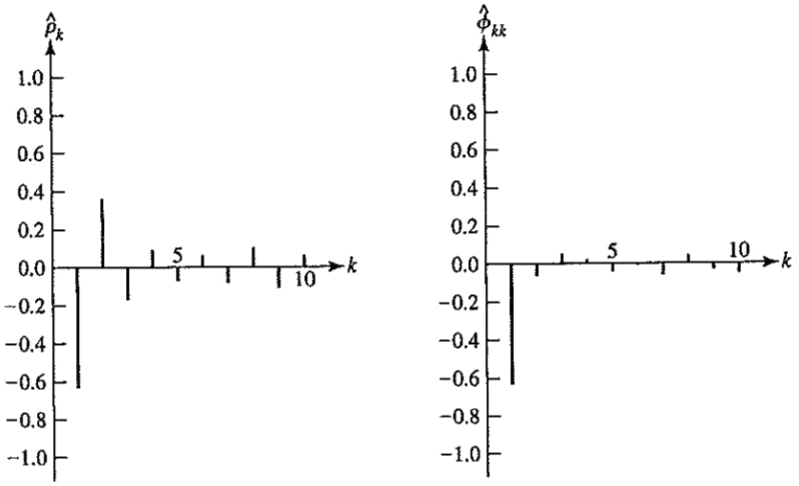


FIGURE 3.5 Sample ACF and sample PACF of a simulated AR(1) series $(1 + .65B)(Z_t - 10) = a_t$.

In discussing stationary autoregressive processes, we have assumed that the zeros of the autoregressive polynomial $\phi_p(B)$ lie outside of the unit circle. In terms of the AR(1) process (3.1.3a or b), it implies that $|\phi_1| < 1$. Thus, when $|\phi_1| \geq 1$, the process is regarded as nonstationary because we have implicitly assumed that the process is expressed as a linear combination of present and past white noise variables. If we also consider a process that is expressed as a linear combination of present and future random shocks, then there exists an AR(1) process with its parameter ϕ_1 greater than 1 in absolute value, which is still stationary in the usual sense of the term as defined in Section 2.1. To see that, consider the process

$$Z_t = \sum_{j=0}^{\infty} (.5)^j a_{t+j}, \tag{3.1.9}$$

where $\{a_t\}$ is a white noise process with mean zero and variance σ_a^2. It is straightforward to verify that the process Z_t in (3.1.9) is indeed stationary in the sense of Section 2.1 with the

ACF, $\rho_k = (.5)^{|k|}$. Now, consider the process (3.1.9) at time $(t-1)$ and multiply both of its sides by 2, i.e.,

$$\begin{aligned} 2Z_{t-1} &= 2\sum_{j=0}^{\infty}(.5)^j a_{t-1+j} \\ &= 2a_{t-1} + \sum_{j=1}^{\infty}(.5)^{j-1}a_{t-1+j} \\ &= 2a_{t-1} + \sum_{j=0}^{\infty}(.5)^j a_{t+j}. \end{aligned} \tag{3.1.10}$$

Thus, (3.1.9) leads to the following equivalent AR(1) model with $\phi_1 = 2$,

$$Z_t - 2Z_{t-1} = b_t, \tag{3.1.11}$$

where $b_t = -2a_{t-1}$. Note, however, that although the b_t in (3.1.11) is a white noise process with mean zero, its variance becomes $4\sigma_a^2$, which is four times larger than the variance of a_t in the following AR(1) model with the same ACF, $\rho_k = (.5)^{|k|}$,

$$Z_t - .5Z_{t-1} = a_t, \tag{3.1.12}$$

which can be written as a linear combination of present and past random shocks, i.e., $Z_t = \sum_{j=0}^{\infty}(.5)^j a_{t-j}$.

In summary, although a process with an ACF of the form $\phi^{|k|}$, where $|\phi| < 1$, can be written either as

$$Z_t - \phi Z_{t-1} = a_t \tag{3.1.13}$$

or

$$Z_t - \phi^{-1}Z_{t-1} = b_t, \tag{3.1.14}$$

where both a_t and b_t are zero mean white noise processes, the variance of b_t in (3.1.14) is larger than the variance of a_t in (3.1.13) by a factor of ϕ^{-2}. Thus, for practical purposes, we will choose the representation (3.1.13). That is, in terms of a stationary AR(1) process, we always refer to the case in which the parameter value is less than 1 in absolute value.

3.1.2 The Second-Order Autoregressive AR(2) Process

For the second-order autoregressive AR(2) process, we have

$$(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t \tag{3.1.15a}$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + a_t. \tag{3.1.15b}$$

The AR(2) process, as a finite autoregressive model, is always invertible. To be stationary, the roots of $\phi(B) = (1 - \phi_1 B - \phi_2 B^2) = 0$ must lie outside of the unit circle. For example, the process $(1 - 1.5B + .56B^2)Z_t = a_t$ is stationary because $(1 - 1.5B + .56B^2) = (1 - .7B)(1 - .8B) = 0$ gives $B = 1/.7$ and $B = 1/.8$ as the two roots, which are larger than 1 in absolute value. Yet $(1 - .2B - .8B^2)Z_t = a_t$ is not stationary because one of the roots of $(1 - .2B - .8B^2) = 0$ is $B = 1$, which is not outside of the unit circle.

The stationarity condition of the AR(2) model can also be expressed in terms of its parameter values. Let B_1 and B_2 be the roots of $(1 - \phi_1 B - \phi_2 B^2) = 0$ or, equivalently, of $\phi_2 B^2 + \phi_1 B - 1 = 0$. We have

$$B_1 = \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

and

$$B_2 = \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

Now,

$$\frac{1}{B_1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

and

$$\frac{1}{B_2} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

The required condition $|B_i| > 1$ implies that $|1/B_i| < 1$ for $i = 1$ and 2. Hence,

$$\left|\frac{1}{B_1} \cdot \frac{1}{B_2}\right| = |\phi_2| < 1$$

and

$$|\phi_1| = \left|\frac{1}{B_1} + \frac{1}{B_2}\right| < 2.$$

Thus, we have the following necessary condition for stationarity regardless of whether the roots are real or complex:

$$\begin{cases} -1 < \phi_2 < 1, \\ -2 < \phi_1 < 2. \end{cases} \tag{3.1.16}$$

For real roots, we need $\phi_1^2 + 4\phi_2 \geq 0$, which implies that

$$-1 < \frac{1}{B_2} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \leq \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} = \frac{1}{B_1} < 1,$$

or, equivalently,

$$\begin{cases} \phi_2 + \phi_1 < 1, \\ \phi_2 - \phi_1 < 1. \end{cases} \tag{3.1.17}$$

For complex roots, we have $\phi_2 < 0$ and $\phi_1^2 + 4\phi_2 < 0$. Thus, in terms of the parameter values, the stationarity condition of the AR(2) model is given by the following triangular region in Figure 3.6 satisfying

$$\begin{cases} \phi_2 + \phi_1 < 1, \\ \phi_2 - \phi_1 < 1, \\ -1 < \phi_2 < 1. \end{cases} \tag{3.1.18}$$

ACF of the AR(2) Process We obtain the autocovariances by multiplying Z_{t-k} on both sides of (3.1.15b) and taking the expectation

$$E(\dot{Z}_{t-k}\dot{Z}_t) = \phi_1 E(\dot{Z}_{t-k}\dot{Z}_{t-1}) + \phi_2 E(\dot{Z}_{t-k}\dot{Z}_{t-2}) + E(\dot{Z}_{t-k}a_t)$$
$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, \qquad k \geq 1.$$

Hence, the autocorrelation function becomes

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \qquad k \geq 1. \tag{3.1.19}$$

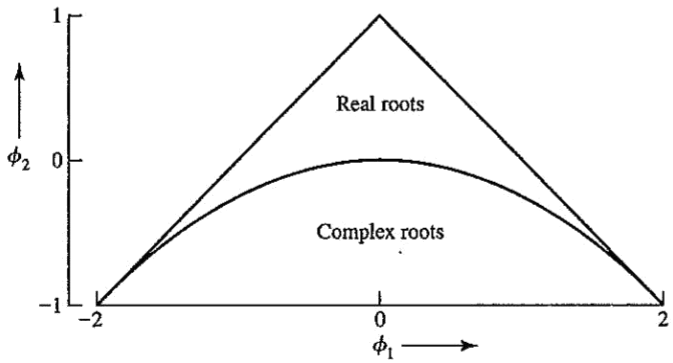


FIGURE 3.6 Stationary regions for the AR(2) model.

Specifically, when $k = 1$ and 2

$$\rho_1 = \phi_1 + \phi_2\rho_1$$
$$\rho_2 = \phi_1\rho_1 + \phi_2,$$

which implies that

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{3.1.20}$$

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2}, \tag{3.1.21}$$

and ρ_k for $k \geq 3$ is calculated recursively through (3.1.19).

The pattern of the ACF is governed by the difference equation given by (3.1.19), namely $(1 - \phi_1 B - \phi_2 B^2)\,\rho_k = 0$. Using Theorem 2.7.1, we obtain

$$\rho_k = \begin{cases} b_1\left[\dfrac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k + b_2\left[\dfrac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k, & \text{if } \phi_1^2 + 4\phi_2 \neq 0, \\ (b_1 + b_2 k)\left[\dfrac{\phi_1}{2}\right]^k, & \text{if } \phi_1^2 + 4\phi_2 = 0, \end{cases} \tag{3.1.22}$$

where the constants b_1 and b_2 can be solved using the initial conditions given in (3.1.20) and (3.1.21). Thus, the ACF will be an exponential decay if the roots of $(1 - \phi_1 B - \phi_2 B^2) = 0$ are real and a damped sine wave if the roots of $(1 - \phi_1 B - \phi_2 B^2) = 0$ are complex.

The AR(2) process was originally used by G. U. Yule in 1921 to describe the behavior of a simple pendulum. Hence, the process is also sometimes called the Yule process.

PACF of the AR(2) Process For the AR(2) process, because

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$$

for $k \geq 1$ as shown in (3.1.19), we have, from (2.3.19),

$$\phi_{11} = \rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{3.1.23a}$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$= \frac{\left(\frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2}\right) - \left(\frac{\phi_1}{1 - \phi_2}\right)^2}{1 - \left(\frac{\phi_1}{1 - \phi_2}\right)^2} \tag{3.1.23b}$$

$$= \frac{\phi_2[(1 - \phi_2)^2 - \phi_1^2]}{(1 - \phi_2)^2 - \phi_1^2} = \phi_2$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{3.1.23c}$$

$$= \frac{\begin{vmatrix} 1 & \rho_1 & \phi_1 + \phi_2\rho_1 \\ \rho_1 & 1 & \phi_1\rho_1 + \phi_2 \\ \rho_2 & \rho_1 & \phi_1\rho_2 + \phi_2\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0$$

because the last column of the numerator is a linear combination of the first two columns. Similarly, we can show that $\phi_{kk} = 0$ for $k \geq 3$. Hence, the PACF of an AR(2) process cuts off after lag 2. Figure 3.7 illustrates the PACF and corresponding ACF for a few selected AR(2) processes.

EXAMPLE 3.3 Table 3.3 and Figure 3.8 show the sample ACF and the sample PACF for a series of 250 values simulated from the AR(2) process $(1 + .5B - .3B^2)Z_t = a_t$, with the a_t being Gaussian $N(0, 1)$ white noise. The oscillating pattern of the ACF is similar to that of an AR(1) model with a negative parameter value. The rate of the decreasing of the autocorrelations, however, rejects the possibility of being an AR(1) model. That $\hat{\phi}_{kk}$ cuts off after lag 2, on the other hand, indicates an AR(2) model.

EXAMPLE 3.4 To consider an AR(2) model with the associated polynomial having complex roots, we simulated a series of 250 values from $(1 - B + .5B^2)Z_t = a_t$, with the a_t being Gaussian $N(0, 1)$ white noise. Table 3.4 and Figure 3.9 show the sample ACF and the sample PACF of this series. The sample ACF exhibits a damped sine wave, and the sample PACF cuts off after lag 2. Both give a fairly clear indication of an AR(2) model.

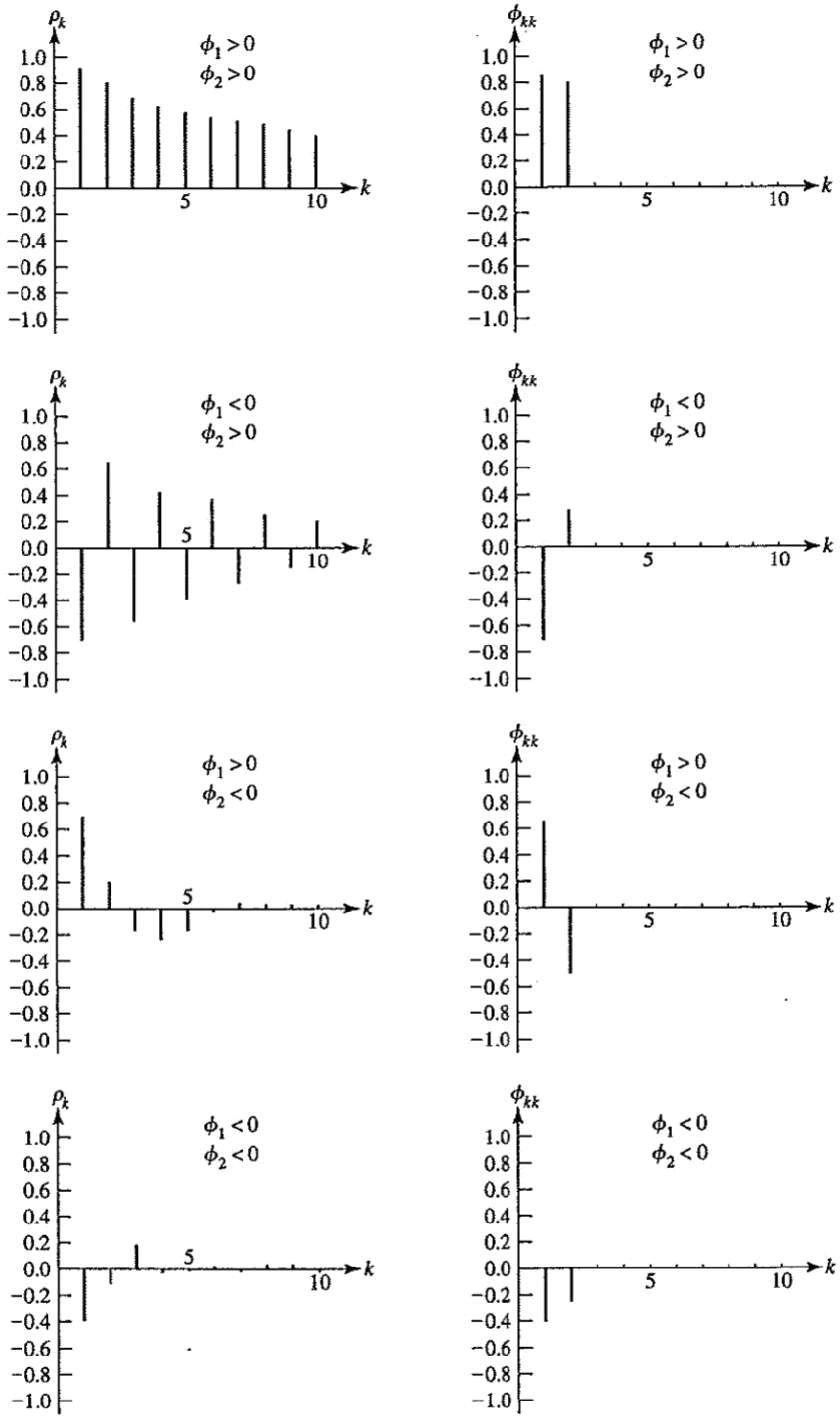


FIGURE 3.7 ACF and PACF of AR(2) process: $(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t$.

TABLE 3.3 Sample ACF and sample PACF for a simulated series from $(1 + .5B - .3B^2)Z_t = a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.70	.62	−.48	.41	−.37	.32	−.30	.27	−.25	.20
St.E.	.06	.09	.11	.11	.12	.12	.13	.13	.13	.13
$\hat{\phi}_{kk}$	−.70	.26	.05	.03	−.08	.00	−.04	.03	−.01	−.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

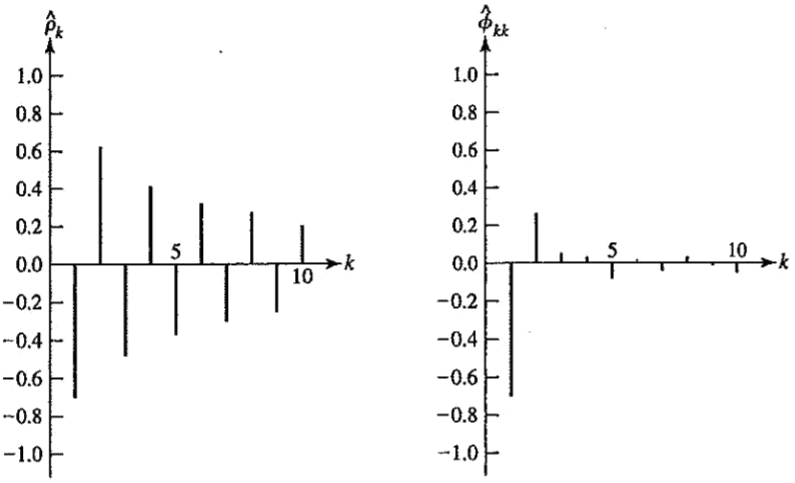


FIGURE 3.8 Sample ACF and sample PACF of a simulated AR(2) series: $(1 + .5B - .3B^2)Z_t = a_t$.

3.1.3 The General pth-Order Autoregressive AR(p) Process

The pth-order autoregressive process AR(p) is

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\dot{Z}_t = a_t \tag{3.1.24a}$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t \tag{3.1.24b}$$

ACF of the General AR(p) Process To find the autocovariance function, we multiply Z_{t-k} on both sides of (3.1.24b)

$$\dot{Z}_{t-k}\dot{Z}_t = \phi_1 \dot{Z}_{t-k}\dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-k}\dot{Z}_{t-p} + \dot{Z}_{t-k}a_t$$

and take the expected value

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}, \qquad k > 0, \tag{3.1.25}$$

TABLE 3.4 Sample ACF and sample PACF for a simulated series from $(1 - B + .5B^2)Z_t = a_t$.

k						$\hat{\rho}_k$					
1–12	.67	.20	−.13	−.26	−.22	−.09	.02	.08	.06	.00	−.10	−.17
St.E.	.06	.09	.09	.09	.09	.09	.09	.09	.10	.10	.10	.10
12–24	−.13	−.04	.07	.13	.10	.03	−.05	−.07	−.09	−.13	−.12	−.09
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
k						$\hat{\phi}_{kk}$					
1–12	.67	−.45	−.04	−.08	.05	−.01	.03	−.01	−.04	−.01	−.13	−.03
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
12–24	.06	−.04	.09	−.02	−.04	.01	−.02	.03	−.12	−.07	−.03	−.03
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

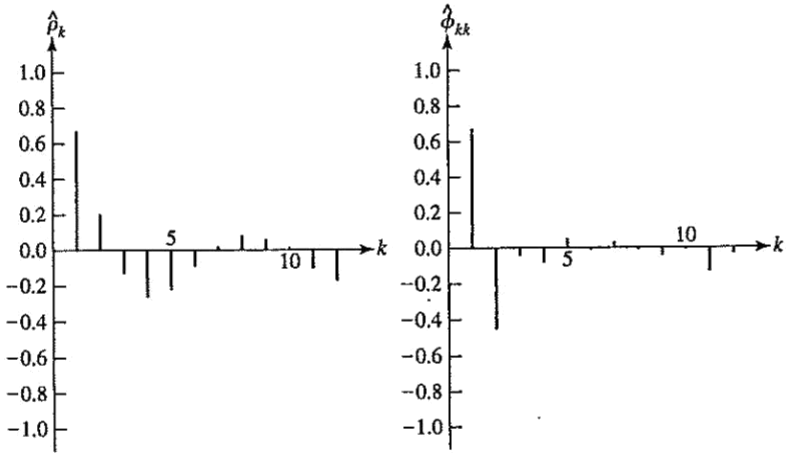


FIGURE 3.9 Sample ACF and sample PACF of a simulated AR(2) series: $(1 - B + .5B^2)Z_t = a_t$.

where we recall that $E(a_t Z_{t-k}) = 0$ for $k > 0$. Hence, we have the following recursive relationship for the autocorrelation function:

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}, \qquad k > 0. \tag{3.1.26}$$

From (3.1.26) we see that the ACF, ρ_k, is determined by the difference equation $\phi_p(B)\rho_k = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\rho_k = 0$ for $k > 0$. Now, we can write

$$\phi_p(B) = \prod_{i=1}^{m} (1 - G_i B)^{d_i},$$

where $\sum_{i=1}^{m} d_i = p$, and G_i^{-1} $(i = 1, 2, \ldots, m)$ are the roots of multiplicity d_i of $\phi_p(B) = 0$. Using the difference equation result in Theorem 2.7.1, we have

$$\rho_k = \sum_{i=1}^{m} \sum_{j=0}^{d_i - 1} b_{ij} k^j G_i^k . \tag{3.1.27}$$

If $d_i = 1$ for all i, then G_i^{-1} are all distinct and the above reduces to

$$\rho_k = \sum_{i=1}^{p} b_i G_i^k , \qquad k > 0. \tag{3.1.28}$$

For a stationary process, $|G_i^{-1}| > 1$ and $|G_i| < 1$. Hence, the ACF ρ_k tails off as a mixture of exponential decays or damped sine waves depending on the roots of $\phi_p(B) = 0$. Damped sine waves appear if some of the roots are complex.

PACF of the General AR(p) Process Because $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$ for $k > 0$, we can easily see that when $k > p$ the last column of the matrix in the numerator of ϕ_{kk} in (2.3.19) can be written as a linear combination of previous columns of the same matrix. Hence, the PACF ϕ_{kk} will vanish after lag p. This property is useful in identifying an AR model as a generating process for a time series discussed in Chapter 6.

3.2 Moving Average Processes

In the moving average representation of a process, if only a finite number of ψ weights are nonzero, i.e., $\psi_1 = -\theta_1$, $\psi_2 = -\theta_2, \ldots, \psi_q = -\theta_q$, and $\psi_k = 0$ for $k > q$, then the resulting process is said to be a moving average process or model of order q and is denoted as MA(q). It is given by

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{3.2.1a}$$

or

$$\dot{Z}_t = \theta(B) a_t, \tag{3.2.1b}$$

where

$$\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q).$$

Because $1 + \theta_1^2 + \cdots + \theta_q^2 < \infty$, a finite moving average process is always stationary. This moving average process is invertible if the roots of $\theta(B) = 0$ lie outside of the unit circle. Moving average processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time. The process arose as a result of the study by Slutzky (1927) on the effect of the moving average of random events. To discuss other properties of the MA(q) process, let us first consider the following simpler cases.

3.2.1 The First-Order Moving Average MA(1) Process

When $\theta(B) = (1 - \theta_1 B)$, we have the first-order moving average MA(1) process

$$\begin{aligned} \dot{Z}_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B)a_t, \end{aligned} \tag{3.2.2}$$

where $\{a_t\}$ is a zero mean white noise process with constant variance σ_a^2. The mean of $\{\dot{Z}_t\}$ is $E(\dot{Z}_t) = 0$, and hence $E(Z_t) = \mu$.

ACF of the MA(1) Process The autocovariance generating function of a MA(1) process is, using (2.6.9),

$$\gamma(B) = \sigma_a^2(1 - \theta_1 B)(1 - \theta_1 B^{-1}) = \sigma_a^2\{-\theta_1 B^{-1} + (1 + \theta_1^2) - \theta_1 B\}.$$

Hence, the autocovariances of the process are

$$\gamma_k = \begin{cases} (1 + \theta_1^2)\sigma_a^2, & k = 0, \\ -\theta_1\sigma_a^2, & k = 1, \\ 0, & k > 1. \end{cases} \tag{3.2.3}$$

The autocorrelation function becomes

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2}, & k = 1, \\ 0, & k > 1, \end{cases} \tag{3.2.4}$$

which cuts off after lag 1, as shown in Figure 3.10.

Because $1 + \theta_1^2$ is always bounded, the MA(1) process is always stationary. For the process to be invertible, however, the root of $(1 - \theta_1 B) = 0$ must lie outside the unit circle. Because the root $B = 1/\theta_1$, we require $|\theta_1| < 1$ for an invertible MA(1) process.

Two remarks are in order.

1. Both the process $\dot{Z}_t = (1 - .4B)a_t$ and the process $\dot{Z}_t = (1 - 2.5B)a_t$ have the same autocorrelation function

$$\rho_k = \begin{cases} \dfrac{-1}{2.9}, & k = 1, \\ 0, & k > 1. \end{cases}$$

   In fact, more generally, for any θ_1, $\dot{Z}_t = (1 - \theta_1 B)a_t$ and $\dot{Z}_t = (1 - 1/\theta_1 B)a_t$ have the same autocorrelations. If the root of $(1 - \theta_1 B)$ lies outside the unit circle, however, then the root of $(1 - 1/\theta_1 B) = 0$ lies inside the unit circle, and vice versa. In other words, among the two processes that produce the same autocorrelations, one and only one is invertible. Thus, for uniqueness and meaningful implication for forecasting discussed in Chapter 5, we restrict ourselves to an invertible process in the model selections.
2. From (3.2.4), it is easy to see that $2|\rho_k| < 1$. Hence, for an MA(1) process, $|\rho_k| < .5$.

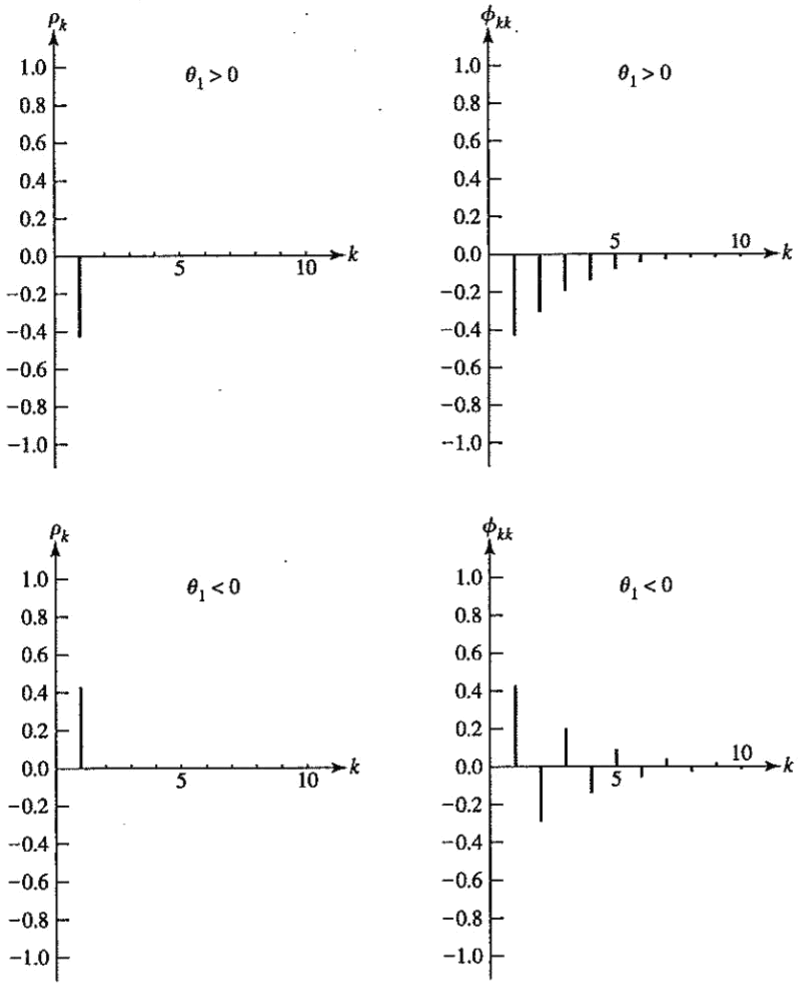


FIGURE 3.10 ACF and PACF of MA(1) processes: $\dot{Z}_t = (1 - \theta_1 B)a_t$.

PACF of the MA(1) Process Using (2.3.19) and (3.2.4), the PACF of an MA(1) process can be easily seen to be

$$\phi_{11} = \rho_1 = \frac{-\theta_1}{1 + \theta_1^2} = \frac{-\theta_1(1 - \theta_1^2)}{1 - \theta_1^4}$$

$$\phi_{22} = -\frac{\rho_1^2}{1 - \rho_1^2} = \frac{-\theta_1^2}{1 + \theta_1^2 + \theta_1^4} = \frac{-\theta_1^2(1 - \theta_1^2)}{1 - \theta_1^6}$$

$$\phi_{33} = \frac{\rho_1^3}{1 - 2\rho_1^2} = \frac{-\theta_1^3}{1 + \theta_1^2 + \theta_1^4 + \theta_1^6} = \frac{-\theta_1^3(1 - \theta_1^2)}{(1 - \theta_1^8)}.$$

In general,

$$\phi_{kk} = \frac{-\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \qquad \text{for } k \geq 1. \tag{3.2.5}$$

Contrary to its ACF, which cuts off after lag 1, the PACF of an MA(1) model tails off exponentially in one of two forms depending on the sign of θ_1 (hence on the sign of ρ_1). If alternating in sign, then it begins with a positive value; otherwise, it decays on the negative side, as shown in Figure 3.10. We also note that $|\phi_{kk}| < \frac{1}{2}$.

EXAMPLE 3.5 The sample ACF and sample PACF are calculated from a series of 250 values simulated from the MA(1) model $Z_t = (1 - .5B)a_t$, using a_t as Gaussian $N(0, 1)$ white noise. They are shown in Table 3.5 and plotted in Figure 3.11. Statistically, only one autocorrelation $\hat{\rho}_1$ and two partial autocorrelations $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ are significant. From the overall pattern, however, $\hat{\rho}_k$ clearly cuts off after lag 1 and $\hat{\phi}_{kk}$ tails off, which indicate a clear MA(1) phenomenon.

TABLE 3.5 Sample ACF and sample PACF for a simulated series from $Z_t = (1 - .5B)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.44	.00	.02	−.03	−.01	−.05	.04	−.03	−.03	.02
St.E.	.06	.07	.07	.07	.07	.07	.07	.07	.08	.08
$\hat{\phi}_{kk}$	−.44	−.24	−.11	−.08	−.07	−.12	−.06	−.07	−.10	−.08
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

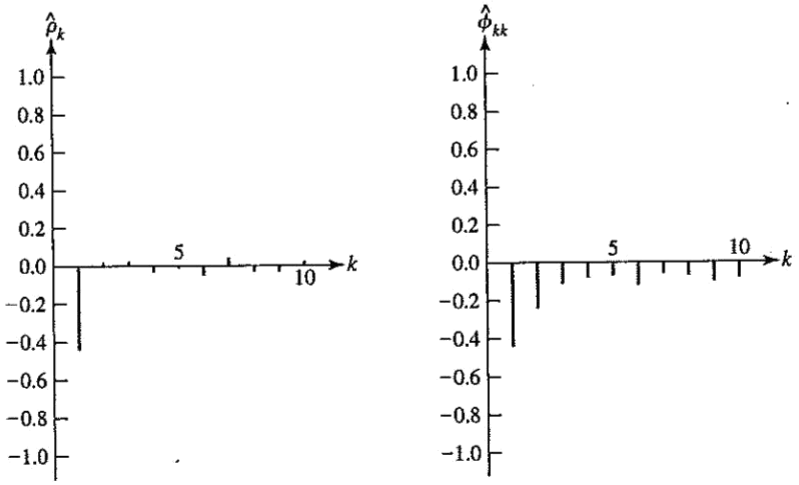


FIGURE 3.11 Sample ACF and sample PACF of a simulated MA(1) series: $Z_t = (1 - .5B)a_t$.

3.2.2 The Second-Order Moving Average MA(2) Process

When $\theta(B) = (1 - \theta_1 B - \theta_2 B^2)$, we have the second-order moving average process

$$\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2)a_t, \tag{3.2.6}$$

where $\{a_t\}$ is a zero mean white noise process. As a finite-order moving average model, the MA(2) process is always stationary. For invertibility, the roots of $(1 - \theta_1 B - \theta_2 B^2) = 0$ must lie outside of the unit circle. Hence,

$$\begin{cases} \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1, \end{cases} \tag{3.2.7}$$

which is parallel to the stationary condition of the AR(2) model, as shown in (3.1.18).

ACF of the MA(2) Process The autocovariance generating function via (2.6.9) is

$$\begin{aligned} \gamma(B) &= \sigma_a^2(1 - \theta_1 B - \theta_2 B^2)(1 - \theta_1 B^{-1} - \theta_2 B^{-2}) \\ &= \sigma_a^2\{-\theta_2 B^{-2} - \theta_1(1 - \theta_2)B^{-1} + (1 + \theta_1^2 + \theta_2^2) - \theta_1(1 - \theta_2)B - \theta_2 B^2\}. \end{aligned}$$

Hence, the autocovariances of the MA(2) model are

$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2 + \theta_2^2)\sigma_a^2, \\ \gamma_1 &= -\theta_1(1 - \theta_2)\sigma_a^2, \\ \gamma_2 &= -\theta_2\sigma_a^2, \end{aligned}$$

and

$$\gamma_k = 0, \qquad k > 2.$$

The autocorrelation function is

$$\rho_k = \begin{cases} \dfrac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, & k = 1, \\ \dfrac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2, \\ 0, & k > 2, \end{cases} \tag{3.2.8}$$

which cuts off after lag 2.

PACF of the MA(2) Process From (2.3.19), using that $\rho_k = 0$ for $k \geq 3$, we obtain

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}$$

$$\vdots$$

The MA(2) process contains the MA(1) process as a special case. Hence, the PACF tails off as an exponential decay or a damped sine wave depending on the signs and magnitudes of θ_1 and θ_2 or, equivalently, the roots of $(1 - \theta_1 B - \theta_2 B^2) = 0$. The PACF will be damped sine wave if the roots of $(1 - \theta_1 B - \theta_2 B^2) = 0$ are complex. They are shown in Figure 3.12 together with the corresponding ACF.

EXAMPLE 3.6 A series of 250 values is simulated from the MA(2) process $Z_t = (1 - .65B - .24B^2)a_t$ with a Gaussian $N(0, 1)$ white noise series a_t. The sample ACF and sample PACF are in Table 3.6 and plotted in Figure 3.13. We see that $\hat{\rho}_k$ clearly cuts off after lag 2 and $\hat{\phi}_{kk}$ tails off as expected for an MA(2) process.

3.2.3 The General qth-Order Moving Average MA(q) Process

The general qth-order moving average process is

$$\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t. \tag{3.2.9}$$

For this general MA(q) process, the variance is

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{q} \theta_j^2, \tag{3.2.10}$$

where $\theta_0 = 1$, and the other autocovariances are

$$\gamma_k = \begin{cases} \sigma_a^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q), & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases} \tag{3.2.11}$$

Hence, the autocorrelation function becomes

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases} \tag{3.2.12}$$

The autocorrelation function of an MA(q) process cuts off after lag q. This important property enables us to identify whether a given time series is generated by a moving average process.

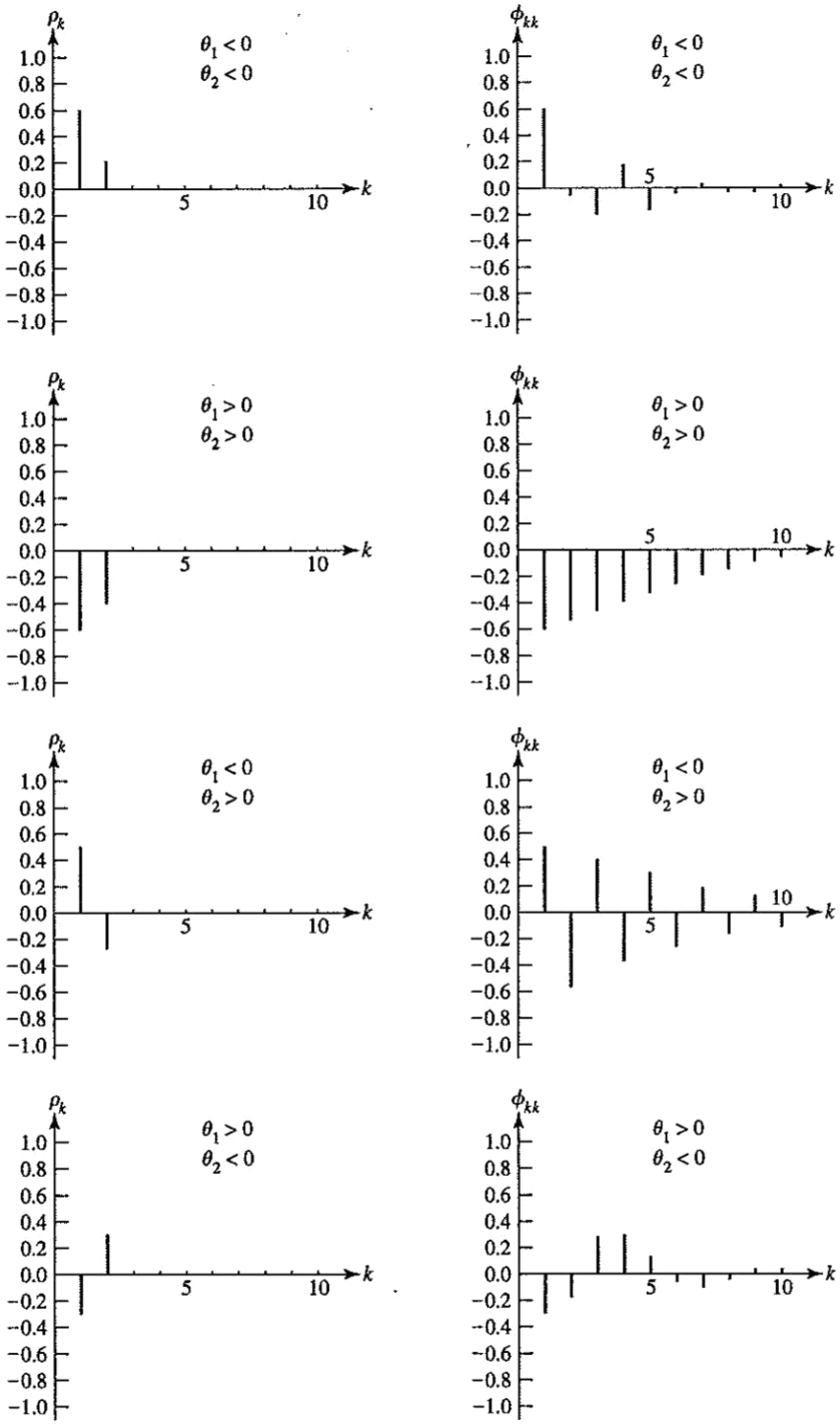


FIGURE 3.12 ACF and PACF of MA(2) processes: $Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$.

TABLE 3.6 Sample ACF and sample PACF for a simulated MA(2) series from $Z_t = (1 - .65B - .24B^2)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.35	−.17	.09	−.06	.01	−.01	−.04	.07	−.07	.09
St.E.	.06	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	−.35	−.34	−.15	−.18	−.11	−.12	−.14	−.05	−.14	.00
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

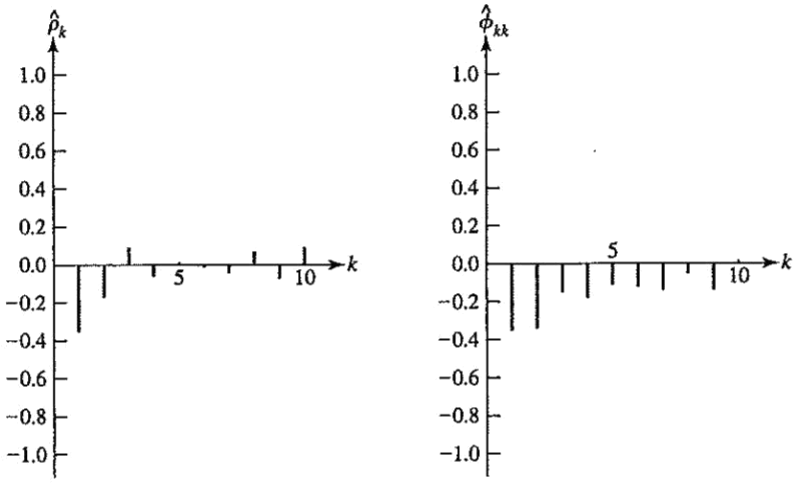


FIGURE 3.13 Sample ACF and sample PACF of a simulated MA(2) series: $Z_t = (1 - .65B - .24B^2)a_t$.

From the discussion of MA(1) and MA(2) processes, we can easily see that the partial autocorrelation function of the general MA(q) process tails off as a mixture of exponential decays and/or damped sine waves depending on the nature of the roots of $(1 - \theta_1 B - \cdots - \theta_q B^q) = 0$. The PACF will contain damped sine waves if some of the roots are complex.

3.3 The Dual Relationship Between AR(p) and MA(q) Processes

For a given stationary AR(p) process,

$$\phi_p(B)\dot{Z}_t = a_t, \tag{3.3.1}$$

where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$, we can write

$$\dot{Z}_t = \frac{1}{\phi_p(B)} a_t = \psi(B)a_t, \tag{3.3.2}$$

with $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots)$ such that

$$\phi_p(B)\psi(B) = 1. \tag{3.3.3}$$

The ψ weights can be derived by equating the coefficients of B^j on both sides of (3.3.3). For example, we can write the AR(2) process as

$$\dot{Z}_t = \frac{1}{(1 - \phi_1 B - \phi_2 B^2)} a_t = (1 + \psi_1 B + \psi_2 B^2 + \cdots) a_t, \tag{3.3.4}$$

which implies that

$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots) = 1,$$

i.e.,

$$\begin{aligned} &1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots \\ &\quad - \phi_1 B - \psi_1 \phi_1 B^2 - \psi_2 \phi_1 B^3 - \cdots \\ &\qquad - \phi_2 B^2 - \psi_1 \phi_2 B^3 - \cdots = 1. \end{aligned}$$

Thus, we obtain the ψ_j's as follows:

$$\begin{aligned} B^1&: \quad \psi_1 - \phi_1 = 0 \longrightarrow \psi_1 = \phi_1 \\ B^2&: \quad \psi_2 - \psi_1\phi_1 - \phi_2 = 0 \longrightarrow \psi_2 = \psi_1\phi_1 + \phi_2 = \phi_1^2 + \phi_2 \\ B^3&: \quad \psi_3 - \psi_2\phi_1 - \psi_1\phi_2 = 0 \longrightarrow \psi_3 = \psi_2\phi_1 + \psi_1\phi_2 \\ &\vdots \end{aligned}$$

Actually, for $j \geq 2$, we have

$$\psi_j = \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2, \tag{3.3.5}$$

where $\psi_0 = 1$. In a special case when $\phi_2 = 0$, we have $\psi_j = \phi_1^j$ for $j \geq 0$. Therefore,

$$\dot{Z}_t = \frac{1}{(1 - \phi_1 B)} a_t = (1 + \phi_1 B + \phi_1^2 B^2 + \cdots) a_t. \tag{3.3.6}$$

This equation implies that a finite-order stationary AR process is equivalent to an infinite-order MA process.

Given a general invertible MA(q) process,

$$\dot{Z}_t = \theta_q(B) a_t \tag{3.3.7}$$

with $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$, we can rewrite it as

$$\pi(B)\dot{Z}_t = \frac{1}{\theta_q(B)}\dot{Z}_t = a_t, \tag{3.3.8}$$

where

$$\begin{aligned}\pi(B) &= 1 - \pi_1 B - \pi_2 B^2 - \cdots \\ &= \frac{1}{\theta_q(B)}.\end{aligned} \tag{3.3.9}$$

For example, we can write the MA(2) process as

$$(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)\dot{Z}_t = \frac{1}{(1 - \theta_1 B - \theta_2 B^2)}\dot{Z}_t = a_t, \tag{3.3.10}$$

where

$$(1 - \theta_1 B - \theta_2 B^2)(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots) = 1,$$

or

$$\begin{aligned}1 &- \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots \\ &- \theta_1 B + \pi_1\theta_1 B^2 + \pi_2\theta_1 B^3 + \cdots \\ &\quad - \theta_2 B^2 + \pi_1\theta_2 B^3 + \cdots = 1.\end{aligned}$$

Thus, the π weights can be derived by equating the coefficients of B^j as follows:

$$\begin{aligned}B^1&: \quad -\pi_1 - \theta_1 = 0 \longrightarrow \pi_1 = -\theta_1 \\ B^2&: \quad -\pi_2 + \pi_1\theta_1 - \theta_2 = 0 \longrightarrow \pi_2 = \pi_1\theta_1 - \theta_2 = -\theta_1^2 - \theta_2 \\ B^3&: \quad -\pi_3 + \pi_2\theta_1 + \pi_1\theta_2 = 0 \longrightarrow \pi_3 = \pi_2\theta_1 + \pi_1\theta_2 \\ &\vdots\end{aligned}$$

In general,

$$\pi_j = \pi_{j-1}\theta_1 + \pi_{j-2}\theta_2, \quad \text{for } j \geq 3. \tag{3.3.11}$$

When $\theta_2 = 0$ and the process becomes the MA(1) process, we have $\pi_j = -\theta_1^j$ for $j \geq 1$, and

$$(1 + \theta_1 B + \theta_1^2 B^2 + \cdots)\dot{Z}_t = \frac{1}{(1 - \theta_1 B)}\dot{Z}_t = a_t. \tag{3.3.12}$$

Thus, in terms of the AR representation, a finite-order invertible MA process is equivalent to an infinite-order AR process.

In summary, a finite-order stationary AR(p) process corresponds to an infinite-order MA process, and a finite-order invertible MA(q) process corresponds to an infinite-order AR process. This dual relationship between the AR(p) and the MA(q) processes also exists in the autocorrelation and partial autocorrelation functions. The AR(p) process has its autocorrelations tailing off and partial autocorrelations cutting off, but the MA(q) process has its autocorrelations cutting off and partial autocorrelations tailing off.

3.4 Autoregressive Moving Average ARMA(p, q) Processes

A natural extension of the pure autoregressive and the pure moving average processes is the mixed autoregressive moving average process, which includes the autoregressive and moving average processes as special cases. The process contains a large class of parsimonious time series models that are useful in describing a wide variety of time series encountered in practice.

3.4.1 The General Mixed ARMA(p, q) Process

As we have shown, a stationary and invertible process can be represented either in a moving average form or in an autoregressive form. A problem with either representation, though, is that it may contain too many parameters, even for a finite-order moving average and a finite-order autoregressive model because a high-order model is often needed for good approximation. In general, a large number of parameters reduces efficiency in estimation. Thus, in model building, it may be necessary to include both autoregressive and moving average terms in a model, which leads to the following useful mixed autoregressive moving average (ARMA) process:

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t, \tag{3.4.1}$$

where

$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$

and

$$\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q.$$

For the process to be invertible, we require that the roots of $\theta_q(B) = 0$ lie outside the unit circle. To be stationary, we require that the roots of $\phi_p(B) = 0$ lie outside the unit circle. Also, we assume that $\phi_p(B) = 0$ and $\theta_q(B) = 0$ share no common roots. Henceforth, we refer to this process as an ARMA(p, q) process or model, in which p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively.

The stationary and invertible ARMA process can be written in a pure autoregressive representation discussed in Section 2.6, i.e.,

$$\pi(B)\dot{Z}_t = a_t, \tag{3.4.2}$$

where

$$\pi(B) = \frac{\phi_p(B)}{\theta_q(B)} = (1 - \pi_1 B - \pi_2 B^2 - \cdots). \tag{3.4.3}$$

This process can also be written as a pure moving average representation,

$$\dot{Z}_t = \psi(B)a_t, \tag{3.4.4}$$

where

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \cdots). \tag{3.4.5}$$

ACF of the ARMA(p, q) Process To derive the autocovariance function, we rewrite (3.4.1) as

$$\dot{Z}_t = \phi_1\dot{Z}_{t-1} + \cdots + \phi_p\dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

and multiply by $\dot{Z}_{t-k}$ on both sides

$$\dot{Z}_{t-k}\dot{Z}_t = \phi_1\dot{Z}_{t-k}\dot{Z}_{t-1} + \cdots + \phi_p\dot{Z}_{t-k}\dot{Z}_{t-p} + \dot{Z}_{t-k}a_t - \theta_1\dot{Z}_{t-k}a_{t-1} - \cdots - \theta_q\dot{Z}_{t-k}a_{t-q}.$$

We now take the expected value to obtain

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} + E(\dot{Z}_{t-k}a_t) - \theta_1 E(\dot{Z}_{t-k}a_{t-1}) - \cdots - \theta_q E(\dot{Z}_{t-k}a_{t-q}).$$

Because

$$E(\dot{Z}_{t-k}a_{t-i}) = 0 \qquad \text{for } k > i,$$

we have

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p}, \qquad k \geq (q + 1), \tag{3.4.6}$$

and hence,

$$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}, \qquad k \geq (q + 1). \tag{3.4.7}$$

Equation (3.4.7) satisfies the pth-order homogeneous difference equation as shown in (3.1.26) for the AR(p) process. Therefore, the autocorrelation function of an ARMA(p, q) model tails off after lag q just like an AR(p) process, which depends only on the autoregressive parameters in the model. The first q autocorrelations $\rho_q, \rho_{q-1}, \ldots, \rho_1$, however, depend on both autoregressive and moving average parameters in the model and serve as initial values for the pattern. This distinction is useful in model identification.

PACF of the ARMA(p, q) Process Because the ARMA process contains the MA process as a special case, its PACF will also be a mixture of exponential decays or damped sine waves depending on the roots of $\phi_p(B) = 0$ and $\theta_q(B) = 0$.

3.4.2 The ARMA(1, 1) Process

$$(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t, \tag{3.4.8a}$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t - \theta_1 a_{t-1}. \tag{3.4.8b}$$

For stationarity, we assume that $|\phi_1| < 1$, and for invertibility, we require that $|\theta_1| < 1$. When $\phi_1 = 0$, (3.4.8a) is reduced to an MA(1) process, and when $\theta_1 = 0$, it is reduced to an AR(1) process. Thus, we can regard the AR(1) and MA(1) processes as special cases of the ARMA(1, 1) process.

In terms of a pure autoregressive representation, we write

$$\pi(B)\dot{Z}_t = a_t,$$

where

$$\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots) = \frac{(1 - \phi_1 B)}{(1 - \theta_1 B)},$$

i.e.,

$$(1 - \theta_1 B)(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots) = (1 - \phi_1 B)$$

or

$$\left[1 - (\pi_1 + \theta_1)B - (\pi_2 - \pi_1\theta_1)B^2 - (\pi_3 - \pi_2\theta_1)B^3 - \cdots\right] = (1 - \phi_1 B).$$

By equating coefficients of B^j on both sides of the above equation, we get

$$\pi_j = \theta_1^{\,j-1}(\phi_1 - \theta_1), \qquad \text{for } j \geq 1. \tag{3.4.9}$$

We can write the ARMA(1, 1) process in a pure moving average representation as

$$Z_t = \psi(B)a_t = \frac{(1 - \theta_1 B)}{(1 - \phi_1 B)}a_t.$$

We note that

$$(1 - \phi_1 B)(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots) = (1 - \theta_1 B),$$

i.e.,

$$\left[1 + (\psi_1 - \phi_1)B + (\psi_2 - \psi_1\phi_1)B^2 + \cdots\right] = (1 - \theta_1 B).$$

Hence,

$$\psi_j = \phi_1^{j-1}(\phi_1 - \theta_1), \qquad \text{for } j \geq 1. \tag{3.4.10}$$

ACF of the ARMA(1, 1) Process To obtain the autocovariance for $\{Z_t\}$, we multiply Z_{t-k} on both sides of (3.4.8b),

$$\dot{Z}_{t-k}\dot{Z}_t = \phi_1\dot{Z}_{t-k}\dot{Z}_{t-1} + \dot{Z}_{t-k}a_t - \theta_1\dot{Z}_{t-k}a_{t-1},$$

and take the expected value to obtain

$$\gamma_k = \phi_1\gamma_{k-1} + E(\dot{Z}_{t-k}a_t) - \theta_1 E(\dot{Z}_{t-k}a_{t-1}). \tag{3.4.11}$$

More specifically, when $k = 0$,

$$\gamma_0 = \phi_1\gamma_1 + E(\dot{Z}_t a_t) - \theta_1 E(\dot{Z}_t a_{t-1}).$$

Recall that $E(\dot{Z}_t a_t) = \sigma_a^2$. For the term $E(\dot{Z}_t a_{t-1})$, we note that

$$\begin{aligned} E(\dot{Z}_t a_{t-1}) &= \phi_1 E(\dot{Z}_{t-1}a_{t-1}) + E(a_t a_{t-1}) - \theta_1 E(a_{t-1}^2) \\ &= (\phi_1 - \theta_1)\sigma_a^2. \end{aligned}$$

Hence,

$$\gamma_0 = \phi_1\gamma_1 + \sigma_a^2 - \theta_1(\phi_1 - \theta_1)\sigma_a^2. \tag{3.4.12}$$

When $k = 1$, we have from (3.4.11)

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2. \tag{3.4.13}$$

Substituting (3.4.13) in (3.4.12), we have

$$\gamma_0 = \phi_1^2\gamma_0 - \phi_1\theta_1\sigma_a^2 + \sigma_a^2 - \phi_1\theta_1\sigma_a^2 + \theta_1^2\sigma_a^2,$$

i.e.,

$$\gamma_0 = \frac{(1 + \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)}\sigma_a^2.$$

Thus,

$$\begin{aligned}\gamma_1 &= \phi_1\gamma_0 - \theta_1\sigma_a^2 \\ &= \frac{\phi_1(1 + \theta_1^2 - 2\phi_1\theta_1)}{(1 - \phi_1^2)}\sigma_a^2 - \theta_1\sigma_a^2 \\ &= \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{(1 - \phi_1^2)}\sigma_a^2.\end{aligned}$$

For $k \geq 2$, we have from (3.4.11)

$$\gamma_k = \phi_1\gamma_{k-1}, \qquad k \geq 2.$$

Hence, the ARMA(1, 1) model has the following autocorrelation function:

$$\rho_k = \begin{cases} 1 & k = 0, \\ \dfrac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}, & k = 1, \\ \phi_1\rho_{k-1}, & k \geq 2. \end{cases} \tag{3.4.14}$$

Note that the autocorrelation function of an ARMA(1, 1) model combines characteristics of both AR(1) and MA(1) processes. The moving average parameter θ_1 enters into the calculation of ρ_1. Beyond ρ_1, the autocorrelation function of an ARMA(1, 1) model follows the same pattern as the autocorrelation function of an AR(1) process.

PACF of the ARMA(1, 1) Process The general form of the PACF of a mixed model is complicated and is not needed. It suffices to note that, because the ARMA(1, 1) process contains the MA(1) process as a special case, the PACF of the ARMA(1, 1) process also tails off exponentially like the ACF, with its shape depending on the signs and magnitudes of ϕ_1 and θ_1. Thus, that both ACF and PACF tail off indicates a mixed ARMA model. Some of the ACF and PACF patterns for the ARMA(1, 1) model are shown in Figure 3.14. By examining Figure 3.14, it can be seen that due to the combined effect of both ϕ_1 and θ_1, the PACF of the ARMA(1, 1) process contains many more different shapes than the PACF of the MA(1) process, which consists of only two possibilities.

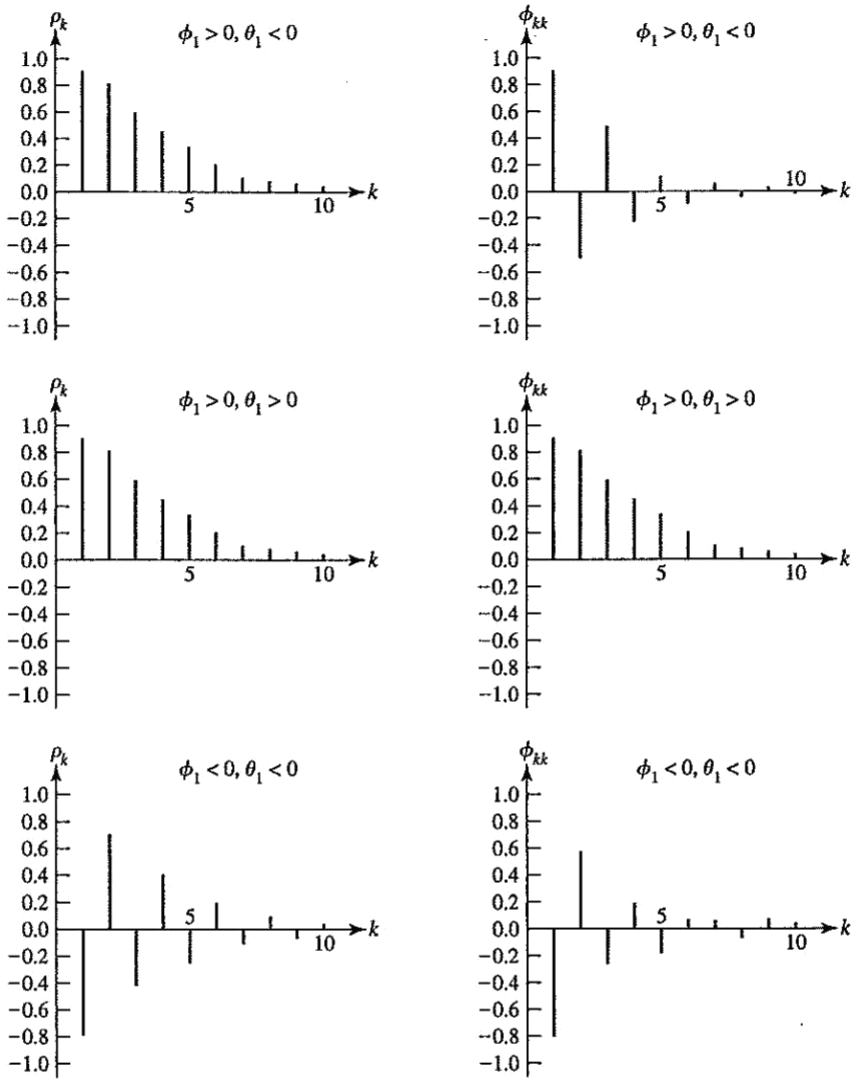


FIGURE 3.14 ACF and PACF of ARMA(1, 1) model $(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t$.

EXAMPLE 3.7 A series of 250 values is simulated from the ARMA(1, 1) process $(1 - .9B)Z_t = (1 - .5B)a_t$, with the a_t being a Gaussian $N(0, 1)$ white noise series. The sample ACF and sample PACF are shown in Table 3.7 and also plotted in Figure 3.15. That both $\hat{\rho}_k$ and $\hat{\phi}_{kk}$ tail off indicates a mixed ARMA model. To decide the proper orders of p and q in a mixed model is a much more difficult and challenging task, sometimes requiring considerable experience and skill. Some helpful methods are discussed in Chapter 6 on model identification. For now, it is sufficient to identify tentatively from the sample ACF and sample PACF whether the phenomenon is a pure AR, pure MA, or mixed ARMA model. Solely based on the sample PACF as shown in Table 3.7, without looking at the sample ACF, we know that the phenomenon

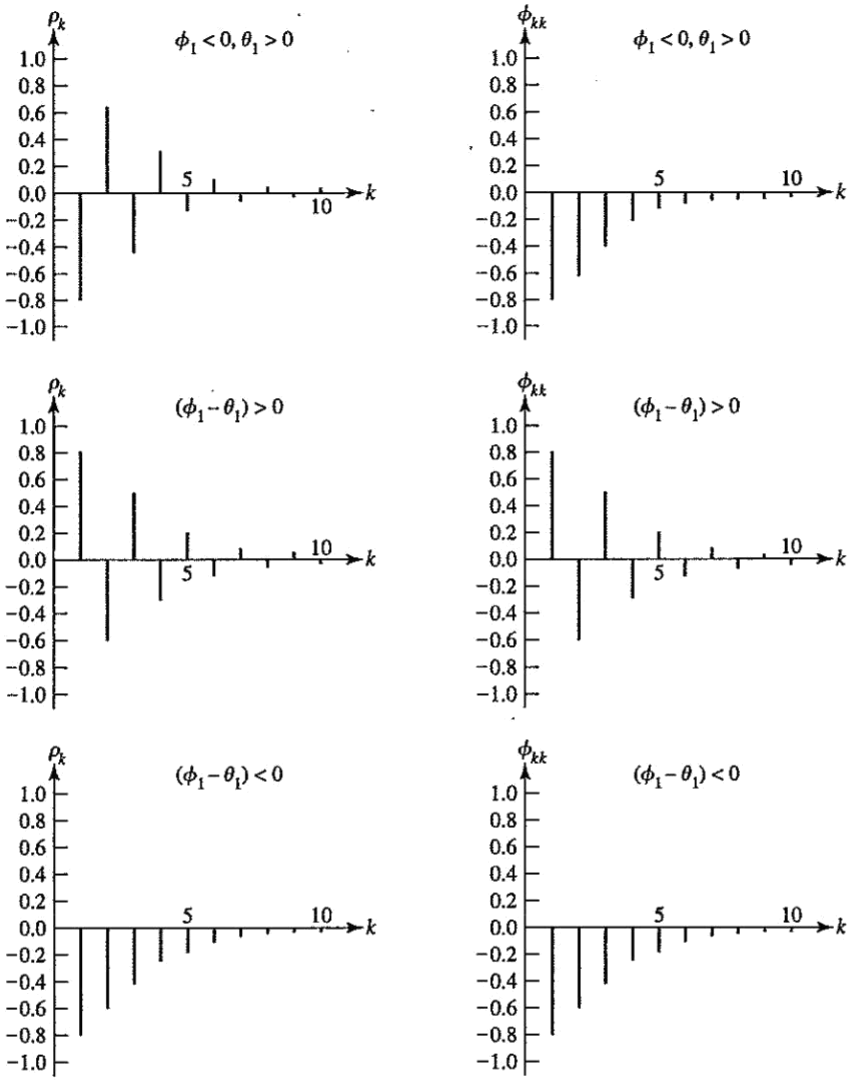


FIGURE 3.14 (*Continued*)

cannot be an MA process because the MA process cannot have a positively exponentially decaying PACF.

EXAMPLE 3.8 The sample ACF and PACF are calculated for a series of 250 values as shown in Table 3.8 and plotted in Figure 3.16. None is statistically significant from 0, which would indicate a white noise phenomenon. In fact, the series is the simulation result from the ARMA(1, 1) process, $(1 - \phi_1 B)Z_t = (1 - \theta_1 B)a_t$ with $\phi_1 = .6$ and $\theta_1 = .5$. The sample ACF and sample PACF are both small because the AR polynomial $(1 - .6B)$ and the MA polynomial $(1 - .5B)$ almost cancel each other out. Recall from (3.4.14) that the ACF of the

TABLE 3.7 Sample ACF and sample PACF for a simulated ARMA(1, 1) series from $(1 - .9B)Z_t = (1 - .5B)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.57	.50	.47	.35	.31	.25	.21	.18	.10	.12
St.E.	.06	.08	.09	.10	.11	.11	.11	.11	.11	.11
$\hat{\phi}_{kk}$	.57	.26	.18	−.03	.01	−.01	.01	.01	−.08	.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

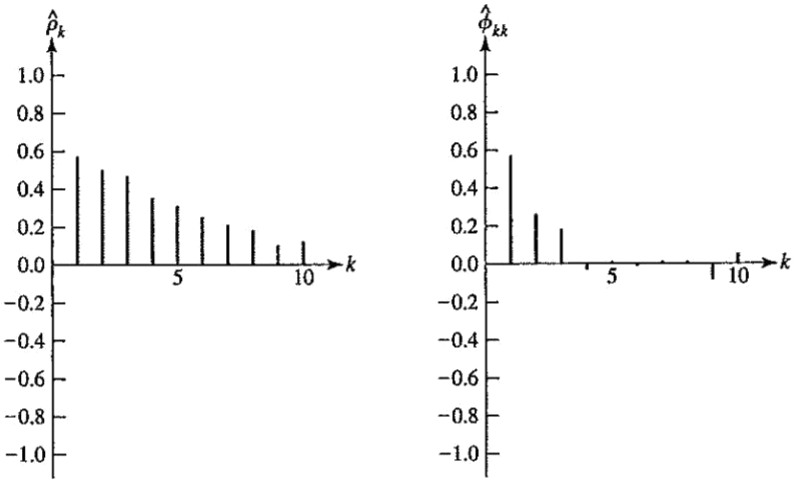


FIGURE 3.15 Sample ACF and sample PACF of a simulated ARMA(1, 1) series: $(1 - .9B)Z_t = (1 - .5B)a_t$.

ARMA(1, 1) process is $\rho_k = \phi_1^{k-1}(\phi_1 - \theta_1)(1 - \phi_1\theta_1)/(1 + \theta_1^2 - 2\phi_1\theta_1)$ for $k \geq 1$, which is approximately equal to zero when $\phi_1 \simeq \theta_1$. Thus, the sample phenomenon of a white noise series implies that the underlying model is either a random noise process or an ARMA process with its AR and MA polynomials being nearly equal. The assumption of no common roots between $\phi_p(B) = 0$ and $\theta_q(B) = 0$ in the mixed model is needed to avoid this confusion.

Before closing this chapter, we note that the ARMA(p, q) model in (3.4.1), i.e.,

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(Z_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t,$$

can also be written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \qquad (3.4.15)$$

TABLE 3.8 Sample ACF and sample PACF for a simulated series of the ARMA(1, 1) process: $(1 - .6B)Z_t = (1 - .5B)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.10	.05	.09	.00	−.02	.02	−.02	.04	−.04	.01
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
$\hat{\phi}_{kk}$	.10	.04	.08	−.02	−.02	.01	−.02	.05	−.05	.02
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

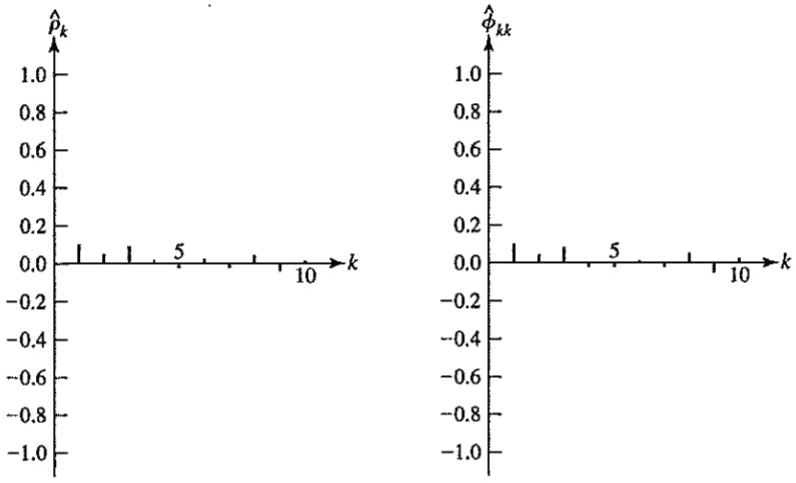


FIGURE 3.16 Sample ACF and sample PACF of a simulated ARMA(1, 1) series: $(1 - .6B)Z_t = (1 - .5B)a_t$.

where

$$\begin{aligned}\theta_0 &= (1 - \phi_1 B - \cdots - \phi_p B^p)\mu \\ &= (1 - \phi_1 - \cdots - \phi_p)\mu. \end{aligned} \tag{3.4.16}$$

In terms of this form, the AR(p) model becomes

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Z_t = \theta_0 + a_t \tag{3.4.17}$$

and the MA(q) model becomes

$$Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t. \tag{3.4.18}$$

It is clear that in the MA(q) process, $\theta_0 = \mu$.

EXERCISES

3.1 Find the ACF and PACF and plot the ACF ρ_k for $k = 0, 1, 2, 3, 4$, and 5 for each of the following models where the a_t is a Gaussian white noise process.

(a) $Z_t - .5Z_{t-1} = a_t$,

(b) $Z_t + .98Z_{t-1} = a_t$,

(c) $Z_t - 1.3Z_{t-1} + .4Z_{t-2} = a_t$,

(d) $Z_t - 1.2Z_{t-1} + .8Z_{t-2} = a_t$.

3.2 Consider the following AR(2) models:

(i) $Z_t - .6Z_{t-1} - .3Z_{t-2} = a_t$,

(ii) $Z_t - .8Z_{t-1} + .5Z_{t-2} = a_t$.

(a) Find the general expression for ρ_k.

(b) Plot the ρ_k, for $k = 0, 1, 2, \ldots, 10$.

(c) Calculate σ_Z^2 by assuming that $\sigma_a^2 = 1$.

3.3 Simulate a series of 100 observations from each of the models with $\sigma_a^2 = 1$ in Exercise 3.1. For each case, plot the simulated series, and calculate and study its sample ACF $\hat{\rho}_k$ and PACF $\hat{\phi}_{kk}$ for $k = 0, 1, \ldots, 20$.

3.4 (a) Show that the ACF ρ_k for the AR(1) process satisfies the difference equation

$$\rho_k - \phi_1 \rho_{k-1} = 0, \qquad \text{for } k \geq 1.$$

(b) Find the general expression for ρ_k.

3.5 Consider the AR(2) process $Z_t = Z_{t-1} - .25Z_{t-2} + a_t$.

(a) Calculate ρ_1.

(b) Use ρ_0, ρ_1 as starting values and the difference equation to obtain the general expression for ρ_k.

(c) Calculate the values ρ_k for $k = 1, 2, \ldots, 10$.

3.6 (a) Find the range of α such that the AR(2) process

$$Z_t = Z_{t-1} + \alpha Z_{t-2} + a_t$$

is stationary.

(b) Find the ACF for the model in part (a) with $\alpha = -\frac{1}{2}$.

3.7 Show that if an AR(2) process is stationary, then

$$\rho_1^2 < \frac{\rho_2 + 1}{2}.$$

3.8 Consider the MA(2) process $Z_t = (1 - 1.2B + .5B^2)a_t$.

(a) Find the ACF using the definition of autocovariance function.

(b) Find the ACF using the autocovariance generating function.

(c) Find the PACF ϕ_{kk} for the process.

3.9 Find an invertible process which has the following ACF:

$$\rho_0 = 1, \qquad \rho_1 = .25, \qquad \text{and} \qquad \rho_k = 0 \qquad \text{for } k \geq 2.$$

3.10 (a) Find a process that has the following autocovariance function:

$$\gamma_0 = 10, \qquad \gamma_1 = 0, \qquad \gamma_2 = -4, \qquad \gamma_k = 0, \qquad \text{for } |k| > 2.$$

(b) Examine stationarity and invertibility for the process obtained in part (a).

3.11 Consider the MA(2) process $Z_t = a_t - .1a_{t-1} + .21a_{t-2}$.

(a) Is the model stationary? Why?

(b) Is the model invertible? Why?

(c) Find the ACF for the above process.

3.12 Simulate a series of 100 observations from the model with $\sigma_a^2 = 1$ in Exercise 3.8. Plot the simulated series and calculate and study its sample ACF, $\hat{\rho}_k$, and PACF, $\hat{\phi}_{kk}$ for k = 0, 1, . . . , 20.

3.13 One can calculate the PACF using either (2.3.19) or (2.5.25). Illustrate the computation of ϕ_{11}, ϕ_{22}, and ϕ_{33} for the MA(1) process using both procedures.

3.14 Consider each of the following models:

(i) $(1 - B)Z_t = (1 - 1.5B)a_t$,

(ii) $(1 - .8B)Z_t = (1 - .5B)a_t$,

(iii) $(1 - 1.1B + .8B^2)Z_t = (1 - 1.7B + .72B^2)a_t$,

(iv) $(1 - .6B)Z_t = (1 - 1.2B + .2B^2)a_t$

(a) Verify whether it is stationary, or invertible, or both.

(b) Express the model in an MA representation if it exists.

(c) Express the model in an AR representation if it exists.

3.15 Consider each of the following processes:

(i) $(1 - .6B)Z_t = (1 - .9B)a_t$,

(ii) $(1 - 1.4B + .6B^2)Z_t = (1 - .8B)a_t$

(a) Find the ACF ρ_k.

(b) Find the PACF ϕ_{kk} for $k = 1, 2, 3$.

(c) Find the autocovariance generating function.

3.16 Simulate a series of 100 observations from each of the models with $\sigma_a^2 = 1$ in Exercise 3.15. For each simulated series, plot the series, calculate, and study its sample ACF $\hat{\rho}_k$ and PACF $\hat{\phi}_{kk}$ for $k = 0, 1, \ldots, 20$.

4

Nonstationary Time Series Models

The time series processes we have discussed so far are all stationary processes, but many applied time series, particularly those arising from economic and business areas, are nonstationary. With respect to the class of covariance stationary processes, nonstationary time series can occur in many different ways. They could have nonconstant means μ_t, time-varying second moments such as nonconstant variance σ_t^2, or both of these properties. For example, the monthly series of unemployed women between ages 16 and 19 in the United States from January 1961 to August 2002 plotted in Figure 4.1 clearly shows that the mean level changes with time. The plot of the yearly U.S. tobacco production between 1871 and

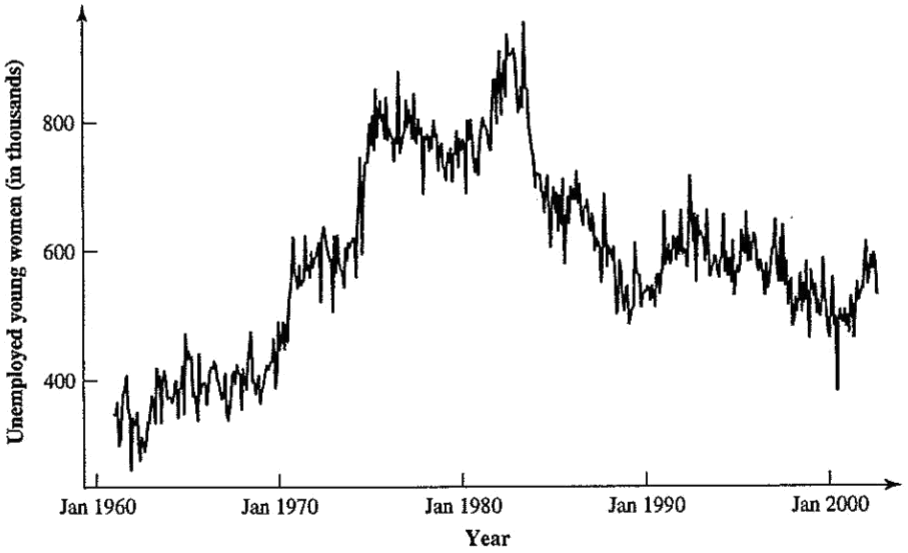


FIGURE 4.1 Monthly series of unemployed women between ages 16 and 19 in the United States from January 1961 to August 2002.

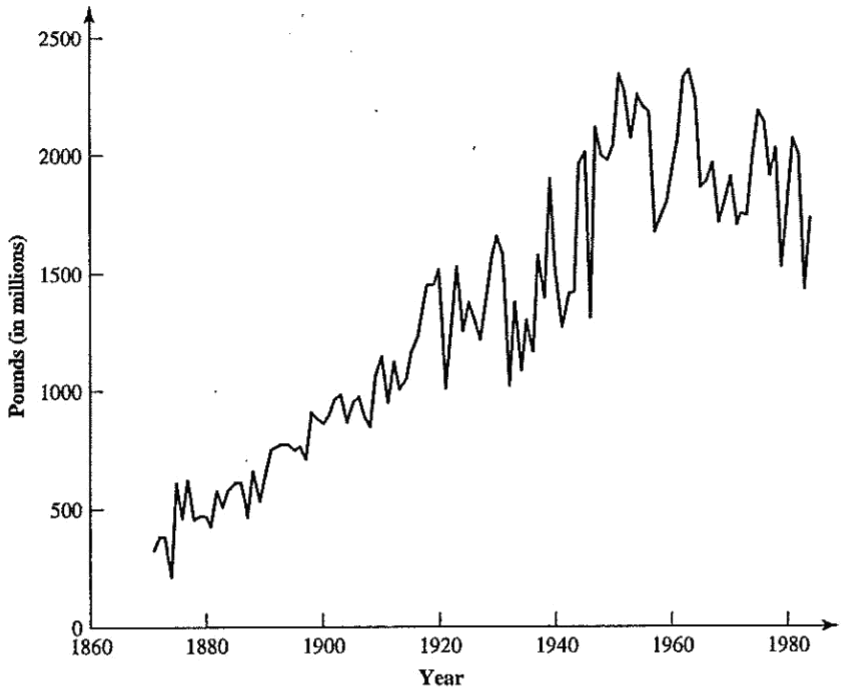


FIGURE 4.2 Yearly U.S. tobacco production between 1871 and 1984.

1984 shown in Figure 4.2 indicates both that the mean level depends on time and that the variance increases as the mean level increases.

In this chapter, we illustrate the construction of a very useful class of homogeneous nonstationary time series models, the autoregressive integrated moving average (ARIMA) models. Some useful differencing and variance stabilizing transformations are introduced to connect the stationary and nonstationary time series models.

4.1 Nonstationarity in the Mean

A process nonstationary in the mean could pose a very serious problem for estimation of the time-dependent mean function without multiple realizations. Fortunately, there are models that can be constructed from a single realization to describe this time-dependent phenomenon. Two such classes of models that are very useful in modeling time series nonstationary in the mean are introduced in this section.

4.1.1 Deterministic Trend Models

The mean function of a nonstationary process could be represented by a deterministic trend of time. In such a case, a standard regression model might be used to describe the

phenomenon. For example, if the mean function μ_t follows a linear trend, $\mu_t = \alpha_0 + \alpha_1 t$, then one can use the deterministic linear trend model

$$Z_t = \alpha_0 + \alpha_1 t + a_t, \tag{4.1.1}$$

with the a_t being a zero mean white noise series. For a deterministic quadratic mean function, $\mu_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, one can use

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + a_t. \tag{4.1.2}$$

More generally, if the deterministic trend can be described by a kth-order polynomial of time, one can model the process by

$$Z_t = \alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k + a_t. \tag{4.1.3}$$

If the deterministic trend can be represented by a sine-cosine curve, one can use

$$Z_t = \nu_0 + \nu \cos(\omega t + \theta) + a_t \tag{4.1.4}$$

$$= \nu_0 + \alpha \cos(\omega t) + \beta \sin(\omega t) + a_t, \tag{4.1.5}$$

where

$$\alpha = \nu \cos \theta, \qquad \beta = -\nu \sin \theta, \tag{4.1.6}$$

$$\nu = \sqrt{\alpha^2 + \beta^2}, \tag{4.1.7}$$

and

$$\theta = \tan^{-1}(-\beta/\alpha). \tag{4.1.8}$$

We call ν the amplitude, ω the frequency, and θ the phase of the curve. More generally, we can have

$$Z_t = \nu_0 + \sum_{j=1}^{m} (\alpha_j \cos \omega_j t + \beta_j \sin \omega_j t) + a_t, \tag{4.1.9}$$

which is often called the model of hidden periodicities. These models can be analyzed using standard regression analysis and are discussed again later in Chapter 13.

4.1.2 Stochastic Trend Models and Differencing

Although many time series are nonstationary, due to some equilibrium forces, different parts of these series behave very much alike except for their difference in the local mean levels. Box and Jenkins (1976, p. 85) refer to this kind of nonstationary behavior as homogeneous nonstationary. In terms of the ARMA models, the process is nonstationary if some roots of its AR polynomial do not lie outside the unit circle. By the nature of homogeneity, however,

the local behavior of this kind of homogeneous nonstationary series is independent of its level. Hence, by letting $\Psi(B)$ be the autoregressive operator describing the behavior, we have

$$\Psi(B)(Z_t + C) = \Psi(B)Z_t \tag{4.1.10}$$

for any constant C. This equation implies that $\Psi(B)$ must be of the form

$$\Psi(B) = \phi(B)(1 - B)^d, \tag{4.1.11}$$

for some $d > 0$, where $\phi(B)$ is a stationary autoregressive operator. Thus, a homogeneous nonstationary series can be reduced to a stationary series by taking a suitable difference of the series. In other words, the series $\{Z_t\}$ is nonstationary, but its dth differenced series, $\{(1 - B)^d Z_t\}$ for some integer $d \geq 1$, is stationary. For example, if the dth differenced series follows a white noise phenomenon, we have

$$(1 - B)^d Z_t = a_t. \tag{4.1.12}$$

To see the implication of this kind of homogeneous nonstationary series, consider $d = 1$ in (4.1.12), i.e.,

$$(1 - B)Z_t = a_t \tag{4.1.13a}$$

or

$$Z_t = Z_{t-1} + a_t. \tag{4.1.13b}$$

Given the past information $Z_{t-1}, Z_{t-2}, \ldots$, the level of the series at time t is

$$\mu_t = Z_{t-1}, \tag{4.1.14}$$

which is subject to the stochastic disturbance at time $(t - 1)$. In other words, the mean level of the process Z_t in $(1 - B)^d Z_t$ for $d \geq 1$ changes through time stochastically, and we characterize the process as having a stochastic trend. This result is different from the deterministic trend model mentioned in the previous section, where the mean level of the process at time t is a pure deterministic function of time.

4.2 Autoregressive Integrated Moving Average (ARIMA) Models

A homogeneous nonstationary time series can be reduced to a stationary time series by taking a proper degree of differencing. The autoregressive moving average models are useful in describing stationary time series, so in this section, we discuss the use of differencing to build a large class of time series models, autoregressive integrated moving average models, which are useful in describing various homogeneous nonstationary time series.

4.2.1 The General ARIMA Model

Obviously, the stationary process resulting from a properly differenced homogeneous nonstationary series is not necessarily white noise as in (4.1.12). More generally, the differenced series $(1 - B)^d Z_t$ follows the general stationary ARMA(p, q) process discussed in (3.4.1) of Chapter 3. Thus, we have

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t, \tag{4.2.1}$$

where the stationary AR operator $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and the invertible MA operator $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ share no common factors. The parameter θ_0 plays very different roles for $d = 0$ and $d > 0$. When $d = 0$, the original process is stationary, and we recall from (3.4.16) that θ_0 is related to the mean of the process, i.e., $\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$. When $d \geq 1$, however, θ_0 is called the deterministic trend term and, as shown in the next section, is often omitted from the model unless it is really needed.

The resulting homogeneous nonstationary model in (4.2.1) has been referred to as the autoregressive integrated moving average model of order (p, d, q) and is denoted as the ARIMA(p, d, q) model. When $p = 0$, the ARIMA(p, d, q) model is also called the integrated moving average model of order (d, q) and is denoted as the IMA(d, q) model. In the following discussion, we illustrate some commonly encountered ARIMA models.

4.2.2 The Random Walk Model

In (4.2.1), if $p = 0$, $d = 1$, and $q = 0$, then we have the well-known random walk model,

$$(1 - B)Z_t = a_t \tag{4.2.2a}$$

or

$$Z_t = Z_{t-1} + a_t. \tag{4.2.2b}$$

This model has been widely used to describe the behavior of the series of a stock price. In the random walk model, the value of Z at time t is equal to its value at time $(t - 1)$ plus a random shock. This behavior is similar to following a drunken man whose position at time t is his position at time $(t - 1)$ plus a step in a random direction at time t.

Note that the random walk model is the limiting process of the AR(1) process $(1 - \phi B)Z_t = a_t$ with $\phi \to 1$. Because the autocorrelation function of the AR(1) process is $\rho_k = \phi^k$, as $\phi \to 1$, the random walk model phenomenon can be characterized by large, nonvanishing spikes in the sample ACF of the original series $\{Z_t\}$ and insignificant zero ACF for the differenced series $\{(1 - B)Z_t\}$.

Next, consider the following simple modification of (4.2.2a) with a nonzero constant term

$$(1 - B)Z_t = \theta_0 + a_t \tag{4.2.3}$$

or

$$Z_t = Z_{t-1} + \theta_0 + a_t. \tag{4.2.4}$$

With reference to a time origin k, by successive substitution, we have

$$\begin{aligned} Z_t &= Z_{t-1} + \theta_0 + a_t \\ &= Z_{t-2} + 2\theta_0 + a_t + a_{t-1} \\ &\vdots \\ &= Z_k + (t-k)\theta_0 + \sum_{j=k+1}^{t} a_j, \qquad \text{for } t > k. \end{aligned} \tag{4.2.5}$$

Thus, it is clear that Z_t contains a deterministic trend with slope or drift θ_0. More generally, for the model involving the dth differenced series $\{(1-B)^d Z_t\}$, the nonzero θ_0 in (4.2.1) can be shown to correspond to the coefficient α_d of t^d in the deterministic trend, $\alpha_0 + \alpha_1 t + \cdots + \alpha_d t^d$. For this reason, when $d > 0$, θ_0 is referred to as the deterministic trend term. For large t, this term can become very dominating so that it forces the series to follow a deterministic pattern. Hence, in general, when $d > 0$, we assume that $\theta_0 = 0$ unless it is clear from the data or the nature of the problem that a deterministic component is really needed.

The process in (4.2.3) with $\theta_0 \neq 0$ is usually called the random walk model with drift. Given $Z_{t-1}, Z_{t-2}, \ldots$, by (4.2.4) the mean level of the series Z_t at time t is

$$\mu_t = Z_{t-1} + \theta_0, \tag{4.2.6}$$

which is influenced by the stochastic disturbance at time $(t-1)$ through the term Z_{t-1} as well as by the deterministic component through the slope θ_0. When $\theta_0 = 0$, we have a model with only a stochastic trend.

EXAMPLE 4.1 To illustrate the results of the random walk model discussed in this section, we simulated 100 observations each from the model $(1-B)Z_t = a_t$ and the model $(1-B)Z_t = 4 + a_t$, where the a_t in both models are i.i.d. normal $N(0, 1)$ white noise. The sample ACF and PACF of the original series are shown in Table 4.1 and Figure 4.3. That the ACF for both series decays very slowly indicates that they are nonstationary. To identify the model properly, we calculate the sample ACF and PACF of the differenced series $(1-B)Z_t$ as shown in Table 4.2 and Figure 4.4. As expected, both exhibit phenomena of a white noise process. In fact, the patterns of $\hat{\rho}_k$ and $\hat{\phi}_{kk}$ are identical for the two models. Then how can we tell the difference between the regular random walk model and the random walk model with drift? We cannot if we rely only on their autocorrelation structures, although the sample ACF of the original series from a random walk model with drift generally decays more slowly. If we look at their behaviors as plotted in Figure 4.5, however, then the difference is striking. The series of the random walk model with drift is clearly dominated by the deterministic linear trend with slope 4. On the other hand, the nonstationarity of the random walk model without drift is shown through a stochastic trend and its values are free to wander.

TABLE 4.1 Sample ACF and sample PACF for the original series Z_t simulated from random walk models.

(a) For Z_t from $(1 - B)Z_t = a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.94	.88	.83	.77	.71	.66	.60	.53	.46	.40
St.E.	.10	.17	.21	.24	.26	.28	.30	.31	.32	.32
$\hat{\phi}_{kk}$	.94	−.07	−.01	−.04	.02	−.07	−.04	−.15	.02	−.04
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10

(b) For Z_t from $(1 - B)Z_t = 4 + a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.94	.94	.91	.88	.85	.82	.79	.76	.73	.70
St.E.	.10	.17	.22	.25	.28	.30	.33	.34	.36	.38
$\hat{\phi}_{kk}$	.97	−.01	−.01	−.02	−.01	−.01	−.02	−.02	−.01	−.01
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10

4.2.3 The ARIMA(0, 1, 1) or IMA(1, 1) Model

When $p = 0$, $d = 1$, and $q = 1$, the model in (4.2.1) becomes

$$(1 - B)Z_t = (1 - \theta B)a_t \tag{4.2.7a}$$

or

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1}, \tag{4.2.7b}$$

where $-1 < \theta < 1$. This IMA(1, 1) model for Z_t is reduced to a stationary MA(1) model for the first differenced series, $(1 - B)Z_t$. The random walk model is a special case of this IMA(1, 1) model with $\theta = 0$. Thus, the basic phenomenon of the IMA(1, 1) model is characterized by the sample ACF of the original series failing to die out and by the sample ACF of the first differenced series exhibiting the pattern of a first-order moving average phenomenon.

For $-1 < \theta < 1$,

$$\begin{aligned}\frac{(1 - B)}{(1 - \theta B)} &= (1 - B)(1 + \theta B + \theta^2 B^2 + \cdots) \\ &= 1 + \theta B + \theta^2 B^2 + \cdots - B - \theta B^2 - \cdots \\ &= 1 - (1 - \theta)B - (1 - \theta)\theta B^2 - (1 - \theta)\theta^2 B^3 - \cdots \\ &= 1 - \alpha B - \alpha(1 - \alpha)B^2 - \alpha(1 - \alpha)^2 B^3 - \cdots,\end{aligned} \tag{4.2.8}$$

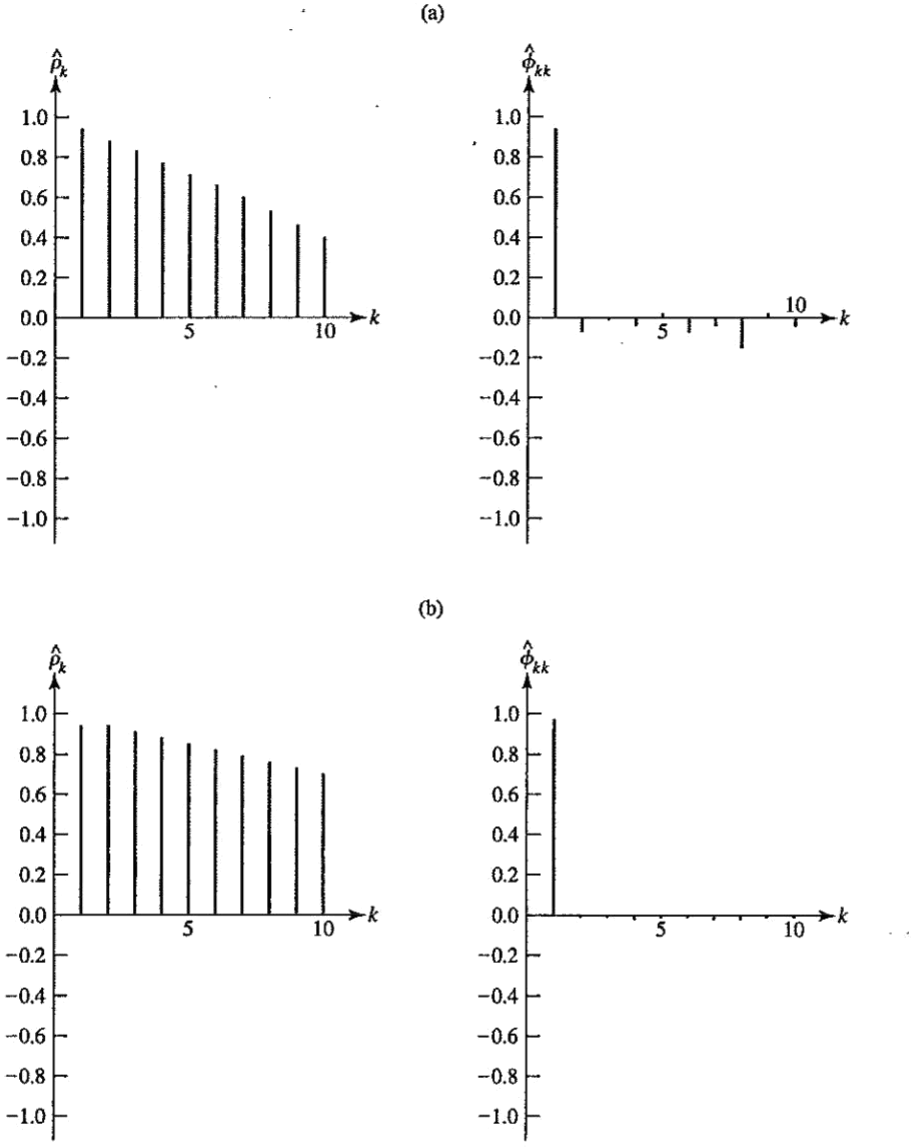


FIGURE 4.3 Sample ACF and PACF of the random walk model. (a) For Z_t from $(1 - B)Z_t = a_t$. (b) For Z_t from $(1 - B)Z_t = 4 + a_t$.

where $\alpha = (1 - \theta)$. Hence,

$$Z_t = \alpha \sum_{j=1}^{\infty} (1 - \alpha)^{j-1} Z_{t-j} + a_t. \tag{4.2.9}$$

TABLE 4.2 Sample ACF and sample PACF for the differenced series $W_t = (1 - B)Z_t$ simulated from random walk models.

(a) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.11	.03	.00	.00	.11	.02	.06	.01	−.02	.06
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
$\hat{\phi}_{kk}$	.11	.02	.00	.00	.11	.00	.05	.00	−.03	.05
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10

(b) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = 4 + a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.11	.03	.00	.00	.11	.02	.06	.01	−.02	.06
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
$\hat{\phi}_{kk}$	.11	.02	.00	.00	.11	.00	.05	.00	−.03	.05
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10

This equation is the AR representation of the model, and from the results of regression analysis the optimal forecast, $\hat{Z}_t$, of Z_t is given by

$$\hat{Z}_t = \alpha \sum_{j=1}^{\infty} (1 - \alpha)^{j-1} Z_{t-j}. \tag{4.2.10}$$

In other words, the optimal forecast of Z_t at time t is an exponentially decreasing weighted moving average of its past values $Z_{t-1}, Z_{t-2}, \ldots$, etc. Moreover, (4.2.10) implies that

$$\begin{aligned}
\hat{Z}_{t+1} &= \alpha \sum_{j=1}^{\infty} (1 - \alpha)^{j-1} Z_{t+1-j} \\
&= \alpha Z_t + (1 - \alpha)\alpha \sum_{j=2}^{\infty} (1 - \alpha)^{j-2} Z_{t+1-j} \\
&= \alpha Z_t + (1 - \alpha)\alpha \sum_{i=1}^{\infty} (1 - \alpha)^{i-1} Z_{t-i} \\
&= \alpha Z_t + (1 - \alpha)\hat{Z}_t,
\end{aligned}$$

which shows that the new forecast of Z at the next time period is equal to the weighted average of the newly available observation and the last forecast. The coefficient is often called the smoothing constant in the method of exponential smoothing. Thus, the general ARIMA(p, d, q) model contains many smoothing methods as special cases. See Abraham and Ledolter (1983) for a more detailed discussion of the relationship between exponential smoothing and the ARIMA models.

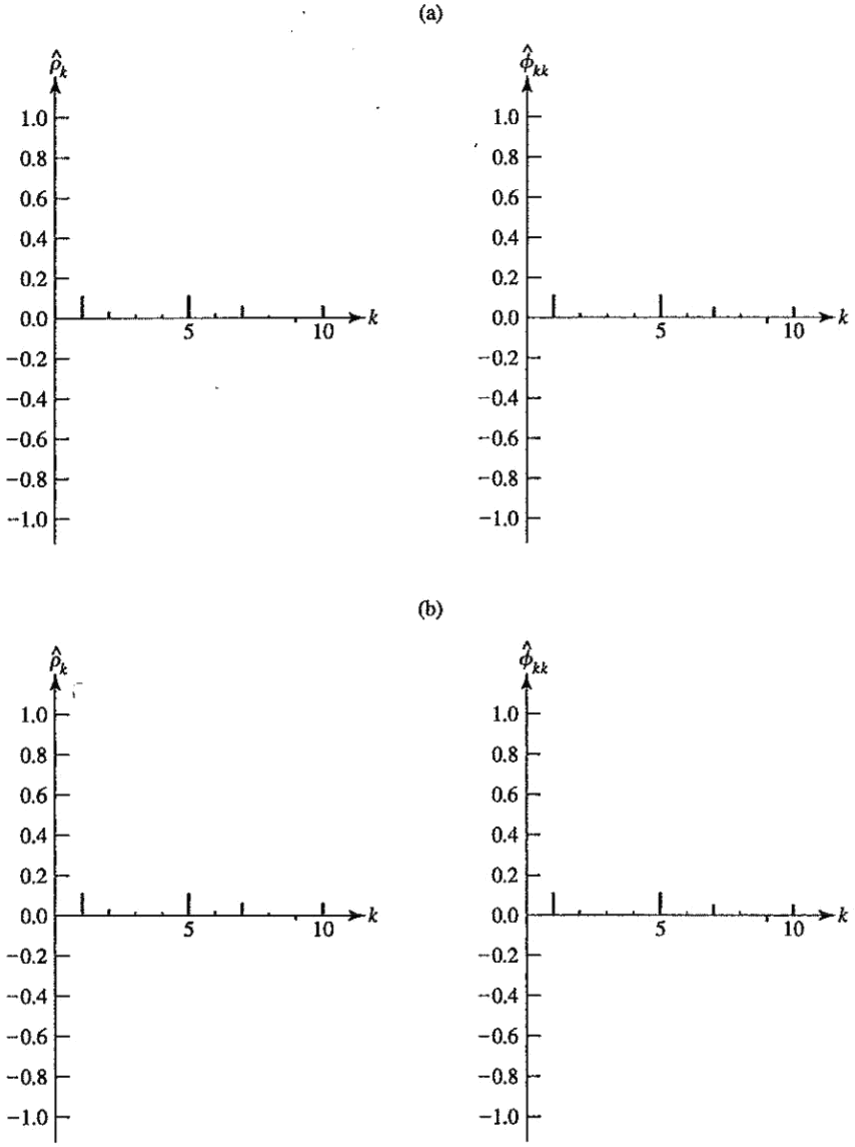


FIGURE 4.4 Sample ACF and PACF of differenced series. (a) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = a_t$. (b) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = 4 + a_t$.

EXAMPLE 4.2 We simulated 250 values for each of the following three ARIMA models: (1) the ARIMA(1, 1, 0) model, $(1 - .8B)(1 - B)Z_t = a_t$; (2) the ARIMA(0, 1, 1) model, $(1 - B)Z_t = (1 - .75B)a_t$; and (3) the ARIMA(1, 1, 1) model, $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$. The series a_t are independent Gaussian $N(0, 1)$ white noise. The sample ACF and PACF of the original three series are computed and shown in Table 4.3 and Figure 4.6. Each

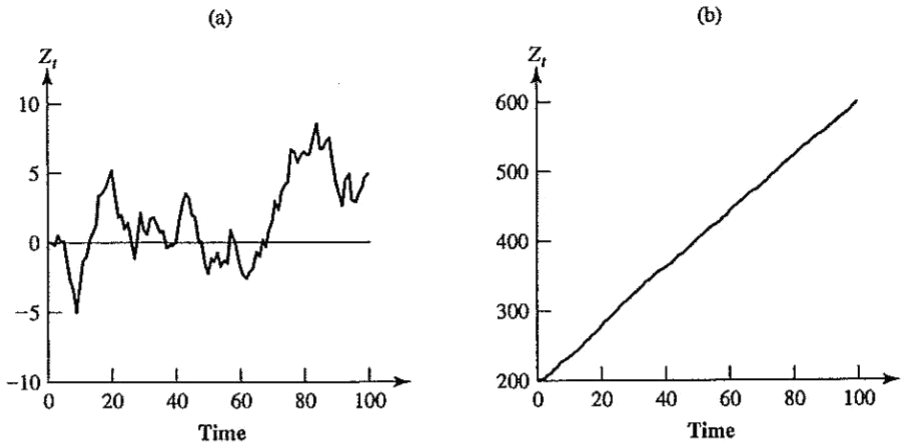


FIGURE 4.5 Simulated series from random walk models. (a) A simulated random walk series: $(1 - B)Z_t = a_t$. (b) A simulated random walk series with drift: $(1 - B)Z_t = 4 + a_t$.

TABLE 4.3 Sample ACF and sample PACF for the original series Z_t simulated from three ARIMA models.

(a) For Z_t from $(1 - .8B)(1 - B)Z_t = a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.97	.95	.92	.89	.86	.83	.80	.77	.74	.72
St.E.	.06	.11	.14	.16	.18	.19	.21	.22	.23	.24
$\hat{\phi}_{kk}$	.97	−.06	−.04	−.02	−.01	−.01	−.01	−.00	.02	.03
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

(b) For Z_t from $(1 - B)Z_t = (1 - .75)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.96	.91	.86	.82	.79	.75	.72	.69	.67	.66
St.E.	.06	.11	.13	.15	.17	.19	.20	.21	.22	.22
$\hat{\phi}_{kk}$	.96	−.13	.04	.01	.03	.01	.00	.13	.07	.06
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

(c) For Z_t from $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.98	.96	.94	.91	.89	.87	.84	.82	.80	.78
St.E.	.06	.11	.14	.16	.18	.20	.21	.23	.24	.25
$\hat{\phi}_{kk}$	.98	−.03	−.03	−.02	−.02	−.02	−.02	−.01	.01	.02
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

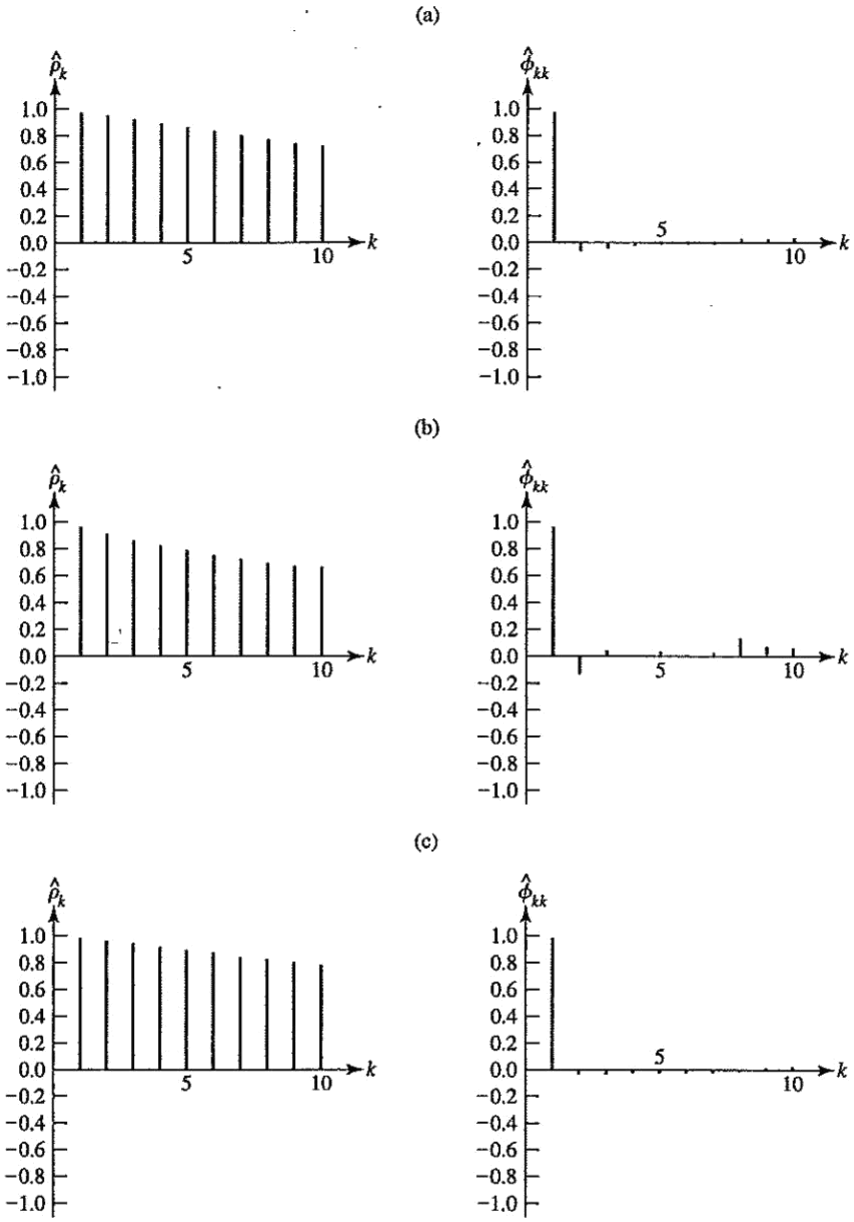


FIGURE 4.6 Sample ACF and PACF from three ARIMA models. (a) For Z_t from $(1 - .8B)(1 - B)Z_t = a_t$. (b) For Z_t from $(1 - B)Z_t = (1 - .75B)a_t$. (c) For Z_t from $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$.

TABLE 4.4 Sample ACF and sample PACF for the differenced series $W_t = (1 - B)Z_t$ simulated from three ARIMA models.

(a) For $W_t = (1 - B)Z_t$ from $(1 - .8B)(1 - B)Z_t = a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.71	.50	.36	.20	.12	.05	.02	.00	−.03	−.01
St.E.	.06	.09	.10	.11	.11	.11	.11	.11	.11	.11
$\hat{\phi}_{kk}$	.71	.01	.01	−.12	.04	−.04	.00	.11	−.03	.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

(b) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = (1 - .75B)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.40	−.07	−.02	−.10	−.08	−.06	−.06	−.14	−.04	.04
St.E.	.06	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	.40	−.28	.16	−.23	.11	−.16	.07	−.10	.02	.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

(c) For $W_t = (1 - B)Z_t$ from $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.41	.34	.33	.17	.18	.12	.08	.11	.05	.09
St.E.	.06	.07	.08	.08	.09	.09	.09	.09	.09	.09
$\hat{\phi}_{kk}$	.41	.20	.17	−.06	.04	−.01	.00	.04	−.03	.06
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

one shows the same phenomenon of the sustained large ACF and an exceptionally large first lag PACF. The dominating phenomenon of nonstationarity overshadows the fine details of the underlying characteristics of these models. To remove the shadow, we take differences for each of these series. The sample ACF and PACF for these differenced series are shown in Table 4.4 and Figure 4.7. Now, all the fine details are evident. The ACF $\hat{\rho}_k$ of $W_t = (1 - B)Z_t$ from the ARIMA(1, 1, 0) model tails off, whereas its PACF $\hat{\phi}_{kk}$ cuts off after lag 1; the ACF $\hat{\rho}_k$ of $W_t = (1 - B)Z_t$ from the ARIMA(0, 1, 1) model cuts off after lag 1, whereas its PACF $\hat{\phi}_{kk}$ tails off; and both the ACF $\hat{\rho}_k$ and PACF $\hat{\phi}_{kk}$ of $W_t = (1 - B)Z_t$ from the ARIMA(1, 1, 1) model tail off as expected from their characteristics discussed in Chapter 3.

4.3 Nonstationarity in the Variance and the Autocovariance

Differencing can be used to reduce a homogeneous nonstationary time series to a stationary time series. Many nonstationary time series, however, are nonhomogeneous. The nonstationarity of these series is not because of their time-dependent means but because of their

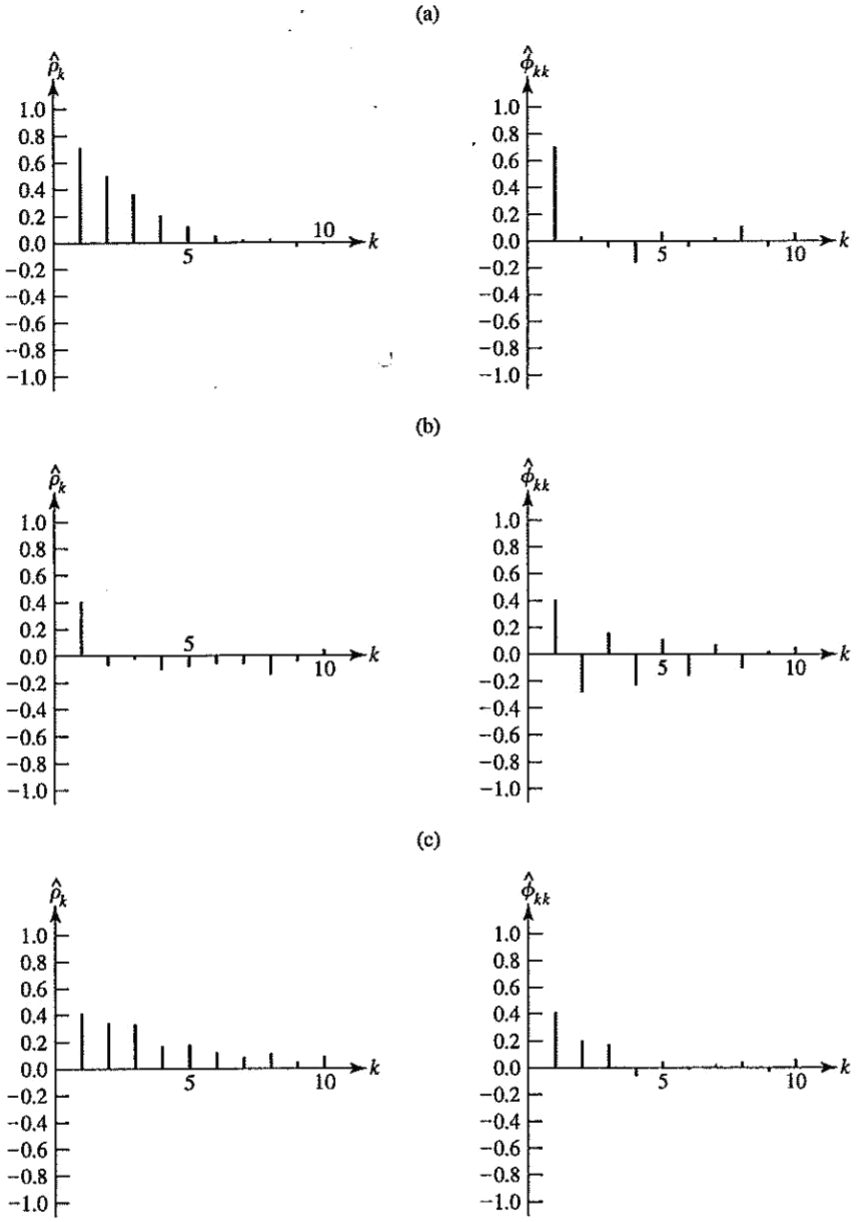


FIGURE 4.7 Sample ACF and PACF for differenced series from three ARIMA models. (a) For $W_t = (1 - B)Z_t$ from $(1 - .8B)(1 - B)Z_t = a_t$. (b) For $W_t = (1 - B)Z_t$ from $(1 - B)Z_t = (1 - .75B)a_t$. (c) For $W_t = (1 - B)Z_t$ from $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$.

time-dependent variances and autocovariances. To reduce these types of nonstationarity, we need transformations other than differencing.

4.3.1 Variance and Autocovariance of the ARIMA Models

A process that is stationary in the mean is not necessarily stationary in the variance and the autocovariance. A process that is nonstationary in the mean, however, will also be nonstationary in the variance and the autocovariance. As we have shown in the previous section, the mean function of the ARIMA model is time dependent. We now show that the ARIMA model is also nonstationary in its variance and autocovariance functions.

First, we note a very fundamental phenomenon about the ARIMA model. That is, although the model is nonstationary, the complete characteristic of the process is determined for all time only by a finite number of parameters, i.e., the ϕ_j, the θ_j, and σ_a^2. Thus, the complete future evolution of the process can be developed from the ARIMA model fitted to a given data set, $\{Z_1, Z_2, \ldots, Z_n\}$. For example, suppose that we fit the IMA(1, 1) model

$$(1 - B)Z_t = (1 - \theta B)a_t \tag{4.3.1a}$$

or

$$Z_t = Z_{t-1} + a_t - \theta a_{t-1} \tag{4.3.1b}$$

to a series of n_0 observations. With reference to this time origin n_0, for $t > n_0$, we can write by successive substitutions,

$$\begin{aligned} Z_t &= Z_{t-1} + a_t - \theta a_{t-1} \\ &= Z_{t-2} + a_t + (1 - \theta)a_{t-1} - \theta a_{t-2} \\ &\vdots \\ &= Z_{n_0} + a_t + (1 - \theta)a_{t-1} + \cdots + (1 - \theta)a_{n_0+1} - \theta a_{n_0}. \end{aligned} \tag{4.3.2}$$

Similarly, for $t - k > n_0$,

$$Z_{t-k} = Z_{n_0} + a_{t-k} + (1 - \theta)a_{t-k-1} + \cdots + (1 - \theta)a_{n_0+1} - \theta a_{n_0}. \tag{4.3.3}$$

Hence, with respect to the time origin n_0,

$$\mathrm{Var}(Z_t) = \left[1 + (t - n_0 - 1)(1 - \theta)^2\right]\sigma_a^2 \tag{4.3.4}$$

$$\mathrm{Var}(Z_{t-k}) = \left[1 + (t - k - n_0 - 1)(1 - \theta)^2\right]\sigma_a^2 \tag{4.3.5}$$

$$\mathrm{Cov}(Z_{t-k}, Z_t) = \left[(1 - \theta) + (t - k - n_0 - 1)(1 - \theta)^2\right]\sigma_a^2, \tag{4.3.6}$$

where we note that Z_{n_0} and a_{n_0} are known values with respect to the time origin n_0, and

$$\begin{aligned}\text{Corr}(Z_{t-k}, Z_t) &= \frac{\text{Cov}(Z_{t-k}, Z_t)}{\sqrt{\text{Var}(Z_{t-k})\,\text{Var}(Z_t)}} \\ &= \frac{(1-\theta) + (t-k-n_0-1)(1-\theta)^2}{\sqrt{\left[1 + (t-k-n_0-1)(1-\theta)^2\right]\left[1 + (t-n_0-1)(1-\theta)^2\right]}}.\end{aligned} \tag{4.3.7}$$

Now, from (4.3.4) to (4.3.7), we have the following important observations:

1. The variance, $\text{Var}(Z_t)$, of the ARIMA process is time dependent, and $\text{Var}(Z_t) \neq \text{Var}(Z_{t-k})$ for $k \neq 0$.
2. The variance $\text{Var}(Z_t)$ is unbounded as $t \to \infty$.
3. The autocovariance $\text{Cov}(Z_{t-k}, Z_t)$ and the autocorrelation $\text{Corr}(Z_{t-k}, Z_t)$ of the process are also time dependent and hence are not invariant with respect to time translation. In other words, they are not only functions of the time difference k but are also functions of both the time origin t and the original reference point n_0.
4. If t is large with respect to n_0, then from (4.3.7), $\text{Corr}(Z_{t-k}, Z_t) \simeq 1$. Because $|\text{Corr}(Z_{t-k}, Z_t)| \leq 1$, it implies that the autocorrelation function vanishes slowly as k increases.

In general, with only a single realization, it is difficult or impossible to make the statistical inferences of a process that is nonstationary in both the mean and the autocovariance or autocorrelation function. Fortunately, for the homogeneous nonstationary process, we can apply a proper differencing to reduce it to stationary process. That is, although the original series Z_t is nonstationary, its properly differenced series $W_t = (1-B)^d Z_t$ is stationary and can be represented as an ARMA process

$$\phi(B)W_t = \theta(B)a_t, \tag{4.3.8}$$

where

$$\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p) \qquad \text{and} \qquad \theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$$

have all their roots outside of the unit circle. Thus, the parameters ϕ_i, θ_j, and σ_a^2 that control the evolution of the nonstationary phenomenon of Z_t can be estimated from the differenced series W_t in exactly the same way as the stationary case discussed in Chapter 7.

4.3.2 Variance Stabilizing Transformations

Not all nonstationary series can be transformed to stationary series by differencing. Many time series are stationary in the mean but are nonstationary in the variance. To overcome this problem, we need a proper variance stabilizing transformation.

It is very common for the variance of a nonstationary process to change as its level changes. Thus,

$$\text{Var}(Z_t) = cf(\mu_t) \tag{4.3.9}$$

for some positive constant c and function f. How do we find a function T so that the transformed series, $T(Z_t)$, has a constant variance? To illustrate the method, we approximate the desired function by a first-order Taylor series about the point μ_t. Let

$$T(Z_t) \simeq T(\mu_t) + T'(\mu_t)(Z_t - \mu_t), \tag{4.3.10}$$

where $T'(\mu_t)$ is the first derivative of $T(Z_t)$ evaluated at μ_t. Now

$$\begin{aligned} \text{Var}\big[T(Z_t)\big] &\cong \big[T'(\mu_t)\big]^2 \text{Var}(Z_t) \\ &= c\big[T'(\mu_t)\big]^2 f(\mu_t). \end{aligned} \tag{4.3.11}$$

Thus, in order for the variance of $T(Z_t)$ to be constant, the variance stabilizing transformation $T(Z_t)$ must be chosen so that

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}}. \tag{4.3.12}$$

Equation (4.3.12) implies that

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t. \tag{4.3.13}$$

For example, if the standard deviation of a series is proportional to the level so that $\text{Var}(Z_t) = c\mu_t^2$, then

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t). \tag{4.3.14}$$

Hence, a logarithmic transformation (the base is irrelevant) of the series, $\ln(Z_t)$, will have a constant variance.

Next, if the variance of the series is proportional to the level so that $\text{Var}(Z_t) = c\mu_t$, then

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t}. \tag{4.3.15}$$

Thus, a square root transformation of the series, $\sqrt{Z_t}$, will have a constant variance.

Third, if the standard deviation of the series is proportional to the square of the level so that $\text{Var}(Z_t) = c\mu_t^4$, then

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^4}} d\mu_t = -\frac{1}{\mu_t}. \tag{4.3.16}$$

Therefore, a desired transformation that has a constant variance will be the reciprocal $1/Z_t$.

More generally, we can use the power transformation

$$T(Z_t) = \frac{Z_t^{\lambda} - 1}{\lambda}, \tag{4.3.17}$$

introduced by Box and Cox (1964). This class of transformations contains many of the previously discussed transformations as special cases. For example, the following table shows some commonly used values of λ and their associated transformations.

Values of λ (lambda)	Transformation
−1.0	$\frac{1}{Z_t}$
−0.5	$\frac{1}{\sqrt{Z_t}}$
0.0	$\ln Z_t$
0.5	$\sqrt{Z_t}$
1.0	Z_t (no transformation)

To see why $\lambda = 0$ corresponds to the logarithmic transformation, we note that

$$\lim_{\lambda \to 0} T(Z_t) = \lim_{\lambda \to 0} \frac{Z_t^{\lambda} - 1}{\lambda} = \ln(Z_t). \tag{4.3.18}$$

One great advantage of using the transformation in (4.3.17) is that we can treat λ as a transformation parameter and estimate its value from data. For example, we can include λ as a parameter in the model, $(1 - \phi_1 B - \cdots - \phi_p B^p)(Z_t^{(\lambda)} - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t$ and choose the value of λ that gives the minimum residual mean square error. Box and Cox (1964) show that the maximum likelihood estimate of λ is the one that minimizes the residual mean square error calculated from the fitted model on the "normalized" data

$$Z_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda(\tilde{Z})^{\lambda - 1}}, \tag{4.3.19}$$

where

$$\tilde{Z} = \left(\prod_{t=1}^{n} Z_t \right)^{1/n} \tag{4.3.20}$$

is the geometric mean of the data coming from the Jacobin of the transformation. Thus, for $\lambda = 0$, the residual mean square error is computed from the fitted model on

$$Z_t^{(0)} = \lim_{\lambda \to 0} Z_t^{(\lambda)} = (\ln Z_t)\tilde{Z}. \tag{4.3.21}$$

In a preliminary data analysis, one can use an AR model as an approximation to obtain the value of λ through an AR fitting that minimizes the residual mean square error. The choice of the optimum value for λ is based on the evaluations of the residual mean square error on a grid of λ values.

Some important remarks are in order.

1. The variance stabilizing transformations introduced above are defined only for positive series. This definition, however, is not as restrictive as it seems because a constant can always be added to the series without affecting the correlation structure of the series.
2. A variance stabilizing transformation, if needed, should be performed before any other analysis such as differencing.
3. Frequently, the transformation not only stabilizes the variance, but also improves the approximation of the distribution by a normal distribution.

EXERCISES

4.1 Consider the model

$$(1 - B)^2 Z_t = (1 - .3B - .5B^2)a_t.$$

(a) Is the model for Z_t stationary? Why?

(b) Let $W_t = (1 - B)^2 Z_t$. Is the model for W_t stationary? Why?

(c) Find the ACF for the second-order differences W_t.

4.2 Consider the following processes:

(a) $(1 - B)^2 Z_t = a_t - .81a_{t-1} + .38a_{t-2}$,

(b) $(1 - B)Z_t = (1 - .5B)a_t$.

Express each of the above processes in the AR representation by actually finding and plotting the π weights.

4.3 (a) Simulate a series of 100 observations from each of the following models:

(i) $(1 - B)Z_t = (1 - .6B)a_t$,

(ii) $(1 - B)Z_t = 5 + (1 - .6B)a_t$,

(iii) $(1 - .9B)(1 - B)Z_t = a_t$,

(iv) $(1 - .9B)(1 - B)Z_t = (1 - .5B)a_t$.

(b) Plot the simulated series.

(c) Calculate and discuss the behavior of the sample ACF, $\hat{\rho}_k$, and PACF, $\hat{\phi}_{kk}$, at $k = 0, 1, \ldots, 20$ for each simulated series and its differences.

4.4 Suppose that Z_t is generated according to $Z_t = a_t + ca_{t-1} + \cdots + ca_1$, for $t \geq 1$, where c is a constant.

(a) Find the mean and covariance for Z_t. Is it stationary?

(b) Find the mean and covariance for $(1 - B)Z_t$. Is it stationary?

4.5 Consider the simple white noise process, $Z_t = a_t$. Discuss the consequence of overdifferencing by examining the ACF, PACF, and AR representation of the differenced series, $W_t = Z_t - Z_{t-1}$.

4.6 Let $Z_1, Z_2, \ldots, Z_n$ be a random sample from a Poisson distribution with mean μ.

(a) Show that the variance of Z_t depends on its mean μ.

(b) Find a proper transformation so that the variance of the transformed variable is independent of μ.

(c) Find the variance of the transformed variable.

4.7 Let r_n be the Pearson correlation coefficient from a sample size of n. It is known that r_n is asymptotically distributed as $N(\rho, (1 - \rho^2)^2/n)$, where ρ is the population correlation coefficient. Show that Fisher's z-transformation $Z = \frac{1}{2}\ln((1 + r_n)/(1 - r_n))$ is actually a variance-stabilizing transformation.

5
Forecasting

Uncertainty is a fact of life for both individuals and organizations. Forecasting is essential for planning and operation control in a variety of areas such as production management, inventory systems, quality control, financial planning, and investment analysis. In this chapter, we develop the minimum mean square error forecasts for a time series following the stationary and nonstationary time series models introduced in Chapters 3 and 4. Formulas will also be developed to update forecasts when new information becomes available. We also discuss the implication of the constructed time series model in terms of its eventual forecast function.

5.1 Introduction

One of the most important objectives in the analysis of a time series is to forecast its future values. Even if the final purpose of time series modeling is for the control of a system, its operation is usually based on forecasting. The term *forecasting* is used more frequently in recent time series literature than the term *prediction*. Most forecasting results, however, are derived from a general theory of linear prediction developed by Kolmogorov (1939, 1941), Wiener (1949), Kalman (1960), Yaglom (1962), and Whittle (1983), among others.

Consider the general ARIMA(p, d, q) model

$$\phi(B)(1 - B)^d Z_t = \theta(B)a_t \tag{5.1.1}$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$, and the series a_t is a Gaussian $N(0, \sigma_a^2)$ white noise process. The deterministic trend parameter θ_0 is omitted for simplicity but no loss of generality. Equation (5.1.1) is one of the most commonly used models in forecasting applications. We discuss the minimum mean square error forecast of Z_t following this model for both cases when $d = 0$ and $d \neq 0$.

5.2 Minimum Mean Square Error Forecasts

In forecasting, our objective is to produce an optimum forecast that has no error or as little error as possible, which leads us to the minimum mean square error forecast. This forecast will produce an optimum future value with the minimum error in terms of the mean square error criterion.

5.2.1 Minimum Mean Square Error Forecasts for ARMA Models

To derive the minimum mean square error forecasts, we first consider the case when $d = 0$ and $\mu = 0$, i.e., the stationary ARMA model

$$\phi(B)\dot{Z}_t = \theta(B)a_t. \tag{5.2.1}$$

Because the model is stationary, we can rewrite it in a moving average representation,

$$\begin{aligned} Z_t &= \psi(B)a_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots, \end{aligned} \tag{5.2.2}$$

where

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \frac{\theta(B)}{\phi(B)} \tag{5.2.3}$$

and $\psi_0 = 1$. For $t = n + l$, we have

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j a_{n+l-j}. \tag{5.2.4}$$

Suppose that at time $t = n$ we have the observations $Z_n, Z_{n-1}, Z_{n-2}, \ldots$ and wish to forecast l-step ahead of future value Z_{n+l} as a linear combination of the observations $Z_n, Z_{n-1}, Z_{n-2}, \ldots$. Because Z_t for $t = n, n-1, n-2, \ldots$ can all be written in the form of (5.2.2), we can let the minimum mean square error forecast $\hat{Z}_n(l)$ of Z_{n+l} be

$$\hat{Z}_n(l) = \psi_l^* a_n + \psi_{l+1}^* a_{n-1} + \psi_{l+2}^* a_{n-2} + \cdots, \tag{5.2.5}$$

where the ψ_j^* are to be determined. The mean square error of the forecast is

$$E(Z_{n+l} - \hat{Z}_n(l))^2 = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} \left[\psi_{l+j} - \psi_{l+j}^*\right]^2,$$

which is easily seen to be minimized when $\psi_{l+j}^* = \psi_{l+j}$. Hence,

$$\hat{Z}_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \cdots. \tag{5.2.6}$$

Using (5.2.4) and that

$$E(a_{n+j} | Z_n, Z_{n-1}, \ldots) = \begin{cases} 0, & j > 0, \\ a_{n+j}, & j \le 0, \end{cases}$$

we have

$$E(Z_{n+l} | Z_n, Z_{n-1}, \ldots) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \cdots.$$

Thus, the minimum mean square error forecast of Z_{n+l} is given by its conditional expectation. That is,

$$\hat{Z}_n(l) = E(Z_{n+l}|Z_n, Z_{n-1}, \ldots). \tag{5.2.7}$$

$\hat{Z}_n(l)$ is usually read as the l-step ahead forecast of Z_{n+l} at the forecast origin n.

The forecast error is

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j}. \tag{5.2.8}$$

Because $E(e_n(l)|Z_t, t \le n) = 0$, the forecast is unbiased with the error variance

$$\text{Var}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2. \tag{5.2.9}$$

For a normal process, the $(1 - \alpha)$ 100% forecast limits are

$$\hat{Z}_n(l) \pm N_{\alpha/2}\left[1 + \sum_{j=1}^{l-1} \psi_j^2\right]^{1/2} \sigma_a, \tag{5.2.10}$$

where $N_{\alpha/2}$ is the standard normal deviate such that $P(N > N_{\alpha/2}) = \alpha/2$.

The forecast error $e_n(l)$ as shown in (5.2.8) is a linear combination of the future random shocks entering the system after time n. Specifically, the one-step ahead forecast error is

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}. \tag{5.2.11}$$

Thus, the one-step ahead forecast errors are independent, which implies that $\hat{Z}_n(1)$ is indeed the best forecast of Z_{n+1}. Otherwise, if one-step ahead forecast errors are correlated, then one can construct a forecast $\hat{a}_{n+1}$ of a_{n+1} from the available errors $a_n, a_{n-1}, a_{n-2}, \ldots$ and hence improve the forecast of Z_{n+1} by simply using $\hat{Z}_n(1) + \hat{a}_{n+1}$ as the forecast. The forecast errors for longer lead times, however, are correlated. This correlation is true for the forecast errors

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} \tag{5.2.12}$$

and

$$e_{n-j}(l) = Z_{n+l-j} - \hat{Z}_{n-j}(l) = a_{n+l-j} + \psi_1 a_{n+l-j-1} + \cdots + \psi_{l-1} a_{n-j+1}, \tag{5.2.13}$$

which are made at the same lead time l but different origins n and $n - j$ for $j < l$. It is also true for the forecast errors for different lead times made from the same time origin. For example,

$$\text{Cov}\left[e_n(2), e_n(1)\right] = E\left[(a_{n+2} + \psi_1 a_{n+1})(a_{n+1})\right] = \psi_1 \sigma_a^2. \tag{5.2.14}$$

5.2.2 Minimum Mean Square Error Forecasts for ARIMA Models

We now consider the general nonstationary ARIMA(p, d, q) model with $d \neq 0$, i.e.,

$$\phi(B)(1 - B)^d Z_t = \theta(B)a_t, \tag{5.2.15}$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^{\phi})$ is a stationary AR operator and $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ is an invertible MA operator, respectively. Although for this process the mean and the second-order moments such as the variance and the autocovariance functions vary over time, as shown in Chapter 4, the complete evolution of the process is completely determined by a finite number of fixed parameters. Hence, we can view the forecast of the process as the estimation of a function of these parameters and obtain the minimum mean square error forecast using a Bayesian argument. It is well known that using this approach with respect to the mean square error criterion, which corresponds to a squared loss function, when the series is known up to time n, the optimal forecast of Z_{n+l} is given by its conditional expectation $E(Z_{n+l}|Z_n, Z_{n-1}, \ldots)$. The minimum mean square error forecast for the stationary ARMA model discussed earlier is, of course, a special case of the forecast for the ARIMA(p, d, q) model with $d = 0$.

To derive the variance of the forecast for the general ARIMA model, we rewrite the model at time $t + l$ in an AR representation that exists because the model is invertible. Thus,

$$\pi(B)Z_{t+l} = a_{t+l}, \tag{5.2.16}$$

where

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1 - B)^d}{\theta(B)}, \tag{5.2.17}$$

or, equivalently,

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} + a_{t+l}. \tag{5.2.18}$$

Following Wegman (1986), we apply the operator

$$1 + \psi_1 B + \cdots + \psi_{l-1} B^{l-1}$$

to (5.2.18) and obtain

$$\sum_{j=0}^{\infty} \sum_{k=0}^{l-1} \pi_j \psi_k Z_{t+l-j-k} + \sum_{k=0}^{l-1} \psi_k a_{t+l-k} = 0, \tag{5.2.19}$$

where $\pi_0 = -1$ and $\psi_0 = 1$. It can be easily shown that

$$\sum_{j=0}^{\infty} \sum_{k=0}^{l-1} \pi_j \psi_k Z_{t+l-j-k} = \pi_0 Z_{t+l} + \sum_{m=1}^{l-1} \sum_{i=0}^{m} \pi_{m-i} \psi_i Z_{t+l-m} + \sum_{j=1}^{\infty} \sum_{i=0}^{l-1} \pi_{l-1+j-i} \psi_i Z_{t-j+1}. \tag{5.2.20}$$

Choosing ψ weights so that

$$\sum_{i=0}^{m} \pi_{m-i}\psi_i = 0, \qquad \text{for } m = 1, 2, \ldots, l-1, \tag{5.2.21}$$

we have

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{t-j+1} + \sum_{i=0}^{l-1} \psi_i a_{t+l-i}, \tag{5.2.22}$$

where

$$\pi_j^{(l)} = \sum_{i=0}^{l-1} \pi_{l-1+j-i}\psi_i. \tag{5.2.23}$$

Thus, given Z_t, for $t \le n$, we have

$$\begin{aligned} \hat{Z}_n(l) &= E(Z_{n+l} \mid Z_t, t \le n) \\ &= \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{n-j+1} \end{aligned} \tag{5.2.24}$$

because $E(a_{n+j} \mid Z_t, t \le n) = 0$, for $j > 0$. The forecast error is

$$\begin{aligned} e_n(l) &= Z_{n+l} - \hat{Z}_n(l) \\ &= \sum_{j=0}^{l-1} \psi_j a_{n+l-j}, \end{aligned} \tag{5.2.25}$$

where the ψ_j weights, by (5.2.21), can be calculated recursively from the π_j weights as follows:

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}\psi_i, \qquad j = 1, \ldots, l-1. \tag{5.2.26}$$

Note that (5.2.25) is exactly the same as (5.2.8). Hence, the results given in (5.2.7) through (5.2.14) hold for both stationary and nonstationary ARMA models.

For a stationary process, $\lim_{l\to\infty} \sum_{j=0}^{l-1} \psi_j^2$ exists. Hence, from (5.2.10), the eventual forecast limits approach two horizontally parallel lines as shown in Figure 5.1(a). For a nonstationary process, $\lim_{l\to\infty} \sum_{j=0}^{l-1} \psi_j^2$ does not exist. In fact, $\sum_{j=0}^{l-1} \psi_j^2$ increases and becomes unbounded as $l \to \infty$. Thus, the forecast limits in this case become wider and wider as the forecast lead l becomes larger and larger as shown in Figure 5.1(b). The practical implication of the latter case is that the forecaster becomes less certain about the result as the forecast lead time gets larger. For more discussion on the properties of the mean square error forecasts, see Shaman (1983).

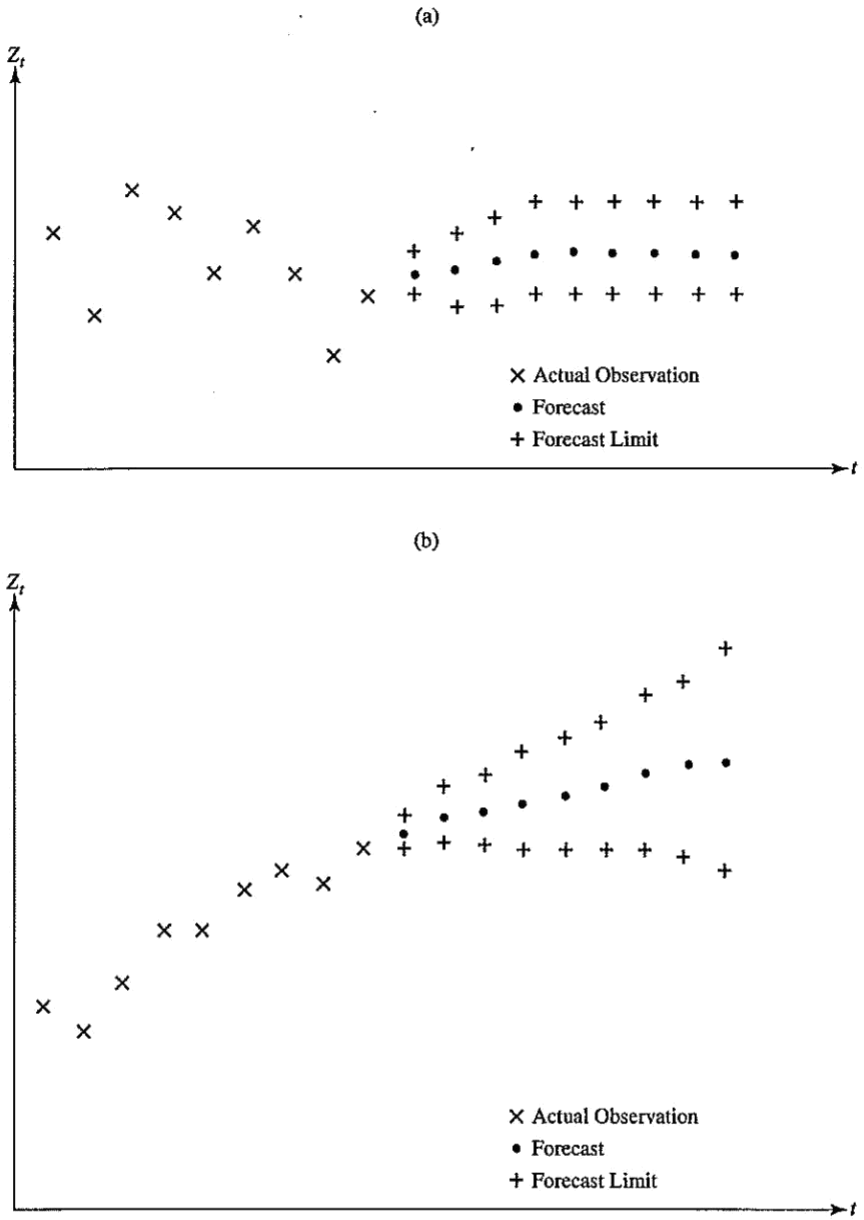


FIGURE 5.1 (a) Forecasts for stationary processes. (b) Forecasts for nonstationary processes.

5.3 Computation of Forecasts

From the result that the minimum mean square error forecasts $\hat{Z}_n(l)$ of Z_{n+l} at the forecast origin n is given by the conditional expectation

$$\hat{Z}_n(l) = E(Z_{n+l}|Z_n, Z_{n-1}, \ldots),$$

we can easily obtain the actual forecasts by directly using the difference equation form of the model. Let

$$\Psi(B) = \phi(B)(1 - B)^d = (1 - \Psi_1 B - \cdots - \Psi_{p+d}B^{p+d}).$$

The general ARIMA(p, d, q) model (5.2.15) can be written as the following difference equation form:

$$(1 - \Psi_1 B - \cdots - \Psi_{p+d}B^{p+d})Z_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t. \tag{5.3.1}$$

For $t = n + l$, we have

$$\begin{aligned} Z_{n+l} = {} & \Psi_1 Z_{n+l-1} + \Psi_2 Z_{n+l-2} + \cdots + \Psi_{p+d} Z_{n+l-p-d} \\ & + a_{n+l} - \theta_1 a_{n+l-1} - \cdots - \theta_q a_{n+l-q}. \end{aligned} \tag{5.3.2}$$

Taking the conditional expectation at time origin n, we obtain

$$\begin{aligned} \hat{Z}_n(l) = {} & \Psi_1 \hat{Z}_n(l-1) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d) \\ & + \hat{a}_n(l) - \theta_1 \hat{a}_n(l-1) - \cdots - \theta_q \hat{a}_n(l-q), \end{aligned} \tag{5.3.3}$$

where

$$\begin{aligned} \hat{Z}_n(j) &= E(Z_{n+j}|Z_n, Z_{n-1}, \ldots), & j \geq 1, \\ \hat{Z}_n(j) &= Z_{n+j}, & j \leq 0, \\ \hat{a}_n(j) &= 0, & j \geq 1, \end{aligned}$$

and

$$\hat{a}_n(j) = Z_{n+j} - \hat{Z}_{n+j-1}(1) = a_{n+j}, \qquad j \leq 0.$$

EXAMPLE 5.1 To illustrate the previous results, we consider the l-step ahead forecast $\hat{Z}_n(l)$ of Z_{n+l} for the following ARIMA(1, 0, 1) or ARMA(1, 1) model:

$$(1 - \phi B)(Z_t - \mu) = (1 - \theta B)a_t. \tag{5.3.4}$$

1. Calculate the forecast $\hat{Z}_n(l)$ as the conditional expectation from the difference equation form.

   For $t = n + l$, we can rewrite the above model in (5.3.4) as

$$Z_{n+l} = \mu + \phi(Z_{n+l-1} - \mu) + a_{n+l} - \theta a_{n+l-1}. \tag{5.3.5}$$

   Hence,

$$\hat{Z}_n(1) = \mu + \phi(Z_n - \mu) - \theta a_n \tag{5.3.6a}$$

   and

$$\begin{aligned} \hat{Z}_n(l) &= \mu + \phi\left[\hat{Z}_n(l-1) - \mu\right] \\ &= \mu + \phi^l(Z_n - \mu) - \phi^{l-1}\theta a_n, \qquad l \geq 2. \end{aligned} \tag{5.3.6b}$$

2. Calculate the forecast error variance $\mathrm{Var}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2$.

   When $|\phi| < 1$, we can calculate the ψ weights from the moving average representation (5.2.2) with $\phi(B) = (1 - \phi B)$ and $\theta(B) = (1 - \theta B)$. That is,

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = (1 - \theta B). \tag{5.3.7}$$

   Equating the coefficients of B^j on both sides gives

$$\psi_j = \phi^{j-1}(\phi - \theta), \qquad j \geq 1. \tag{5.3.8}$$

   Hence, the forecast error variance becomes

$$\mathrm{Var}(e_n(l)) = \sigma_a^2 \left\{ 1 + \sum_{j=1}^{l-1} \left[\phi^{j-1}(\phi - \theta)\right]^2 \right\}, \tag{5.3.9}$$

   which converges to $\sigma_a^2\left[1 + (\phi - \theta)^2/(1 - \phi^2)\right]$ as $l \to \infty$.

When $\phi = 1$, which corresponds to taking the first difference, the model in (5.3.4) becomes an IMA(1, 1) process

$$(1 - B)Z_t = (1 - \theta B)a_t, \tag{5.3.10}$$

where we note that $(1 - B)\mu = 0$. To calculate the ψ weights needed for the forecast variance, because the MA representation does not exist, we first rewrite (5.3.10) in an AR form when time equals $t + l$, i.e.,

$$\pi(B)Z_{t+l} = a_{t+l}$$

where

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots - = \frac{(1 - B)}{(1 - \theta B)}$$

or

$$(1 - B) = 1 - (\pi_1 + \theta)B - (\pi_2 - \pi_1\theta)B^2 - (\pi_3 - \pi_2\theta)B^3 - \cdots.$$

Equating the coefficients of B^j on both sides gives

$$\pi_j = (1 - \theta)\theta^{j-1}, \qquad j \geq 1. \tag{5.3.11}$$

Now, by applying (5.2.26) we obtain

$$\begin{aligned}
\psi_1 &= \pi_1 = (1 - \theta), \\
\psi_2 &= \pi_2 + \pi_1\psi_1 = (1 - \theta)\theta + (1 - \theta)^2 = (1 - \theta), \\
&\vdots
\end{aligned}$$

That is, we have $\psi_j = (1 - \theta)$, $1 \leq j \leq l - 1$. Thus, the variance of $e_n(l)$, from (5.2.9), becomes

$$\text{Var}(e_n(l)) = \sigma_a^2\left[1 + (l - 1)(1 - \theta)^2\right], \tag{5.3.12}$$

which approaches $+\infty$ as $l \to \infty$.

As expected, (5.3.12) is the limiting case of (5.3.9) when $\phi \to 1$. Thus, when ϕ is close to 1, the choice between a stationary ARMA(1, 1) model and a nonstationary IMA(1, 1) model has very different implications for forecasting. It can be seen even more clearly from the limiting behavior of the l-step ahead forecast $\hat{Z}_n(l)$ in (5.3.6b). For $|\phi| < 1$, $\hat{Z}_n(l)$ approaches the mean, μ, of the process as $l \to \infty$. When $\phi \to 1$, the first equation of (5.3.6b) implies that $\hat{Z}_n(l) = \hat{Z}_n(l - 1)$ for all l. That is, the forecast is free to wander, with no tendency for the values to remain clustered around a fixed level.

5.4 The ARIMA Forecast as a Weighted Average of Previous Observations

Recall that we can always represent an invertible ARIMA model in an autoregressive representation. In this representation, Z_t is expressed as an infinite weighted sum of previous observations plus a random shock, i.e.,

$$Z_{n+l} = \sum_{j=1}^{\infty} \pi_j Z_{n+l-j} + a_{n+l}, \tag{5.4.1}$$

or, equivalently,

$$\pi(B)Z_{n+l} = a_{n+l},$$

where

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)}. \tag{5.4.2}$$

Thus,

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j \hat{Z}_n(l-j), \qquad l \geq 1. \tag{5.4.3}$$

By repeated substitutions, we see that $\hat{Z}_n(l)$ can be expressed as a weighted sum of current and past observations Z_t, for $t \leq n$. For example,

$$\begin{aligned}
\hat{Z}_n(1) &= \pi_1 Z_n + \pi_2 Z_{n-1} + \pi_3 Z_{n-2} + \cdots \\
&= \sum_{j=1}^{\infty} \pi_j Z_{n+1-j}, \\
\hat{Z}_n(2) &= \pi_1 \hat{Z}_n(1) + \pi_2 Z_n + \pi_3 Z_{n-1} + \cdots \\
&= \pi_1 \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} + \sum_{j=1}^{\infty} \pi_{j+1} Z_{n+1-j} \\
&= \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j},
\end{aligned}$$

where

$$\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1}, \qquad j \geq 1.$$

More generally, it can be shown by successive substitutions that

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{n+1-j}, \tag{5.4.4}$$

where

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{i=1}^{l-1} \pi_i \pi_j^{(l-i)}, \qquad l > 1, \tag{5.4.5}$$

and

$$\pi_j^{(1)} = \pi_j.$$

Thus, many smoothing results, such as moving averages and exponential smoothing, are special cases of ARIMA forecasting. ARIMA models provide a very natural and optimal way to obtain the required weights for forecasting. The user does not have to specify either the number or the form of weights as required in the moving average method and the exponential smoothing method. Note also that the ARIMA forecasts are minimum mean square error forecasts. This optimal property is not shared in general by the moving average and the exponential smoothing methods.

For an invertible process, the π weights in (5.4.3) or (5.4.4) form a convergent series, which implies that for a given degree of accuracy, $\hat{Z}_n(l)$ depends only on a finite number of recent observations. The associated π weights provide very useful information for many important managerial decisions.

EXAMPLE 5.2 For the ARMA(1, 1) model in (5.3.4) with $|\theta| < 1$, we have, from (5.4.2), with $d = 0$,

$$(1 - \phi B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots)(1 - \theta B) \tag{5.4.6}$$

or

$$(1 - \phi B) = 1 - (\pi_1 + \theta)B - (\pi_2 - \pi_1\theta)B^2 - (\pi_3 - \pi_2\theta)B^3 - \cdots.$$

Equating the coefficients of B^j on both sides gives

$$\pi_j = (\phi - \theta)\theta^{j-1}, \qquad j \geq 1. \tag{5.4.7}$$

Assuming that $\mu = 0$, from (5.4.3), we have

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} (\phi - \theta)\theta^{j-1}\hat{Z}_n(l - j). \tag{5.4.8}$$

When $l = 1$, we get

$$\hat{Z}_n(1) = (\phi - \theta)\sum_{j=1}^{\infty} \theta^{j-1} Z_{n+1-j}. \tag{5.4.9}$$

For $l = 2$, from (5.4.4) and (5.4.5),

$$\begin{aligned}
\hat{Z}_n(2) &= \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j} \\
&= \sum_{j=1}^{\infty} \left[\pi_{j+1} + \pi_1\pi_j\right] Z_{n+1-j} \\
&= \sum_{j=1}^{\infty} \left[(\phi - \theta)\theta^j + (\phi - \theta)^2\theta^{j-1}\right] Z_{n+1-j} \\
&= \phi(\phi - \theta)\sum_{j=1}^{\infty} \theta^{j-1} Z_{n+1-j}.
\end{aligned} \tag{5.4.10}$$

Again, $\hat{Z}_n(2)$ is a weighted average of the previous observations with the $\pi_j^{(2)}$ weights given by $\pi_j^{(2)} = \phi(\phi - \theta)\theta^{j-1}$ for $j \geq 1$. Note that by comparing (5.4.9) and (5.4.10), we see that

$$\hat{Z}_n(2) = \phi\hat{Z}_n(1), \tag{5.4.11}$$

which, as expected, agrees with (5.3.6b) with $l = 2$ and $\mu = 0$.

To see the implications of these weights, let us examine (5.4.9) more carefully. For $|\theta| < 1$, $\pi_j = \theta^{j-1}(\phi - \theta)$ converges to 0 as $j \to \infty$, which implies that the more recent observations have a greater influence on the forecast. For $|\theta| \geq 1$, although the model is still stationary, its AR representation does not exist. To see the trouble, note that if $|\theta| > 1$, then the π weights rapidly approach $-\infty$ or $+\infty$ as j increases, which implies that the more remote past observations have a much greater influence on the forecast. When $|\theta| = 1$, the π weights become $\pi_j = (\phi - 1)$ for $\theta = 1$ and $\pi_j = (-1)^{j-1}(1 + \phi)$ for $\theta = -1$, which have the same absolute value for all j. Hence, all past and present observations are equally important in their effect on the forecast. Thus, a meaningful forecast can be derived only from an invertible process. The models corresponding to $|\theta| > 1$ and $|\theta| = 1$ are both noninvertible.

5.5 Updating Forecasts

Recall that when a time series Z_t is available for $t \leq n$, the l-step ahead minimum mean square forecast error for the forecast origin n using the general ARIMA model is obtained in (5.2.25), which, for convenience, is listed again in the following:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j}. \tag{5.5.1}$$

In particular, the one-step ahead forecast error is

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}. \tag{5.5.2}$$

Clearly, the forecast error at the forecast origin $(n - 1)$ is

$$Z_n - \hat{Z}_{n-1}(1) = a_n. \tag{5.5.3}$$

From (5.5.1), it is clear that

$$e_{n-1}(l + 1) = e_n(l) + \psi_l a_n, \tag{5.5.4}$$

where

$$e_{n-1}(l + 1) = Z_{n+l} - \hat{Z}_{n-1}(l + 1)$$

and

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l).$$

Hence, after substituting, rearranging, and using (5.5.3), we have the following updating equation:

$$\hat{Z}_n(l) = \hat{Z}_{n-1}(l + 1) + \psi_l\big[Z_n - \hat{Z}_{n-1}(1)\big], \tag{5.5.5}$$

or, equivalently,

$$\hat{Z}_{n+1}(l) = \hat{Z}_n(l+1) + \psi_l\big[Z_{n+1} - \hat{Z}_n(1)\big]. \tag{5.5.6}$$

The updated forecast is obtained by adding to the previous forecast a constant multiple ψ_l of the one-step ahead forecast error $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$, which is certainly sensible. For example, when the value Z_{n+1} becomes available and is found to be higher than the previous forecast (resulting in a positive forecast error $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$), we will naturally modify the forecast $\hat{Z}_n(l+1)$ made earlier by proportionally adding a constant multiple of this error.

5.6 Eventual Forecast Functions

Let the ARIMA model be

$$\psi(B)Z_t = \theta(B)a_t,$$

where $\psi(B) = \phi(B)(1-B)^d$. Recall from Equation (5.3.3) that

$$\begin{aligned}\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-1) + \Psi_2\hat{Z}_n(l-2) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d)\\ + \hat{a}_n(l) - \theta_1\hat{a}_n(l-1) - \cdots - \theta_q\hat{a}_n(l-q).\end{aligned}$$

When $l > q$, $\hat{Z}_n(l)$ becomes

$$\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-1) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d)$$

or

$$\Psi(B)\hat{Z}_n(l) = 0, \tag{5.6.1}$$

where B is now operating on l such that $B\hat{Z}_n(l) = \hat{Z}_n(l-1)$. That is, $\hat{Z}_n(l)$ for $l > q$ satisfies the homogeneous difference equation of order $(p+d)$. Let $\Psi(B) = \prod_{i=1}^{N}(1-R_iB)^{m_i}$ with $\sum_{i=1}^{N} m_i = (p+d)$. Then, from Theorem 2.7.1, the general solution is given by

$$\hat{Z}_n(l) = \sum_{i=1}^{N}\left(\sum_{j=0}^{m_i-1} C_{ij}^{(n)}l^j\right)R_i^l, \tag{5.6.2}$$

which, as seen shortly, holds actually for all $l \geq (q-p-d+1)$ and hence is often referred to as the eventual forecast function. The $C_{ij}^{(n)}$ are constants that depend on time origin n. For a given forecast origin n, they are fixed constants for all forecast lead times l. The constants change when the forecast origin is changed.

The reason that Equation (5.6.2) holds for all $l \geq (q-p-d+1)$ follows from (5.6.2) being the solution of Equation (5.6.1) for $l > q$. Thus, when $l = q+1$, the points $\hat{Z}_n(q), \ldots, \hat{Z}_n(q-p-d+1)$ actually lie on the curve represented by Equation (5.6.2).

In other words, the function in (5.6.2) is the unique curve that passes through the $(p + d)$ values given by $\hat{Z}_n(q), \hat{Z}_n(q-1), \ldots, \hat{Z}_n(q-p-d+1)$, where $\hat{Z}_n(-j) = Z_{n-j}$ for $j \geq 0$. These $(p + d)$ values can serve as initial conditions for solving the constants in (5.6.2).

EXAMPLE 5.3 For the ARIMA(1, 0, 1) model given in (5.3.4),

$$(1 - \phi B)(Z_t - \mu) = (1 - \theta B)a_t,$$

the forecast $\hat{Z}_n(l)$ satisfies the difference equation $(1 - \phi B)(\hat{Z}_t(l) - \mu) = 0$ for $l > 1$. The eventual forecast function is given by

$$\hat{Z}_n(l) - \mu = C_1^{(n)}\phi^l$$

or

$$\hat{Z}_n(l) = \mu + C_1^{(n)}\phi^l, \tag{5.6.3}$$

which holds for $l \geq (q - p - d + 1) = 1$. Thus, the function passes through $\hat{Z}_n(1)$, i.e., $\hat{Z}_n(1) = \mu + C_1^{(n)}\phi$ and $C_1^{(n)} = (\hat{Z}_n(1) - \mu)\phi^{-1}$. Thus, $\hat{Z}_n(l) = \mu + \left[\hat{Z}_n(1) - \mu\right]\phi^{l-1}$. As $l \to \infty$, $\hat{Z}_n(l)$ approaches the mean μ of the stationary process as expected.

EXAMPLE 5.4 Consider the ARIMA(1, 1, 1) model

$$(1 - \phi B)(1 - B)Z_t = (1 - \theta B)a_t.$$

The eventual forecast function is the solution of $(1 - \phi B)(1 - B)\hat{Z}_n(l) = 0$ and is given by

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)}\phi^l, \tag{5.6.4}$$

which holds for $l \geq (q - p - d + 1) = 0$. Thus, the function passes through the first forecast $\hat{Z}_n(1)$ and the last observation Z_n, and they can be used to solve the constants. Because $\hat{Z}_n(1) = C_1^{(n)} + C_2^{(n)}\phi$ and $Z_n = \hat{Z}_n(0) = C_1^{(n)} + C_2^{(n)}$, we have $C_1^{(n)} = (\hat{Z}_n(1) - \phi Z_n)/(1 - \phi)$, $C_2^{(n)} = (Z_n - \hat{Z}_n(1))/(1 - \phi)$; hence,

$$\hat{Z}_n(l) = \frac{\left(\hat{Z}_n(1) - \phi Z_n\right)}{(1 - \phi)} + \frac{\left(Z_n - \hat{Z}_n(1)\right)}{(1 - \phi)}\phi^l. \tag{5.6.5}$$

EXAMPLE 5.5 Consider the ARIMA(0, 2, 1) model

$$(1 - B)^2 Z_t = (1 - \theta B)a_t.$$

The eventual forecast function is the solution of $(1 - B)^2\hat{Z}_n(l) = 0$ for $l > 1$, and from (5.6.2) it is given by

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)} l, \tag{5.6.6}$$

which holds for $l \geq (q - p - d + 1) = 0$. Thus, the function is a straight line passing through $\hat{Z}_n(1)$ and Z_n. Because $\hat{Z}_n(1) = C_1^{(n)} + C_2^{(n)}$ and $Z_n = \hat{Z}_n(0) = C_1^{(n)}$, we have $C_1^{(n)} = Z_n$, $C_2^{(n)} = \hat{Z}_n(1) - Z_n$, and $\hat{Z}_n(l) = Z_n + (\hat{Z}_n(1) - Z_n)l$.

EXAMPLE 5.6 Consider the following ARIMA(1, 1, 1) model with a deterministic trend term

$$(1 - \phi B)(1 - B)Z_t = \theta_0 + (1 - \theta B)a_t. \tag{5.6.7}$$

The eventual forecast function is the solution of $(1 - \phi B)(1 - B)\hat{Z}_n(l) = \theta_0$ for $l > 1$. The solution for the homogeneous equation is given in Equation (5.6.4). For a particular solution, we note from the discussion in Section 4.2.2 that the deterministic term θ_0 in this case is related to the coefficient of a first-order deterministic time trend, $\alpha_1^{(n)} + \alpha_2^{(n)} l$. Thus,

$$(1 - \phi B)(1 - B)(\alpha_1^{(n)} + \alpha_2^{(n)} l) = \theta_0. \tag{5.6.8}$$

Because $(1 - B)(\alpha_1^{(n)} + \alpha_2^{(n)} l) = \alpha_2^{(n)}$, it follows that $(1 - \phi B)\alpha_2^{(n)} = \theta_0$ and $\alpha_2^{(n)} = \theta_0/(1 - \phi)$, and a particular solution becomes

$$\alpha_1^{(n)} + \theta_0 \frac{l}{(1 - \phi)}, \tag{5.6.9}$$

which holds also for $l \geq (q - p - d + 1) = 0$. Combining the solution (5.6.4) of the homogeneous equation and a particular solution given in (5.6.9), we get the eventual forecast function

$$\hat{Z}_n(l) = b_1^{(n)} + b_2^{(n)}\phi^l + \frac{\theta_0}{(1 - \phi)} l, \tag{5.6.10}$$

which passes through $\hat{Z}_n(1)$ and Z_n, where $b_1^{(n)}$ and $b_2^{(n)}$ are combined new constants to be determined. Because $\hat{Z}_n(1) = b_1^{(n)} + b_2^{(n)}\phi + \theta_0/(1 - \phi)$ and $Z_n = \hat{Z}_n(0) = b_1^{(n)} + b_2^{(n)}$, we have $b_1^{(n)} = [(\hat{Z}_n(1) - \phi Z_n)(1 - \phi) - \theta_0]/(1 - \phi)^2$, $b_2^{(n)} = [(Z_n - \hat{Z}_n(1))(1 - \phi) + \theta_0]/(1 - \phi)^2$, and

$$\hat{Z}_n(l) = \frac{[(\hat{Z}_n(1) - \phi Z_n)(1 - \phi) - \theta_0]}{(1 - \phi)^2} + \frac{[(Z_n - \hat{Z}_n(1))(1 - \phi) + \theta_0]}{(1 - \phi)^2}\phi^l + \frac{\theta_0}{(1 - \phi)} l. \tag{5.6.11}$$

The difference between Equations (5.6.11) and (5.6.5) is drastic. When the forecast lead time l is large, the forecast from model (5.6.7) is clearly dominated by the last term related to a deterministic trend.

5.7 A Numerical Example

As a numerical example, consider the AR(1) model

$$(1 - \phi B)(Z_t - \mu) = a_t$$

with $\phi = .6$, $\mu = 9$, and $\sigma_a^2 = .1$. Suppose that we have the observations $Z_{97} = 9.6$, $Z_{98} = 9$, $Z_{99} = 9$, $Z_{100} = 8.9$, and and want to forecast Z_{101}, Z_{102}, Z_{103}, and Z_{104} with their associated 95% forecast limits. We proceed as follows:

1. The AR(1) model can be written as

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + a_t,$$

and the general form of the forecast equation is

$$\begin{aligned} \hat{Z}_t(l) &= \mu + \phi(\hat{Z}_t(l-1) - \mu) \\ &= \mu + \phi^l(Z_t - \mu), \qquad l \geq 1. \end{aligned} \tag{5.7.1}$$

Thus,

$$\begin{aligned} \hat{Z}_{100}(1) &= 9 + .6(8.9 - 9) = 8.94, \\ \hat{Z}_{100}(2) &= 9 + (.6)^2(8.9 - 9) = 8.964, \\ \hat{Z}_{100}(3) &= 9 + (.6)^3(8.9 - 9) = 8.9784, \\ \hat{Z}_{100}(4) &= 9 + (.6)^4(8.9 - 9) = 8.98704. \end{aligned}$$

2. To obtain the forecast limits, we obtain the ψ weights from the relationship

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1. \tag{5.7.2}$$

That is,

$$\psi_j = \phi^j, \qquad j \geq 0. \tag{5.7.3}$$

The 95% forecast limits for Z_{101} from (5.2.10) are

$$8.94 \pm 1.96\sqrt{.1} \qquad \text{or} \qquad 8.320 < Z_{101} < 9.560.$$

The 95% forecast limits for Z_{102} are

$$8.964 \pm 1.96\sqrt{1 + (.6)^2}\sqrt{.1} \qquad \text{or} \qquad 8.241 < Z_{102} < 9.687.$$

The 95% forecast limits for Z_{103} and Z_{104} can be obtained similarly. The results are shown in Figure 5.2.

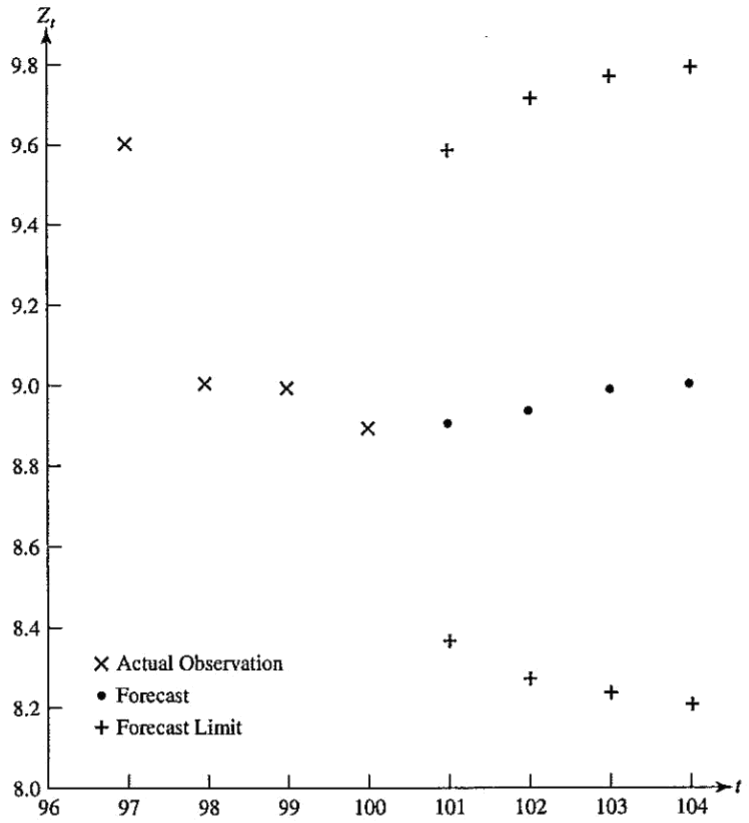


FIGURE 5.2 Forecasts with 95% forecast limits for an AR(1) process.

3. Suppose now that the observation at $t = 101$ turns out to be $Z_{101} = 8.8$. Because $\psi_l = \phi^l = (.6)^l$, we can update the forecasts for Z_{102}, Z_{103}, and Z_{104} by using (5.5.5) as follows:

$$\begin{aligned}
\hat{Z}_{101}(1) &= \hat{Z}_{100}(2) + \psi_1[Z_{101} - \hat{Z}_{100}(1)] \\
&= 8.964 + .6(8.8 - 8.94) = 8.88, \\
\hat{Z}_{101}(2) &= \hat{Z}_{100}(3) + \psi_2[Z_{101} - \hat{Z}_{100}(1)] \\
&= 8.9784 + (.6)^2(8.8 - 8.94) = 8.928, \\
\hat{Z}_{101}(3) &= \hat{Z}_{100}(4) + \psi_3[Z_{101} - \hat{Z}_{100}(1)] \\
&= 8.98704 + (.6)^3(8.8 - 8.94) = 8.9568.
\end{aligned}$$

The earlier forecasts for Z_{102}, Z_{103}, and Z_{104} made at $t = 100$ are adjusted downward due to the negative forecast error made for Z_{101}.

This forecasting is based on the assumption that the parameters are known in the model. In practice, the parameters are, of course, unknown and have to be estimated from the given

observations $\{Z_1, Z_2, \ldots, Z_n\}$. With respect to the forecasting origin $t = n$, however, the estimates are known constants; hence, the results remain the same under this conditional sense. Estimation of the model parameters is discussed in Chapter 7.

EXERCISES

5.1 Consider each of the following models:

(i) $(1 - \phi_1 B)(Z_t - \mu) = a_t$,

(ii) $(1 - \phi_1 B - \phi_2 B^2)(Z_t - \mu) = a_t$,

(iii) $(1 - \phi_1 B)(1 - B)Z_t = (1 - \theta_1 B)a_t$.

(a) Find the l-step ahead forecast $\hat{Z}_n(l)$ of Z_{n+l}.

(b) Find the variance of the l-step ahead forecast error for $l = 1, 2$, and 3.

5.2 (a) Show that the covariance between forecast errors from different origins is given by

$$\text{Cov}[e_n(l), e_{n-j}(l)] = \sigma_a^2 \sum_{i=j}^{l-1} \psi_i \psi_{i-j}, \qquad l > j.$$

(b) Show that the covariance between forecast errors from the same origin but with different lead times is given by

$$\text{Cov}[e_n(l), e_n(l + j)] = \sigma_a^2 \sum_{i=0}^{l-1} \psi_i \psi_{i+j}.$$

5.3 Consider the model

$$(1 - .68B)(1 - B)^2 Z_t = (1 - .75B + .34B^2)a_t.$$

(a) Compute and plot the correlations between the error of the forecast $\hat{Z}_t(5)$ and those of the forecasts $\hat{Z}_{t-j}(5)$ for $j = 1, 2, \ldots, 5$.

(b) Compute and plot the correlations between the error of the forecast $\hat{Z}_t(3)$ and those of $\hat{Z}_t(l)$ for $l = 1, 2, \ldots, 5$.

5.4 Verify Equation (5.2.20).

5.5 A sales series was fitted by the ARIMA(2, 1, 0) model

$$(1 - 1.4B + .48B^2)(1 - B)Z_t = a_t,$$

where $\hat{\sigma}_a^2 = 58{,}000$ and the last three observations are $Z_{n-2} = 640$, $Z_{n-1} = 770$, and $Z_n = 800$.

(a) Calculate the forecasts of the next three observations.

(b) Find the 95% forecast limits for the forecasts in part (a).

(c) Find the eventual forecast function.

5.6 Consider the IMA(1, 1) model

$$(1 - B)Z_t = (1 - \theta B)a_t.$$

(a) Write down the forecast equation that generates the forecasts, $\hat{Z}_n(l)$.
(b) Find the 95% forecast limits for $\hat{Z}_n(l)$.
(c) Express the forecasts as a weighted average of previous observations.
(d) Discuss the connection of this forecast $\hat{Z}_n(l)$ with the one generated by the simple exponential smoothing method.

5.7 (a) Show that (5.2.23) and (5.4.5) are equivalent.
(b) Illustrate the equivalence of (5.2.23) and (5.4.5) using the model in Exercise 5.6.

5.8 Consider the ARIMA(0, 1, 1) model

$$(1 - B)Z_t = (1 - .8B)a_t.$$

(a) Find the π weights for the AR representation

$$Z_t = \hat{Z}_t + a_t,$$

where $\hat{Z}_t = \sum_{j=1}^{\infty}\pi_j Z_{t-j}$, and show that $\sum_{j=1}^{\infty}\pi_j = 1$.
(b) Let $\hat{Z}_t(2) = \sum_{j=1}^{\infty}\pi_j^{(2)}Z_{t-j+1}$ be the two-step ahead forecasts of Z_{t+2} at time t. Express $\pi_j^{(2)}$ in terms of the weights π_j.

5.9 Consider an AR(2) model $(1 - \phi_1 B - \phi_2 B^2)(Z_t - \mu) = a_t$, where $\phi_1 = 1.2$, $\phi_2 = -.6$, $\mu = 65$, and $\sigma_a^2 = 1$. Suppose that we have the observations $Z_{76} = 60.4$, $Z_{77} = 58.9$, $Z_{78} = 64.7$, $Z_{79} = 70.4$, and $Z_{80} = 62.6$.
(a) Forecast Z_{81}, Z_{82}, Z_{83}, and Z_{84}.
(b) Find the 95% forecast limits for the forecasts in part (a).
(c) Suppose that the observation at $t = 81$ turns out to be $Z_{81} = 62.2$. Find the updated forecasts for Z_{82}, Z_{83}, and Z_{84}.

5.10 Consider the model

$$(1 - .43B)(1 - B)Z_t = a_t$$

and the observations $Z_{49} = 33.4$, $Z_{50} = 33.9$, and $\sigma_a^2 = 2$.
(a) Compute the forecast $Z_{50}(l)$, for $l = 1, 2, 3$ and their 90% forecast limits.
(b) What is the eventual forecast function for the forecasts made at $t = 50$?
(c) At time $t = 51$, Z_{51} became known and equaled 34.1. Update the forecasts obtained in part (a).

5.11 Obtain the eventual forecast function for the following models:
(a) $(1 - .6B)Z_t = (1 - .8B + .3B^2)a_t$,
(b) $(1 - .3B)(1 - B)Z_t = .4 + a_t$,

(c) $(1 - 1.2B + .6B^2)(Z_t - 65) = a_t$,

(d) $(1 - B)^2 Z_t = \theta_0 + (1 - .2B - .3B^2)a_t$,

(e) $(1 - .6B)(1 - B)^2 Z_t = (1 - .75B + .34B^2)a_t$.

5.12 Consider the AR(2) model, $Z_t = .2 + 1.8Z_{t-1} - .81Z_{t-2} + a_t$, where the a_t is the Gaussian white noise with mean 0 and variance 4, and $Z_{47} = 19$, $Z_{48} = 22$, $Z_{49} = 17$, and $Z_{50} = 21$.

(a) Find the mean of the process.

(b) Find the variance of the process.

(c) Find the forecasts, $\hat{Z}_{50}(\ell)$, for $l = 1, 2$, and 3.

(d) Find the 95% forecast limits in part (c).

(e) Find the eventual forecast function for the forecast made at $t = 50$ and its limiting value.

6

Model Identification

In time series analysis, the most crucial steps are to identify and build a model based on the available data. These steps require a good understanding of the processes discussed in Chapters 3 and 4, particularly the characteristics of these processes in terms of their ACF, ρ_k, and PACF, ϕ_{kk}. In practice, these ACF and PACF are unknown, and for a given observed time series $Z_1, Z_2, \ldots$, and Z_n, they have to be estimated by the sample ACF, $\hat{\rho}_k$, and PACF, $\hat{\phi}_{kk}$, discussed in Section 2.5. Thus, in model identification, our goal is to match patterns in the sample ACF, $\hat{\rho}_k$, and sample PACF, $\hat{\phi}_{kk}$, with the known patterns of the ACF, ρ_k, and PACF, ϕ_{kk}, for the ARMA models. For example, because we know that the ACF cuts off at lag 1 for an MA(1) model, a large single significant spike at lag 1 for $\hat{\rho}_k$ will indicate an MA(1) model as a possible underlying process.

After introducing systematic and useful steps for model identification, we give illustrative examples of identifying models for a wide variety of actual time series data. We also discuss some other identification tools such as the inverse autocorrelation and extended sample autocorrelation functions.

6.1 Steps for Model Identification

To illustrate the model identification, we consider the general ARIMA(p, d, q) model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t. \quad (6.1.1)$$

Model identification refers to the methodology in identifying the required transformations, such as variance stabilizing transformations and differencing transformations, the decision to include the deterministic parameter θ_0 when $d \geq 1$, and the proper orders of p and q for the model. Given a time series, we use the following useful steps to identify a tentative model.

Step 1. Plot the time series data and choose proper transformations.

In any time series analysis, the first step is to plot the data. Through careful examination of the plot, we usually get a good idea about whether the series contains a trend, seasonality, outliers, nonconstant variances, and other nonnormal and nonstationary phenomena. This understanding often provides a basis for postulating a possible data transformation.

In time series analysis, the most commonly used transformations are variance-stabilizing transformations and differencing. Because variance-stabilizing transformations such as the power transformation require non-negative values and differencing may create some negative values, we should always apply variance-stabilizing transformations before taking differences. A series with nonconstant variance often needs a logarithmic transformation. More generally, to stabilize the variance, we can apply Box–Cox's power transformation discussed in Section 4.3.2. Because a variance-stabilizing transformation, if necessary, is often chosen before we do any further analysis, we refer to this transformed data as the original series in the following discussion unless mentioned otherwise.

Step 2. Compute and examine the sample ACF and the sample PACF of the original series to further confirm a necessary degree of differencing so that differenced series is stationary. Some general rules are the following:

1. If the sample ACF decays very slowly (the individual ACF may not be large) and the sample PACF cuts off after lag 1, then it indicates that differencing is needed. Try taking the first differencing $(1 - B)Z_t$. One can also use the unit root test proposed by Dickey and Fuller (1979). In a borderline case, differencing is generally recommended (see Dickey, Bell, and Miller [1986]).
2. More generally, to remove nonstationarity we may need to consider a higher-order differencing $(1 - B)^d Z_t$ for $d > 1$. In most cases, d is either 0, 1, or 2. Note that if $(1 - B)^d Z_t$ is stationary, then $(1 - B)^{d+i} Z_t$ for $i = 1, 2, \ldots$ are also stationary. Some authors argue that the consequences of unnecessary differencing are much less serious than those of underdifferencing, but do beware of the artifacts created by overdifferencing so that unnecessary overparameterization can be avoided.

Step 3. Compute and examine the sample ACF and PACF of the properly transformed and differenced series to identify the orders of p and q where we recall that p is the highest order in the autoregressive polynomial $(1 - \phi_1 B - \cdots - \phi_p B^p)$, and q is the highest order in the moving average polynomial $(1 - \theta_1 B - \cdots - \theta_q B^q)$. Usually, the needed orders of these p and q are less than or equal to 3. Table 6.1 summarizes the important results for selecting p and q.

A strong duality exists between the AR and the MA models in terms of their ACFs and PACFs. To identify a reasonably appropriate ARIMA model, ideally, we need a minimum of $n = 50$ observations, and the number of sample lag-k autocorrelations and partial autocorrelations to be calculated should be about $n/4$, although occasionally for data of good quality one may be able to identify an adequate model with a smaller sample size.

TABLE 6.1 Characteristics of theoretical ACF and PACF for stationary processes.

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p, q)	Tails off after lag $(q - p)$	Tails off after lag $(p - q)$

We identify the orders p and q by matching the patterns in the sample ACF and PACF with the theoretical patterns of known models. The art of a time series analyst's model identification is very much like the method of an FBI agent's criminal search. Most criminals disguise themselves to avoid being recognized, which is also true of ACF and PACF. The sampling variation and the correlation among the sample ACF and PACF as shown in Section 2.5 often disguise the theoretical ACF and PACF patterns. Hence, in the initial model identification we always concentrate on the general broad features of these sample ACF and PACF without focusing on the fine details. Model improvement can be easily achieved at a later stage of diagnostic checking.

The estimated variances of both the sample ACF and PACF given in (2.5.21) and (2.5.27) are very rough approximations. Some authors recommend that a conservative threshold of 1.5 standard deviations be used in checking the significance of the short-term lags of these ACF and PACF at the initial model identification phase, especially for a relatively short series.

Step 4. Test the deterministic trend term θ_0 when $d > 0$.

As discussed in Section 4.2, for a nonstationary model, $\phi(B)(1 - B)^d Z_t = \theta_0 + \theta(B)a_t$, the parameter θ_0 is usually omitted so that it is capable of representing series with random changes in the level, slope, or trend. If there is reason to believe that the differenced series contains a deterministic trend mean, however, then we can test for its inclusion by comparing the sample mean $\overline{W}$ of the differenced series $W_t = (1 - B)^d Z_t$ with its approximate standard error $S_{\overline{W}}$. To derive $S_{\overline{W}}$, we note from Exercise 2.11 that $\lim_{n\to\infty} n \operatorname{Var}(\overline{W}) = \sum_{j=-\infty}^{\infty} \gamma_j$; hence,

$$\sigma_{\overline{W}}^2 = \frac{\gamma_0}{n} \sum_{j=-\infty}^{\infty} \rho_j = \frac{1}{n} \sum_{j=-\infty}^{\infty} \gamma_j = \frac{1}{n}\gamma(1), \tag{6.1.2}$$

where $\gamma(B)$ is the autocovariance generating function defined in (2.6.8) and $\gamma(1)$ is its value at $B = 1$. Thus, the variance and hence the standard error for $\overline{W}$ is model dependent. For example, for the ARIMA(1, d, 0) model, $(1 - \phi B)W_t = a_t$, we have, from (2.6.9),

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

so that

$$\begin{aligned} \sigma_{\overline{W}}^2 &= \frac{\sigma_a^2}{n} \frac{1}{(1 - \phi)^2} = \frac{\sigma_W^2}{n} \frac{1 - \phi^2}{(1 - \phi)^2} \\ &= \frac{\sigma_W^2}{n}\left(\frac{1 + \phi}{1 - \phi}\right) = \frac{\sigma_W^2}{n}\left(\frac{1 + \rho_1}{1 - \rho_1}\right), \end{aligned} \tag{6.1.3}$$

where we note that $\sigma_W^2 = \sigma_a^2/(1 - \phi^2)$. The required standard error is

$$S_{\overline{W}} = \sqrt{\frac{\hat{\gamma}_0}{n}\left(\frac{1 + \hat{\rho}_1}{1 - \hat{\rho}_1}\right)}. \tag{6.1.4}$$

Expressions of $S_{\overline{W}}$ for other models can be derived similarly. At the model identification phase, however; because the underlying model is unknown, most available software use the approximation

$$S_{\overline{W}} = \left[\frac{\hat{\gamma}_0}{n}(1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \cdots + 2\hat{\rho}_k)\right]^{1/2}, \tag{6.1.5}$$

where $\hat{\gamma}_0$ is the sample variance and $\hat{\rho}_1, \ldots, \hat{\rho}_k$ are the first k significant sample ACFs of $\{W_t\}$. Under the null hypothesis $\rho_k = 0$ for $k \geq 1$, Equation (6.1.5) reduces to

$$S_{\overline{W}} = \sqrt{\hat{\gamma}_0/n}. \tag{6.1.6}$$

Alternatively, one can include θ_0 initially and discard it at the final model estimation if the preliminary estimation result is not significant. Parameter estimation is discussed in Chapter 7.

6.2 Empirical Examples

In this section, we present a variety of real-world examples to illustrate the method of model identification. Many computer programs such as AUTOBOX, EVIEWS, MINITAB, RATS, SAS, SCA, SPLUS, and SPSS are available to facilitate the necessary computations. Some programs are available for both the mainframe and personal computers.

EXAMPLE 6.1 Figure 6.1 shows Series W1, which is the daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. The data consist of 45 daily observations of consecutive business days between

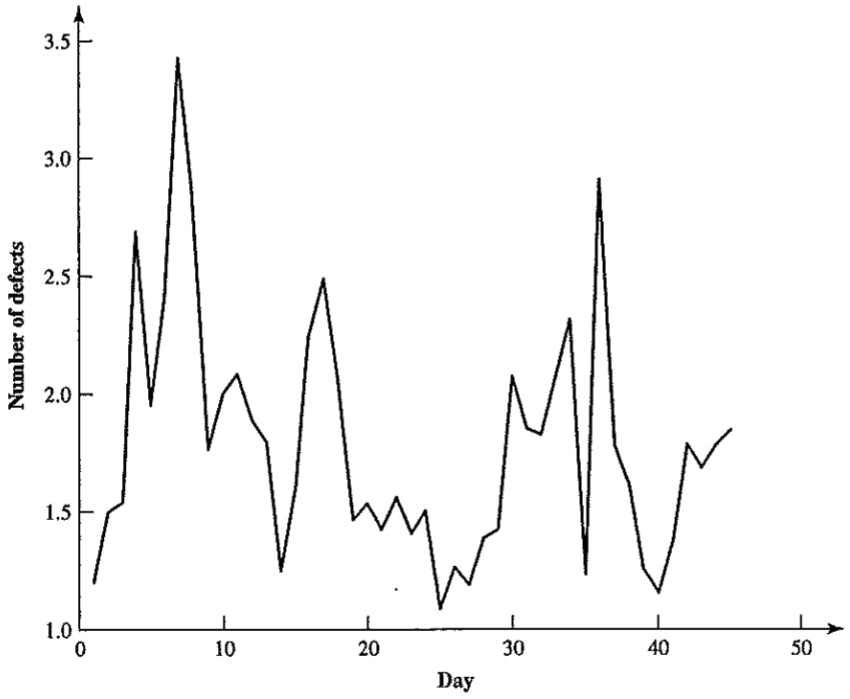


FIGURE 6.1 Daily average number of truck manufacturing defects (Series W1).

TABLE 6.2 Sample ACF and sample PACF for the daily average number of truck manufacturing defects (Series W1).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.43	.26	.14	.08	−.09	−.07	−.21	−.11	−.05	−.01
St.E.	.15	.15	.17	.18	.19	.19	.19	.19	.19	.19
$\hat{\phi}_{KK}$	.43	.09	.00	.00	−.16	.00	−.18	.07	.05	.01
St.E.	.15	.15	.15	.15	.15	.15	.15	.15	.15	.15

November 4 to January 10, as reported in Burr (1976, p. 134). The figure suggests a stationary process with constant mean and variance. The sample ACF and sample PACF are reported in Table 6.2 and plotted in Figure 6.2. That the sample ACF decays exponentially and the sample PACF has a single significant spike at lag 1 indicates that the series is likely to be generated by an AR(1) process,

$$(1 - \phi B)(Z_t - \mu) = a_t. \tag{6.2.1}$$

EXAMPLE 6.2 Series W2 is the classic series of the Wolf yearly sunspot numbers between 1700 and 2001, giving a total of $n = 302$ observations. Scientists believe that the sunspot numbers affect the weather of Earth and hence human activities such as agriculture, telecommunications, and others. The Wolf sunspot numbers have been discussed extensively in time series literature, for example, Yule (1927), Bartlett (1950), Whittle (1954), and Brillinger and Rosenblatt (1967). This series is also known as the Wolfer sunspot numbers, named after Wolfer, who was a student of Wolf. For an interesting account of the history of the series, see Izenman (1985).

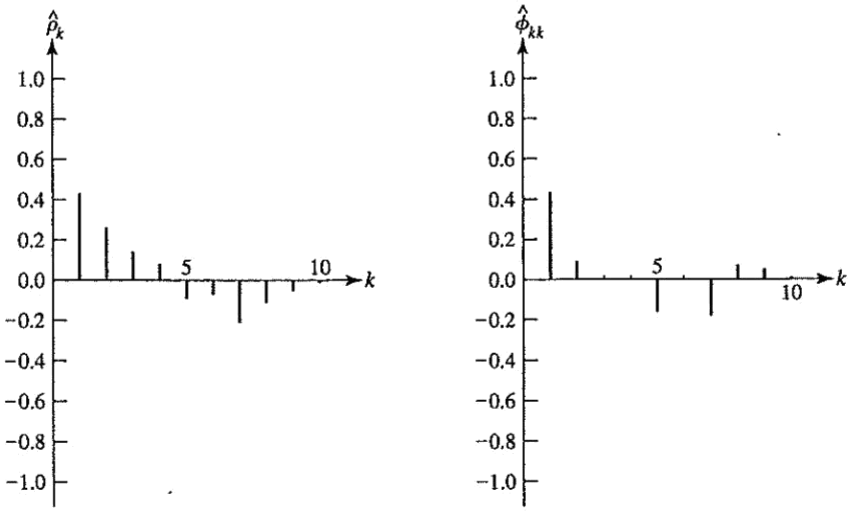


FIGURE 6.2 Sample ACF and sample PACF of the daily average number of truck manufacturing defects (Series W1).

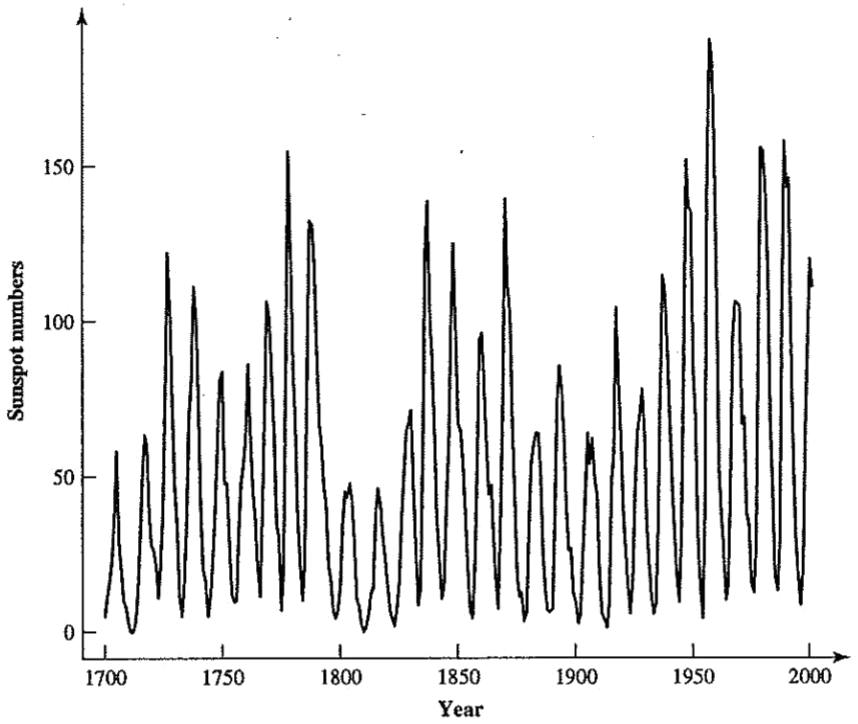


FIGURE 6.3 Wolf yearly sunspot numbers, 1700–2001 (Series W2).

The plot of the data is given in Figure 6.3. It indicates that the series is stationary in the mean but may not be stationary in variance. To investigate the required transformation for variance stabilization, we use the power transformation analysis discussed in Section 4.3.2 with the result given in Table 6.3. It suggests that a square root transformation be applied to the data.

The sample ACF and sample PACF of the transformed data are shown in Table 6.4 and Figure 6.4. The sample ACF shows a damping sine-cosine wave, and the sample PACF has relatively large spikes at lags 1, 2, and 9, suggesting that a tentative model may be an AR(2),

TABLE 6.3 Residual mean square errors in the power transformation.

λ	Residual mean square error
1.0	277.81
0.5	241.26
0.0	256.43
−0.5	6,964.83
−1.0	2,040,887.01

TABLE 6.4 Sample ACF and sample PACF for the square root transformed sunspot numbers (Series W2).

k	$\hat{\rho}_k$									
1–10	.82	.46	.05	−.26	−.42	−.39	−.19	.12	.44	.65
St.E.	.06	.09	.10	.10	.10	.10	.11	.11	.11	.12
11–20	.67	.50	.21	−.08	−.29	−.38	−.33	−.18	.05	.26
St.E.	.13	.14	.14	.15	.15	.15	.15	.15	.15	.15
k	$\hat{\phi}_{kk}$									
1–10	.82	−.68	−.12	.01	−.02	.20	.19	.20	.27	.00
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
11–20	−.02	.00	−.05	.11	−.04	−.11	−.08	−.13	.05	−.04
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

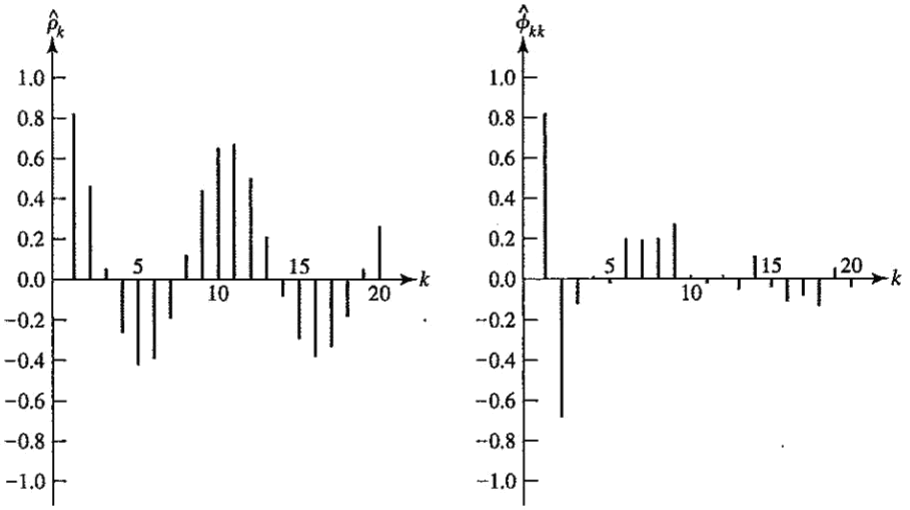


FIGURE 6.4 Sample ACF and sample PACF for the square root transformed sunspot numbers (Series W2).

$$\left(1 - \phi_1 B - \phi_2 B^2\right)\left(\sqrt{Z_t} - \mu\right) = a_t, \tag{6.2.2}$$

or an AR(9),

$$\left(1 - \phi_1 B - \cdots - \phi_9 B^9\right)\left(\sqrt{Z_t} - \mu\right) = a_t. \tag{6.2.3}$$

By ignoring the values of $\hat{\phi}_{kk}$ beyond lag 3, Box, Jenkins, and Reinsel (1994) suggest that an AR(3) model, $\left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3\right)\left(\sqrt{Z_t} - \mu\right) = a_t$, is also possible, even though

their analysis was based on the nontransformed data between 1770 and 1869. Because of the large autocorrelation of .67 at lag 11, many scientists believe that the series is cyclical with a period of 11 years. We examine this series more carefully in later chapters.

EXAMPLE 6.3 Series W3 is a laboratory series of blowfly data taken from Nicholson (1950). A fixed number of adult blowflies with balanced sex ratios were kept inside a cage and given a fixed amount of food daily. The blowfly population was then enumerated every other day for approximately two years, giving a total of $n = 364$ observations. Brillinger et al. (1980) first applied time series analysis on the data set. Later Tong (1983) considered the following two subseries:

Blowfly A: for Z_t between $20 \leq t \leq 145$,

Blowfly B: for Z_t between $218 \leq t \leq 299$,

and argued that the series Blowfly A is possibly generated by a nonlinear model. Series W3 used in our analysis is the series Blowfly B of 82 observations as shown in Figure 6.5.

The data plot suggests that the series is stationary in the mean but possibly nonstationary in variance. The power transformation analysis as shown in Table 6.5 indicates that no transformation is needed and the series is stationary in variance. The sample ACF and PACF are shown in Table 6.6 and Figure 6.6. The sample $\hat{\rho}_k$ decays exponentially and $\hat{\phi}_{kk}$ cuts off after lag 1. Thus, the following AR(1) model is entertained:

$$(1 - \phi B)(Z_t - \mu) = a_t. \tag{6.2.4}$$

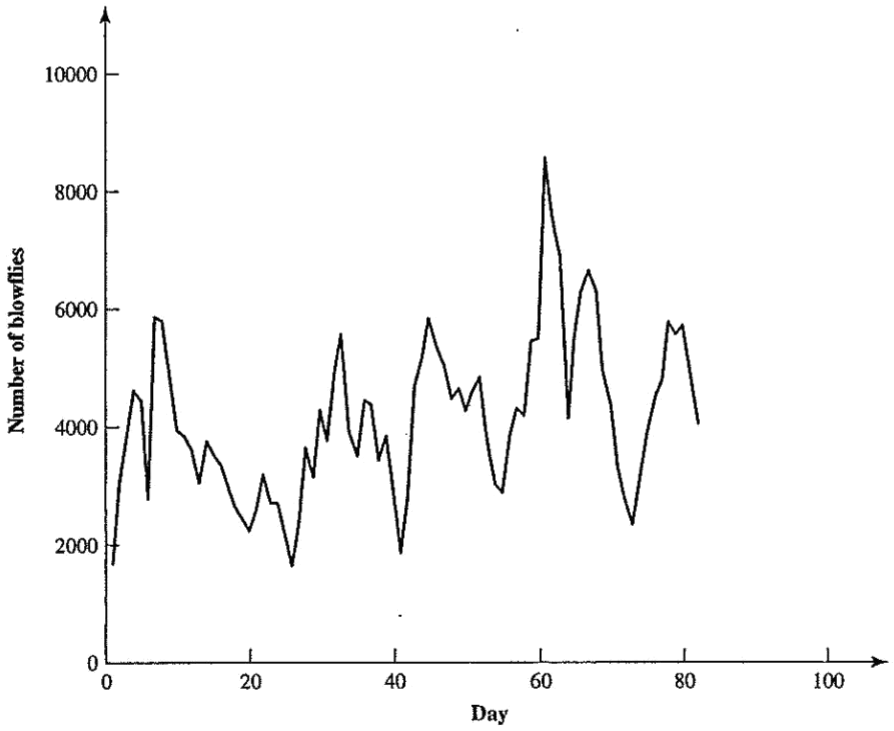


FIGURE 6.5 Blowfly data (Series W3).

TABLE 6.5 Results of the power transformation on blowfly data.

λ	Residual mean square error
1.0	811,704.20
0.5	816,233.82
0.0	825,749.54
−0.5	853,698.72
−1.0	997,701.34

TABLE 6.6 Sample ACF and sample PACF for the blowfly data (Series W3).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.73	.49	.30	.20	.12	.02	−.01	−.04	−.01	−.03
St.E.	.11	.16	.18	.18	.19	.19	.19	.19	.19	.19
$\hat{\phi}_{kk}$	.73	−.09	−.04	.04	−.03	−.12	.07	−.05	.07	−.08
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11

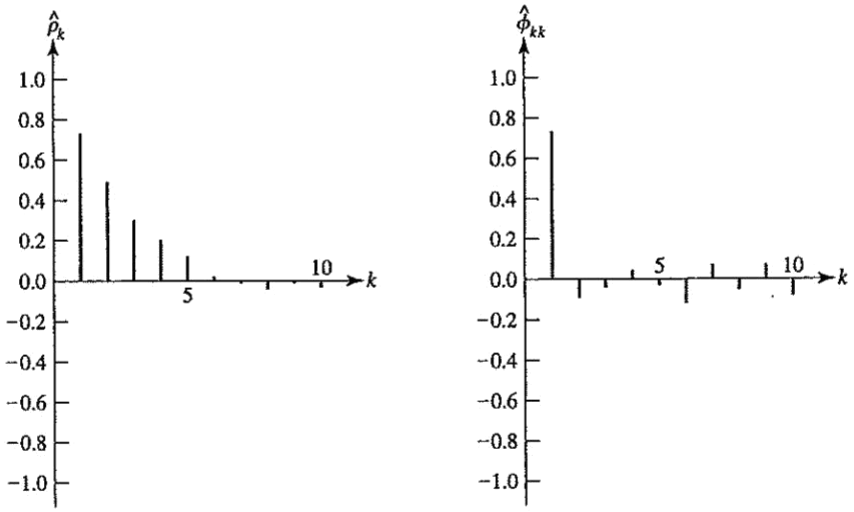


FIGURE 6.6 Sample ACF and sample PACF for the blowfly data (Series W3).

EXAMPLE 6.4 Series W4 is the monthly number of unemployed young women between ages 16 and 19 in the United States from January 1961 to August 2002 as shown in Figure 4.1. It is clearly nonstationary in the mean, suggesting differencing is needed. This conclusion is further confirmed by the sustained large spikes of the sample ACF shown in Table 6.7 and Figure 6.7. The differenced series as shown in Figure 6.8 seems to be stationary.

TABLE 6.7 Sample ACF and sample PACF of Series W4: the U.S. monthly series of unemployed young women between ages 16 and 19 from January 1961 to August 2002.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.96	.95	.94	.94	.93	.92	.91	.90	.89	.89
St.E.	.05	.08	.10	.11	.13	.14	.15	.16	.17	.18
$\hat{\phi}_{kk}$	.96	.43	.17	.13	.05	.03	−.02	.02	−.03	−.01
St.E.	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

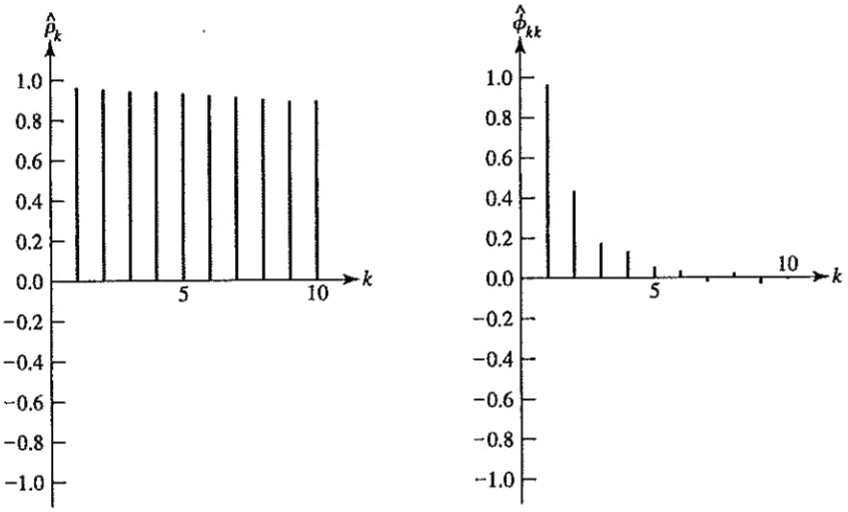


FIGURE 6.7 Sample ACF and sample PACF of Series W4: the U.S. monthly series of unemployed young women between ages 16 and 19 from January 1961 to August 2002.

The sample ACF and sample PACF for this differenced series are reported in Table 6.8 and also plotted in Figure 6.9. The sample ACF now cuts off after lag 1 while the sample PACF tails off. This pattern is very similar to the ACF and PACF for MA(1) with positive θ_1 as shown in Figure 3.10, suggesting an MA(1) model for the differenced series and hence an IMA(1, 1) model for the original series. To test whether a deterministic trend parameter θ_0 is needed, we calculate the t-ratio $\overline{W}/S_{\overline{W}} = .3707/1.925 = .1926$, which is not significant. Thus, our proposed model is

$$(1 - B)Z_t = (1 - \theta_1 B)a_t. \tag{6.2.5}$$

EXAMPLE 6.5 The death rate is of major interest for many state and federal governments. Figure 6.10 shows Series W5, which is the yearly cancer death rate (all forms, per 100,000 population) of Pennsylvania between 1930 and 2000 published in the 2000 Pennsylvania Vital Statistics Annual Report by the Pennsylvania Department of Health. The series is clearly nonstationary with an increasing trend. This nonstationarity is also shown by the slowly decaying ACF in Table 6.9 and Figure 6.11. Both Figure 6.11 and the evaluation of power

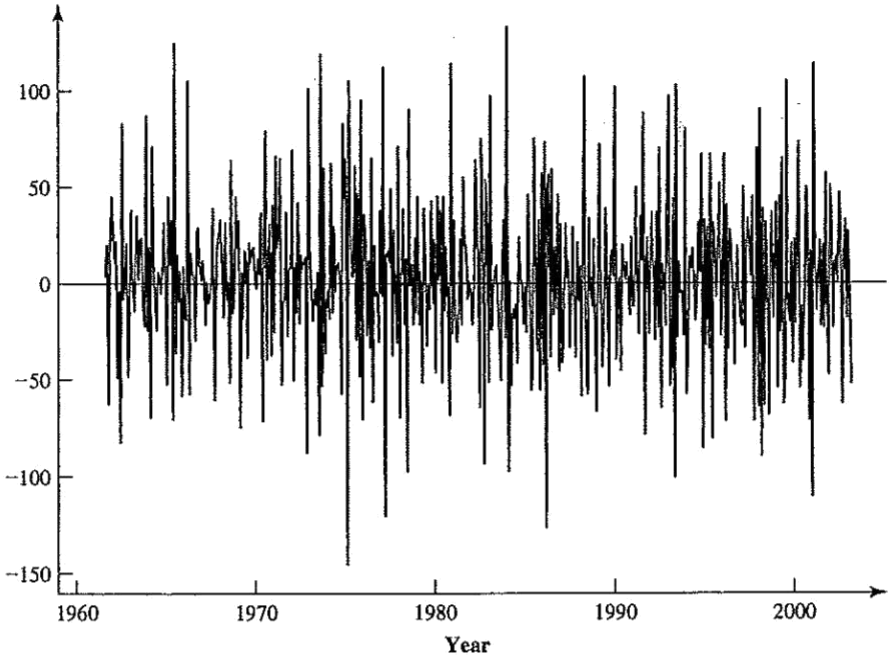


FIGURE 6.8 The differenced monthly series, $W_t = (1 - B)Z_t$, of the U.S. monthly series of unemployed young women between ages 16 and 19 from January 1961 to August 2002 (Series W4).

TABLE 6.8 Sample ACF and sample PACF of the differenced U.S. monthly series, $W_t = (1 - B)Z_t$, of unemployed young women between ages 16 and 19 (Series W4).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.47	.06	−.07	.04	.00	.04	−.04	.06	−.05	.01
St.E.	.04	.05	.05	.11	.05	.05	.05	.05	.05	.05
$\hat{\phi}_{kk}$	−.47	−.21	−.18	−.10	−.05	.02	−.01	.06	.01	.00
St.E.	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04

$\overline{W} = .3707, S_{\overline{W}} = 1.9246.$

transformations indicate no transformation other than differencing is needed. The sample ACF and PACF of the differenced data shown in Table 6.10 and Figure 6.12 suggest a white noise process for the differenced series. The t-ratio, $\overline{W}/S_{\overline{W}} = 2.0543/.3336 = 6.158$, implies that a deterministic trend term is recommended. Hence, the following random walk model with drift is entertained:

$$(1 - B)Z_t = \theta_0 + a_t. \tag{6.2.6}$$

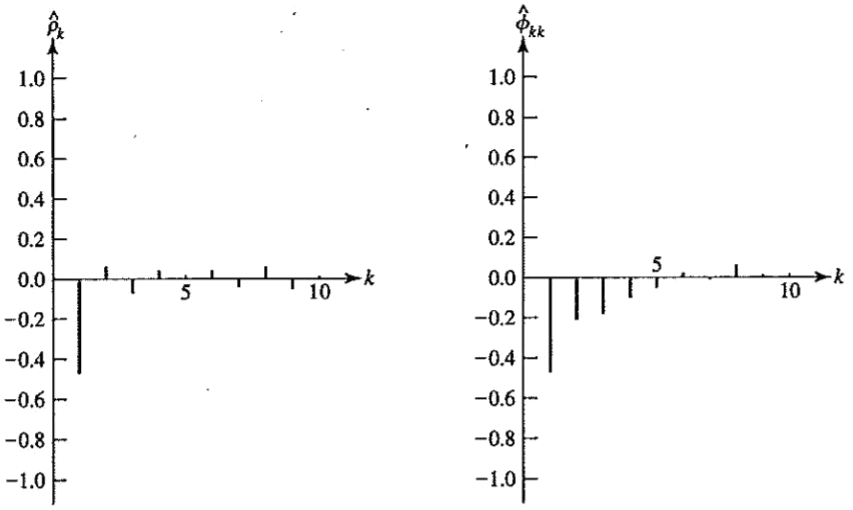


FIGURE 6.9 Sample ACF and sample PACF of the differenced U.S. monthly series, $W_t = (1 - B)Z_t$, of unemployed young women between ages 16 and 19 (Series W4).

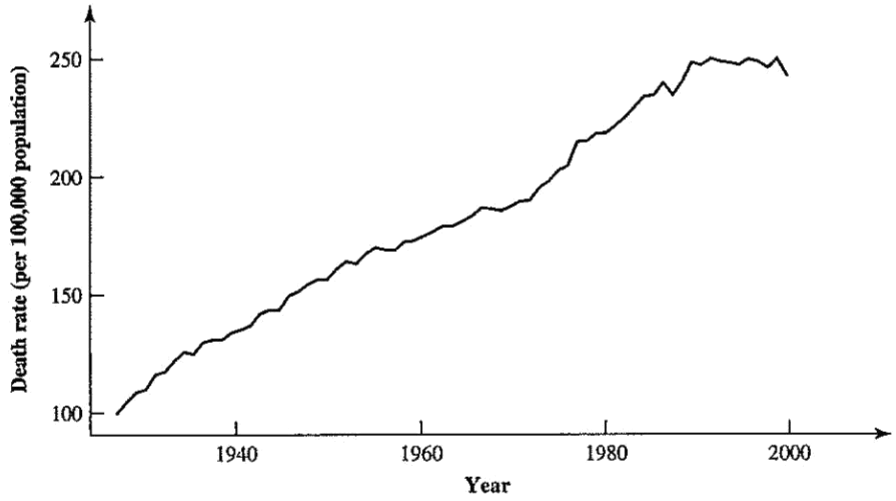


FIGURE 6.10 Series W5, the yearly Pennsylvania cancer death rate between 1930 and 2000.

Based on the sample ACF and PACF of the original nondifferenced data in Table 6.9, one may also suggest an alternative AR(1) model

$$(1 - \phi_1 B)(Z_t - \mu) = a_t. \tag{6.2.7}$$

The clear upward trend, however, suggests ϕ_1 to be close to 1. We will investigate both models in Chapter 7 when we discuss parameter estimation.

TABLE 6.9 Sample ACF and sample PACF of the Pennsylvania cancer death rate between 1930 and 2000 (Series W5).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.96	.92	.88	.84	.79	.75	.71	.67	.62	.58
St.E.	.12	.20	.25	.29	.32	.35	.37	.39	.41	.42
$\hat{\phi}_{kk}$	.96	−.04	.00	−.06	.00	−.03	−.02	−.03	−.06	−.01
St.E.	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12

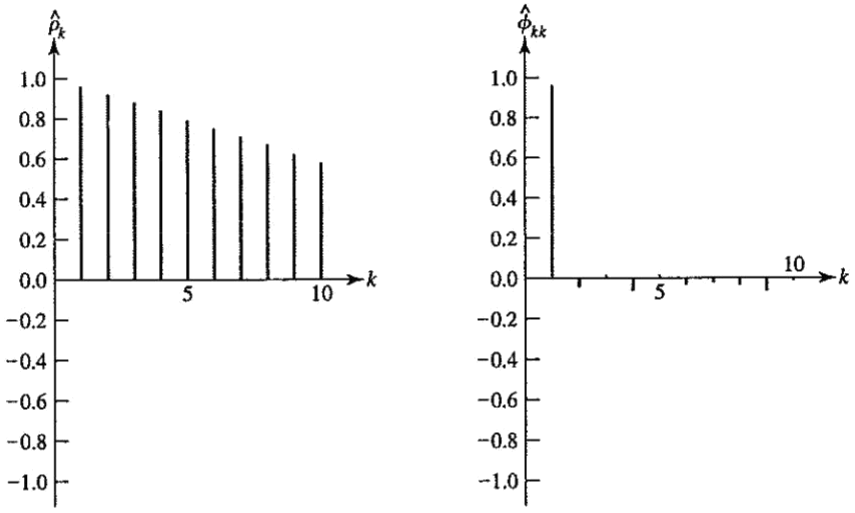


FIGURE 6.11 Sample ACF and sample PACF of Series W5, the yearly Pennsylvania cancer death rate between 1930 and 2000.

EXAMPLE 6.6 We now examine Series W6, which is the yearly U.S. tobacco production from 1871 to 1984 published in the 1985 Agricultural Statistics by the United States Department of Revenue and shown in Figure 4.2. The plot indicates that the series is nonstationary both in the mean and the variance. In fact, the standard deviation of the series is changing in time roughly in proportion to the level of the series. Hence, from the results of Section 4.3.2, a logarithmic transformation is suggested. These transformed data are plotted in Figure 6.13 and show an upward trend with a constant variance.

The very slowly decaying ACF as shown in Table 6.11 and Figure 6.14 further supports the need for differencing. Hence, the sample ACF and PACF for the differenced data, $W_t = (1 - B)\ln Z_t$, are reported in Table 6.12 with their plots in Figure 6.15. The ACF cuts off after lag 1, and the PACF tails off exponentially, which looks very similar to Figure 3.10 with $\theta_1 > 0$. It suggests that an IMA(1, 1) is a possible model. To determine whether a deterministic trend term θ_0 is needed, we examine the t-ratio, $t = \overline{W}/S_{\overline{W}} = .0147/.0186 = .7903$, which is not significant. Hence, we entertain the following IMA(1, 1) model as our tentative model:

$$(1 - B)\ln Z_t = (1 - \theta_1 B)a_t. \tag{6.2.8}$$

TABLE 6.10 Sample ACF and sample PACF for the differenced series of the Pennsylvania cancer death rate between 1930 and 2000 (Series W5).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.19	.12	.21	−.05	.07	.13	.07	−.13	.11	−.24
St.E.	.12	.12	.13	.13	.13	.13	.13	.13	.13	.13
$\hat{\phi}_{kk}$	−.19	.09	.25	.02	.01	.11	.13	−.16	−.03	−.25
St.E.	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12

$\overline{W} = 2.0543$, $S_{\overline{W}} = .3336$.

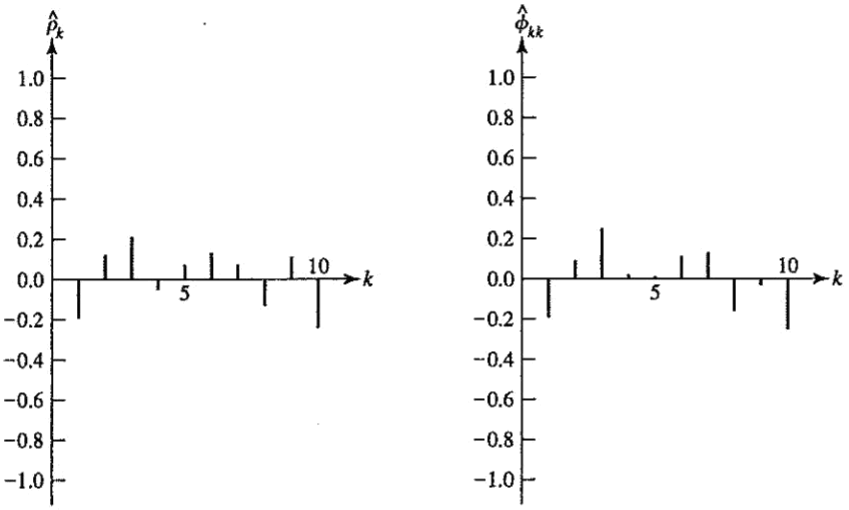


FIGURE 6.12 Sample ACF and sample PACF for the differenced series of the Pennsylvania cancer death rate between 1930 and 2000 (Series W5).

EXAMPLE 6.7 Figure 6.16(a) shows Series W7, the yearly number of lynx pelts sold by the Hudson's Bay Company in Canada between 1857 and 1911 as reported in Andrews and Herzberg (1985). The result of the power transformation in Table 6.13 indicates that a logarithmic transformation is required. The natural logarithm of the series is stationary and is plotted in Figure 6.16(b).

The sample ACF in Table 6.14 and Figure 6.17 show a clear sine-cosine phenomenon indicating an AR(p) model with $p \geq 2$. The three significant $\hat{\phi}_{kk}$ strongly suggest $p = 3$. Thus, our entertained model is

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(\ln Z_t - \mu) = a_t. \tag{6.2.9}$$

A related series that was studied by many time series analysts is the number of Canadian lynx trapped for the years from 1821 to 1934. References include Campbell and Walker (1977), Tong (1977), Priestley (1981, Section 5.5), and Lin (1987). The series of lynx pelt sales that we analyze here is much shorter and has received much less attention in the literature. We use this series extensively for various illustrations in this book.

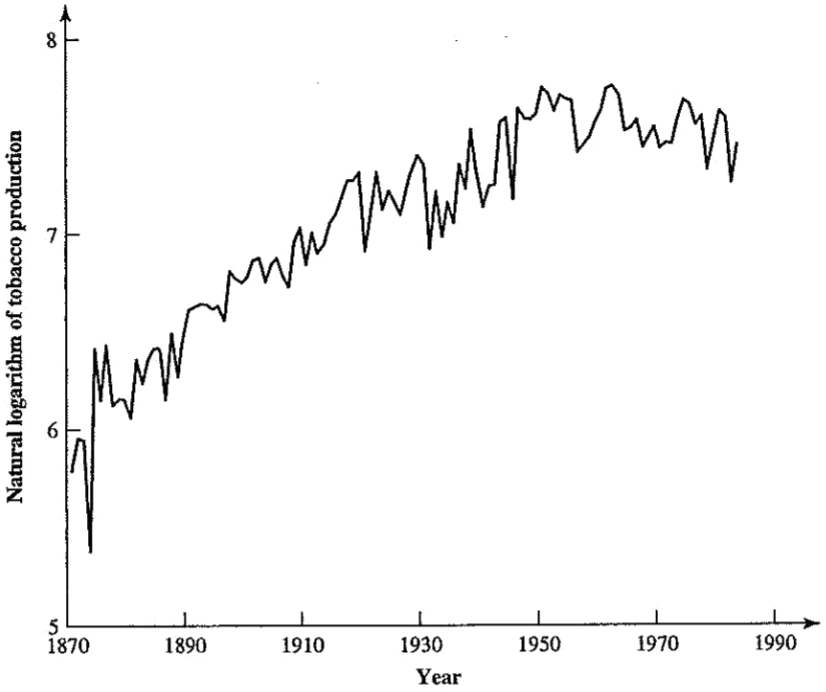


FIGURE 6.13 Natural logarithms of the U.S. yearly tobacco production in million pounds (Series W6).

TABLE 6.11 Sample ACF and sample PACF for natural logarithms of the U.S. yearly tobacco production (Series W6).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.90	.88	.84	.79	.78	.76	.75	.72	.69	.66
St.E.	.15	.19	.22	.24	.27	.28	.30	.32	.33	.34
$\hat{\phi}_{kk}$	.90	.37	.05	−.11	.15	.14	.08	−.11	−.12	.00
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

6.3 The Inverse Autocorrelation Function (IACF)

Let

$$\phi_p(B)(Z_t - \mu) = \theta_q(B)a_t \tag{6.3.1}$$

be an ARMA(p, q) model where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ is a stationary autoregressive operator, $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ is an invertible moving average

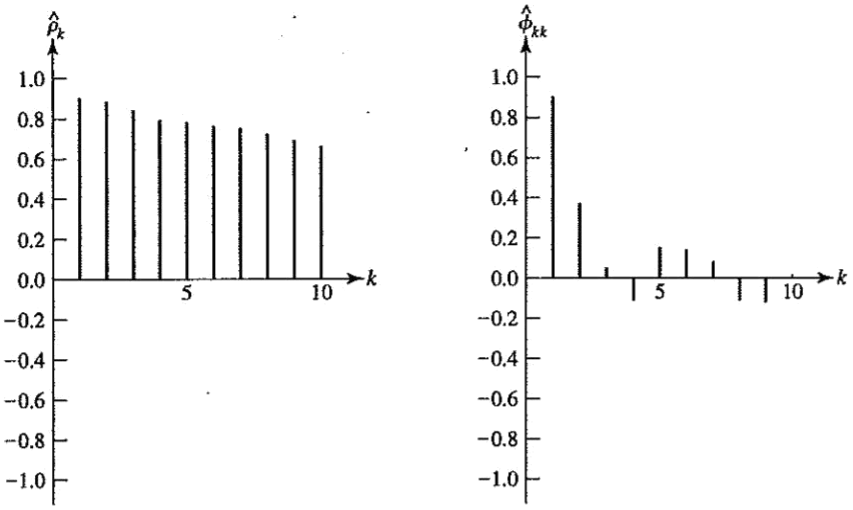


FIGURE 6.14 Sample ACF and sample PACF for natural logarithms of the U.S. yearly tobacco production (Series W6).

TABLE 6.12 Sample ACF and sample PACF for the differenced series of natural logarithms of the U.S. yearly tobacco production (Series W6).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.51	.11	−.09	.02	−.03	.00	.04	.04	−.05	−.01
St.E.	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12
$\hat{\phi}_{kk}$	−.51	−.20	−.17	−.14	−.13	−.12	−.04	.06	.02	−.03
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

$\overline{W} = .0147, S_{\overline{W}} = .0186.$

operator, and the a_t is a white noise series with a zero mean and a constant variance σ_a^2. We can rewrite Equation (6.3.1) as the moving average representation

$$(Z_t - \mu) = \frac{\theta_q(B)}{\phi_p(B)} a_t = \psi(B) a_t, \tag{6.3.2}$$

where $\psi(B) = \theta_q(B)/\phi_p(B)$. From (2.6.9), the autocovariance generating function of this model is given by

$$\begin{aligned} \gamma(B) &= \sum_{k=-\infty}^{\infty} \gamma_k B^k = \sigma_a^2 \psi(B)\psi(B^{-1}) \\ &= \sigma_a^2 \frac{\theta_q(B)\theta_q(B^{-1})}{\phi_p(B)\phi_p(B^{-1})}. \end{aligned} \tag{6.3.3}$$

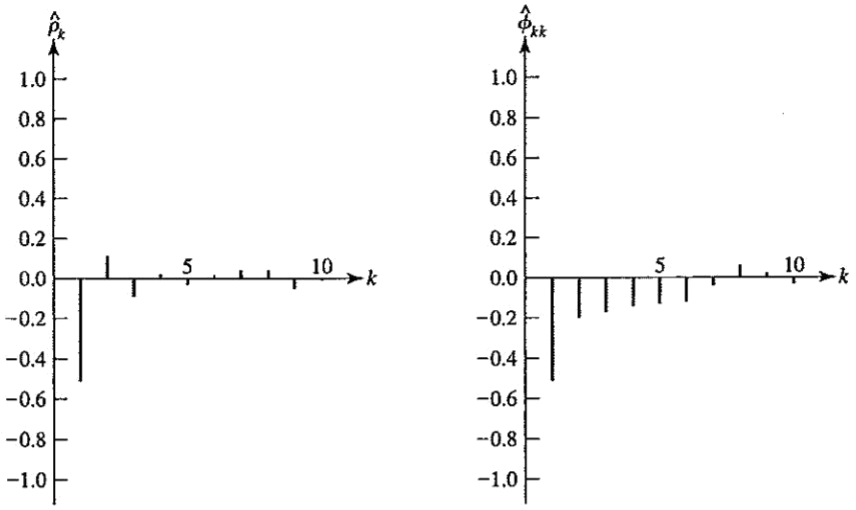


FIGURE 6.15 Sample ACF and sample PACF for the differenced natural logarithms of the tobacco data (Series W6).

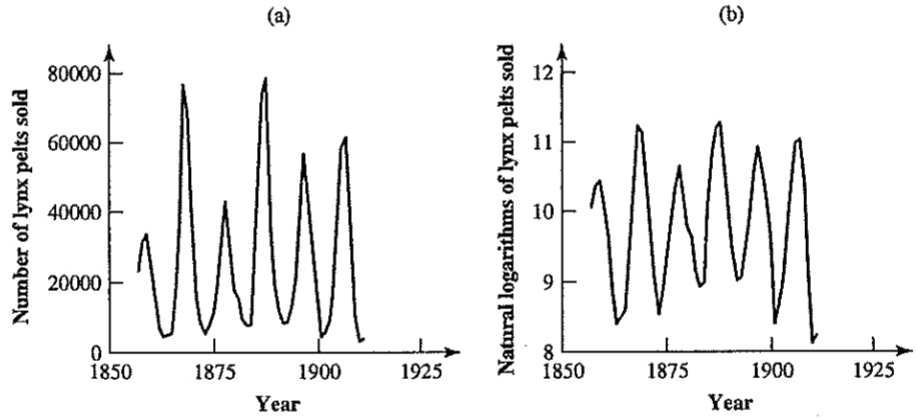


FIGURE 6.16 The Canadian lynx pelt sales data (Series W7). (a) The yearly number of lynx pelts sold in Canada between 1857 and 1911. (b) The natural logarithms of yearly numbers of lynx pelts sold in Canada between 1857 and 1911.

Assume that $\gamma(B) \neq 0$ for all $|B| = 1$ and let

$$\gamma^{(I)}(B) = \frac{1}{\gamma(B)} = \sum_{k=-\infty}^{\infty} \gamma_k^{(I)} B^k$$
$$= \frac{1}{\sigma_a^2} \frac{\phi_p(B)\phi_p(B^{-1})}{\theta_q(B)\theta_q(B^{-1})}. \tag{6.3.4}$$

TABLE 6.13 Results of the power transformation on the lynx pelt sales.

λ	Residual mean square error
1.0	113,502,248.25
0.5	78,458,357.01
0.0	54,193,701.90
−0.5	635,636,192.04
−1.0	9.2929725E14

TABLE 6.14 Sample ACF and sample PACF for natural logarithms of the yearly number of lynx pelts sold (Series W7).

(a) ACF, $\hat{\rho}_k$

1–10	.73	.22	−.32	−.69	−.76	−.53	−.08	.35	.61	.59
St.E.	.13	.19	.20	.21	.25	.29	.30	.30	.31	.33
11–20	.31	−.06	−.41	−.58	−.49	−.21	.16	.44	.54	.40
St.E.	.35	.36	.37	.37	.39	.40	.40	.41	.41	.43

(b) PACF, $\hat{\phi}_{kk}$

1–10	.73	−.68	−.36	−.20	−.09	−.08	.13	−.08	.06	−.07
St.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
11–20	−.13	.05	−.19	.07	−.02	−.04	.14	−.04	.10	−.09
St.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13

Clearly, from the assumptions on $\phi_p(B)$ and $\theta_q(B)$, the coefficients $\{\gamma_k^{(I)}\}$ in (6.3.4) form a positive definite, square summable sequence that is symmetric about the origin. Hence, they are proper autocovariances for a process. With respect to the model given in (6.3.1), the process having $\gamma^{(I)}(B)$ as its autocovariance generating function is referred to as the inverse process. Hence, $\gamma^{(I)}(B)$ is also referred to as the inverse autocovariance generating function of $\{Z_t\}$. Naturally, the inverse autocorrelation generating function of $\{Z_t\}$, from (2.6.10), is given by

$$\rho^{(I)}(B) = \frac{\gamma^{(I)}(B)}{\gamma_0^{(I)}}. \tag{6.3.5}$$

The kth lag inverse autocorrelation is defined as

$$\rho_k^{(I)} = \frac{\gamma_k^{(I)}}{\gamma_0^{(I)}}, \tag{6.3.6}$$

which, of course, is equal to the coefficient of B^k or B^{-k} in $\rho^{(I)}(B)$. As a function of k, $\rho_k^{(I)}$ is called the inverse autocorrelation function (IACF).

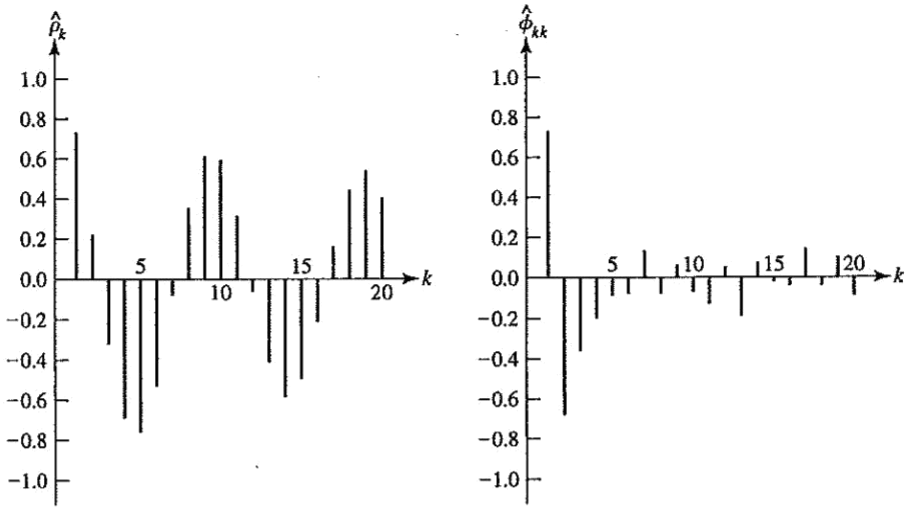


FIGURE 6.17 Sample ACF and sample PACF for natural logarithms of the yearly number of lynx pelts sold (Series W7).

From Equations (6.3.3) and (6.3.4), it is clear that if $\{Z_t\}$ is an ARMA (p, q) process, then the inverse process will be an ARMA(q, p) process. Specifically, if $\{Z_t\}$ is an AR(p) process with autocorrelations tailing off, then the inverse process will be an MA(p) process with its autocorrelations cutting off at lag p. In other words, the inverse autocorrelation function of an AR(p) process will cut off at lag p. Similarly, an MA(q) process with autocorrelations cutting off at lag q will have its inverse autocorrelations tailing off. Hence, the inverse autocorrelation function of a process exhibits characteristics similar to the partial autocorrelation function and can be used as an identification tool in model identification.

For an AR(p) process, it is easily seen that the inverse autocorrelation function is given by

$$\rho_k^{(I)} = \begin{cases} \dfrac{-\phi_k + \phi_1\phi_{k+1} + \cdots + \phi_{p-k}\phi_p}{1 + \phi_1^2 + \cdots + \phi_p^2}, & k = 1, \ldots, p, \\ 0, & k > p. \end{cases} \tag{6.3.7}$$

Because one can approximate an ARMA model by an AR(p) model with a sufficiently large p, one method to obtain a sample inverse autocorrelation $\hat{\rho}_k^{(I)}$ is to replace the AR parameters of the approximate AR model by their estimates in (6.3.7); i.e.,

$$\hat{\rho}_k^{(I)} = \begin{cases} \dfrac{-\hat{\phi}_k + \hat{\phi}_1\hat{\phi}_{k+1} + \cdots + \hat{\phi}_{p-k}\hat{\phi}_p}{1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_p^2}, & k = 1, \ldots, p, \\ 0, & k > p. \end{cases} \tag{6.3.8}$$

The parameter estimation is discussed in Chapter 7. Under the null hypothesis of a white noise process, the standard error of $\hat{\rho}_k^{(I)}$ is given by

$$S_{\hat{\rho}_k^{(I)}} = \sqrt{\frac{1}{n}}. \tag{6.3.9}$$

Thus, one can use the limits $\pm 2/\sqrt{n}$ to assess the significance of the sample inverse autocorrelations.

EXAMPLE 6.8 As illustrations, Table 6.15 shows sample inverse autocorrelations for two time series examined in Section 6.2. Table 6.15(a) is the sample inverse autocorrelation for the daily average series of truck manufacturing defects discussed in Example 6.1 in which we entertained an AR(1) model for the series. Note that although the inverse autocorrelation function seems to cut off at lag 1, it is not statistically significant at $\alpha = .05$. Table 6.15(b) is the sample inverse autocorrelation for the natural logarithms of the Canadian lynx pelt sales that was identified as an AR(3) model based on the three significant PACF examined in Example 6.7. In terms of the IACF, however, the maximum AR order will be 2. In fact, because the standard error of the sample IACF is .14, the model implied by the IACF may be an AR(1), which is clearly not appropriate from the sample ACF in Table 6.14(a).

It is not unusual that the values of the IACF are smaller than those of PACF in the above examples. If the underlying model is an AR(p), then we know that $\phi_{kk} = \rho_k^{(I)} = 0$ for $k > p$. Equation (6.3.7), however, implies that

$$\rho_p^{(I)} = \frac{-\phi_p}{1 + \phi_1^2 + \cdots + \phi_p^2}, \tag{6.3.10}$$

and from the discussion in the closing paragraph of Section 2.3, we have

$$\phi_{pp} = \phi_p. \tag{6.3.11}$$

TABLE 6.15 The sample inverse autocorrelation function (SIACF).

(a) The daily average number of truck manufacturing defects (Series W1)

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k^{(I)}$	−.27	.05	−.03	−.07	.10	−.11	.22	.01	−.05	.02
St.E.	.15	.15	.15	.15	.15	.15	.15	.15	.15	.15

(b) The natural logarithms of Canadian lynx pelt sales (Series W7)

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k^{(I)}$	−.67	.23	.01	−.03	.09	−.11	.15	−.11	.07	−.13
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14

Hence, $|\phi_{pp}| > |\rho_p^{(I)}|$, and values of the sample IACF in general are smaller than values of the sample PACF, particularly at lower lags. In one study, Abraham and Ledolter (1984) conclude that, as an identification aid, the PACF generally outperforms the IACF. Some computer programs such as SAS (1999) and SCA (1992) provide both PACF and IACF options for analysis.

The inverse autocorrelation function was first introduced by Cleveland (1972) through the inverse of a spectrum and the relationship between the spectrum and the autocorrelation function, leading to another method to obtain a sample inverse autocorrelation function. We return to this point in Chapter 12.

6.4 Extended Sample Autocorrelation Function and Other Identification Procedures

The autocorrelation function, the partial autocorrelation function, and the inverse autocorrelation function are useful in identifying the orders of AR and MA models. In this section, we introduce the extended sample autocorrelation function and other procedures that will help us identify the orders of mixed ARMA models.

6.4.1 The Extended Sample Autocorrelation Function (ESACF)

From the previous empirical examples, it seems clear that, due to the cutting off property of the PACF and IACF for AR models and the same cutting off property of the ACF for MA models, the identification of the order p of an AR model and the order q of an MA model through the sample ACF, PACF, and IACF are relatively simple. For a mixed ARMA process, however, the ACF, PACF, and IACF all exhibit tapering off behavior, which makes the identification of the orders p and q much more difficult. One commonly used method considers that if Z_t follows an ARMA(p, q) model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \tag{6.4.1a}$$

or, equivalently,

$$Z_t = \theta_0 + \sum_{i=1}^{p} \phi_i Z_{t-1} - \sum_{i=1}^{q} \theta_i a_{t-1} + a_t, \tag{6.4.1b}$$

then

$$\begin{aligned} Y_t &= (1 - \phi_1 B - \cdots - \phi_p B^p)Z_t \\ &= Z_t - \sum_{i=1}^{p} \phi_i Z_{t-i} \end{aligned} \tag{6.4.2}$$

follows an MA(q) model

$$Y_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \tag{6.4.3}$$

where without loss of generality we assume $\theta_0 = 0$. Thus, some authors such as Tiao and Box (1981) suggest using the sample ACF of the estimated residuals

$$\hat{Y}_t = (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p) Z_t \tag{6.4.4}$$

from an ordinary least squares (OLS) AR fitting to identify q and hence obtaining the orders p and q for the ARMA(p, q) model. For example, an MA(2) residual process from an AR(1) fitting implies a mixed ARMA(1, 2) model. But as we show in Section 7.4, however, because the OLS estimates of the AR parameter ϕ_i's in (6.4.4) are not consistent when the underlying model is a mixed ARMA(p, q) with $q > 0$, the procedure may lead to incorrect identification.

To derive consistent estimates of ϕ_i, suppose that n observations adjusted for mean are available from the ARMA(p, q) process in (6.4.1a). If an AR(p) model is fitted to the data, i.e., if

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + e_t, \qquad t = p + 1, \ldots, n \tag{6.4.5}$$

where e_t represents the error term, then the OLS estimates $\hat{\phi}_i^{(0)}$ of ϕ_i, $i = 1, \ldots, p$, will be inconsistent and the estimated residuals

$$\hat{e}_t^{(0)} = Z_t - \sum_{i=1}^{p} \hat{\phi}_i^{(0)} Z_{t-i}$$

will not be white noise. In fact, if $q \geq 1$, then the lagged values $\hat{e}_{t-i}^{(0)}$, $i = 1, \ldots, q$, will contain some information about the process Z_t, which leads to the following iterated regressions. First, we consider an AR(p) regression plus an added term $\hat{e}_{t-1}^{(0)}$, i.e.,

$$Z_t = \sum_{i=1}^{p} \phi_i^{(1)} Z_{t-i} + \beta_1^{(1)} \hat{e}_{t-1}^{(0)} + e_t^{(1)}, \qquad t = p + 2, \ldots, n \tag{6.4.6}$$

where the superscript (1) refers to the first iterated regression and $e_t^{(1)}$ represents the corresponding error term. The OLS estimates $\hat{\phi}_i^{(1)}$ will be consistent if $q = 1$. If $q > 1$, however, then the $\hat{\phi}_i^{(1)}$ will again be inconsistent, the estimated residuals $\hat{e}_t^{(1)}$ are not white noise, and the lagged values $\hat{e}_{t-i}^{(1)}$ will contain some information about Z_t. We thus consider the second iterated AR(p) regression

$$Z_t = \sum_{i=1}^{p} \phi_i^{(2)} Z_{t-i} + \beta_1^{(2)} \hat{e}_{t-1}^{(1)} + \beta_2^{(2)} \hat{e}_{t-2}^{(0)} + e_t^{(2)}, \qquad t = p + 3, \ldots, n. \tag{6.4.7}$$

The OLS estimates $\hat{\phi}_i^{(2)}$ will be consistent if $q = 2$. For $q > 2$, the $\hat{\phi}_i^{(2)}$ will again be inconsistent. Consistent estimates can be obtained by repeating the above iteration, however. That is, the OLS estimates $\hat{\phi}_i^{(q)}$ obtained from the following qth iterated AR(p) regression will be consistent:

$$Z_t = \sum_{i=1}^{p} \phi_i^{(q)} Z_{t-i} + \sum_{i=1}^{q} \beta_i^{(q)} \hat{e}_{t-i}^{(q-i)} + e_t^{(q)}, \qquad t = p + q + 1, \ldots, n \tag{6.4.8}$$

where $\hat{e}_t^{(j)} = Z_t - \sum_{i=1}^{p}\hat{\phi}_i^{(j)}Z_{t-i} - \sum_{i=1}^{q}\hat{\beta}_i^{(j)}\hat{e}_{t-i}^{(j-i)}$ is the estimated residual of the jth iterated AR(p) regression and the $\hat{\phi}_i^{(j)}$ and $\hat{\beta}_i^{(j)}$ are the corresponding least squares estimates.

In practice, the true order p and q of the ARMA(p, q) model are usually unknown and have to be estimated. Based on the preceding consideration, however, Tsay and Tiao (1984) suggest a general set of iterated regressions and introduce the concept of the extended sample autocorrelation function (ESACF) to estimate the orders p and q. Specifically, for $m = 0, 1, 2, \ldots$, let $\hat{\phi}_i^{(j)}$, $i = 1, \ldots, m$, be the OLS estimates from the jth iterated AR(m) regression of the ARMA process Z_t. They define the mth ESACF at lag j, $\hat{\rho}_j^{(m)}$, of Z_t as the sample autocorrelation function for the transformed series

$$Y_t^{(j)} = (1 - \hat{\phi}_1^{(j)}B - \cdots - \hat{\phi}_m^{(j)}B^m)Z_t. \tag{6.4.9}$$

It is useful to arrange $\hat{\rho}_j^{(m)}$ in a two-way table as shown in Table 6.16 where the first row corresponding to $\hat{\rho}_j^{(0)}$ gives the standard sample ACF of Z_t, the second row gives the first ESACF $\hat{\rho}_j^{(1)}$, and so on. The rows are numbered 0, 1, ... to specify the AR order, and the columns are numbered in a similar way for the MA order.

Note that the ESACF $\hat{\rho}_j^{(m)}$ is a function of n, the number of observations, even though it is not explicitly shown. In fact, it can be shown (see Tsay and Tiao [1984]) that for an ARMA(p, q) process, we have the following convergence in probability, i.e., for $m = 1, 2, \ldots$, and $j = 1, 2, \ldots$, we have

$$\hat{\rho}_j^{(m)} \xrightarrow{P} \begin{cases} 0, & 0 \le m - p < j - q, \\ X \ne 0, & \text{otherwise.} \end{cases} \tag{6.4.10}$$

Thus, by (6.4.10), the asymptotic ESACF table for an ARMA(1, 1) model becomes the one shown in Table 6.17. The zeros can be seen to form a triangle with the vertex at the (1, 1) position. More generally, for an ARMA(p, q) process, the vertex of the zero triangle in the asymptotic ESACF will be at the (p, q) position. Hence, the ESACF can be a useful tool in model identification, particularly for a mixed ARMA model.

Of course, in practice, we have finite samples, and the $\hat{\rho}_j^{(m)}$ for $0 \le m - p < j - q$ may not be exactly zero. However, the asymptotic variance of $\hat{\rho}_j^{(m)}$ can be approximated using Bartlett's formula or more crudely by $(n - m - j)^{-1}$ on the hypothesis that the transformed

TABLE 6.16 The ESACF table.

	MA					
AR	0	1	2	3	4	⋯
0	$\hat{\rho}_1^{(0)}$	$\hat{\rho}_2^{(0)}$	$\hat{\rho}_3^{(0)}$	$\hat{\rho}_4^{(0)}$	$\hat{\rho}_5^{(0)}$	⋯
1	$\hat{\rho}_1^{(1)}$	$\hat{\rho}_2^{(1)}$	$\hat{\rho}_3^{(1)}$	$\hat{\rho}_4^{(1)}$	$\hat{\rho}_5^{(1)}$	⋯
2	$\hat{\rho}_1^{(2)}$	$\hat{\rho}_2^{(2)}$	$\hat{\rho}_3^{(2)}$	$\hat{\rho}_4^{(2)}$	$\hat{\rho}_5^{(2)}$	⋯
3	$\hat{\rho}_1^{(3)}$	$\hat{\rho}_2^{(3)}$	$\hat{\rho}_3^{(3)}$	$\hat{\rho}_4^{(3)}$	$\hat{\rho}_5^{(3)}$	⋯
⋮			⋮			

TABLE 6.17 The asymptotic ESACF for an ARMA(1, 1) model.

	MA					
AR	0	1	2	3	4	⋯
0	X	X	X	X	X	⋯
1	X	0	0	0	0	⋯
2	X	X	0	0	0	⋯
3	X	X	X	0	0	⋯
4	X	X	X	X	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱

series $Y_t^{(j)}$ of (6.4.9) is white noise. The ESACF table can then be constructed using indicator symbols with X referring to values greater than or less than ±2 standard deviations and 0 for values within ±2 standard deviations.

EXAMPLE 6.9 To illustrate the method, we use SCA to compute the ESACF for the natural logarithms of Canadian lynx pelt sales discussed in Example 6.7. Table 16.18(a) shows the ESACF and Table 6.18(b) the corresponding indicator symbols for the ESACF of the series. The vertex of the triangle suggests a mixed ARMA(2, 1) model. This model is different from an AR(3) model that we entertained earlier based on the characteristic of the PACF. We further examine the series in Chapter 7.

TABLE 6.18 (a) The ESACF for natural logarithms of Canadian lynx pelt sales.

	MA									
AR	0	1	2	3	4	5	6	7	8	9
					(a) The ESACF					
0	.73	.22	−.32	−.69	−.76	−.53	−.08	.33	.61	.59
1	.68	.22	−.28	−.60	−.65	.51	−.09	.31	.53	.57
2	−.54	.01	.20	−.15	.17	−.26	.25	−.15	−.05	.10
3	−.53	.06	.22	.09	−.02	−.10	.15	−.16	−.20	.06
4	.00	.13	.37	−.12	−.02	−.09	−.01	−.18	−.16	−.03
					(b) The ESACF with indicator symbols					
0	X	0	X	X	X	X	0	X	X	X
1	X	0	0	X	X	X	0	X	X	X
2	X	0	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0	0	0
4	0	X	X	0	0	0	0	0	0	0

TABLE 6.19 The ESACF of Series C in Box, Jenkins, and Reinsel.

	MA								
AR	0	1	2	3	4	5	6	7	8
0	X	X	X	X	X	X	X	X	X
1	X	X	X	X	X	X	X	X	X
2	0	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0
5	X	X	X	0	0	0	0	0	0

EXAMPLE 6.10 As another example, Table 6.19 shows the ESACF for Series C of 226 temperature readings in Box, Jenkins, and Reinsel (1994). Because the vertex of the zero triangle occurs at the (2, 0) position, the ESACF approach suggests an AR(2) model for this series.

OLS estimation is used in the iterated regression. Because the OLS regression estimation can always be applied to a regression model regardless of whether a series is stationary or nonstationary, invertible or noninvertible, Tsay and Tiao (1984) allow the roots of AR and MA polynomials to be on or outside the unit circle in their definition of the ESACF. Hence, the ESACF is computed from the original nondifferenced data. As a result, the ARIMA(p, d, q) process is identified as the ARMA(P, q) process with $P = p + d$. For example, based on the sample ACF and PACF of Series C as given in Table 6.20, Box, Jenkins, and Reinsel (1994) suggested an ARIMA(1, 1, 0) model or an ARIMA(0, 2, 0) model. Both these ARIMA(1, 1, 0) and ARIMA(0, 2, 0) models, however, are identified as an ARMA(2, 0) model in terms of the ESACF.

Because the ESACF proposed by Tsay and Tiao (1984) is defined on the original series, they argue that the use of ESACF eliminates the need for differencing and provides a unified identification procedure to both stationary and nonstationary processes. To see whether a series is nonstationary, they suggest that for given specified values of p and q, the iterated AR estimates can be examined to see whether the AR polynomial contains a nonstationary factor with roots on the unit circle. For example, in the previous Series C in Box, Jenkins, and Reinsel (1994), the iterated AR estimates are $\hat{\phi}_1^{(0)} = 1.81$ and $\hat{\phi}_2^{(0)} = -.82$. Because $\hat{\phi}^{(0)}(B) \simeq (1 - B)(1 - .8B)$, it indicates that the AR polynomial contains the factor $(1 - B)$. Other than a few nice exceptions, however, the task of identifying nonstationarity through this approach is generally difficult. The real advantage of the ESACF is for the identification of p and q for the mixed ARMA models. This advantage can be used much better if the ESACF is used for a properly transformed stationary series, particularly because a tentatively identified model will be subjected to more efficient estimation procedures (such as the maximum likelihood estimation), which generally require stationarity.

Due to sampling variations and correlations among sample ACF, the pattern in the ESACF table from most time series may not be as clear-cut as those shown in the above examples. From the author's experience, though, models can usually be identified without much difficulty through a joint study of ACF, PACF, and ESACF.

TABLE 6.20 Sample ACF and PACF of Series C in Box, Jenkins, and Reinsel (1994).

(a) Z_t

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.98	.94	.90	.85	.80	.75	.69	.64	.58	.52
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	.98	−.81	−.03	−.02	−.10	−.07	−.01	−.03	.04	−.04
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

(b) $(1 - B)Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.80	.65	.53	.44	.38	.32	.26	.19	.14	.14
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	.80	−.01	−.01	.06	.03	−.03	−.01	−.08	.00	.10
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

(c) $(1 - B)^2Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.08	−.07	−.12	−.06	.01	−.02	.05	−.05	−.12	.12
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	−.08	−.08	−.14	−.10	−.03	−.05	.02	−.06	−.16	.09
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

Some computer programs such as AUTOBOX (1987) and SCA (1992) provide the option to compute the ESACF in the model identification phase.

6.4.2 Other Identification Procedures

Other model identification procedures include the information criterion (AIC) proposed by Akaike (1974b); the *R*-and-*S*-array introduced by Gray, Kelley, and McIntire (1978); and the corner method suggested by Beguin, Gourieroux, and Monfort (1980). The statistical properties of the statistics used in the *R*-and-*S*-array approach and the corner method are still largely unknown, and the software needed for these methods is not easily available. Interested readers are referred to their original research papers listed in the reference section of this book. The information criterion is discussed in Chapter 7.

At this point, it is appropriate to say that model identification is both a science and an art. One should not use one method to the exclusion of others. Through careful examination of the ACF, PACF, IACF, ESACF, and other properties of time series, model identification truly becomes the most interesting aspect of time series analysis.

EXERCISES

6.1 Identify appropriate ARIMA models from the sample ACF below. Justify your choice using the knowledge of the theoretical ACF for ARIMA models.

(a) $n = 121$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.15	−.08	.04	.08	.08	.03	.02	.05	.04	−.11

(b) $n = 250$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.63	.36	−.17	.09	−.07	.06	−.08	.10	−.11	.06

(c) $n = 250$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.35	−.17	.09	−.06	.01	−.01	−.04	.07	−.07	.09

(d) $n = 100$, data $= Z_t$, $W_t = (1 - B)Z_t$, $\overline{W} = 2.5$, $S_W^2 = 20$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_Z(k)$	.99	.98	.98	.97	.94	.91	.89	.86	.85	.83
$\hat{\rho}_W(k)$	.45	−.04	.12	.06	−.18	.16	−.07	.05	.10	.09

(e) $n = 100$, data $= Z_t$, $W_t = (1 - B)Z_t$, $\overline{W} = 35$, $S_W^2 = 1500$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_Z(k)$	.94	.93	.90	.89	.87	.86	.84	.81	.80	.80
$\hat{\rho}_W(k)$	.69	.50	.33	.19	.10	.08	.03	.01	.01	.00

6.2 Identify proper ARIMA models for the following data sets (read across):

(a)

−2.401	−.574	.382	−.535	−1.639	−.960	−1.118
−.719	−1.236	.117	−.493	−2.282	−1.823	.645
−.179	.589	1.413	.370	.082	−.531	−1.891
−.961	−.865	−.790	−1.476	−2.491	−4.479	−2.809
−2.154	−1.532	−2.119	−3.349	−1.588	.740	.907
1.540	.557	2.259	2.622	.701	2.463	2.714
2.089	3.750	4.322	3.186	3.192	2.939	3.263
3.279	.295	.227	1.356	1.912	1.060	.370
−.195	.340	1.084	1.237	.610	2.126	3.960
3.317	2.167	1.292	.595	.140	−.082	−.769
.870	1.551	2.610	2.193	1.353	−.600	−.455
.203	1.472	1.367	1.875	2.082	1.604	2.033
3.746	2.954	.676	1.163	1.368	.343	−.334
1.041	1.328	1.325	.968	1.970	2.296	2.896
1.918	1.569					

(b)

−1.453	.867	.727	−.765	−1.317	.024	−.542
−.048	−.805	.858	−.563	−1.986	−.454	1.738
−.566	.697	1.060	−.478	−.140	−.581	−1.572
.174	−.289	−.270	−1.002	−1.605	−2.984	−.122
.469	−.239	−1.200	−2.077	.421	1.693	.463

	.996	−.367	1.925	1.267	−.872	2.043	1.236
	.461	2.497	2.072	.593	1.281	1.023	1.500
	1.321	−1.673	.050	1.219	1.098	−.087	−.266
	−.417	.457	.880	.586	−.132	1.760	2.684
	.941	.177	−.008	−.180	−.217	−.165	−.720
	1.332	1.029	1.679	.627	.038	−1.412	−.095
	.476	1.350	.484	1.055	.957	.355	1.071
	2.526	.707	−1.096	.757	.670	−.477	−.540
	1.241	.704	.528	.173	1.389	1.115	1.519
	.180	.419					
(c)	3.485	5.741	5.505	3.991	3.453	4.773	4.142
	4.598	3.796	5.430	3.960	2.541	4.054	6.155
	3.778	5.066	5.422	3.908	4.302	3.876	2.888
	4.613	4.075	4.054	3.288	2.654	1.215	3.979
	3.452	3.569	2.523	1.584	3.998	5.135	3.842
	4.404	3.077	5.432	4.795	2.747	5.767	4.988
	4.311	6.456	6.114	4.785	5.646	5.516	6.121
	6.059	3.196	5.050	6.231	6.119	4.988	4.885
	4.777	5.666	6.081	5.801	5.126	7.067	8.015
	6.358	5.752	5.700	5.614	5.629	5.705	5.155
	7.204	6.871	7.555	6.565	6.081	4.719	6.090
	6.637	7.492	6.635	7.264	7.221	6.694	7.493
	9.012	7.274	5.622	7.593	7.533	6.432	6.424
	8.219	7.668	7.534	7.232	8.501	8.266	8.748
	7.501	7.856					
(d)	.315	−.458	−.488	−.170	.565	−.344	−1.176
	−1.054	−.826	.710	−.341	−1.809	−1.242	−.667
	−.999	2.812	1.286	−1.084	−1.505	−2.556	−.144
	−1.749	−3.032	−2.958	−2.827	−3.392	−2.431	−2.757
	−2.822	−3.314	−2.738	−1.979	−1.671	−2.977	−.709
	.718	.736	.879	1.642	2.180	1.963	.716
	.769	.973	.334	1.309	.878	.062	.169
	.677	1.851	.242	.828	−.317	−1.042	−2.093
	.653	.261	2.020	2.136	1.635	−.141	−1.747
	−2.047	−.752	−.211	−1.062	−1.565	.232	.015
	−.935	−.338	.853	.888	3.069	3.364	3.854
	4.419	2.145	2.291	1.753	1.058	1.048	.200
	1.424	.590	.356	.476	.684	−2.260	−.569
	−1.014	−.207	.638	−.664	−.469	−.215	−.296
	−1.561	.246					

6.3 Identify models for Series W1 through W7 using ESACF and compare them with the models suggested in the book using ACF and PACF.

7

Parameter Estimation, Diagnostic Checking, and Model Selection

After identifying a tentative model, the next step is to estimate the parameters in the model. With full generality, we consider the general ARMA(p, q) model. That is, we discuss the estimation of parameters $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)'$, $\mu = E(Z_t)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)$, and $\sigma_a^2 = E(a_t^2)$ in the model

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

where $\dot{Z}_t = Z_t - \mu$, Z_t ($t = 1, 2, \ldots, n$) are n observed stationary or properly transformed stationary time series and $\{a_t\}$ are i.i.d. $N(0, \sigma_a^2)$ white noise. Several widely used estimation procedures are discussed.

Once parameters have been estimated, we check on the adequacy of the model for the series. Very often several models can adequately represent a given series. Thus, after introducing diagnostic checking, we also present some criteria that are commonly used for model selection in time series model building.

7.1 The Method of Moments

The method of moments consists of substituting sample moments such as the sample mean, $\bar{Z}$, sample variance $\hat{\gamma}_0$, and sample ACF $\hat{\rho}_i$ for their theoretical counterparts and solving the resultant equations to obtain estimates of unknown parameters. For example, in the AR(p) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t, \tag{7.1.1}$$

the mean $\mu = E(Z_t)$ is estimated by $\bar{Z}$. To estimate $\boldsymbol{\phi}$, we first use that $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$ for $k \geq 1$ to obtain the following system of Yule–Walker equations:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p.
\end{aligned} \tag{7.1.2}$$

Then, replacing ρ_k by $\hat{\rho}_k$, we obtain the moment estimators $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ by solving the above linear system of equations. That is,

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-2} & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-3} & \hat{\rho}_{p-2} \\ \vdots & & & & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}. \tag{7.1.3}$$

These estimators are usually called Yule–Walker estimators.

Having obtained $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$, we use the result

$$\begin{aligned}
\gamma_0 &= E(\dot{Z}_t\dot{Z}_t) = E\left[\dot{Z}_t(\phi_1\dot{Z}_{t-1} + \phi_2\dot{Z}_{t-2} + \cdots + \phi_p\dot{Z}_{t-p} + a_t)\right] \\
&= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_a^2
\end{aligned} \tag{7.1.4}$$

and obtain the moment estimator for σ_a^2 as

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1 - \hat{\phi}_2\hat{\rho}_2 - \cdots - \hat{\phi}_p\hat{\rho}_p). \tag{7.1.5}$$

EXAMPLE 7.1 For the AR(1) model

$$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + a_t, \tag{7.1.6}$$

the Yule–Walker estimator for ϕ_1, from (7.1.3), is

$$\hat{\phi}_1 = \hat{\rho}_1. \tag{7.1.7}$$

The moment estimators for μ and σ_a^2 are given by

$$\hat{\mu} = \bar{Z} \tag{7.1.8}$$

and

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1), \tag{7.1.9}$$

respectively, where $\hat{\gamma}_0$ is the sample variance of the Z_t series.

Next, let us consider a simple MA(1) model

$$\dot{Z}_t = a_t - \theta_1 a_{t-1}. \tag{7.1.10}$$

Again, μ is estimated by $\bar{Z}$. For θ_1, we use that

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$$

and solve the above quadratic equation for θ_1 after replacing ρ_1 by $\hat{\rho}_1$. The result leads to

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}. \tag{7.1.11}$$

If $\hat{\rho}_1 = \pm.5$, then we have a unique solution $\hat{\theta}_1 = \pm 1$, which gives a noninvertible estimated model. If $|\hat{\rho}_1| > .5$, then the real-valued moment estimator $\hat{\theta}_1$ does not exist. This result is expected, because a real-valued MA(1) model always has $|\rho_1| < .5$, as discussed in Section 3.2.1. For $|\hat{\rho}_1| < .5$, there exist two distinct real-valued solutions and we always choose the one that satisfies the invertibility condition. After having obtained $\hat{\theta}_1$, we calculate the moment estimator for σ_a^2 as

$$\hat{\sigma}_a^2 = \frac{\hat{\gamma}_0}{1 + \hat{\theta}_1^2}. \tag{7.1.12}$$

This example of the MA(1) model shows that the moment estimators for MA and mixed ARMA models are complicated. More generally, regardless of AR, MA, or ARMA models, the moment estimates are very sensitive to rounding errors. They are usually used to provide initial estimates needed for a more efficient nonlinear estimation discussed later in this chapter, particularly for the MA and mixed ARMA models. The moment estimators are not recommended for final estimation results and should not be used if the process is close to being nonstationary or noninvertible.

7.2 Maximum Likelihood Method

Because of its many nice properties, the maximum likelihood method has been widely used in estimation. In this section, we discuss several alternative maximum likelihood estimates that are used in time series analysis.

7.2.1 Conditional Maximum Likelihood Estimation

For the general stationary ARMA(p, q) model

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{7.2.1}$$

where $\dot{Z}_t = Z_t - \mu$ and $\{a_t\}$ are i.i.d. $N(0, \sigma_a^2)$ white noise, the joint probability density of $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$ is given by

$$P(\mathbf{a} \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right]. \tag{7.2.2}$$

Rewriting (7.2.1) as

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + \dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \cdots - \phi_p \dot{Z}_{t-p}, \tag{7.2.3}$$

we can write down the likelihood function of the parameters $(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2)$.

Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ and assume the initial conditions $\mathbf{Z}_* = (Z_{1-p}, \ldots, Z_{-1}, Z_0)'$ and $\mathbf{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$ are known. The conditional log-likelihood function is

$$\ln L_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})}{2\sigma_a^2}, \tag{7.2.4}$$

where

$$S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}) = \sum_{t=1}^{n} a_t^2(\boldsymbol{\phi}, \mu, \boldsymbol{\theta} \mid \mathbf{Z}_*, \mathbf{a}_*, \mathbf{Z}), \tag{7.2.5}$$

is the conditional sum of squares function. The quantities of $\hat{\boldsymbol{\phi}}$, $\hat{\mu}$, and $\hat{\boldsymbol{\theta}}$, which maximize Function (7.2.4), are called the conditional maximum likelihood estimators. Because $\ln L_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2)$ involves the data only through $S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$, these estimators are the same as the conditional least squares estimators obtained from minimizing the conditional sum of squares function $S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$, which, we note, does not contain the parameter σ_a^2.

There are a few alternatives for specifying the initial conditions $\mathbf{Z}_*$ and $\mathbf{a}_*$. Based on the assumptions that $\{Z_t\}$ is stationary and $\{a_t\}$ is a series of i.i.d. $N(0, \sigma_a^2)$, random variables, we can replace the unknown Z_t by the sample mean $\bar{Z}$ and the unknown a_t by its expected value of 0. For the model in (7.2.1), we may also assume $a_p = a_{p-1} = \cdots = a_{p+1-q} = 0$ and calculate a_t for $t \geq (p+1)$ using (7.2.3). The conditional sum of squares function in (7.2.5) thus becomes

$$S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}) = \sum_{t=p+1}^{n} a_t^2(\boldsymbol{\phi}, \mu, \boldsymbol{\theta} \mid \mathbf{Z}), \tag{7.2.6}$$

which is also the form used by most computer programs.

After obtaining the parameter estimates $\hat{\boldsymbol{\phi}}$, $\hat{\mu}$, and $\hat{\boldsymbol{\theta}}$, the estimate $\hat{\sigma}_a^2$ of σ_a^2 is calculated from

$$\hat{\sigma}_a^2 = \frac{S_*(\hat{\boldsymbol{\phi}}, \hat{\mu}, \hat{\boldsymbol{\theta}})}{\text{d.f.}}, \tag{7.2.7}$$

where the number of degrees of freedom d.f. equals the number of terms used in the sum of $S_*(\hat{\boldsymbol{\phi}}, \hat{\mu}, \hat{\boldsymbol{\theta}})$ minus the number of parameters estimated. If (7.2.6) is used to calculate the

sum of squares, d.f. $= (n - p) - (p + q + 1) = n - (2p + q + 1)$. For other models, the d.f. should be adjusted accordingly.

7.2.2 Unconditional Maximum Likelihood Estimation and Backcasting Method

As seen from Chapter 5, one of the most important functions of a time series model is to forecast the unknown future values. Naturally, one asks whether we can back-forecast or backcast the unknown values $\mathbf{Z}_* = (Z_{1-p}, \ldots, Z_{-1}, Z_0)'$ and $\mathbf{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$ needed in the computation of the sum of squares and likelihood functions. Indeed, that is possible because any ARMA model can be written in either the forward form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\dot{Z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \tag{7.2.8}$$

or the backward form

$$(1 - \phi_1 F - \cdots - \phi_p F^p)\dot{Z}_t = (1 - \theta_1 F - \cdots - \theta_q F^q)e_t, \tag{7.2.9}$$

where $F^j Z_t = Z_{t+j}$. Because of the stationarity, (7.2.8) and (7.2.9) should have exactly the same autocovariance structure, which implies that $\{e_t\}$ is also a white noise series with mean zero and variance σ_e^2. Thus, in the same way as we use the forward form (7.2.8) to forecast the unknown future values Z_{n+j} for $j > 0$ based on the data $(Z_1, Z_2, \ldots, Z_n)$, we can also use the backward form (7.2.9) to backcast the unknown past values Z_j and hence compute a_j for $j \leq 0$ based on the data $\{Z_n, Z_{n-1}, \ldots, Z_1\}$. Therefore, for a further improvement in estimation, Box, Jenkins, and Reinsel (1994) suggest the following unconditional log-likelihood function:

$$\ln L(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})}{2\sigma_a^2}, \tag{7.2.10}$$

where $S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$ is the unconditional sum of squares function given by

$$S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}) = \sum_{t=-\infty}^{n} \left[E(a_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \mathbf{Z})\right]^2 \tag{7.2.11}$$

and $E(a_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \mathbf{Z})$ is the conditional expectation of a_t given $\boldsymbol{\phi}$, μ, $\boldsymbol{\theta}$, and $\mathbf{Z}$. Some of these terms have to be calculated using backcasts illustrated in Example 7.2.

The quantities $\hat{\boldsymbol{\phi}}$, $\hat{\mu}$, and $\hat{\boldsymbol{\theta}}$ that maximize Function (7.2.10) are called unconditional maximum likelihood estimators. Again because $\ln L(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2)$ involves the data only through $S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$, these unconditional maximum likelihood estimators are equivalent to the unconditional least squares estimators obtained by minimizing $S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$. In practice, the summation in (7.2.11) is approximated by a finite form

$$S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}) = \sum_{t=-M}^{n} \left[E(a_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \mathbf{Z})\right]^2, \tag{7.2.12}$$

where M is a sufficiently large integer such that the backcast increment $\left|E(Z_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, Z) - E(Z_{t-1} \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, Z)\right|$ is less than any arbitrary predetermined small ε value for $t \leq -(M + 1)$.

This expression implies that $E(Z_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \mathbf{Z}) \simeq \mu$; hence, $E(a_t \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \mathbf{Z})$ is negligible for $t \leq -(M+1)$.

After obtaining the parameter estimates $\hat{\boldsymbol{\phi}}$, $\hat{\mu}$, and $\hat{\boldsymbol{\theta}}$, the estimate $\hat{\sigma}_a^2$ of σ_a^2 can then be calculated as

$$\hat{\sigma}_a^2 = \frac{S(\hat{\boldsymbol{\phi}}, \hat{\mu}, \hat{\boldsymbol{\theta}})}{n}. \tag{7.2.13}$$

For efficiency, the use of backcasts for parameter estimation is important for seasonal models (discussed in Chapter 8), for models that are close to being nonstationary, and especially for series that are relatively short. Most computer programs have implemented this option.

EXAMPLE 7.2 To illustrate the backcasting method, consider the AR(1) model that can be written in the forward form

$$a_t = Z_t - \phi Z_{t-1} \tag{7.2.14}$$

or, equivalently, in the backward form

$$e_t = Z_t - \phi Z_{t+1}, \tag{7.2.15}$$

where, without loss of generality, we assume that $E(Z_t) = 0$. Consider a very simple example with ten observations, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{10})$ from the process, which are listed in Table 7.1

TABLE 7.1 Calculation of $S(\phi = .3)$ for $(1 - \phi B)Z_t = a_t$ using the backcasting method.

t	$E(a_t\|\mathbf{Z})$	$-.3E(Z_{t-1}\|\mathbf{Z})$	$E(Z_t\|\mathbf{Z})$	$.3E(Z_{t+1}\|\mathbf{Z})$	$E(e_t\|\mathbf{Z})$
−3			−.0016	−.0016	0
−2	−.0049	.0005	−.0054	−.0054	0
−1	−.0164	.0016	−.018	−.018	0
0	−.0546	.0054	−.06	−.06	0
1	−.182	.018	−.2		
2	−.34	.06	−.4		
3	−.38	.12	−.5		
4	−.35	.15	−.5		
5	−.45	.15	−.6		
6	−.32	.18	−.5		
7	−.25	.15	−.4		
8	−.08	.12	−.2		
9	−.04	.06	−.1		
10	−.17	.03	−.2		

under the column $E(Z_t \mid \mathbf{Z})$ for $t = 1, 2, \ldots, 10$. Suppose that $\phi = .3$ and we want to calculate the unconditional sum of squares

$$S(\phi = .3) = \sum_{t=-M}^{10} \left[E(a_t \mid \phi = .3, \mathbf{Z})\right]^2, \tag{7.2.16}$$

where M is chosen so that $\left|E(Z_t \mid \phi = .3, \mathbf{Z}) - E(Z_{t-1} \mid \phi = .3, \mathbf{Z})\right| < .005$ for $t \leq -(M+1)$. To simplify the notations for this example we write $E(a_t \mid \phi = .3, \mathbf{Z})$ as $E(a_t \mid \mathbf{Z})$ and $E(Z_t \mid \phi = .3, \mathbf{Z})$ as $E(Z_t \mid \mathbf{Z})$.

To obtain $E(a_t \mid \mathbf{Z})$ we use (7.2.14) and compute

$$E(a_t \mid \mathbf{Z}) = E(Z_t \mid \mathbf{Z}) - \phi E(Z_{t-1} \mid \mathbf{Z}). \tag{7.2.17}$$

This computation of $E(a_t \mid \mathbf{Z})$ for $t \leq 1$, however, involves the unknown Z_t values for $t \leq 0$, which need to be backcasted. To do so, we use the backward form in (7.2.15), i.e.,

$$E(Z_t \mid \mathbf{Z}) = E(e_t \mid \mathbf{Z}) + \phi E(Z_{t+1} \mid \mathbf{Z}). \tag{7.2.18}$$

First, we note that in terms of the backward form, e_t for $t \leq 0$ are unknown future random shocks with respect to the observations $Z_n, Z_{n-1}, \ldots, Z_2$, and Z_1. Hence,

$$E(e_t \mid \mathbf{Z}) = 0, \qquad \text{for } t \leq 0. \tag{7.2.19}$$

Therefore, for $\phi = .3$, we have from (7.2.18)

$$\begin{aligned}
E(Z_0 \mid \mathbf{Z}) &= E(e_0 \mid \mathbf{Z}) + .3E(Z_1 \mid \mathbf{Z}) \\
&= 0 + (.3)(-.2) = -.06 \\
E(Z_{-1} \mid \mathbf{Z}) &= E(e_{-1} \mid \mathbf{Z}) + .3E(Z_0 \mid \mathbf{Z}) \\
&= 0 + (.3)(-.06) = -.018 \\
E(Z_{-2} \mid \mathbf{Z}) &= E(e_{-2} \mid \mathbf{Z}) + .3E(Z_{-1} \mid \mathbf{Z}) \\
&= (.3)(-.018) = -.0054 \\
E(Z_{-3} \mid \mathbf{Z}) &= E(e_{-3} \mid \mathbf{Z}) + .3E(Z_{-2} \mid \mathbf{Z}) \\
&= (.3)(-.0054) = -.00162.
\end{aligned}$$

Because $\left|E(Z_{-3} \mid \mathbf{Z}) - E(Z_{-2} \mid \mathbf{Z})\right| = .00378 < .005$, the predetermined ε value, we choose $M = 2$.

Now, with these backcasted values Z_t for $t \leq 0$, we can return to the forward form in (7.2.17) to compute $E(a_t \mid \mathbf{Z})$ for $\phi = .3$ from $t = -2$ to $t = 10$ as follows:

$$\begin{aligned}
E(a_{-2} \mid \mathbf{Z}) &= E(Z_{-2} \mid \mathbf{Z}) - .3E(Z_{-3} \mid \mathbf{Z}) \\
&= -.0054 - (.3)(-.00162) = -.0049
\end{aligned}$$

$$
\begin{aligned}
E(a_{-1} \mid \mathbf{Z}) &= E(Z_{-1} \mid \mathbf{Z}) - .3E(Z_{-2} \mid \mathbf{Z}) \\
&= -.018 - (.3)(-.0054) = -.0164 \\
E(a_0 \mid \mathbf{Z}) &= E(Z_0 \mid \mathbf{Z}) - .3E(Z_{-1} \mid \mathbf{Z}) \\
&= -.06 - (.3)(-.018) = -.0546 \\
E(a_1 \mid \mathbf{Z}) &= E(Z_1 \mid \mathbf{Z}) - .3E(Z_0 \mid \mathbf{Z}) \\
&= -.2 - (.3)(-.06) = -.182 \\
E(a_2 \mid \mathbf{Z}) &= E(Z_2 \mid \mathbf{Z}) - .3E(Z_1 \mid \mathbf{Z}) \\
&= -.4 - (.3)(-.2) = -.34 \\
&\vdots \\
E(a_{10} \mid \mathbf{Z}) &= E(Z_{10} \mid \mathbf{Z}) - .3E(Z_9 \mid \mathbf{Z}) \\
&= -.2 - (.3)(-.1) = -.17.
\end{aligned}
$$

All the above computations can be carried out systematically as shown in Table 7.1 and we obtain

$$S(\phi = .3) = \sum_{t=-2}^{10} \left[E(a_t \mid \phi = .3, \mathbf{Z})\right]^2 = .8232.$$

Similarly, we can obtain $S(\phi)$ for other values of ϕ and hence find its minimum.

For the AR(1) model we do not need the value $E(e_t \mid \mathbf{Z})$ for $t \geq 1$. For other models, they may be required, but the procedure is the same. For more detailed examples, see Box, Jenkins, and Reinsel (1994, p. 233).

7.2.3 Exact Likelihood Functions

Both the conditional and unconditional likelihood functions (7.2.4) and (7.2.10) are approximations. To illustrate the derivation of the exact likelihood function for a time series model, consider the AR(1) process

$$(1 - \phi B)\dot{Z}_t = a_t \tag{7.2.20}$$

or

$$\dot{Z}_t = \phi \dot{Z}_{t-1} + a_t,$$

where $\dot{Z}_t = (Z_t - \mu)$, $|\phi| < 1$ and the a_t are i.i.d. $N(0, \sigma_a^2)$. Rewriting the process in the moving average representation, we have

$$\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j}. \tag{7.2.21}$$

Clearly, the $\dot{Z}_t$ will be distributed as $N\left(0, \sigma_a^2/(1-\phi^2)\right)$. The $\dot{Z}_t$, however, are highly correlated. To derive the joint probability density function $P(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ of $(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ and hence the likelihood function for the parameters, we consider

$$\begin{aligned} e_1 &= \sum_{j=0}^{\infty} \phi^j a_{1-j} = \dot{Z}_1 \\ a_2 &= \dot{Z}_2 - \phi \dot{Z}_1 \\ a_3 &= \dot{Z}_3 - \phi \dot{Z}_2 \\ &\vdots \\ a_n &= \dot{Z}_n - \phi \dot{Z}_{n-1}. \end{aligned} \tag{7.2.22}$$

Note that e_1 follows the normal distribution $N\left(0, \sigma_a^2/(1-\phi^2)\right)$, a_t, for $2 \leq t \leq n$, follows the normal distribution $N(0, \sigma_a^2)$, and they are all independent of one another. Hence, the joint probability density of $(e_1, a_2, \ldots, a_n)$ is

$$\begin{aligned} &p(e_1, a_2, \ldots, a_n) \\ &\quad = \left[\frac{(1-\phi^2)}{2\pi\sigma_a^2}\right]^{1/2} \exp\left[\frac{-e_1^2(1-\phi^2)}{2\sigma_a^2}\right]\left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{n} a_t^2\right]. \end{aligned} \tag{7.2.23}$$

Now consider the following transformation:

$$\begin{aligned} \dot{Z}_1 &= e_1 \\ \dot{Z}_2 &= \phi \dot{Z}_1 + a_2 \\ \dot{Z}_3 &= \phi \dot{Z}_2 + a_3 \\ &\vdots \\ \dot{Z}_n &= \phi \dot{Z}_{n-1} + a_n. \end{aligned} \tag{7.2.24}$$

The Jacobian for the transformation, from (7.2.22), is

$$J = \begin{vmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -\phi & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & -\phi & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \vdots & & & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & -\phi & 1 \end{vmatrix} = 1,$$

It follows that

$$P(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n) = P(e_1, a_2, \ldots, a_n)$$
$$= \left[\frac{(1-\phi^2}{2\pi\sigma_a^2)}\right]^{1/2} \exp\left[\frac{-\dot{Z}_1{}^2(1-\phi^2)}{2\sigma_a^2}\right]$$
$$\cdot \left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{n}(\dot{Z}_t - \phi\dot{Z}_{t-1})^2\right]. \quad (7.2.25)$$

Hence, for a given series $(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ we have the following exact log-likelihood function:

$$\ln L(\dot{Z}_1, \ldots, \dot{Z}_n \mid \phi, \mu, \sigma_a^2) = -\frac{n}{2}\ln 2\pi + \frac{1}{2}\ln(1-\phi^2) - \frac{n}{2}\ln\sigma_a^2 - \frac{S(\phi,\mu)}{2\sigma_a^2}, \quad (7.2.26)$$

where

$$S(\phi, \mu) = (Z_1 - \mu)^2(1-\phi^2) + \sum_{t=2}^{n}\left[(Z_t - \mu) - \phi(Z_{t-1} - \mu)\right]^2 \quad (7.2.27)$$

is the sum of squares term that is a function of only ϕ and μ.

The exact closed form of the likelihood function of a general ARMA model is complicated. Tiao and Ali (1971) derived it for an ARMA(1, 1) model. Newbold (1974) derived it for a general ARMA(p, q) model. Interested readers are advised also to see Ali (1977), Ansley (1979), Hillmer and Tiao (1979), Ljung and Box (1979), and Nicholls and Hall (1979), among others for additional references.

7.3 Nonlinear Estimation

It is clear that the maximum likelihood estimation and the least squares estimation involve minimizing either the conditional sum of squares $S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$ or the unconditional sum of squares $S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$, which are the sums of squares of the error terms a_t's. For an AR(p) process,

$$a_t = \dot{Z}_t - \phi_1\dot{Z}_{t-1} - \phi_2\dot{Z}_{t-2} - \cdots - \phi_p\dot{Z}_{t-p} \quad (7.3.1)$$

and a_t is clearly linear in parameters. For a model containing an MA factor, however, the a_t is nonlinear in parameters. To see that, consider a simple ARMA(1, 1) model

$$\dot{Z}_t - \phi_1\dot{Z}_{t-1} = a_t - \theta_1 a_{t-1}. \quad (7.3.2)$$

To calculate a_t, we note that

$$\begin{aligned} a_t &= \dot{Z}_t - \phi_1 \dot{Z}_{t-1} + \theta_1 a_{t-1} \\ &= \dot{Z}_t - \phi_1 \dot{Z}_{t-1} + \theta_1(\dot{Z}_{t-1} - \phi_1 \dot{Z}_{t-2} + \theta_1 a_{t-2}) \\ &= \dot{Z}_t - (\phi_1 - \theta_1)\dot{Z}_{t-1} - \phi_1\theta_1 \dot{Z}_{t-2} + \theta_1^2 a_{t-2} \\ &\vdots \end{aligned} \tag{7.3.3}$$

This equation is clearly nonlinear in the parameters. Hence, for a general ARMA model, a nonlinear least squares estimation procedure must be used to obtain estimates.

The nonlinear least squares procedure involves an iterative search technique. Because a linear model is a special case of the nonlinear model, we can illustrate the main ideas of the nonlinear least squares using the following linear regression model:

$$\begin{aligned} Y_t &= E(Y_t \mid X_{ti}'s) + e_t \\ &= \alpha_1 X_{t1} + \alpha_2 X_{t2} + \cdots + \alpha_p X_{tp} + e_t \end{aligned} \tag{7.3.4}$$

for $t = 1, 2, \ldots, n$, where the e_t's are i.i.d. $N(0, \sigma_e^2)$ independent of all the X_{ti}. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ and $\underline{\overline{\mathbf{X}}}$ be the corresponding matrix for the independent variables X_{ti}'s. From results in linear regression analysis, we know that the least squares estimators are given by

$$\hat{\boldsymbol{\alpha}} = (\underline{\overline{\mathbf{X}}}'\underline{\overline{\mathbf{X}}})^{-1}\underline{\overline{\mathbf{X}}}'\mathbf{Y}, \tag{7.3.5}$$

which follows a multivariate normal distribution $\mathrm{MN}(\boldsymbol{\alpha}, V(\hat{\boldsymbol{\alpha}}))$ with

$$V(\hat{\boldsymbol{\alpha}}) = \sigma_a^2(\underline{\overline{\mathbf{X}}}'\underline{\overline{\mathbf{X}}})^{-1}. \tag{7.3.6}$$

The minimum residual (error) sum of squares is

$$S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n}(Y_t - \hat{\alpha}_1 X_{t1} - \hat{\alpha}_2 X_{t2} - \cdots - \hat{\alpha}_p X_{tp})^2. \tag{7.3.7}$$

The least squares estimates in (7.3.5) can also be obtained by the following two-step procedure discussed in Miller and Wichern (1977).

Let $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_p)'$ be an initial guess value of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$. We can rewrite the minimum residual sum of squares in (7.3.7) as

$$\begin{aligned} S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n}\big[& Y_t - \tilde{\alpha}_1 X_{t1} - \cdots - \tilde{\alpha}_p X_{tp} \\ & - (\hat{\alpha}_1 - \tilde{\alpha}_1)X_{t1} - \cdots - (\hat{\alpha}_p - \tilde{\alpha}_p)X_{tp}\big]^2 \end{aligned} \tag{7.3.8}$$

or

$$S(\boldsymbol{\delta}) = S(\hat{\boldsymbol{\alpha}}) = \sum_{t=1}^{n}(\tilde{e}_t - \delta_1 X_{t1} - \cdots - \delta_p X_{tp})^2, \tag{7.3.9}$$

where the $\tilde{e}_t$'s are estimated residuals based on the initial given values $\tilde{\alpha}$ and $\delta = (\hat{\alpha} - \tilde{\alpha})$. Now, $S(\delta)$ in Equation (7.3.9) and $S(\hat{\alpha})$ in (7.3.7) are in the same form. Hence, the least squares value of δ is

$$\delta = (\underline{\tilde{X}}'\underline{\tilde{X}})^{-1}\underline{\tilde{X}}'\tilde{e}, \tag{7.3.10}$$

where $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_n)'$. Once the values $\delta = (\delta_1, \delta_2, \ldots, \delta_p)$ are calculated, the least squares estimates are given by

$$\hat{\alpha} = \tilde{\alpha} + \delta. \tag{7.3.11}$$

Note that the residual $\tilde{e}_t$ is calculated as $Y_t - \tilde{Y}_t$, where

$$\tilde{Y}_t = \tilde{\alpha}_1 X_{t1} + \cdots + \tilde{\alpha}_p X_{tp}$$

represents a guess of the regression equation obtained by using the original model and the given initial values $\tilde{\alpha}_i$'s. Moreover, it is clear from Equation (7.3.4) that

$$\frac{\partial E(Y_t \mid X_{ti}\text{'s})}{\partial \alpha_i} = X_{ti} \tag{7.3.12}$$

for $i = 1, 2, \ldots, p$ and $t = 1, 2, \ldots, n$. Hence, the $\underline{\tilde{X}}$ matrix used in the equation of the least squares estimates in (7.3.5) and (7.3.10) is actually the matrix of the partial derivatives of the regression function with respect to each of the parameters.

Now, consider the following model (linear or nonlinear):

$$Y_t = f(\mathbf{X}_t, \alpha) + e_t, \qquad t = 1, 2, \ldots, n, \tag{7.3.13}$$

where $\mathbf{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tp})$ is a set of independent variables corresponding to the observations, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ is a vector of parameters, and e_t is a white noise series having zero mean and constant variance σ_e^2 independent of $\mathbf{X}_t$. Let $Y = (Y_1, Y_2, \ldots, Y_n)'$ and $\mathbf{f}(\alpha) = [f(\mathbf{X}_1, \alpha), f(\mathbf{X}_2, \alpha), \ldots, f(\mathbf{X}_n, \alpha)]'$. From the previous discussion, the least squares estimators (linear or nonlinear) can always be calculated iteratively as follows.

Step 1. Given any vector of initial guess values $\tilde{\alpha}$, compute the residual $\tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})$ and the residual sum of squares

$$S(\tilde{\alpha}) = \tilde{\mathbf{e}}'\tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})'(\mathbf{Y} - \tilde{\mathbf{Y}}), \tag{7.3.14}$$

where $\tilde{\mathbf{Y}} = \mathbf{f}(\tilde{\alpha})$ is a vector of predicted values obtained by replacing the unknown parameters by the initial guess values. Approximate the model $f(\mathbf{X}_t, \alpha)$ with the first-order Taylor series expansion about the initial value $\tilde{\alpha}$. That is,

$$\mathbf{f}(\alpha) = \mathbf{f}(\tilde{\alpha}) + \underline{\tilde{\mathbf{X}}}_{\tilde{\alpha}}\delta \tag{7.3.15}$$

where $\boldsymbol{\delta} = (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}})$ and $\underline{\overline{\mathbf{X}}}_{\tilde{\alpha}} = [X_{ij}]$ is the $n \times p$ matrix of the partial derivatives at $\tilde{\boldsymbol{\alpha}}$ in the above linear approximation. That is,

$$X_{ij} = \left.\frac{\partial f(X_t, \boldsymbol{\alpha})}{\partial \alpha_j}\right|_{\alpha=\tilde{\alpha}}, \qquad \begin{matrix} t = 1, 2, \ldots, n, \\ j = 1, 2, \ldots, p. \end{matrix} \tag{7.3.16}$$

Then we calculate

$$\boldsymbol{\delta} = (\underline{\overline{\mathbf{X}}}'_{\tilde{\alpha}} \underline{\overline{\mathbf{X}}}_{\tilde{\alpha}})^{-1} \underline{\overline{\mathbf{X}}}'_{\tilde{\alpha}} \tilde{\mathbf{e}} = (\delta_1, \delta_2, \ldots, \delta_p)'. \tag{7.3.17}$$

Note that for a linear model the $\underline{\overline{\mathbf{X}}}_{\tilde{\alpha}}$ is fixed and equals $\underline{\overline{\mathbf{X}}}$; for a nonlinear model, this $\underline{\overline{\mathbf{X}}}_{\tilde{\alpha}}$ changes from iteration to iteration.

Step 2. Obtain the updated least squares estimates

$$\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\delta} \tag{7.3.18}$$

and the corresponding residual sum of squares $S(\hat{\boldsymbol{\alpha}})$. We note that δ_i in $\boldsymbol{\delta}$ represents the difference or change in the parameter values. For a linear model, step 2 gives the final least squares estimates. For a nonlinear model, step 2 only leads to new initial values for further iterations.

In summary, for a given general ARMA(p, q) model, we can use the nonlinear least squares procedure to find the least squares estimates that minimize the error sum of squares $S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$ or $S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$. The nonlinear least squares routine starts with initial guess values of the parameters. It monitors these values in the direction of the smaller sum of squares and updates the initial guess values. The iterations continue until some specified convergence criteria are reached. Some convergence criteria that have been used are the relative reduction in the sum of squares, the maximum change in the parameter values less than a specified level, or the number of iterations greater than a certain number. To achieve a proper and faster convergence, many search algorithms are developed. One of the algorithms commonly used is due to Marquardt (1963). It is a compromise between the Gauss–Newton method and the method of steepest descent. For more discussions on nonlinear estimation, see Draper and Smith (1981), among others.

Properties of the Parameter Estimates Let $\boldsymbol{\alpha} = (\boldsymbol{\phi}, \mu, \boldsymbol{\theta})$, $\hat{\boldsymbol{\alpha}}$ be the estimate of $\boldsymbol{\alpha}$ and $\underline{\overline{\mathbf{X}}}_{\hat{\alpha}}$ be the matrix of the partial derivatives in the final iteration of the nonlinear least squares procedure. We know that asymptotically $\hat{\boldsymbol{\alpha}}$ is distributed as a multivariate normal distribution $\text{MN}(\boldsymbol{\alpha}, V(\hat{\boldsymbol{\alpha}}))$. The estimated variance-covariance matrix $\hat{V}(\hat{\boldsymbol{\alpha}})$ of $\hat{\boldsymbol{\alpha}}$ is

$$\begin{aligned} \hat{V}(\hat{\boldsymbol{\alpha}}) &= \hat{\sigma}_a^2 (\underline{\overline{\mathbf{X}}}'_{\hat{\alpha}} \underline{\overline{\mathbf{X}}}_{\hat{\alpha}})^{-1} \\ &= (\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}), \end{aligned} \tag{7.3.19}$$

where $\hat{\sigma}_a^2$ is estimated as in (7.2.7) or (7.2.13) and $\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}$ is the sample covariance between $\hat{\alpha}_i$ and $\hat{\alpha}_j$. We can test the hypothesis H_0: $\alpha_i = \alpha_{i0}$ using the following t statistic:

$$t = \frac{\hat{\alpha}_i - \alpha_{i0}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i}}} \tag{7.3.20}$$

with the degrees of freedom equaling $n - (p + q + 1)$ for the general ARMA model in (7.2.1). (More generally, the degrees of freedom equals the sample size used in estimation minus the number of parameters estimated in the model.) The estimated correlation matrix of these estimates is

$$\hat{R}(\boldsymbol{\alpha}) = (\hat{\rho}_{\hat{\alpha}_i \hat{\alpha}_j}), \tag{7.3.21}$$

where

$$\hat{\rho}_{\hat{\alpha}_i \hat{\alpha}_j} = \frac{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_j}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i} \hat{\sigma}_{\hat{\alpha}_j \hat{\alpha}_j}}}.$$

A high correlation among estimates indicates overparameterization, which should be avoided as it often causes difficulties in the convergence of the nonlinear least squares.

The results presented in this section are perfectly general. They hold for both linear and nonlinear models. We illustrate their uses in the following simple example.

EXAMPLE 7.3 Consider the AR(1) model

$$Z_t = \phi Z_{t-1} + a_t. \tag{7.3.22}$$

To put the estimation of ϕ in the context of this section, we use (7.3.13) to rewrite (7.3.22) as

$$Y_t = f(Z_{t-1}, \phi) + e_t, \tag{7.3.23}$$

where $Y_t = Z_t, f(Z_{t-1}, \phi) = \phi Z_{t-1}$, and $e_t = a_t$, for $t = 2, \ldots, n$. Thus,

$$\mathbf{Y} = \begin{bmatrix} Z_2 \\ \vdots \\ Z_n \end{bmatrix} \quad \text{and} \quad \mathbf{f}(\phi) = \begin{bmatrix} f(Z_1, \phi) \\ \vdots \\ f(Z_{n-1}, \phi) \end{bmatrix}.$$

Applying the nonlinear estimation to (7.3.23), we choose, with no loss of generality, the initial guess value $\tilde{\phi}$ of ϕ to be 0. It follows from steps 1 and 2 of the procedure that

$$\tilde{\mathbf{e}} = \mathbf{Y} - \tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{f}(\tilde{\phi}) = \begin{bmatrix} Z_2 \\ \vdots \\ Z_n \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{X}}_{\tilde{\phi}} = [X_{t1}],$$

where $\underline{\mathbf{X}}_{\tilde{\phi}}$ is the $(n-1) \times 1$ matrix of partial derivatives. That is,

$$X_{t1} = \left. \frac{\partial f(Z_{t-1}, \phi)}{\partial \phi} \right|_{\phi=0} = Z_{t-1}, \qquad t = 2, \ldots, n$$

and

$$\underline{\overline{\mathbf{X}}}_{\tilde{\phi}} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_{n-1} \end{bmatrix}.$$

Thus,

$$\hat{\phi} = (\underline{\overline{\mathbf{X}}}'_{\tilde{\phi}} \underline{\overline{\mathbf{X}}}_{\tilde{\phi}})^{-1} \underline{\overline{\mathbf{X}}}'_{\tilde{\phi}} \tilde{\mathbf{e}} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2}. \tag{7.3.24}$$

Properties of $\hat{\phi}$ The estimator $\hat{\phi}$ is distributed asymptotically as $N(\phi, V(\hat{\phi}))$, where $V(\hat{\phi}) = \sigma_a^2(\underline{\overline{\mathbf{X}}}'_{\tilde{\phi}} \underline{\overline{\mathbf{X}}}_{\tilde{\phi}})^{-1} = \sigma_a^2\left(\sum_{t=2}^{n} Z_{t-1}^2\right)^{-1}$ is the variance-covariance matrix of $\hat{\phi}$. The estimated variance-covariance matrix is

$$\hat{V}(\hat{\phi}) = \hat{\sigma}_a^2(\underline{\overline{\mathbf{X}}}'_{\tilde{\phi}} \underline{\overline{\mathbf{X}}}_{\tilde{\phi}})^{-1} = \hat{\sigma}_a^2\left(\sum_{t=2}^{n} Z_{t-1}^2\right)^{-1}, \tag{7.3.25}$$

where $\hat{\sigma}_a^2 = \sum_{t=2}^{n}\left(Z_t - \hat{\phi} Z_{t-1}\right)^2/\left(n-1\right)$. Conditional on $\{Z_t, t \leq n\}$, we can test the hypothesis H_0: $\phi = \phi_0$ using the standard normal statistic if σ_a^2 is known,

$$N = \frac{\hat{\phi} - \phi_0}{\sqrt{\sigma_a^2\left(\sum_{t=2}^{n} Z_{t-1}^2\right)^{-1}}}, \tag{7.3.26}$$

or the t statistic if σ_a^2 is unknown and estimated,

$$t = \frac{\hat{\phi} - \phi_0}{\sqrt{\hat{\sigma}_a^2\left(\sum_{t=2}^{n} Z_{t-1}^2\right)^{-1}}}, \tag{7.3.27}$$

which follows a t-distribution with $(n-1)$ degrees of freedom.

For the asymptotical distribution of $\hat{\phi}$, we note that $\hat{\phi} \xrightarrow{D} N(\phi, V(\hat{\phi}))$ or $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(0, nV(\hat{\phi}))$. Now, if $|\phi| < 1$, then we have

$$nV(\hat{\phi}) = \sigma_a^2\left(\frac{\sum_{t=2}^{n} Z_{t-1}^2}{n}\right)^{-1} \xrightarrow{P} \sigma_a^2\left(\frac{\sigma_a^2}{(1-\phi^2)}\right)^{-1} = (1-\phi^2). \tag{7.3.28}$$

Thus, asymptotically, when $|\phi| < 1$, we have $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(0, (1-\phi^2))$ or, equivalently, $\hat{\phi} \xrightarrow{D} N(\phi, (1-\phi^2)/n)$.

7.4 Ordinary Least Squares (OLS) Estimation in Time Series Analysis

Regression analysis is possibly the most commonly used statistical method in data analysis. As a result, the ordinary least squares (OLS) estimation developed for linear regression models is perhaps also the most frequently used estimation procedure in statistics. In this section, we discuss some problems of OLS estimation in time series analysis.

Consider the simple linear regression model

$$Z_t = \phi X_t + e_t, \qquad t = 1, \ldots, n. \tag{7.4.1}$$

Under the following basic assumptions on the error term e_t:

1. Zero mean: $E(e_t) = 0$
2. Constant variance: $E(e_t^2) = \sigma_e^2$
3. Nonautocorrelation: $E(e_t e_k) = 0$ for $t \neq k$
4. Uncorrelated with explanatory variable X_t: $E(X_t e_t) = 0$

it is well known that the OLS estimator

$$\hat{\phi} = \frac{\sum_{t=1}^{n} X_t Z_t}{\sum_{t=1}^{n} X_t^2} \tag{7.4.2}$$

is a consistent and the best linear unbiased estimator of ϕ. Note however, that assumption 4 is crucial for this result to hold. Assumption 4 automatically follows if the explanatory variables are nonstochastic. In a noncontrollable study, however, particularly when time series data are involved, the explanatory variables are usually also random variables.

Now, consider the following time series model:

$$Z_t = \phi Z_{t-1} + e_t, \qquad t = 1, \ldots, n. \tag{7.4.3}$$

The OLS estimator of ϕ, based on available data, is

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2}. \tag{7.4.4}$$

We would like to ask whether $\hat{\phi}$ is still unbiased and consistent in this case when the explanatory variable is a lagged dependent variable. The answer depends on the stochastic nature of the error term e_t. To see that, we rewrite $\hat{\phi}$ as

$$\begin{aligned} \hat{\phi} &= \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2} = \frac{\sum_{t=2}^{n} Z_{t-1}(\phi Z_{t-1} + e_t)}{\sum_{t=2}^{n} Z_{t-1}^2} \\ &= \phi + \frac{\sum_{t=2}^{n} Z_{t-1} e_t}{\sum_{t=2}^{n} Z_{t-1}^2} \end{aligned} \tag{7.4.5}$$

and consider the following two cases.

Case 1: $e_t = a_t$. That is, the e_t is a zero mean white noise series of constant variance σ_a^2. In this case, it is easy to see that $\hat{\phi}$ in (7.4.4) is equivalent to the first lag sample autocorrelation $\hat{\rho}_1$ for the series Z_t. If $|\phi| < 1$ and hence Z_t becomes an AR(1) process with an absolutely summable autocorrelation function, then by Section 2.5 $\hat{\rho}_1$ is a consistent estimator of ρ_1, which is equal to ϕ. Thus, $\hat{\phi}$ in (7.4.4) is an asymptotically unbiased and consistent estimator of ϕ.

Case 2: $e_t = (1 - \theta B)a_t$, where the a_t is a zero mean white noise series of constant variance σ_a^2; hence e_t is an MA(1) process. Under this condition, the series Z_t becomes an ARMA(1, 1) process

$$Z_t = \phi Z_{t-1} + a_t - \theta a_{t-1}, \tag{7.4.6}$$

and

$$E(Z_{t-1}e_t) = E\left[Z_{t-1}(a_t - \theta a_{t-1})\right] = -\theta\sigma_a^2. \tag{7.4.7}$$

This result shows that autocorrelation in the error term not only violates assumption 3 but also causes a violation of assumption 4 when the explanatory variables contain a lagged dependent variable. In this case, $\hat{\phi}$ is no longer a consistent estimator of ϕ because $\hat{\phi} \simeq \hat{\rho}_1$ is a consistent estimator for ρ_1, and for an ARMA(1, 1) process, by (3.4.14),

$$\rho_1 = \frac{(\phi - \theta)(1 - \phi\theta)}{1 + \theta^2 - 2\phi\theta} \neq \phi.$$

In summary, the OLS estimator for the parameter of an explanatory variable in a regression model will be inconsistent unless the error term is uncorrelated with the explanatory variable. For ARMA(p, q) models, this condition usually does not hold except when $q = 0$. Estimation methods discussed in Sections 7.2 and 7.3 are more efficient and commonly used in time series analysis.

7.5 Diagnostic Checking

Time series model building is an iterative procedure. It starts with model identification and parameter estimation. After parameter estimation, we have to assess model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the $\{a_t\}$ are white noise. That is, the a_t's are uncorrelated random shocks with zero mean and constant variance. For any estimated model, the residuals $\hat{a}_t$'s are estimates of these unobserved white noise a_t's. Hence, model diagnostic checking is accomplished through a careful analysis of the residual series $\{\hat{a}_t\}$. Because this residual series is the product of parameter estimation, the model diagnostic checking is usually contained in the estimation phase of a time series package.

To check whether the errors are normally distributed, one can construct a histogram of the standardized residuals $\hat{a}_t/\hat{\sigma}_a$ and compare it with the standard normal distribution using

the chi-square goodness of fit test or even Tukey's simple five-number summary. To check whether the variance is constant, we can examine the plot of residuals. To check whether the residuals are approximately white noise, we compute the sample ACF and sample PACF (or IACF) of the residuals to see whether they do not form any pattern and are all statistically insignificant, i.e., within two standard deviations if $\alpha = .05$.

Another useful test is the portmanteau lack of fit test. This test uses all the residual sample ACFs as a unit to check the joint null hypothesis

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_K = 0,$$

with the test statistic

$$Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}_k^2. \tag{7.5.1}$$

This test statistic is the modified Q statistic originally proposed by Box and Pierce (1970). Under the null hypothesis of model adequacy, Ljung and Box (1978) and Ansley and Newbold (1979b) show that the Q statistic approximately follows the $\chi^2(K-m)$ distribution, where $m = p + q$.

Based on the results of these residual analyses, if the entertained model is inadequate, a new model can be easily derived. For example, assume that the entertained AR(1) model

$$(1 - \phi_1 B)(Z_t - \mu) = b_t \tag{7.5.2}$$

produces an MA(1) residual series instead of a white noise series, i.e.,

$$b_t = (1 - \theta_1 B)a_t. \tag{7.5.3}$$

Then we should re-identify an ARMA(1, 1) model

$$(1 - \phi_1 B)(Z_t - \mu) = (1 - \theta_1 B)a_t \tag{7.5.4}$$

and go through the iterative stages of the model building until a satisfactory model is obtained. As mentioned earlier, if OLS estimation is used and the model should be indeed a mixed model, then the OLS estimates of the AR parameters based on a misidentified model are inconsistent. Although that may sometimes cause problems, the above procedure of using the residuals to modify models usually works fine.

7.6 Empirical Examples for Series W1–W7

For an illustration, we estimated the AR(3) model identified in Example 6.7 for Series W7—the yearly numbers of lynx pelt sales—and obtained the following result:

$$\underset{(.122)}{(1 - .97B} + \underset{(.184)}{.12B^2} + \underset{(.128)}{.50B^3})(\ln Z_t - 9.8) = a_t, \quad \underset{(.08)}{} \tag{7.6.1}$$

and $\hat{\sigma}_a^2 = .124$, where the values in the parentheses under the estimates refer to the standard errors of those estimates. They are all significant except for ϕ_2, and the model can be refitted with ϕ_2 removed if necessary.

To check model adequacy Table 7.2 gives the residual ACF and PACF and the Q statistics. The residual ACF and PACF are all small and exhibit no patterns. For $K = 24$, the Q statistic is $Q = 26.7$, which is not significant as $\chi^2_{.05}(21) = 32.7$, the chi-square value at the significance level $\alpha = .05$ for the degrees of freedom $= K - m = 24 - 3 = 21$. Thus, we conclude that the AR(3) model is adequate for the data.

Similarly, we use the nonlinear estimation procedure discussed in Section 7.3 to fit the models identified in Section 6.2 for Series W1 to W6. The results are summarized in Table 7.3.

Diagnostic checking similar to the one for Series W7 was performed for each model fitted in Table 7.3. All models except the AR(2) for Series W2 are adequate. Related tables are not shown here. Instead, we recall that in Example 6.9, Section 6.4.1, Series W7 was alternatively identified as an ARMA(2, 1) model using the ESACF. The estimation of this model gives

$$\underset{(.063)\ \ (.058)}{(1 - 1.55B + .94B^2)}(\ln Z_t - 9.8) = \underset{(.05)\ \ \ (.121)}{(1 - .59B)}a_t \tag{7.6.2}$$

with all parameters being significant and $\hat{\sigma}_a^2 = .116$. The result was also presented in Table 7.3. The residual autocorrelations from this ARMA(2, 1) model shown in Table 7.4 also indicate the adequacy of the model. In fact, both AR(3) and ARMA(2, 1) models fit the data almost equally well, raising the question of model selection to be discussed next.

TABLE 7.2 Residual ACF and PACF for the AR(3) model for Series W7.

(a) ACF $\hat{\rho}_k$

1–12	−.18	−.17	.27	−.00	−.01	−.15	.14	−.09	−.09	.05	.02	.03
St.E.	.14	.14	.15	.16	.16	.16	.16	.16	.16	.16	.16	.16
Q.	1.8	3.5	7.6	7.6	7.6	9.0	10.3	10.8	11.3	11.5	11.5	11.6
13–24	−.25	.18	.02	−.12	.22	−.05	.04	−.03	−.00	.03	−.09	−.15
St.E.	.16	.17	.17	.17	.18	.18	.18	.18	.18	.18	.18	.18
Q.	16.0	18.4	18.5	19.5	23.3	23.5	23.6	23.7	23.7	23.8	24.5	26.7

(b) PACF $\hat{\phi}_{kk}$

1–12	−.18	−.21	.21	.06	.09	−.21	.10	−.15	.02	−.08	.10	.01
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14
13–24	−.20	.07	−.04	.07	.17	.02	.04	−.06	−.06	−.04	.01	−.04
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14

TABLE 7.3 Summary of models fitted to Series W1–W7 (values in the parentheses under estimates refer to the standard errors of those estimates).

Series	No. of observations	Fitted models	$\hat{\sigma}_a^2$
W1	45	$(1 - .43B)Z_t = 1.04 + a_t$ (.134) (.25)	.2
W2	302	$(1 - 1.41B + .7B^2)\sqrt{Z_t} = 1.88 + a_t$ (.04) (.04) (.17)	1.37
		$(1 - 1.21B + .49B^2 + .12B^3 - .24B^4 + .23B^5 - .01B^6$ (.06) (.09) (.09) (.09) (.09) (.09) $-.17B^7 + .22B^8 - .31B^9)\sqrt{Z_t} = .80 + a_t$ (.09) (.09) (.06) (.28)	1.216
		$(1 - 1.23B + .52B^2 - .2B^9)\sqrt{Z_t} = .56 + a_t$ (.04) (.04) (.02) (.23)	1.222
W3	82	$(1 - .73B)Z_t = 1127.1 + a_t$ (.07) (313.36)	773,982.936
W4	500	$(1 - B)Z_t = (1 - .60B)a_t$ (.04)	1,324.727
W5	71	$(1 - B)Z_t = 2.05 + a_t$ (.33)	7.678
		$(1 - .98)Z_t = 4.63 + a_t$ (.01) (1.36)	7.283
W6	114	$(1 - B)\ln Z_t = (1 - .60B)a_t$ (.07)	.028
W7	55	$(1 - .97B + .12B^2 + .5B^3)\ln Z_t = 6.41 + a_t$ (.12) (.18) (.13) (.81)	.124
		$(1 - 1.55B + .94B^2)\ln Z_t = 3.9 + (1 - .59B)a_t$ (.06) (.06) (.41) (.12)	.116

TABLE 7.4 Residual autocorrelations, $\hat{\rho}_k$, from the ARMA(2, 1) model for Series W7.

1–8	−.14	.07	.24	−.02	.10	−.17	.16	−.11
St.E.	(.14)	(.14)	(.14)	(.15)	(.15)	(.15)	(.15)	(.16)
Q.	1.1	1.4	4.6	4.6	5.3	7.1	8.7	9.5
9–16	−.06	−.04	−.07	.10	−.26	.17	.02	−.09
St.E.	(.16)	(.16)	(.16)	(.16)	(.16)	(.17)	(.17)	(.17)
Q.	9.8	9.9	10.3	10.9	16.0	18.2	18.2	18.9

7.7 Model Selection Criteria

In time series analysis or more generally in any data analysis, several models may adequately represent a given data set. Sometimes, the best choice is easy; other times the choice can be very difficult. Thus, numerous criteria for model comparison have been introduced in the literature for model selection. They are different from the model identification methods discussed in Chapter 6. Model identification tools such as ACF, PACF, IACF, and ESACF are used only for identifying likely adequate models. Residuals from all adequate models are approximately white noise and are, in general, indistinguishable in terms of these functions. For a given data set, when there are multiple adequate models, the selection criterion is normally based on summary statistics from residuals computed from a fitted model or on forecast errors calculated from the out-sample forecasts. The latter is often accomplished by using the first portion of the series for model construction and the remaining portion as a holdout period for forecast evaluation. In this section, we introduce some model selection criteria based on residuals. Criteria based on out-sample forecast errors are discussed in Chapter 8.

1. Akaike's AIC and BIC Assume that a statistical model of M parameters is fitted to data. To assess the quality of the model fitting, Akaike (1973, 1974b) introduced an information criterion. The criterion has been called AIC (Akaike's information criterion) in the literature and is defined as

$$\text{AIC}(M) = -2\ln[\text{maximum likelihood}] + 2M, \tag{7.7.1}$$

where M is the number of parameters in the model. For the ARMA model and n effective number of observations, recall from (7.2.10) that the log-likelihood function is

$$\ln L = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{1}{2\sigma_a^2}S(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}). \tag{7.7.2}$$

Maximizing (7.7.2) with respect to $\boldsymbol{\phi}$, μ, $\boldsymbol{\theta}$, and σ_a^2, we have, from (7.2.13),

$$\ln \hat{L} = -\frac{n}{2}\ln \hat{\sigma}_a^2 - \frac{n}{2}(1 + \ln 2\pi). \tag{7.7.3}$$

Because the second term in (7.7.3) is a constant, the AIC criterion reduces to

$$\text{AIC}(M) = n\ln \hat{\sigma}_a^2 + 2M. \tag{7.7.4}$$

The optimal order of the model is chosen by the value of M, which is a function of p and q, so that AIC(M) is minimum.

Shibata (1976) has shown that the AIC tends to overestimate the order of the autoregression. Akaike (1978, 1979) has developed a Bayesian extension of the minimum AIC procedure, called the Bayesian information criterion (BIC), which takes the form

$$\text{BIC}(M) = n\ln \hat{\sigma}_a^2 - (n - M)\ln\left(1 - \frac{M}{n}\right) + M\ln n + M\ln\left[\left(\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - 1\right)/M\right], \tag{7.7.5}$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, M is the number of parameters, and $\hat{\sigma}_Z^2$ is the sample variance of the series. Through a simulation study Akaike (1978) has claimed that the BIC is less likely to overestimate the order of the autoregression. For further discussion on the properties of AIC, see Findley (1985).

2. *Schwartz's SBC* Similar to Akaike's BIC, Schwartz (1978) suggested the following Bayesian criterion of model selection, which has been called SBC (Schwartz's Bayesian criterion):

$$\text{SBC}(M) = n \ln \hat{\sigma}_a^2 + M \ln n. \tag{7.7.6}$$

Again in (7.7.6), $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, M is the number of parameters in the model, and n is the effective number of observations that is equivalent to the number of residuals that can be calculated from the series.

3. *Parzen's CAT* Parzen (1977) has suggested the following model selection criterion, which he called CAT (criterion for autoregressive transfer functions):

$$\text{CAT}(p) = \begin{cases} -\left(1 + \dfrac{1}{n}\right), & p = 0, \\ \dfrac{1}{n}\displaystyle\sum_{j=1}^{p} \frac{1}{\hat{\sigma}_j^2} - \frac{1}{\hat{\sigma}_p^2}, & p = 1, 2, 3, \ldots, \end{cases} \tag{7.7.7}$$

where $\hat{\sigma}_j^2$ is the unbiased estimate of σ_a^2 when an AR(j) model is fitted to the series and n is the number of observations. The optimal order of p is chosen so that CAT(p) is minimum.

We have introduced only some commonly used model selection criteria. There are many other criteria introduced in the literature. Interested readers are referred to Hannan and Quinn (1979), Stone (1979), Hannan (1980), and others.

EXAMPLE 7.4 The AIC has become a standard tool in time series model fitting, and its computation is available in many time series programs. In Table 7.5, we use the SAS/ETS (1999) software to compute AIC for Series W7, Canadian lynx pelt sales. From the table, it is clear that the minimum AIC occurs for $p = 2$ and $q = 1$. Hence, based on the AIC, an ARMA(2, 1) model should be selected for the data. Note that a competitive AR(3) model that we fitted earlier to the same data set gives the second smallest value of AIC.

TABLE 7.5 AIC values for Series W7.

p \ q	0	1	2	3	4
0	142.1200	93.6185	70.0174	57.8203	55.3785
1	94.7730	69.9294	62.2191	56.4811	58.4985
2	31.9081	*23.7781	41.5118	49.7238	48.1051
3	24.0529	25.2286	43.7349	31.1166	47.2080
4	25.6708	27.4769	88.6398	60.0407	75.5548

EXERCISES

7.1 Assume that 100 observations from an ARMA(1, 1) model

$$Z_t - \phi_1 Z_{t-1} = a_t - \theta_1 a_{t-1}$$

gave the following estimates: $\hat{\sigma}_z^2 = 10$, $\hat{\rho}_1 = .523$, and $\hat{\rho}_2 = .418$. Find method of moments estimates for ϕ_1, θ_1, and σ_a^2.

7.2 Assume that 100 observations from an AR(2) model

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

gave the following sample ACF: $\hat{\rho}_1 = .8$, $\hat{\rho}_2 = .5$, and $\hat{\rho}_3 = .4$. Estimate ϕ_1 and ϕ_2.

7.3 Given the set of observations 2.2, 4.5, 2.5, 2.3, 1.1, 3.0, 2.1, and 1.0, calculate the conditional sum of squares $S(\theta_1, \theta_2)$ for the MA(2) process with $\theta_1 = -.5$ and $\theta_2 = .2$.

7.4 Given the set of observations 6, 2, 4, 5, 3, 4, 2, and 1, illustrate how to calculate the conditional sum of squares function $S(\phi_1, \theta_1)$ for the ARMA(1, 1) model.

7.5 Consider the following observations from an ARIMA(0, 1, 1) model with $\theta = .4$:

t	Z_t	$W_t = (1 - B)Z_t$
0	59	
1	62	3
2	58	−4
3	63	5
4	79	16
5	90	11
6	88	−2

(a) Calculate the conditional sum of squares (with $a_0 = 0$).

(b) Calculate the unconditional sum of squares using the backcasting method as shown in Table 7.1.

7.6 Simulate 100 observations from an ARMA(1, 1) model.

(a) Fit the simulated series with an AR(1) or an MA(1) model. Carry out diagnostic checking, and modify your fitted model from the result of residual analysis.

(b) Estimate the parameters of your modified model, and compare with the true parameter values of the model.

7.7 A summary of models fitted for the series W1 to W7 is given in Table 7.3. Perform residual analysis and model checking for each of the fitted models.

7.8 Use AIC to find a model for each of the series W1 to W7, and compare it with the fitted model given in Table 7.3.

7.9 Suppose that $(1 - \phi B)Z_t = (1 - \theta B)a_t$ is a tentatively entertained model for a process. Given

t	1	2	3	4	5	6	7	8	9	10	11	12
Z_t	−3.1	−.8	1.2	.6	2.8	−.9	.3	−1.4	−2.5	−1.1	.9	1.4

calculate the unconditional sum of squares for $\phi = .4$ and $\theta = .8$.

7.10 Consider the AR(1) model

$$(1 - \phi B)(Z_t - \mu) = a_t.$$

(a) For $\mu = 0$, find the maximum likelihood estimator for ϕ and its associated variance.

(b) Find the maximum likelihood estimators for ϕ and μ when $\mu \neq 0$.

(c) Discuss the relationship between the ordinary least square estimator and the maximum likelihood estimator for ϕ in the given model.

7.11 Illustrate the nonlinear estimation procedure discussed in Section 7.3 using the following models:

(a) The MA(1) model $Z_t = (1 - \theta B)a_t$.

(b) The AR(2) model $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$.

7.12 Derive the joint asymptotical distribution of the least squares estimators, $\left[\hat{\phi}_1, \hat{\phi}_2\right]'$, for the AR(2) model

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t.$$

7.13 Consider the yearly data of lumber production (in billions of board feet) in the United States given as follows:

Year	Production									
1921–1930	29.0	35.2	41.0	39.5	41.0	39.8	37.3	36.8	38.7	29.4
1931–1940	20.0	13.5	17.2	18.8	22.9	27.6	29.0	24.8	28.8	31.2
1941–1950	36.5	36.3	34.3	32.9	28.1	34.1	35.4	37.0	32.2	38.0
1951–1960	37.2	37.5	36.7	36.4	37.4	38.2	32.9	33.4	37.2	32.9
1961–1970	32.0	33.2	34 7	36.6	36.8	36.6	34.7	36.5	35.8	34.7
1071–1980	37.0	37.7	38.6	34.6	32.6	36.3	39.4	40.5	40.6	35.4
1981–1982	31.7	30.0								

(a) Plot the data and perform necessary analysis to construct an appropriate model for the series.

(b) Find and plot the forecasts for the next four years, and calculate 95% forecast limits.

(c) Update your forecasts when the 1983 observation became available and equaled 34.6.

8

Seasonal Time Series Models

Because of their common occurrence in our daily activities, we devote a separate chapter to seasonal time series. After a brief introduction of some basic concepts and conventional methods, we extend the autoregressive integrated moving average models to represent seasonal series. Detailed examples are given to illustrate the methods.

8.1 General Concepts

Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitive phenomenon is called the seasonal period. For example, the quarterly series of ice-cream sales is high each summer, and the series repeats this phenomenon each year, giving a seasonal period of 4. Similarly, monthly auto sales and earnings tend to decrease during August and September every year because of the changeover to new models, and the monthly sales of toys rise every year in the month of December. The seasonal period in these later cases is 12. Seasonal phenomena may stem from factors such as weather, which affects many business and economic activities like tourism and home building; cultural events like Christmas, which is closely related to sales such as jewelry, toys, greeting cards, and stamps; and graduation ceremonies in the summer months, which are directly related to the labor force status in these months.

As an illustration, Figure 8.1 shows the U.S. monthly employment figures (in thousands) for young men aged between 16 and 19 years from 1971 to 1981. The seasonal nature of the series is apparent. The numbers increase dramatically in the summer months, with peaks occurring in the month of June when schools are not in session, and decrease in the fall months when schools reopen. The phenomenon repeats itself every 12 months, and thus the seasonal period is 12.

More generally, suppose that the series $\{Z_t\}$ is seasonal with seasonal period s. To analyze the data, it is helpful to arrange the series in a two-dimensional table as shown in Table 8.1 according to the period and season and including the totals and averages. Wold (1938) credits these arrangements of the table to Buys-Ballot (1847); hence, the table has also been called the Buys-Ballot table in the literature.

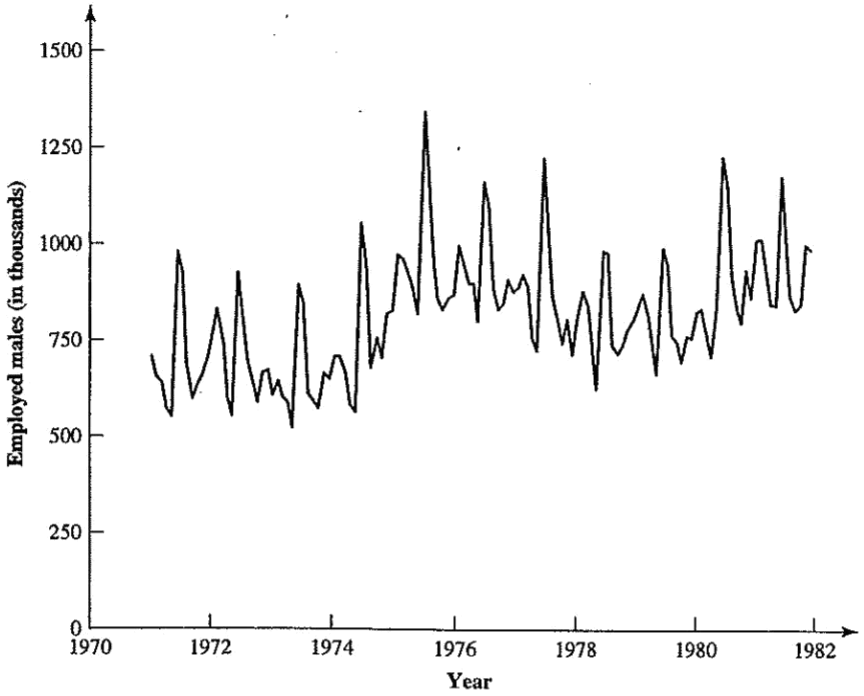


FIGURE 8.1 U.S. monthly employment figures (in thousands) for young men aged between 16 and 19 years from 1971 to 1981.

TABLE 8.1 Buys-Ballot table for a seasonal time series.

	1	2	3	⋯	s	Total	Average
	Z_1	Z_2	Z_3	⋯	Z_s	$T_{\cdot 1}$	$\bar{Z}_{\cdot 1}$
	Z_{s+1}	Z_{s+2}	Z_{s+3}	⋯	Z_{2s}	$T_{\cdot 2}$	$\bar{Z}_{\cdot 2}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$Z_{(n-1)s+1}$	$Z_{(n-1)s+2}$	$Z_{(n-1)s+3}$	⋯	Z_{ns}	$T_{\cdot n}$	$\bar{Z}_{\cdot n}$
Total	$T_{1\cdot}$	$T_{2\cdot}$	$T_{3\cdot}$	⋯	$T_{n\cdot}$	T	T/s
Average	$\bar{Z}_{1\cdot}$	$\bar{Z}_{2\cdot}$	$\bar{Z}_{3\cdot}$	⋯	$\bar{Z}_{n\cdot}$	T/n	T/ns

$T_{j\cdot}$ = the total for the jth season,
$\bar{Z}_{j\cdot}$ = the average for the jth season,
$T_{\cdot j}$ = the total for the jth period,
$\bar{Z}_{\cdot j}$ = the average for the jth period,
T = the grand total of the series.

8.2 Traditional Methods

Conventionally, time series have been thought to consist of a mixture of trend-cycle (P_t), seasonal (S_t), and irregular components (e_t). If these components are assumed to be additive, then one can write the time series Z_t as

$$Z_t = P_t + S_t + e_t. \tag{8.2.1}$$

To estimate these components, several decomposition methods have been introduced in the literature.

8.2.1 Regression Method

In the regression method, an additive seasonal time series is written as the following regression model:

$$\begin{aligned} Z_t &= P_t + S_t + e_t \\ &= \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it} + \sum_{j=1}^{k} \beta_j V_{jt} + e_t, \end{aligned} \tag{8.2.2}$$

where $P_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i U_{it}$, and the U_{it} are the trend-cycle variables; $S_t = \sum_{j=1}^{k} \beta_j V_{jt}$, and the V_{jt} are the seasonal variables. For example, a linear trend-cycle component P_t can be written as

$$P_t = \alpha_0 + \alpha_1 t. \tag{8.2.3}$$

More generally, a trend-cycle component can be written as an mth-order polynomial in time as

$$P_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i. \tag{8.2.4}$$

Similarly, the seasonal component S_t can be described as a linear combination of seasonal dummy (indicator) variables or as a linear combination of sine-cosine functions of various frequencies. For example, a seasonal series of period s can be written as

$$S_t = \sum_{j=1}^{s-1} \beta_j D_{jt}, \tag{8.2.5}$$

where $D_{jt} = 1$ if t corresponds to the seasonal period j and 0 otherwise. Note that when the seasonal period equals s, we need only $(s - 1)$ seasonal dummy variables. In other words, β_s is set to be 0 so that the coefficient β_j, $j \neq s$, represents the seasonal effect of the jth period as compared with the period s. Alternatively, S_t can be written as

$$S_t = \sum_{j=1}^{[s/2]} \left[\beta_j \sin\left(\frac{2\pi jt}{s}\right) + \gamma_j \cos\left(\frac{2\pi jt}{s}\right) \right], \tag{8.2.6}$$

where $[s/2]$ is the integer portion of $s/2$. A model of this type is discussed further in Chapter 13. Thus, the model (8.2.2) becomes

$$Z_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{s-1} \beta_j D_{jt} + e_t \tag{8.2.7}$$

or

$$Z_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i t^i + \sum_{j=1}^{[s/2]} \left[\beta_j \sin\left(\frac{2\pi jt}{s}\right) + \gamma_j \cos\left(\frac{2\pi jt}{s}\right) \right] + e_t. \tag{8.2.8}$$

For a given data set Z_t and specified values of m and s, the standard least squares regression method can be used to obtain estimates $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_j$ of the parameters α_i, β_j, and γ_j. The estimates of P_t, S_t, and e_t for Equation (8.2.7) are then given by

$$\hat{P}_t = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i t^i, \tag{8.2.9a}$$

$$\hat{S}_t = \sum_{j=1}^{s-1} \hat{\beta}_j D_{jt}, \tag{8.2.9b}$$

and

$$\hat{e}_t = Z_t - \hat{P}_t - \hat{S}_t. \tag{8.2.9c}$$

For Equation (8.2.8) they are given by

$$\hat{P}_t = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i t^i, \tag{8.2.10a}$$

$$\hat{S}_t = \sum_{j=1}^{[s/2]} \left[\hat{\beta}_j \sin\left(\frac{2\pi jt}{s}\right) + \hat{\gamma}_j \cos\left(\frac{2\pi jt}{s}\right) \right], \tag{8.2.10b}$$

and

$$\hat{e}_t = Z_t - \hat{P}_t - \hat{S}_t. \tag{8.2.10c}$$

8.2.2 Moving Average Method

The moving average method is developed based on the assumption that an annual sum of a seasonal time series possesses little seasonal variation. Thus, by letting $N_t = P_t + e_t$ be

the nonseasonal component of the series, an estimate of the nonseasonal component can be obtained by using a symmetric moving average operator. That is,

$$\hat{N}_t = \sum_{i=-m}^{m} \lambda_i Z_{t-i}, \tag{8.2.11}$$

where m is a positive integer and the λ_i are constants such that $\lambda_i = \lambda_{-i}$ and $\sum_{i=-m}^{m} \lambda_i = 1$. An estimate of the seasonal component is derived by subtracting $\hat{N}_t$ from the original series, i.e.,

$$\hat{S}_t = Z_t - \hat{N}_t. \tag{8.2.12}$$

These estimates may be obtained iteratively by repeating various moving average operators. A successful example of using this moving average procedure is the Census X-12 method, which has been widely employed in government and industry.

The series with seasonal component removed, i.e., $Z_t - \hat{S}_t$, is referred to as the seasonally adjusted series. Hence, the decomposition procedures are also known as seasonal adjustment. Based on prevailing belief that the seasonal component is a well-behaved phenomenon that can be predicted with a reasonable accuracy, there is great interest in government and industry to adjust the series seasonally. The topic is presented only briefly here. Interested readers are referred to an excellent collection of articles edited by Zellner (1978) and to some research papers on the topic such as Dagum (1980), Pierce (1980), Hillmer and Tiao (1982), Bell and Hillmer (1984), and Cupingood and Wei (1986).

8.3 Seasonal ARIMA Models

Traditional methods presented in Section 8.2 are based on assumptions that the seasonal component is deterministic and independent of other nonseasonal components. Many time series, however, are not as well behaved. More likely, the seasonal component may be stochastic and correlated with nonseasonal components. In this section, we extend the stochastic ARIMA models discussed in previous chapters to model seasonal time series.

For an illustration, let us examine the U.S. monthly employment figures (in thousands) for young men aged between 16 and 19 years from 1971 to 1981 using the Buys-Ballot table shown in Table 8.2. The table indicates that the employment figures are not only related from month to month but also related from year to year. Thus, to forecast the employment level for the month of June, one should examine not only the employment levels in the neighboring months such as May and July, but also the employment levels for the month of June in the previous years.

In general, the Buys-Ballot table implies that $\{Z_t\}$ contains within periods and between periods relationships. Within periods relationships represent the correlation among $\ldots Z_{t-2}$, Z_{t-1}, Z_t, Z_{t+1}, Z_{t+2}, $\ldots$ and between periods relationships represent correlation among $\ldots Z_{t-2s}$, Z_{t-s}, Z_t, Z_{t+s}, Z_{t+2s}, $\ldots$.

Suppose that we do not know that the series $\{Z_t\}$ contains the between period seasonal variation and fit a nonseasonal ARIMA model for the series, i.e.,

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B) b_t. \tag{8.3.1}$$

TABLE 8.2 The Buys-Ballot table for the U.S. monthly employment figures (in thousands) for young men aged between 16 and 19 years.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total	Avg.
1971	707	655	638	574	552	980	926	680	597	637	660	704	8,310	692.5
1972	758	835	747	617	554	929	815	702	640	588	669	675	8,529	710.75
1973	610	651	605	592	527	898	839	614	594	576	672	651	7,829	652.42
1974	714	715	672	588	567	1,057	949	683	771	708	824	835	9,083	756.92
1975	980	969	931	892	828	1,350	1,218	977	863	838	866	877	11,589	965.75
1976	1,007	951	906	911	812	1,172	1,101	900	841	853	922	886	11,262	938.5
1977	896	936	902	765	735	1,234	1,052	868	798	751	820	725	10,482	873.5
1978	821	895	851	734	636	994	990	750	727	754	792	817	9,761	813.42
1979	856	886	833	733	675	1,004	956	777	761	709	777	771	9,738	811.5
1980	840	847	774	720	898	1,240	1,168	936	853	910	953	874	10,963	913.58
1981	1,026	1,030	946	860	856	1,190	1,038	883	843	857	1,016	1,003	11,548	962.33
Total	9,215	9,310	8,805	7,986	7,590	12,048	11,052	8,770	8,288	8,181	8,971	8,818	109,094	9,091.17
Avg.	837.73	851.82	800.45	726	690	1,095.27	1,004.73	797.27	753.36	743.73	815.55	801.64	9,917.64	826.47

Then obviously the series $\{b_t\}$ will not be white noise as it contains unexplained between period (seasonal) correlations. Let

$$\rho_{j(s)} = \frac{E(b_{t-js} - \mu_b)(b_t - \mu_b)}{\sigma_b^2}, \qquad j = 1, 2, 3, \ldots \tag{8.3.2}$$

be the autocorrelation function for $\{b_t\}$ representing the unexplained between periods relationship. It can be easily seen that this between periods relationship can also be represented as an ARIMA model

$$\Phi_P(B^s)(1 - B^s)^D b_t = \Theta_Q(B^s)a_t, \tag{8.3.3}$$

where

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

and

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

are polynomials in B^s with no common roots. The roots of these polynomials lie outside of the unit circle, and $\{a_t\}$ is a zero mean white noise process.

For illustration, suppose that in (8.3.3) $P = 1$, $s = 12$, $D = 0$, and $Q = 0$. Then

$$(1 - \Phi B^{12})b_t = a_t. \tag{8.3.4}$$

If $\Phi = .9$, then the autocorrelation function of $\{b_t\}$ becomes $\rho_{j(12)} = (.9)^j$ as shown in Figure 8.2.

Similarly, if $P = 0$, $s = 12$, $D = 0$, and $Q = 1$, we have

$$b_t = (1 - \Theta B^{12})a_t. \tag{8.3.5}$$

If $\Theta = .8$, then the autocorrelation function becomes

$$\rho_{j(12)} = \begin{cases} \dfrac{-.8}{1.64}, & j = 1, \\ 0, & j \neq 1, \end{cases}$$

which is shown in Figure 8.3.

Combining (8.3.1) and (8.3.3) we get the well-known Box–Jenkins multiplicative seasonal ARIMA model,

$$\Phi_P(B^s)\phi_p(B)(1 - B)^d(1 - B^s)^D \dot{Z}_t = \theta_q(B)\Theta_Q(B^s)a_t, \tag{8.3.6}$$

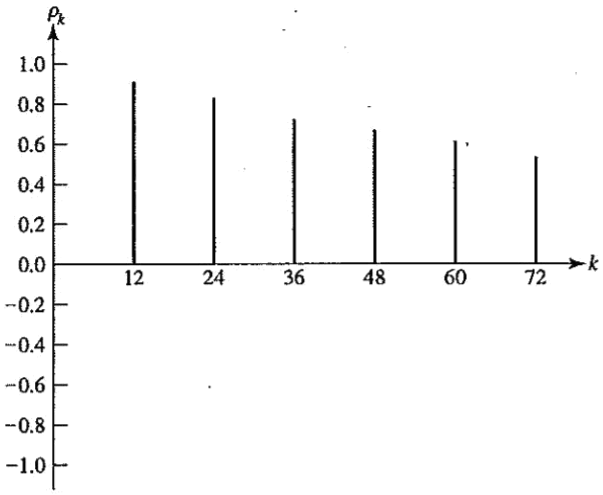


FIGURE 8.2 ACF for $(1 - .9B^{12})b_t = a_t$.

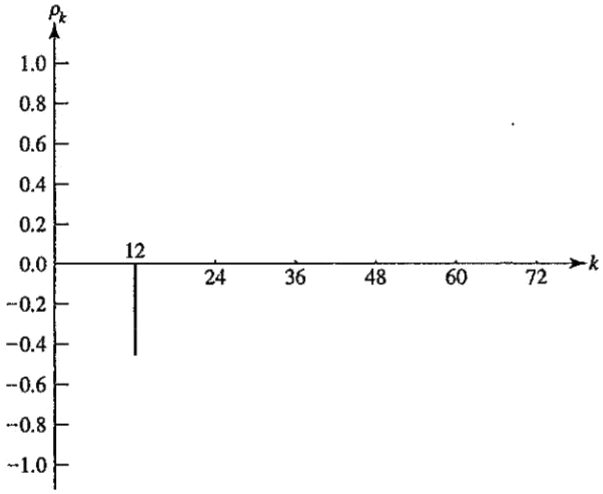


FIGURE 8.3 ACF for $b_t = (1 - .8B^{12})a_t$.

where

$$\dot{Z}_t = \begin{cases} Z_t - \mu, & \text{if } d = D = 0, \\ Z_t, & \text{otherwise.} \end{cases}$$

For convenience, we often call $\phi_p(B)$ and $\theta_q(B)$ the regular autoregressive and moving average factors (polynomials) and $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ the seasonal autoregressive and moving average factors (or polynomials), respectively. The model in (8.3.6) is often denoted as ARIMA$(p, d, q) \times (P, D, Q)_s$, where the subindex s refers to the seasonal period.

EXAMPLE 8.1 Let us consider the ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ model

$$(1 - B)(1 - B^{12})Z_t = (1 - \theta B)(1 - \Theta B^{12})a_t. \tag{8.3.7}$$

This model has been found very useful to represent a variety of seasonal time series such as airline data and trade series. The model was first introduced by Box and Jenkins to represent the international air travel data. Hence, it is also referred to as the airline model in the literature.

The autocovariances of $W_t = (1 - B)(1 - B^{12})Z_t$ can be easily found to be

$$\begin{aligned}
\gamma_0 &= (1 + \theta^2)(1 + \Theta^2)\sigma_a^2, \\
\gamma_1 &= -\theta(1 + \Theta^2)\sigma_a^2, \\
\gamma_{11} &= \theta\Theta\sigma_a^2, \\
\gamma_{12} &= -\Theta(1 + \theta^2)\sigma_a^2, \\
\gamma_{13} &= \theta\Theta\sigma_a^2, \\
\gamma_j &= 0, \quad \text{otherwise.}
\end{aligned} \tag{8.3.8}$$

Hence, the ACF becomes

$$\begin{aligned}
\rho_1 &= \frac{-\theta}{(1 + \theta^2)}, \\
\rho_{11} &= \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)} = \rho_{13}, \\
\rho_{12} &= \frac{-\Theta}{(1 + \Theta^2)}, \\
\rho_j &= 0, \quad \text{otherwise.}
\end{aligned} \tag{8.3.9}$$

For $\theta = .4$ and $\Theta = .6$, the plot of ρ_k is shown in Figure 8.4.

The General ARIMA Models More generally, we can write a general ARIMA model as follows:

$$\prod_{j=1}^{M} \phi_j(B) \prod_{i=1}^{K} (1 - B^{s_i})^{d_i} \dot{Z}_t = \theta_0 + \prod_{k=1}^{N} \theta_k(B) a_t. \tag{8.3.10}$$

Thus, the model may contain K differencing factors, M autoregressive factors, and N moving average factors. This extension is useful in describing many nonstandard time series that, for example, may contain a mixture of seasonal phenomena of different periods. Because it is this general form that most time series software use, we now explain this general model in more detail.

The ith differencing factor is

$$(1 - B^{s_i})^{d_i}$$

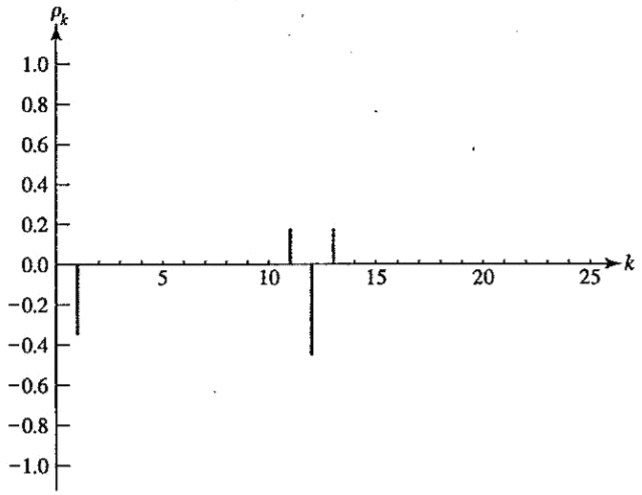


FIGURE 8.4 ACF for $W_t = (1 - .4B)(1 - .6B^{12})a_t$.

with the order s_i (the power of B) and the degree d_i. If $K = 0$, then $\dot{Z}_t = Z_t - \mu$. Otherwise, $\dot{Z}_t = Z_t$, as the mean of the series, does not exist. The parameter θ_0 represents the deterministic trend and is considered only when $K \neq 0$. The jth autoregressive factor

$$\phi_j(B) = (1 - \phi_{j1}B - \phi_{j2}B^2 - \cdots - \phi_{jp_j}B^{p_j})$$

contains one or more autoregressive parameters, ϕ_{jm}. The kth moving average factor is

$$\theta_k(B) = (1 - \theta_{k1}B - \theta_{k2}B^2 - \cdots - \theta_{kq_k}B^{q_k})$$

and contains one or more moving average parameters, θ_{kn}. In most applications, the values of K, M, and N are usually less than or equal to 2. Parameters in autoregressive and moving average factors beyond the first factor of each kind are normally considered seasonal parameters in the model. Clearly, one may use the factors in any desired order—for example, making the first factors in each case as the "seasonal factors"—if so desired. The same is true for the differencing factors.

PACF, IACF, and ESACF for Seasonal Models The PACF and IACF for seasonal models are more complicated. In general, the seasonal and nonseasonal autoregressive components have their PACF and IACF cutting off at the seasonal and nonseasonal lags. On the other hand, the seasonal and nonseasonal moving average components produce PACF and IACF that show exponential decays or damped sine waves at the seasonal and nonseasonal lags. The computation of the ESACF for seasonal models is time consuming, and the patterns are in general very complicated. Besides, because the ESACF provides only the information about the maximum order of p and q, its use in modeling seasonal time series is very limited. We have found that in identifying seasonal time series, the standard ACF analysis is still the most useful method.

Model Building and Forecasting for Seasonal Models Because seasonal models are special forms of the ARIMA models introduced in Chapters 3 and 4, the model identification, parameter estimation, diagnostic checking, and forecasting for these models follow the same general methods introduced in Chapters 5, 6, and 7. The description of these methods is not repeated here. Instead, we illustrate the application of these methods to several seasonal time series in the next section.

8.4 Empirical Examples

EXAMPLE 8.2 This example shows a simulation of 150 values from an ARIMA(0, 1, 1) × $(0, 1, 1)_4$ model

$$(1 - B)(1 - B^4)Z_t = (1 - \theta B)(1 - \Theta B^4)a_t \tag{8.4.1}$$

with $\theta = .8$, $\Theta = .6$, and a_t being Gaussian $N(0, 1)$ white noise. The series is listed as Series W8 in the appendix. As shown in Figure 8.5, it is clearly seasonal with an upward trend.

Table 8.3 and Figure 8.6 show the sample ACF and PACF for the series. The ACF is large and decays very slowly, whereas the PACF has a single large spike at lag 1. These values indicate that the series is nonstationary and that differencing is called for.

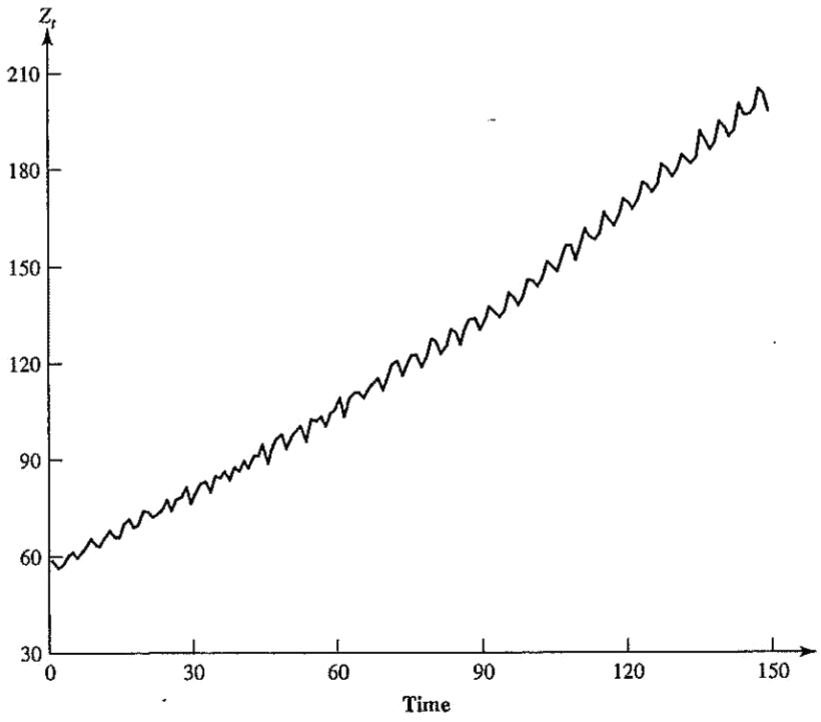


FIGURE 8.5 A simulated ARIMA(0, 1, 1) × $(0, 1, 1)_4$ series $(1 - B)(1 - B^4)Z_t = (1 - .8B)(1 - .6B^4)a_t$.

TABLE 8.3 Sample ACF and sample PACF for a simulated series from $(1 - B)(1 - B^4)Z_t = (1 - .8B)(1 - .6B^4)a_t$.

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	.98	.96	.94	.92	.90	.88	.87	.88
St.E.	.08	.14	.18	.21	.23	.26	.28	.29
$\hat{\phi}_{kk}$	.98	.04	.11	−.04	−.12	−.00	.11	−.03
St.E.	.08	.08	.08	.08	.08	.08	.08	.08

$\overline{Z} = 174.87, S_Z = 47.93, n = 150.$

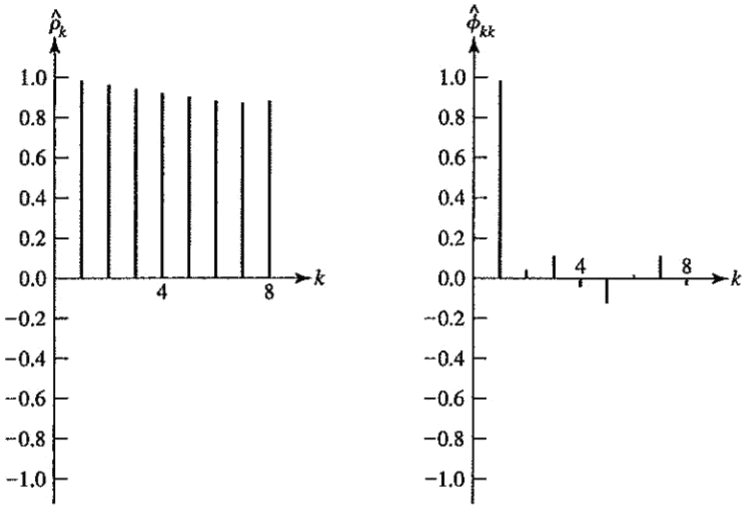


FIGURE 8.6 Sample ACF and sample PACF for a simulated series from $(1 - B)(1 - B^4)Z_t = (1 - .8B)(1 - .6B^4)a_t$.

To remove nonstationarity, the series is differenced and the sample ACF and PACF of the differenced series $(1 - B)Z_t$ are computed as shown in Table 8.4(a) and Figure 8.7(a). That the ACF decays very slowly at multiples of seasonal period 4 implies that a seasonal differencing $(1 - B^4)$ is also needed so as to achieve stationarity. Thus, we examine the sample ACF and sample PACF for $(1 - B)(1 - B^4)Z_t$ as given in Table 8.4(b) and Figure 8.7(b). It is known that the ACF of an ARIMA(0, 0, 1) × $(0, 0, 1)_4$ model is zero except at lags 1, 3, 4, and 5. Hence, the sample ACF of $(1 - B)(1 - B^4)Z_t$ in Table 8.4(b) and Figure 8.7(b) implies an ARIMA(0, 1, 1) × $(0, 1, 1)_4$ model for the original series $\{Z_t\}$. That is,

$$(1 - B)(1 - B^4)Z_t = (1 - \theta B)(1 - \Theta B^4)a_t.$$

Parameter estimation procedures can now be used to estimate the values of θ and Θ.

TABLE 8.4 Sample ACF and sample PACF for the differenced series simulated from $(1 - B^4)(1 - B)Z_t = (1 - .8B)(1 - .6B^4)a_t$.

(a) For $W_t = (1 - B)Z_t$ ($\overline{W} = 1.10$, $S_W = 3.87$, $n = 149$)

k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$	−.11	−.77	−.06	.89	−.04	−.77	−.06	.87	−.04	−.76	−.04	.82
St.E.	.08	.08	.12	.12	.16	.16	.18	.18	.21	.21	.23	.23
$\hat{\phi}_{kk}$	−.11	−.79	−.80	.14	.19	.01	−.20	.19	.07	.01	.09	.11
St.E.	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08

(b) For $W_t = (1 - B)(1 - B^4)Z_t$ ($\overline{W} = -.01$, $S_W = 1.57$, $n = 145$)

k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$	−.56	.07	.29	−.53	.27	−.05	−.06	.10	.03	−.06	.05	.01
St.E.	.08	.11	.11	.11	.13	.13	.13	.13	.13	.13	.13	.13
$\hat{\phi}_{kk}$	−.56	−.36	.23	−.34	−.34	−.27	.04	−.19	−.06	−.07	.05	.00
St.E.	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08

EXAMPLE 8.3 We now examine the U.S. monthly employment figures for young men between 16 and 19 years of age from 1971 to 1981, which is listed as Series W9 in the appendix. The series was shown earlier in Figure 8.1.

Model Identification The columns and the associated marginal column totals of the Buys-Ballot table in Table 8.2 indicate a strong seasonal variation with seasonal period 12. The row totals over the years, on the other hand, imply that there is a stochastic trend in the series. This trend has to be removed by differencing before a model can be identified. The sample ACF for the original and differenced series are computed and given in Table 8.5. The need for both regular differencing $(1 - B)$ and seasonal differencing $(1 - B^{12})$ is clear from Table 8.5(b) and (c). Now, the ACF for $\{W_t = (1 - B)(1 - B^{12})Z_t\}$ as shown in Table 8.5(d) has only one significant spike at lag 12. Because $\overline{W} = .66$, $S_w = 71.85$, $n = 119$, the t-value of $\overline{W} = .66/(71.85/\sqrt{119}) = .1$, which is not significant, and thus the deterministic trend term θ_0 is not needed. Hence, we identify the following ARIMA(0, 1, 0) × $(0, 1, 1)_{12}$ process as a tentative model for the series

$$(1 - B)(1 - B^{12})Z_t = (1 - \Theta B^{12})a_t. \tag{8.4.2}$$

Parameter Estimation and Diagnostic Checking Using the standard nonlinear estimation procedure, we obtain the following results:

$$(1 - B)(1 - B^{12})Z_t = \underset{(.066)}{(1 - .79B^{12})}a_t \tag{8.4.3}$$

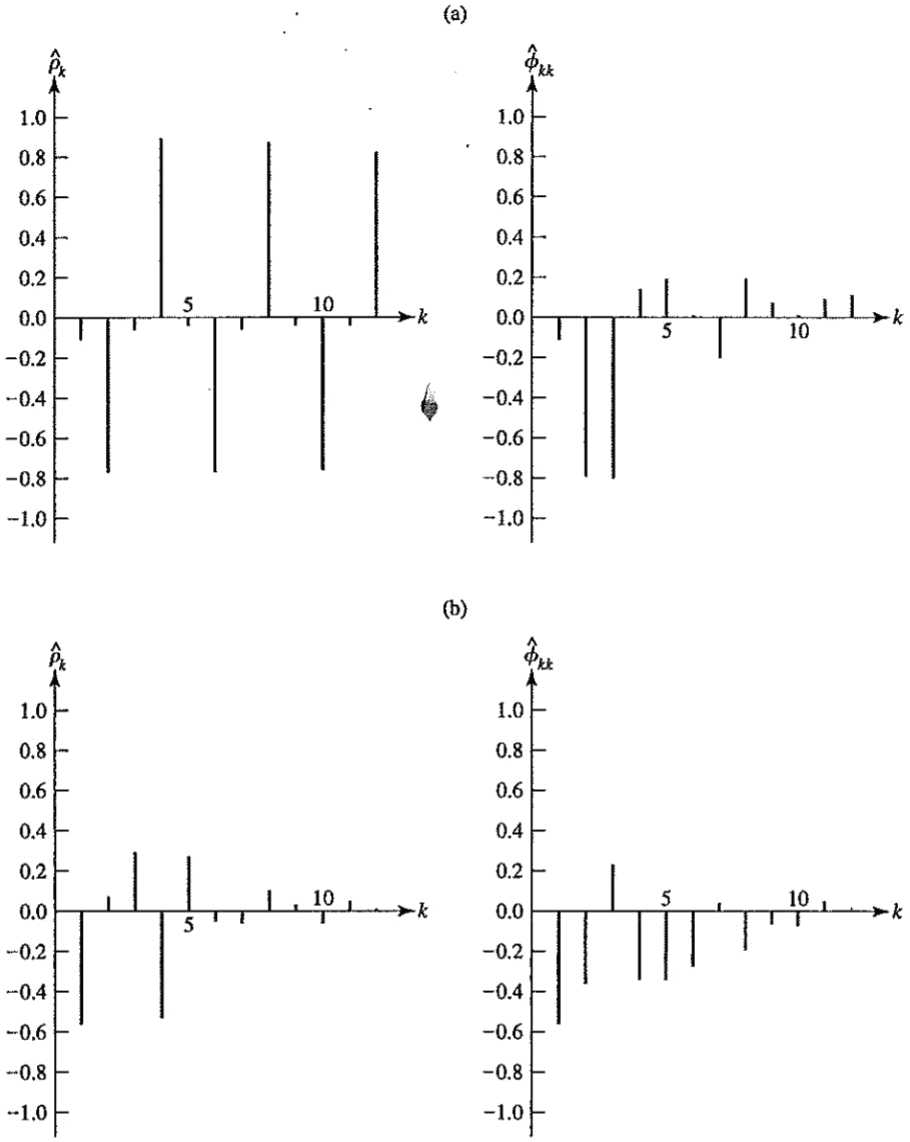


FIGURE 8.7 Sample ACF and sample PACF of the differenced series. (a) For $W_t = (1 - B)Z_t$. (b) For $W_t = (1 - B)(1 - B^4)Z_t$.

with $\hat{\sigma}_a^2 = 3327.3$. The residual ACF for the fitted model as shown in Table 8.6, however, has a significant spike at lag 1. Thus, we modify the model to an ARIMA $(0, 1, 1) \times (0, 1, 1)_{12}$, i.e.,

$$(1 - B)(1 - B^{12})Z_t = (1 - \theta B)(1 - \Theta B^{12})a_t. \tag{8.4.4}$$

TABLE 8.5 Sample ACF for Series W9.

(a) $\hat{\rho}_k$ for $\{Z_t\}(\bar{Z} = 826.47, S_Z = 161.81, n = 132)$

1–12	−.57	.22	.23	.36	.46	.40	.42	.30	.14	.09	.34	.65
St.E.	.09	.11	.12	.12	.13	.14	.15	.16	.16	.16	.16	.17
13–24	.29	−.00	.00	.13	.21	.14	.14	.02	−.11	−.13	.11	.38
St.E.	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19
25–36	.04	−.22	−.19	−.08	−.02	−.06	−.03	−.13	−.24	−.27	−.02	.24
St.E.	.20	.20	.20	.20	.20	.20	.20	.20	.20	.21	.21	.21

(b) $\hat{\rho}_k$ for $\{W_t = (1 - B)Z_t\}(\bar{W} = 2.26, S_W = 148.89, n = 131)$

1–12	−.09	−.42	−.15	.04	.19	−.10	.18	.05	−.14	−.36	−.07	.81
St.E.	.09	.09	.10	.10	.10	.11	.11	.11	.11	.11	.12	.12
13–24	−.08	−.35	−.15	.05	.18	−.09	.17	−.00	−.12	−.31	−.05	.74
St.E.	.16	.16	.16	.16	.16	.17	.17	.17	.17	.17	.17	.17
25–36	−.10	−.34	−.10	−.05	.13	−.09	.16	.11	−.11	−.32	−.02	.68
St.E.	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.21	.21

(c) $\hat{\rho}_k$ for $\{W_t = (1 - B^{12})Z_t\}(\bar{W} = 26.98, S_W = 114.22, n = 120)$

1–12	.80	.66	.59	.48	.41	.39	.36	.31	.22	.12	.01	−.11
St.E.	.09	.14	.16	.18	.19	.20	.20	.21	.21	.21	.21	.21
13–24	−.06	−.04	−.08	−.05	−.05	−.08	−.11	−.19	−.21	−.20	−.25	−.28
St.E.	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22
25–36	−.29	−.30	−.29	−.35	−.38	−.38	−.38	−.34	−.29	−.30	−.23	−.21
St.E.	.23	.23	.23	.24	.24	.25	.25	.26	.26	.26	.26	.27

(d) $\hat{\rho}_k$ for $\{W_t = (1 - B)(1 - B^{12})Z_t\}(\bar{W} = .66, S_W = 71.85, n = 119)$

1–12	−.15	−.16	.11	−.11	−.09	−.00	.03	.09	.01	.04	.02	−.43
St.E.	.09	.09	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
13–24	.09	.13	−.17	.07	.07	−.00	.14	−.14	−.05	.17	−.07	−.06
St.E.	.11	.11	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12
25–36	.01	−.06	.17	−.05	−.09	.01	−.11	−.00	.14	−.21	.12	.06
St.E.	.12	.12	.12	.13	.13	.13	.13	.13	.13	.13	.13	.13

Parameter estimation gives

$$(1 - B)(1 - B^{12})Z_t = \underset{(.088)}{(1 - .33B)}\underset{(.062)}{(1 - .82B^{12})}a_t \tag{8.4.5}$$

with $\hat{\sigma}_a^2 = 3068$. Values of the residual ACF of this modified model as shown in Table 8.7 are all small and exhibit no patterns. For $K = 24$, the value 20.7 of the Q statistic is not significant

TABLE 8.6 Residual ACF, $\hat{\rho}_k$, from the fitted model $(1 - B)(1 - B^{12})Z_t = (1 - .79B^{12})a_t$.

1–12	−.23	−.12	.11	−.12	−.03	.02	.12	−.05	.06	.03	−.14	.01
St.E.	.09	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
13–24	−.03	.07	−.10	.02	.04	.02	.14	−.19	.04	.08	−.08	−.05
St.E.	.10	.10	.10	.10	.11	.11	.11	.11	.11	.11	.11	.11
25–36	−.02	.02	.07	−.02	−.08	−.03	−.05	−.04	.12	−.18	.10	.00
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11	.12	.12

because $\chi^2_{.05}(22) = 33.9$. Thus, the fitted ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model in (8.4.5) is judged adequate for the series.

Forecasting Because the model in (8.4.5) is adequate, we can use it to forecast the future employment figures. As discussed in Chapter 5, for a given forecast origin, say $t = 132$, forecasts can be calculated directly from the difference equation. For Equation (8.4.5) we have

$$Z_{t+l} = Z_{t+l-1} + Z_{t+l-12} - Z_{t+l-13} + a_{t+l} - .33a_{t+l-1} - .82a_{t+l-12} + .27a_{t+l-13}.$$

Thus, the l-step ahead forecast from the time origin $t = 132$ is given by

$$\begin{aligned}\hat{Z}_{132}(l) = {} & \hat{Z}_{132}(l-1) + \hat{Z}_{132}(l-12) - \hat{Z}_{132}(l-13) + E(a_{132+l} \mid Z_{132}, Z_{131}, \ldots) \\ & - .33E(a_{132+l-1} \mid Z_{132}, Z_{131}, \ldots) \\ & - .82E(a_{132+l-12} \mid Z_{132}, Z_{131}, \ldots) \\ & + .27E(a_{132+l-13} \mid Z_{132}, Z_{131}, \ldots),\end{aligned}$$

where

$$\hat{Z}_{132}(j) = Z_{132+j}, \qquad j \leq 0,$$

and

$$E(a_{132+j} \mid Z_{132}, Z_{131}, \cdots) = \begin{cases} \hat{a}_{132+j}, & j \leq 0, \\ 0, & j > 0. \end{cases}$$

TABLE 8.7 Residual ACF, $\hat{\rho}_k$, from the fitted model $(1 - B)(1 - B^{12})Z_t = (1 - .33B)(1 - .82B^{12})a_t$.

1–12	.03	−.10	.04	−.13	−.06	.05	.13	−.00	.07	.00	−.15	−.03
St.E.	.09	.09	.09	.09	.09	.09	.09	.10	.10	.10	.10	.10
Q	.01	1.4	1.5	3.8	4.2	4.6	6.8	6.8	7.5	7.5	10.5	10.6
13–24	−.04	.03	−.08	.01	.06	.07	.12	−.15	.01	.04	−.08	−.06
St.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
Q	10.8	11.0	11.9	11.9	12.4	13.2	15.2	18.7	18.7	19.0	20.1	20.7

To derive the forecast variance, we first rewrite the model in the following AR representation because the model is nonstationary but invertible:

$$\pi(B)Z_t = a_t, \tag{8.4.6}$$

where

$$\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots) = \frac{(1 - B)(1 - B^{12})}{(1 - .33B)(1 - .82B^{12})},$$

and, hence,

$$(1 - \pi_1 B - \pi_2 B^2 - \cdots)(1 - .33B - .82B^{12} + .27B^{13}) = (1 - B)(1 - B^{12}). \tag{8.4.7}$$

By equating the coefficients of B^j on both sides of (8.4.7), we have

$$\begin{cases} \pi_j = (.33)^{j-1}(.67), & 1 \le j \le 11, \\ \pi_{12} = (.33)^{11}(.67) - .82 + 1, & \\ \pi_{13} = (.33)\pi_{12} + .82\pi_1 + .27 - 1, & \\ \pi_j = (.33)\pi_{j-1} + .82\pi_{j-12} - .27\pi_{j-13}, & j \ge 14. \end{cases} \tag{8.4.8}$$

Now, from (5.2.26), the ψ_j weights that are needed for calculating the forecast variance can be easily obtained as follows:

$$\psi_1 = \pi_1 = .67,$$
$$\psi_2 = \pi_2 + \pi_1\psi_1 = (.33)(.67) + (.67)^2 = .67,$$
$$\vdots$$

and

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}\psi_i, \qquad j = 1, \ldots, l - 1,$$

where the π_j weights are given in (8.4.8). By (5.2.9), the variance of the forecast error is

$$\text{Var}[e_{132}(l)] = 3068 \sum_{j=0}^{l-1} \psi_j^2 \tag{8.4.9}$$

and the 99% forecast limits, by (5.2.10), are given by

$$\hat{Z}_{132}(l) \pm 1.96\left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} (3068)^{1/2}. \tag{8.4.10}$$

The first 12 forecasts, i.e., $\hat{Z}_{132}(l)$, for $l = 1, 2, \ldots, 12$, and their standard errors are given in Table 8.8. These forecasts and their 95% forecast limits are also plotted in Figure 8.8. Because the process is nonstationary, the limits become wider as the forecast lead time l becomes larger.

TABLE 8.8 Forecasts for youth employment at the origin of December 1981.

Lead time	Forecasts $\hat{Z}_{132}(l)$	Standard error
1	1068.74	55.39
2	1077.53	66.70
3	1024.00	76.35
4	946.74	84.91
5	927.79	92.69
6	1319.45	99.86
7	1230.29	106.55
8	1024.07	112.84
9	974.65	118.80
10	977.97	124.47
11	1051.58	129.90
12	1035.20	135.11

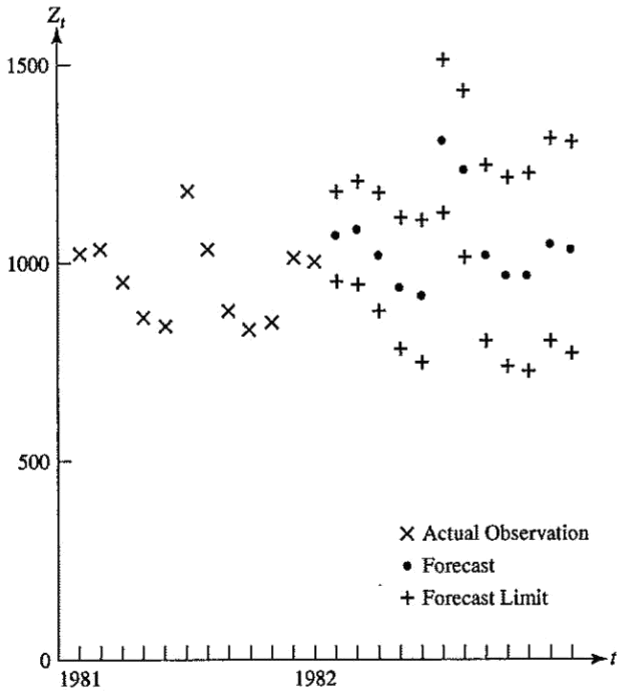


FIGURE 8.8 Forecasts and their 95% confidence limits for youth employment at the origin of December 1981.

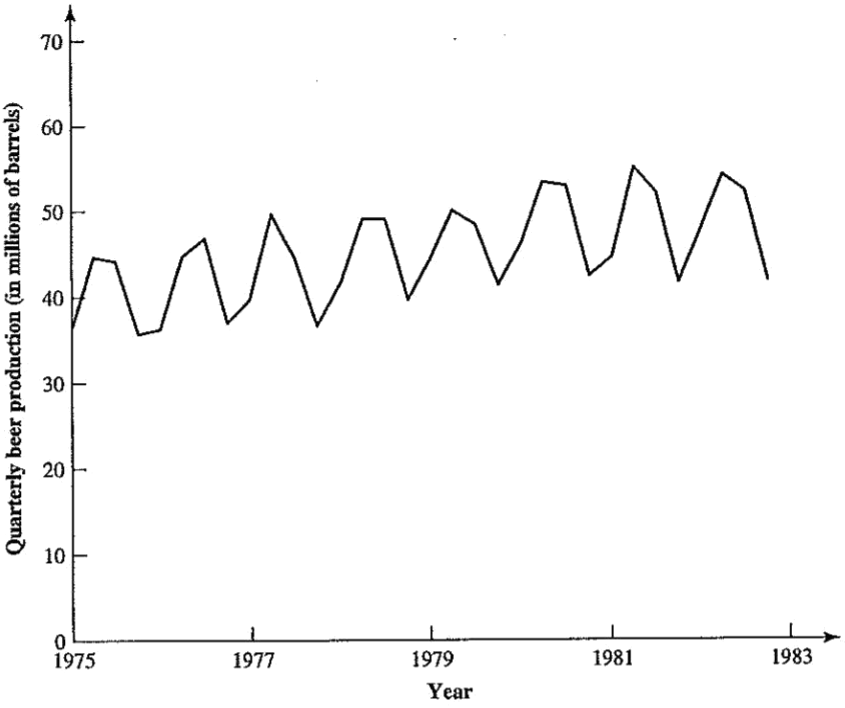


FIGURE 8.9 Quarterly beer production, in millions of barrels, between 1975 and 1982 (Series W10).

EXAMPLE 8.4 Figure 8.9 shows the 32 consecutive quarters of U.S. beer production, in millions of barrels, from the first quarter of 1975 to the fourth quarter of 1982. The series is listed as Series W10 in the appendix. The numbers of sample ACF and PACF values used to identify a seasonal model with a seasonal period s should be at least 3s. Series W10 may be too short. We choose it because sometimes we do need to construct a model from a relatively short series.

The First Tentative Model Based on the Original Series To assess the forecasting performance of a model, in this example, we use only the first 30 observations of the series for model construction. The power transformation analysis of these 30 observations indicates that there is no need for a power transformation. The sample ACF and sample PACF as shown in Table 8.9 demonstrate a strong seasonal pattern of period 4. Although the sample ACF at multiples of 4 are larger than their neighboring values, they are decreasing and not large. It is not evident that we need a seasonal differencing. Thus, we start with a tentative model of the form

$$(1 - \Phi B^4)(Z_t - \mu) = a_t. \tag{8.4.11}$$

The estimation of the above model gives $\hat{\Phi} = .93$ with the standard error of .07. The 95% confidence interval of Φ clearly contains 1 as a possible value. The residual ACF has a significant spike at lag 4 (table not shown); hence, the model is inadequate.

TABLE 8.9 Sample ACF and sample PACF for U.S. quarterly beer production (Series W10).

(a) $\hat{\rho}_k$ for $\{Z_t\}$ ($\bar{Z} = 44.92, S_z = 5.67, n = 30$)				
1–4	.24	−.32	.35	.75
St.E.	.18	.19	.21	.23
5–8	.08	−.32	.24	.51
St.E.	.30	.30	.32	.32
9–12	−.05	−.36	.08	.31
St.E.	.34	.34	.36	.36
13–16	−.14	−.40	−.01	.17
St.E.	.36	.37	.38	.38
(b) $\hat{\phi}_{kk}$ for $\{Z_t\}$				
1–4	.24	−.39	.68	.45
St.E.	.18	.18	.18	.18
5–8	.00	−.25	−.27	−.08
St.E.	.18	.18	.18	.18
9–12	−.06	−.04	−.18	−.12
St.E.	.18	.18	.18	.18
13–16	−.02	−.01	.10	−.13
St.E.	.18	.18	.18	.18

Models Based on the Seasonally Differenced Series As a result of the previous analysis, we consider the seasonally differenced series $W_t = (1 - B^4)Z_t$ with its sample ACF and sample PACF shown in Table 8.10.

That the sample ACF of the series $(1 - B^4)Z_t$ as shown in Table 8.10(a) has only one significant spike at lag 4 suggests the following tentative model:

$$(1 - B^4)Z_t = \theta_0 + (1 - \Theta B^4)a_t, \tag{8.4.12}$$

where θ_0 is included because $\overline{W}/S_{\overline{W}} = 1.36/\left(2.03/\sqrt{26}\right) = 3.42$, which is significant. Parameter estimation gives

$$\underset{\quad}{(1 - B^4)Z_t} = \underset{(.09)}{1.49} + \underset{\quad}{(1 - .87B^4)a_t}, \tag{8.4.13}$$
$$\underset{(.16)}{\quad}$$

and $\hat{\sigma}_a^2 = 2.39$, where values in parentheses below the parameter estimates are the associated standard errors. The residual ACF as shown in Table 8.11 indicates no model inadequacy.

Careful readers may notice, however, that the sample PACF of $(1 - B^4)Z_t$ as shown in Table 8.10(b) also has only one significant spike at lag 4, with its magnitude approximately

TABLE 8.10 Sample ACF and sample PACF for the seasonally differenced series of U.S. quarterly beer production.

(a) $\hat{\rho}_k$ for $\{W_t = (1 - B^4)Z_t\}$ ($\overline{W} = 1.36$, $S_W = 2.03$, $n = 26$)				
1–4	−.14	−.17	.22	−.54
St.E.	.20	.20	.21	.21
5–8	.02	.31	.04	.15
St.E.	.26	.26	.27	.27
9–12	−16	−.21	.02	−.20
St.E.	.27	.28	.29	.29
13–16	.17	.23	−.30	.15
St.E.	.29	.30	.30	.31
(b) $\hat{\phi}_{kk}$ for $\{W_t = (1 - B^4)Z_t\}$				
1–4	−.14	−.19	.17	−.56
St.E.	.20	.20	.20	.20
5–8	−.00	.08	.36	−.08
St.E.	.20	.20	.20	.20
9–12	−.26	−.12	.22	−.26
St.E.	.20	.20	.20	.20
13–16	−.15	−.02	−.07	.10
St.E.	.20	.20	.20	.20

TABLE 8.11 Residual ACF, $\hat{\rho}_k$, from the fitted model $(1 - B^4)Z_t = 1.49 + (1 - .87B^4)a_t$.

1–4	−.12	−.28	.21	−.13
St.E.	.20	.20	.21	.22
5–8	−.12	.14	−.00	.10
St.E.	.22	.23	.23	.23
9–12	−.13	−.10	−.07	−.08
St.E.	.23	.23	.24	.24
13–16	.09	.10	−.26	.16
St.E.	.24	.24	.24	.25

equal to the sample ACF at the same lag. This fact suggests that the following purely seasonal AR process could also be a good tentative model:

$$(1 - \Phi B^4)(1 - B^4)Z_t = \theta_0 + a_t. \tag{8.4.14}$$

TABLE 8.12 Residual ACF, $\hat{\rho}_k$, from the fitted model $(1 + .66B^4)(1 - B^4)\,Z_t = 2.51 + a_t$.

1–4	−.04	−.31	.20	−.02
St.E.	.21	.21	.23	.24
5–8	−.17	.09	.22	−.28
St.E.	.24	.25	.25	.26
9–12	−.17	.07	−.05	−.11
St.E.	.27	.27	.28	.28
13–16	.17	.08	−.30	.11
St.E.	.28	.28	.28	.30

The estimation results are

$$\underset{(.18)}{(1 + .66B^4)}(1 - B^4)Z_t = \underset{(.05)}{2.51} + a_t \tag{8.4.15}$$

with $\hat{\sigma}_a^2 = 3.06$. The residual ACF as shown in Table 8.12 also indicates model adequacy.

Model Selection Based on Forecast Errors It is not uncommon that several adequate models can be found to represent a series. The ultimate choice of a model may depend on the goodness of fit, such as the residual mean square or criteria discussed in Chapter 7. If the main purpose of a model is to forecast future values, then alternative criteria for model selection can be based on forecast errors. Let the l-step ahead forecast error be

$$e_l = Z_{n+l} - \hat{Z}_n(l), \tag{8.4.16}$$

where n is the forecast origin larger than or equal to the length of the series so that the evaluation is based on outsample forecasts. The comparison is usually based on the following summary statistics.

1. The mean percentage error, which is also referred to as bias because it measures forecast bias, is

$$\text{MPE} = \left(\frac{1}{M}\sum_{l=1}^{M}\frac{e_l}{Z_{n+l}}\right)100\%.$$

2. The mean square error is

$$\text{MSE} = \frac{1}{M}\sum_{l=1}^{M}e_l^2.$$

TABLE 8.13 Comparison of the forecasts between models (8.4.13) and (8.4.15).

		Model (8.4.13)		Model (8.4.15)	
Lead time	Actual value	Forecast value	Error	Forecast value	Error
1	52.31	54.38	−2.07	55.26	−2.95
2	41.83	45.37	−3.54	44.71	−2.88
MPE		−6.21%		−6.26%	
MSE		8.41		8.50	
MAE		2.81		2.92	
MAPE		6.21%		6.26%	

3. The mean absolute error is

$$\text{MAE} = \frac{1}{M}\sum_{l=1}^{M}|e_l|.$$

4. The mean absolute percentage error is

$$\text{MAPE} = \left(\frac{1}{M}\sum_{l=1}^{M}\left|\frac{e_l}{Z_{n+l}}\right|\right)100\%.$$

For illustration, we calculate the one-step ahead and two-step ahead forecasts $\hat{Z}_{30}(l)$ for $l = 1$ and 2 from the forecast origin 30 for the two competing models (8.4.13) and (8.4.15). The forecast errors and the corresponding summary statistics are shown in Table 8.13. The results indicate that Equation (8.4.13) slightly outperforms Equation (8.4.15) in terms of forecasts.

EXERCISES

8.1 Find the ACF for the following seasonal models:

(a) $Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^s)a_t$,

(b) $(1 - \Phi_1 B^s)Z_t = (1 - \theta_1 B)a_t$,

(c) $(1 - \Phi_1 B^s)(1 - \phi_1 B)Z_t = a_t$.

8.2 Identify appropriate time series models for the data from the following sample autocorrelations:

(a) $n = 56$, data $= Z_t$, $\nabla Z_t = (1 - B)Z_t$, $\nabla\nabla_4 Z_t = (1 - B)(1 - B^4)Z_t$

k	1	2	3	4	5	6	7	8	9	10	$\overline{W}$	S_W^2
$\hat{\rho}_Z(k)$	.92	.83	.81	.80	.71	.63	.60	.58	.50	.42	1965.6	376.6
$\hat{\rho}_{\nabla Z}(k)$	−.05	−.86	.04	.79	−.02	−.77	.00	.78	−.07	−.75	22.1	102.9
$\hat{\rho}_{\nabla\nabla_4 Z}(k)$	−.40	−.11	.43	−.61	.22	.15	−.26	.15	.01	−.10	−.16	53.77

where W is used to refer to the corresponding series Z_t, ∇Z_t, or $\nabla\nabla_4 Z_t$.

(b) $n = 102$, data $= (1 - B)(1 - B^{12})\ln Z_t$, $\overline{W} = .24$, $s_W^2 = 102.38$.

k	$\hat{\rho}_k$											
1–12	−.39	−.24	.17	.21	−.27	−.03	.26	−.10	.20	.07	.44	−.58
13–24	−.09	.17	.01	−.24	.16	.04	−.12	−.01	.11	.08	.33	.28
25–36	.01	−.14	−.02	.18	−.13	.04	−.01	.10	−.13	−.09	.27	−.22

8.3 Consider the following model:

$$(1 - B^{12})(1 - B)Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$$

with $\theta_1 = .2$, $\Theta_1 = .8$, and $\sigma_a^2 = 1$.

(a) Express the model in terms of the AR form. Compute and plot the π weights.

(b) Compute and plot the ψ weights that are needed to evaluate the forecast variance.

(c) Find the forecasts and their 95% forecast limits for the next 12 periods.

8.4 Consider the following model:

$$(1 - \Phi_1 B^4)(1 - \phi_1 B)(1 - B^4)(1 - B)Z_t = a_t.$$

(a) Find the one-step ahead forecast in terms of past observations.

(b) Find the eventual forecast function.

8.5 Consider the number of visitors (in thousands) to a Colorado ski resort given as follows:

Year	Winter	Spring	Summer	Fall
1981	33.63	36.46	41.18	43.16
1982	46.45	50.63	54.41	58.66
1983	62.52	65.55	69.62	72.92
1984	74.64	80.31	80.97	87.75
1985	88.07	94.00	96.16	96.98
1986	103.90	107.77	110.42	114.91

(a) Plot the series, and examine carefully the trend and seasonal phenomenon contained in the data.

(b) Give a preliminary evaluation of the series using Buys-Ballot table.

(c) Fit a proper regression model such as Equation (8.2.8) to the data using only observations between 1981 and 1985.

(d) Fit a seasonal ARIMA model to the data using only observations between 1981 and 1985.

(e) Calculate the forecasts for the next 4 periods using the models obtained in parts (c) and (d); compare two models based on forecast errors.

8.6 Consider the following U.S. liquor sales (in millions of dollars):

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1970	580	514	555	563	627	596	632	639	577	611	639	875
1971	650	594	650	668	712	731	779	712	708	738	758	1073
1972	669	652	743	709	751	774	803	760	749	757	779	1066
1973	734	707	785	762	838	876	878	871	807	834	877	1236
1974	789	744	827	831	895	889	955	983	976	929	989	1294
1975	860	799	899	866	1016	978	1042	1026	944	1002	1009	1368
1976	908	849	916	958	1008	1033	1129	1019	984	1045	1049	1459
1977	910	927	981	1011	1041	1080	1138	1072	1033	1072	1111	1591
1978	950	932	1049	1021	1097	1151	1194	1174	1160	1135	1209	1692
1979	1071	1044	1158	1122	1209	1334	1360	1368	1297	1283	1375	1974
1980	1294	1258	1301	1297	1425	1378	1429	1452	1305	1377	1439	1958

(a) Build a seasonal ARIMA model for the series.

(b) Forecast the next 12 observations, and find their 95% forecast limits.

8.7 Consider the following U.S. personal consumption of gasoline and oil (in billions of dollars) between 1967 and 1982:

Year	I	II	III	IV
1967	16.6	16.9	17.1	17.5
1968	18.1	18.3	19.0	19.1
1969	19.8	20.8	20.9	21.4
1970	21.8	22.3	22.5	23.1
1971	23.5	23.4	24.0	24.6
1972	24.8	24.7	25.5	26.6
1973	27.6	27.9	28.4	30.6
1974	32.4	36.9	38.2	38.9
1975	38.6	39.3	41.3	42.4
1976	42.8	43.0	44.1	45.9
1977	47.2	48.6	48.3	48.5
1978	49.3	49.9	51.5	54.3
1979	57.7	62.2	70.5	75.9
1980	80.7	84.4	85.1	88.9
1981	94.1	95.1	94.6	94.7
1982	93.4	88.6	89.9	89.6

(a) Build a seasonal ARIMA model for the series.

(b) Forecast the next 4 observations and find their associated 95% forecast limits.

8.8 Consider the following monthly employment figures for young men between ages 16 and 19 in the United States from January 1982 to November 2002 (in thousands):

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1982	1052	1076	1039	999	1010	1306	1232	1064	1024	1035	1137	1113
1983	1071	1055	1059	943	925	1310	1175	1025	898	869	862	847
1984	856	831	830	728	719	1025	995	713	748	745	763	792
1985	814	816	756	678	729	1001	1033	754	721	845	763	762
1986	721	800	750	736	770	1044	926	752	747	703	724	676
1987	755	784	752	684	759	879	776	726	654	674	670	671
1988	690	654	704	572	623	844	832	647	643	649	569	583
1989	767	690	578	559	647	848	694	593	590	637	676	613
1990	663	650	605	576	650	811	784	645	627	651	685	665
1991	720	700	776	636	800	992	905	673	696	683	729	698
1992	752	816	820	645	791	1164	958	771	809	660	776	713
1993	754	781	775	757	775	1055	915	725	665	683	707	628
1994	793	735	731	755	784	952	906	696	665	647	575	642
1995	673	743	654	699	749	959	916	755	685	712	696	684
1996	744	695	734	708	734	904	962	690	625	696	681	624
1997	741	710	723	674	641	972	864	688	622	624	594	478
1998	680	678	686	535	658	889	836	644	687	668	650	615
1999	736	618	645	574	575	817	737	543	582	573	600	597
2000	646	662	545	541	529	824	725	639	536	504	538	564
2001	684	662	579	561	555	859	757	663	606	642	692	642
2002	689	702	712	636	668	974	883	753	657	552	669	

(a) Build two alternative seasonal models for the series using observations from January 1982 to December 2001.

(b) Compare your models with the model discussed in Example 8.3.

(c) Forecast the next 12 observations, and comment on a model selection in terms of forecasting performance.

9

Testing for a Unit Root

As discussed in Chapter 4, a nonstationary time series can often be reduced to a stationary series by differencing. To decide whether to take the difference, so far we have relied on our own judgement after examining the plot of the series and the pattern of its autocorrelation function. For a formal treatment, we present several unit root tests for nonseasonal and seasonal time series in this chapter. Empirical examples are used to illustrate the tests.

9.1 Introduction

In Chapter 4, we introduced a homogeneous nonstationary time series that can be transformed to a stationary series by taking a proper order of differencing. Specifically, if a time series Z_t is nonstationary and its dth difference, $\Delta^d Z_t = (1 - B)^d Z_t$, is stationary and also can be represented as a stationary ARMA(p, q) process, then the Z_t is said to follow an ARIMA(p, d, q) model. The Z_t in this case is referred to as an integrated process or series.

In model identification discussed in Chapter 6, the decision for differencing is based on an informal procedure relying on a subjective judgment of whether the sample ACF of a time series fails to decay rapidly. If we decided that was the case, then we would fit the series with an ARIMA model. Because an ARIMA model can be regarded as a more general ARMA model with a unit root in the associated AR polynomial, for a more formal procedure we can introduce tests to test for a unit root in our model fitting. Most of these tests can be performed by simply using the output from a standard multiple regression procedure.

9.2 Some Useful Limiting Distributions

To develop a desired test statistic and discuss its properties, we will use the approach employed by Chan and Wei (1988) and begin with the following useful notions. Alternative proofs of some results presented in this chapter can also be found in Fuller (1996, Chapter 10).

A process, $W(t)$, is continuous if its time index t belongs to an interval of a real line. For distinction, we often write a continuous process as $W(t)$ rather than W_t. A process $W(t)$ is

said to be a Wiener process (also known as Brownian motion process) if it contains the following properties:

1. $W(0) = 0$;
2. $E[W(t)] = 0$;
3. $W(t)$ follows a nondegenerate normal distribution for each t; and
4. $W(t)$ has independent increments, i.e. $[W(t_2) - W(t_1)]$ and $[W(t_4) - W(t_3)]$ are independent for any nonoverlapping time intervals (t_1, t_2) and (t_3, t_4).

With no loss of generality, we consider t in the closed interval between 0 and 1, i.e., $t \in [0, 1]$. Furthermore, if for any t, $W(t)$ is distributed as $N(0, t)$, then the process is also called standard Brownian motion.

Given the i.i.d. random variables a_t, for $t = 1, \ldots, n$, with mean 0 and variance σ_a^2, define

$$F_n(x) = \begin{cases} 0, & \text{for } 0 \le x < 1/n, \\ a_1/\sqrt{n}, & \text{for } 1/n \le x < 2/n, \\ (a_1 + a_2)/\sqrt{n}, & \text{for } 2/n \le x < 3/n, \\ \quad \vdots & \\ (a_1 + a_2 + \cdots + a_n)/\sqrt{n}, & \text{for } x = 1. \end{cases}$$

That is,

$$F_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[xn]} a_t, \tag{9.2.1}$$

where $x \in [0, 1]$ and $[xn]$ represents the integer portion of (xn). Now,

$$F_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[xn]} a_t = \frac{\sqrt{[xn]}}{\sqrt{n}} \frac{1}{\sqrt{[xn]}} \sum_{t=1}^{[xn]} a_t. \tag{9.2.2}$$

Because, as $n \longrightarrow \infty$, $\left[\sqrt{[xn]}/\sqrt{n}\right] \longrightarrow \sqrt{x}$ and $\sum_{t=1}^{[xn]} a_t/\sqrt{[xn]}$ converges to a $N(0, \sigma_a^2)$ random variable by the central limit theorem, it follows that $F_n(x)$ converges in distribution to $\sqrt{x}N(0, \sigma_a^2) = N(0, x\sigma_a^2)$ as $n \longrightarrow \infty$, which will be denoted by $F_n(x) \xrightarrow{D} N(0, x\sigma_a^2)$. In the following discussion, we also use the notation $X_n \xrightarrow{P} X$ to indicate the convergence of X_n to X in probability as $n \longrightarrow \infty$.

It can be easily seen that the limit of the sequence of the random variable $F_n(x)/\sigma_a$ can be described by a Wiener process, i.e.,

$$\frac{F_n(x)}{\sigma_a} \xrightarrow{D} W(x) \tag{9.2.3}$$

or

$$F_n(x) \xrightarrow{D} \sigma_a W(x), \tag{9.2.4}$$

where $W(x)$ at time $t = x$ follows an $N(0, x)$. Specifically,

$$F_n(1) = \sum_{t=1}^{n} \frac{a_t}{\sqrt{n}} \xrightarrow{D} \sigma_a W(1), \tag{9.2.5}$$

where $W(1)$ follows an $N(0, 1)$.

Let $Z_t = a_1 + a_2 + \cdots + a_t$ and $Z_0 = 0$. We can rewrite $F_n(x)$ in (9.2.1) as

$$F_n(x) = \begin{cases} 0, & \text{for } 0 \le x < 1/n, \\ Z_1/\sqrt{n}, & \text{for } 1/n \le x < 2/n, \\ Z_2/\sqrt{n}, & \text{for } 2/n \le x < 3/n, \\ \vdots & \\ Z_n/\sqrt{n}, & \text{for } x = 1. \end{cases} \tag{9.2.6}$$

Then, (9.2.5) implies that

$$\frac{Z_n}{\sqrt{n}} \xrightarrow{D} \sigma_a W(1) \tag{9.2.7}$$

and

$$\frac{Z_n^2}{n} \xrightarrow{D} \sigma_a^2 \big[W(1)\big]^2. \tag{9.2.8}$$

Also, it is clear that the integral $\int_0^1 F_n(x)\,dx$ is simply the sum of the area defined in (9.2.6), i.e.,

$$\int_0^1 F_n(x)\,dx = \frac{1}{n}\frac{Z_1}{\sqrt{n}} + \frac{1}{n}\frac{Z_2}{\sqrt{n}} + \cdots + \frac{1}{n}\frac{Z_{n-1}}{\sqrt{n}} = n^{-3/2}\sum_{t=1}^{n} Z_{t-1}.$$

Thus, by (9.2.4),

$$n^{-3/2}\sum_{t=1}^{n} Z_{t-1} = \int_0^1 F_n(x)\,dx \xrightarrow{D} \sigma_a \int_0^1 W(x)\,dx. \tag{9.2.9}$$

Similarly,

$$\int_0^1 \big[F_n(x)\big]^2 dx = \frac{1}{n}\left(\frac{Z_1}{\sqrt{n}}\right)^2 + \frac{1}{n}\left(\frac{Z_2}{\sqrt{n}}\right)^2 + \cdots + \frac{1}{n}\left(\frac{Z_{n-1}}{\sqrt{n}}\right)^2 = n^{-2}\sum_{t=1}^{n} Z_{t-1}^2$$

and

$$n^{-2}\sum_{t=1}^{n} Z_{t-1}^2 = \int_0^1 \big[F_n(x)\big]^2 dx \xrightarrow{D} \sigma_a^2 \int_0^1 \big[W(x)\big]^2 dx. \tag{9.2.10}$$

Next,

$$Z_t^2 = (Z_{t-1} + a_t)^2 = Z_{t-1}^2 + 2Z_{t-1}a_t + a_t^2,$$
$$Z_{t-1}a_t = \tfrac{1}{2}\big[Z_t^2 - Z_{t-1}^2 - a_t^2\big],$$

and summing from 1 to n gives

$$\sum_{t=1}^{n} Z_{t-1}a_t = \frac{1}{2}[Z_n^2 - Z_0^2] - \frac{1}{2}\sum_{t=1}^{n} a_t^2 = \frac{1}{2}Z_n^2 - \frac{1}{2}\sum_{t=1}^{n} a_t^2.$$

Therefore,

$$n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t = \frac{1}{2}\left[\frac{Z_n^2}{n}\right] - \frac{1}{2}\left[\sum_{t=1}^{n}\frac{a_t^2}{n}\right]$$
$$\xrightarrow{D} \frac{1}{2}\sigma_a^2\big[W(1)\big]^2 - \frac{1}{2}\sigma_a^2 = \frac{1}{2}\sigma_a^2\big\{\big[W(1)\big]^2 - 1\big\}, \tag{9.2.11}$$

which follows from (9.2.8) and that $\left[\sum_{t=1}^{n} a_t^2/n\right] \xrightarrow{P} \sigma_a^2$.

9.3 Testing for a Unit Root in the AR(1) Model

The simplest model that may contain a unit root is the AR(1) model. We discuss three cases for this model: a model without a constant term, a model with a constant term, and a model with a linear time trend.

9.3.1 Testing the AR(1) Model without a Constant Term

For the simple AR(1) model

$$Z_t = \phi Z_{t-1} + a_t \tag{9.3.1}$$

with $t = 1, \ldots, n$ and $Z_0 = 0$, where the a_t is the Gaussian $N(0, \sigma_a^2)$ white noise process, the unit root test implies a test for a random walk model, $H_0 : \phi = 1$. The alternative is that the series is stationary, i.e., $H_1 : \phi < 1$. Thus, if the test is based on the statistic $\hat{\phi}$ used to estimate ϕ, the null hypothesis will be rejected if $\hat{\phi}$ is much less than 1 or if $(\hat{\phi} - 1)$ is too negative. With the assumed Z_0 and following (7.4.4) we obtain the OLS estimator of ϕ in (9.3.1) as

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1}Z_t}{\sum_{t=1}^{n} Z_{t-1}^2}. \tag{9.3.2}$$

It is tempting to use the normal distribution given in (7.3.26) or the t-distribution given in (7.3.27) to test the above hypothesis. These distributions, however, only hold when $|\phi| < 1$. Under the hypothesis $H_0 : \phi = 1$, we note that

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1}Z_t}{\sum_{t=1}^{n} Z_{t-1}^2} = 1 + \frac{\sum_{t=1}^{n} Z_{t-1}a_t}{\sum_{t=1}^{n} Z_{t-1}^2}.$$

Thus,

$$n(\hat{\phi} - 1) = \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t}{n^{-2}\sum_{t=1}^{n} Z_{t-1}^2} \tag{9.3.3}$$

$$\xrightarrow{D} \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int_0^1 [W(x)]^2 dx}. \tag{9.3.4}$$

The result follows from (9.2.10), (9.2.11) and that under H_0, (9.3.1) becomes a random walk model that can be written as

$$Z_t = a_1 + a_2 + \cdots + a_t. \tag{9.3.5}$$

Recall from (7.3.26) and (7.3.27) that when $|\phi| < 1$, we can test the hypothesis $H_0 : \phi = \phi_0$ using either the normal or the t-distribution. When $\phi = 1$, however, we can no longer use these distributions. In fact, in (9.3.4), $W(1)$ is known to be an $N(0, 1)$ random variable. Hence, $[W(1)]^2$ follows the chi-square distribution with one degree of freedom, i.e., $\chi^2(1)$. The probability that a $\chi^2(1)$ random variable is less than 1 is .6827. Because the denominator is always positive, the probability that $n(\hat{\phi} - 1) < 0$ approaches .6827 as n becomes large. The OLS estimator $\hat{\phi}$ clearly underestimates the true value in this case, and the limiting distribution of $n(\hat{\phi} - 1)$ is clearly skewed to the left. As a result, the null hypothesis is rejected only when $n(\hat{\phi} - 1)$ is really too negative, i.e., much less than the rejection limit when the normal or the t-distribution is used.

For any given probability density function $f(x)$, if we wish to calculate

$$I(f) = \int_0^1 f(x)dx$$

and the integration cannot be solved algebraically, we can use the Monte Carlo method to approximate the integral by a sum in the following way. First, generate n independent uniform random variables $X_1, \ldots, X_n$ on $[0, 1]$. Then compute

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

By the law of large numbers, this sum converges to $E[f(X)]$ as n becomes large, which is simply

$$E\big[f(X)\big] = \int_0^1 f(x)dx = I(f).$$

The percentiles for the empirical distribution of $n(\hat{\phi} - 1)$ given in (9.3.4) were constructed by Dickey (1976) using the Monte Carlo method and reported in Fuller (1996, p. 641). They are reproduced in Table F(a) of the appendix.

The commonly used t-statistic under H_0 is

$$T = \frac{\hat{\phi} - 1}{S_{\hat{\phi}}} = \frac{\hat{\phi} - 1}{\left[\hat{\sigma}_a^2\left(\sum_{t=1}^{n} Z_{t-1}^2\right)^{-1}\right]^{1/2}}, \tag{9.3.6}$$

where

$$S_{\hat{\phi}} = \left[\hat{\sigma}_a^2\left(\sum_{t=1}^{n} Z_{t-1}^2\right)^{-1}\right]^{1/2} \quad \text{and} \quad \hat{\sigma}_a^2 = \sum_{t=1}^{n} \frac{(Z_t - \hat{\phi} Z_{t-1})^2}{(n-1)}.$$

It can be easily seen from (9.3.3) that

$$\begin{aligned} T &= \frac{n(\hat{\phi} - 1)\left[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2}}{[\hat{\sigma}_a^2]^{1/2}} \\ &= \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1} a_t}{\left[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2}[\hat{\sigma}_a^2]^{1/2}} \\ &\xrightarrow{D} \frac{\frac{1}{2}\sigma_a^2\left\{\left[W(1)\right]^2 - 1\right\}}{\left\{\sigma_a^2 \int_0^1 \left[W(x)\right]^2 dx\right\}^{1/2}[\sigma_a^2]^{1/2}} = \frac{\frac{1}{2}\left\{\left[W(1)\right]^2 - 1\right\}}{\left\{\int_0^1 \left[W(x)\right]^2 dx\right\}^{1/2}}, \end{aligned} \tag{9.3.7}$$

which follows from (9.2.10), (9.2.11) and that $\hat{\sigma}_a^2$ is a consistent estimator of $\hat{\sigma}_a^2$.

The nature of the distribution of T is similar to that of $n(\hat{\phi} - 1)$. We reject H_0 if T is too large negatively. The percentiles for the empirical distribution of T were also constructed by Dickey (1976) using the Monte Carlo method and reported in Fuller (1996, p. 642). They are reproduced in Table G(a) of the appendix.

The use of the assumed initial value $Z_0 = 0$ is purely for the convenience of deriving a limiting distribution. In actual data analysis, we normally use only real data points; hence (9.3.1) is computed only for $t = 2, \ldots, n$. As a result, we often use only $(n - 1)$ sums of squares and products when estimating the regression parameter in (9.3.2). With that in mind, the test statistic $n(\hat{\phi} - 1)$ given in (9.3.3) can be replaced by $(n - 1)(\hat{\phi} - 1)$. It is clear that the large sample results of $(n - 1)(\hat{\phi} - 1)$ and $n(\hat{\phi} - 1)$ are exactly the same.

EXAMPLE 9.1 As shown in Table 7.3 of Chapter 7, we fit Series W5 with two alternative models, a random walk model and an AR(1) model with a large value of ϕ. Now let us more rigorously test for a unit root. To facilitate the test, we consider the series Z_t, which is the mean adjusted data set of the 71 annual cancer death rates of Pennsylvanians between 1930 and 2000. That is, $Z_t = Y_t - \overline{Y}$, where Y_t is the annual cancer death rate and $\overline{Y}$ is the mean of the series. To test whether the underlying process contains a unit root, we fit the following regression model

$$Z_t = \phi Z_{t-1} + a_t. \tag{9.3.8}$$

The OLS regression equation is

$$\hat{Z}_t = \underset{(.009)}{.985Z_{t-1}},$$

where the value given in the parenthesis is the standard error of $\hat{\phi}$. With $n = 71$, we have $(n-1)(\hat{\phi}-1) = 70(.985-1) = -1.05$, which is not less than the 5% critical value -7.7 for $n = 50$ or -7.9 for $n = 100$ given in Table F(a). Thus, the null hypothesis, $H_0: \phi = 1$, cannot be rejected, and we conclude that the underlying process for the data contains a unit root.

The t statistic is given by

$$T = \frac{\hat{\phi}-1}{S_{\hat{\phi}}} = \frac{.985-1}{.009} = -1.67,$$

which again is not less than the 5% critical value -1.95 given in Table G(a). Hence, we cannot reject the null hypothesis, $H_0: \phi = 1$, and obtain the same conclusion that the underlying model contains a unit root.

9.3.2 Testing the AR(1) Model with a Constant Term

The mean, $\mu = E(Z_t)$, of the AR(1) process given in (9.3.1) was assumed to be 0. For a nonzero mean AR(1) process, we consider the following model with a constant term

$$Z_t = \alpha + \phi Z_{t-1} + a_t, \tag{9.3.9}$$

for $t = 1, \ldots, n$, where we recall from (3.4.16) that $\alpha = \mu(1-\phi)$.

First, note that in terms of the matrix form of a standard regression model, we can rewrite (9.3.9) as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & Z_0 \\ \vdots & \vdots \\ 1 & Z_{n-1} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \phi \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

The OLS estimator, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Thus,

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\phi}_\mu \end{bmatrix} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= \begin{bmatrix} n & \sum_{t=1}^{n} Z_{t-1} \\ \sum_{t=1}^{n} Z_{t-1} & \sum_{t=1}^{n} Z_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} Z_t \\ \sum_{t=1}^{n} Z_{t-1}Z_t \end{bmatrix}, \end{aligned}$$

where following Dickey and Fuller (1979), we use the subscript μ in $\hat{\phi}_\mu$ to denote that the estimate is based on (9.3.9), with a constant term related to the mean of the process. Because $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$ and under $H_0 : \phi = 1$, we have $\alpha = 0$, it follows that under the null hypothesis we have

$$\begin{bmatrix} \hat{\alpha} - 0 \\ \hat{\phi}_\mu - 1 \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} Z_{t-1} \\ \sum_{t=1}^{n} Z_{t-1} & \sum_{t=1}^{n} Z_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} a_t \\ \sum_{t=1}^{n} Z_{t-1} a_t \end{bmatrix},$$

or

$$\begin{bmatrix} \sqrt{n} & 0 \\ 0 & n \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\phi}_\mu - 1 \end{bmatrix} = \left\{ \begin{bmatrix} (\sqrt{n})^{-1} & 0 \\ 0 & n^{-1} \end{bmatrix} \begin{bmatrix} n & \sum_{t=1}^{n} Z_{t-1} \\ \sum_{t=1}^{n} Z_{t-1} & \sum_{t=1}^{n} Z_{t-1}^2 \end{bmatrix} \begin{bmatrix} (\sqrt{n})^{-1} & 0 \\ 0 & n^{-1} \end{bmatrix} \right\}^{-1}$$

$$\left\{ \begin{bmatrix} (\sqrt{n})^{-1} & 0 \\ 0 & n^{-1} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{n} a_t \\ \sum_{t=1}^{n} Z_{t-1} a_t \end{bmatrix} \right\}.$$

Thus,

$$\begin{bmatrix} \sqrt{n}\hat{\alpha} \\ n(\hat{\phi}_\mu - 1) \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2} \sum_{t=1}^{n} Z_{t-1} \\ n^{-3/2} \sum_{t=1}^{n} Z_{t-1} & n^{-2} \sum_{t=1}^{n} Z_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} \frac{a_t}{\sqrt{n}} \\ \sum_{t=1}^{n} \frac{Z_{t-1} a_t}{n} \end{bmatrix}$$

$$\xrightarrow{D} \begin{bmatrix} 1 & \sigma_a \int_0^1 W(x)\,dx \\ \sigma_a \int_0^1 W(x)\,dx & \sigma_a^2 \int_0^1 [W(x)]^2\,dx \end{bmatrix}^{-1} \begin{bmatrix} \sigma_a W(1) \\ \frac{1}{2}\sigma_a^2 \{[W(1)]^2 - 1\} \end{bmatrix} \tag{9.3.10}$$

by (9.2.5), (9.2.9), (9.2.10), and (9.2.11). It follows that

$$n(\hat{\phi}_\mu - 1) \xrightarrow{D} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1) \int_0^1 W(x)\,dx}{\int_0^1 [W(x)]^2\,dx - \left[\int_0^1 W(x)\,dx\right]^2}. \tag{9.3.11}$$

The corresponding t statistic is

$$T_\mu = \frac{\hat{\phi}_\mu - 1}{S_{\hat{\phi}_\mu}}, \tag{9.3.12}$$

where from (7.3.19),

$$S_{\hat{\phi}_\mu} = \left\{ \hat{\sigma}_a^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} n & \sum_{t=1}^{n} Z_{t-1} \\ \sum_{t=1}^{n} Z_{t-1} & \sum_{t=1}^{n} Z_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}^{1/2},$$

$$\hat{\sigma}_a^2 = \left\{ \sum_{t=1}^{n} \frac{(Z_t - \hat{\alpha} - \hat{\phi}_\mu Z_{t-1})^2}{n-2} \right\}.$$

For the sampling distribution of T_μ, we first note that

$$n^2 S_{\hat{\phi}}^2 = \frac{\hat{\sigma}_a^2}{n^{-2}\sum_{t=1}^{n} Z_{t-1}^2 - n^{-3}\left[\sum_{t=1}^{n} Z_{t-1}\right]^2};$$

hence, from (9.2.9) and (9.2.10)

$$n^2 S_{\hat{\phi}}^2 \xrightarrow{D} \frac{1}{\int_0^1 [W(x)]^2\,dx - \left[\int_0^1 W(x)\,dx\right]^2}. \tag{9.3.13}$$

Rewriting T_μ as

$$T_\mu = \frac{n(\hat{\phi}_\mu - 1)}{[n^2 S_{\hat{\phi}_\mu}^2]^{1/2}} \tag{9.3.14}$$

and using (9.3.11) and (9.3.13), we have that

$$T_\mu = \frac{n(\hat{\phi}_\mu - 1)}{[n^2 S_{\hat{\phi}_\mu}^2]^{1/2}} \xrightarrow{D} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int_0^1 W(x)\,dx}{\left\{\int_0^1 [W(x)]^2\,dx - \left[\int_0^1 W(x)\,dx\right]^2\right\}^{1/2}}. \tag{9.3.15}$$

Again, the subscript μ is used in T_μ to denote that (9.3.9), with a constant term related to the mean of the series, is used. The empirical percentiles of $n(\hat{\phi}_\mu - 1)$ and T_μ were constructed by Dickey (1976) and reported in Fuller (1996, pp. 641–642). They are reproduced in Table F(b) and Table G(b), respectively.

EXAMPLE 9.2 Let us consider Series W6, the 114 yearly observations of U.S. tobacco production from January 1871 to 1984. To test for a unit root, we fit the following regression model

$$Z_t = \alpha + \phi Z_{t-1} + a_t. \tag{9.3.16}$$

The OLS regression equation is

$$\hat{Z}_t = \underset{(51.46)}{125.34} + \underset{(.0358)}{.9143} Z_{t-1},$$

where the values in the parentheses are the standard errors. With $n = 114$, we have $(n-1)(\hat{\phi}_\mu - 1) = 113(.9143 - 1) = -9.6841$, which is not less than the 5% critical value -13.7 for n $=$ 100 and -13.9 for $n = 250$ given in Table F(b). Thus, the null hypothesis, $H_0 : \phi = 1$, cannot be rejected, and we conclude that the underlying process for the data contains a unit root with $\alpha = 0$.

The t statistic in this case becomes

$$T_\mu = \frac{\hat{\phi}_\mu - 1}{S_{\hat{\phi}_\mu}} = \frac{.9143 - 1}{.0358} = -2.3939,$$

which is not less than the 5% critical value -2.90 for $n = 100$ and -2.88 for $n = 250$ given in Table G(b). Thus, again the null hypothesis, $H_0 : \phi = 1$, cannot be rejected.

9.3.3 Testing the AR(1) Model with a Linear Time Trend

Sometimes a time series process may contain a deterministic linear trend. One example is

$$Z_t = \alpha + \delta t + \phi Z_{t-1} + a_t. \tag{9.3.17}$$

The test statistic based on the OLS estimator of ϕ, denoted by $\hat{\phi}_t$, and its limiting distribution can be derived using the same procedure as in previous cases. The empirical percentiles of the distributions of $n(\hat{\phi}_t - 1)$ and $T_t = (\hat{\phi}_t - 1)/S_{\hat{\phi}_t}$ under the hypothesis that $\phi = 1$ and $\delta = 0$ were constructed by Dickey (1976) and reported in Fuller (1996, pp. 641–642). They are reproduced in Table F(c) and Table G(c), respectively. Again, the subscript t is used in $\hat{\phi}_t$ and T_t to denote that (9.3.17) with a linear time trend is used.

EXAMPLE 9.3 Let us again consider Series W5, the 71 annual cancer death rates of Pennsylvanians between 1930 and 2000. It was shown in Example 9.1 that the underlying process for the series contains a unit root. Recall from Table 7.3 that a random walk model with drift was used to fit the series. As shown in Section 4.2.2, the implication is that the model may contain a deterministic component. Thus, we will now try to fit the series that is presently denoted as Z_t with the model

$$Z_t = \alpha + \delta t + \phi Z_{t-1} + a_t.$$

The OLS regression equation is

$$\hat{Z}_t = \underset{(7.214)}{13.26} + \underset{(.1543)}{.188t} + \underset{(.0698)}{.9013Z_{t-1}}, \tag{9.3.18}$$

where the values in the parentheses are the standard errors. Now the value of the test statistic becomes $(n-1)(\hat{\phi}_t - 1) = 70(.9013 - 1) = -6.909$ which is not less than the 5% critical value -19.7 for $n = 50$ or -20.6 for $n = 100$ given in Table F(c). Thus, the null hypothesis, $H_0 : \phi = 1$, cannot be rejected, and we conclude that the model for the series of annual cancer death rates contains a unit root.

The t statistic in this case becomes

$$T_t = \frac{\hat{\phi}_t - 1}{S_{\hat{\phi}_t}} = \frac{.9013 - 1}{.0698} = -1.41,$$

which is not less than the 5% critical value -3.50 for $n = 50$ or -3.45 for $n = 100$ given in Table G(c). Thus, again the null hypothesis, $H_0 : \phi = 1$, cannot be rejected.

It is important to note the difference between (9.3.9) and (9.3.17). Not rejecting $H_0 : \phi = 1$ in (9.3.9) simply implies that the model contains a unit root because $\phi = 1$ also indicates that $\alpha = 0$. In (9.3.17), however, not rejecting the null hypothesis, $H_0 : \phi = 1$, does not necessarily imply that $\alpha = 0$. Thus, the conclusion in this case is that the model contains a unit root with or without a deterministic term, depending on whether the intercept term is statistically significant. As shown in (9.3.18), the constant term is statistically significant at the 10% level. In a preliminary analysis, a random walk model with drift could be used to fit the series.

9.4 Testing for a Unit Root in a More General Model

Note that the AR(1) model without a constant term in (9.3.1) is reduced under the null hypothesis of a unit root to a random walk model

$$(1 - B)Z_t = a_t,$$

where a_t is the Gaussian white noise process. As shown in (9.3.5), Z_t in this case becomes the sum of t independent and identically distributed random variables, i.e., $Z_t = a_1 + a_2 + \cdots + a_t$. More generally, we may have

$$(1 - B)Z_t = X_t, \tag{9.4.1}$$

where X_t is a nonwhite noise stationary process. For the following discussion, we assume that X_t is a general linear process

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B)a_t. \tag{9.4.2}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, $\psi_0 = 0$, and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$.

To test for a unit root in this general case, we can fit the following OLS regression

$$Z_t = \phi Z_{t-1} + X_t, \tag{9.4.3}$$

and consider the estimator

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1} Z_t}{\sum_{t=1}^{n} Z_{t-1}^2}. \tag{9.4.4}$$

Under the null hypothesis, H_0: $\phi = 1$, we have

$$\hat{\phi} = \frac{\sum_{t=1}^{n} Z_{t-1} Z_t}{\sum_{t=1}^{n} Z_{t-1}^2} = 1 + \frac{\sum_{t=1}^{n} Z_{t-1} X_t}{\sum_{t=1}^{n} Z_{t-1}^2}$$

and

$$n(\hat{\phi} - 1) = \frac{n^{-1} \sum_{t=1}^{n} Z_{t-1} X_t}{n^{-2} \sum_{t=1}^{n} Z_{t-1}^2}. \tag{9.4.5}$$

For the sampling distribution of (9.4.5), we note that by substitution (9.4.1) becomes

$$Z_t = X_1 + X_2 + \cdots + X_t + Z_0. \tag{9.4.6}$$

Now,

$$\begin{aligned}
X_1 &= a_1 + \psi_1 a_0 + \psi_2 a_{-1} + \cdots \\
X_2 &= a_2 + \psi_1 a_1 + \psi_2 a_0 + \psi_3 a_{-1} + \cdots \\
&\vdots \\
X_{t-1} &= a_{t-1} + \psi_1 a_{t-2} + \psi_2 a_{t-3} + \cdots + \psi_{t-2} a_1 + \psi_{t-1} a_0 + \psi_t a_{-1} + \cdots \\
X_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots + \psi_{t-1} a_1 + \psi_t a_0 + \psi_{t+1} a_{-1} + \cdots.
\end{aligned}$$

Hence,

$$\begin{aligned}
X_1 + X_2 + \cdots + X_t = a_t &+ (1 + \psi_1) a_{t-1} + (1 + \psi_1 + \psi_2) a_{t-2} + \cdots \\
&+ (1 + \psi_1 + \psi_2 + \cdots + \psi_{t-1}) a_1 + (\psi_1 + \psi_2 + \cdots + \psi_t) a_0 \\
&+ (\psi_2 + \cdots + \psi_{t+1}) a_{-1} + \cdots
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{j=0}^{\infty}\psi_j\right)(a_t + a_{t-1} + \cdots + a_1) - \sum_{j=0}^{\infty}\psi_{j+1}a_t - \sum_{j=0}^{\infty}\psi_{j+2}a_{t-1} - \cdots \\
&\quad - \sum_{j=0}^{\infty}\psi_{j+t}a_1 - \sum_{j=0}^{\infty}\psi_{j+t+1}a_0 + \sum_{j=0}^{\infty}\psi_{j+1}a_0 \\
&\quad - \sum_{j=0}^{\infty}\psi_{j+t+2}a_{-1} + \sum_{j=0}^{\infty}\psi_{j+2}a_{-1} + \cdots \\
&= \psi(1)[a_1 + a_2 + \cdots + a_t] + Y_t - Y_0
\end{aligned}
$$

where $Y_t = -\sum_{i=0}^{\infty}\left(\sum_{j=0}^{\infty}\psi_{j+1+i}\right)a_{t-i} = \sum_{i=0}^{\infty}\alpha_i a_{t-i}$, and $\alpha_i = -\sum_{j=0}^{\infty}\psi_{j+1+i}$. Thus,

$$
\begin{aligned}
Z_t &= X_1 + X_2 + \cdots + X_t + Z_0 \\
&= \psi(1)[a_1 + a_2 + \cdots + a_t] + Y_t - Y_0 + Z_0.
\end{aligned} \tag{9.4.7}
$$

Because

$$
\begin{aligned}
\sum_{j=0}^{\infty}|\alpha_j| &= |\alpha_0| + |\alpha_1| + |\alpha_2| + \cdots \\
&= |\psi_1 + \psi_2 + \psi_3 + \cdots| + |\psi_2 + \psi_3 + \psi_4 + \cdots| + |\psi_3 + \psi_4 + \psi_5 + \cdots| + \cdots \\
&\le \sum_{j=0}^{\infty} j|\psi_j| < \infty,
\end{aligned}
$$

it follows that the process Y_t is stationary. Clearly, $|\psi(1)| < \infty$. Thus, the nonstationary process Z_t given in (9.4.1) is actually the sum of a random walk, a stationary process, and the initial conditions.

Let

$$
H_n(x) = \begin{cases} 0, & \text{for } 0 \le x < 1/n, \\ Z_1/\sqrt{n}, & \text{for } 1/n \le x < 2/n, \\ Z_2/\sqrt{n}, & \text{for } 2/n \le x < 3/n, \\ \vdots & \\ Z_n/\sqrt{n}, & \text{for } x = 1, \end{cases} \tag{9.4.8}
$$

where $Z_t = \psi(1)[a_1 + a_2 + \cdots + a_t] + Y_t - Y_0 + Z_0$. Because $[Y_n - Y_0 + Z_0]/\sqrt{n} \xrightarrow{P} 0$, comparing (9.4.8) with (9.2.1) or (9.2.6), we see from (9.2.4) that under (9.4.1), the limiting distribution of $H_n(x)$ is given by

$$
H_n(x) \xrightarrow{D} \sigma_a \psi(1) W(x). \tag{9.4.9}
$$

Specifically,

$$\frac{Z_n}{\sqrt{n}} = H_n(1) \xrightarrow{D} \sigma_a \psi(1) W(1). \tag{9.4.10}$$

It is easily seen from (9.4.7) that the limiting distributions of $Z_n/\sqrt{n}$ and $\left(\frac{1}{\sqrt{n}}\right)\sum_{j=1}^{n} X_t$ are the same. Thus, we also have

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_t \xrightarrow{D} \sigma_a \psi(1) W(1), \tag{9.4.11}$$

or, equivalently,

$$\sqrt{n}\,\bar{X} \xrightarrow{D} N\big(0, \sigma_a^2 [\psi(1)]^2\big), \tag{9.4.12}$$

which is often known as the central limit theorem for stationary processes.

It follows similarly from (9.2.9) and (9.2.10) that

$$n^{-3/2} \sum_{t=1}^{n} Z_{t-1} = \int_0^1 H_n(x)\, dx \xrightarrow{D} \sigma_a \psi(1) \int_0^1 W(x) dx, \tag{9.4.13}$$

and

$$n^{-2} \sum_{t=1}^{n} Z_{t-1}^2 = \int_0^1 \big[H_n(x)\big]^2 dx \xrightarrow{D} \big[\sigma_a \psi(1)\big]^2 \int_0^1 \big[W(x)\big]^2 dx. \tag{9.4.14}$$

Because

$$\begin{aligned} Z_t^2 &= (Z_{t-1} + X_t)^2 = Z_{t-1}^2 + 2Z_{t-1}X_t + X_t^2, \\ Z_{t-1}X_t &= \tfrac{1}{2}[Z_t^2 - Z_{t-1}^2 - X_t^2], \end{aligned}$$

summing from 1 to n we get

$$\sum_{t=1}^{n} Z_{t-1}X_t = \frac{1}{2}[Z_n^2 - Z_0^2] - \frac{1}{2}\sum_{t=1}^{n} X_t^2.$$

Hence,

$$n^{-1}\sum_{t=1}^{n} Z_{t-1}X_t = \frac{1}{2}\left[\frac{Z_n^2}{n}\right] - \frac{1}{2}[Z_0^2/n] - \frac{1}{2}\left[\frac{\sum_{t=1}^{n} X_t^2}{n}\right] \xrightarrow{D} \frac{1}{2}\big[\sigma_a \psi(1) W(1)\big]^2 - \frac{1}{2}\sigma_X^2, \tag{9.4.15}$$

which follows from (9.4.10) and that $Z_0^2/n \xrightarrow{P} 0$ and $\left[\sum_{t=1}^{n} X_t^2/n\right] \xrightarrow{P} \sigma_X^2$.

It follows from (9.4.14) and (9.4.15) that

$$n(\hat{\phi} - 1) = \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1}X_t}{n^{-2}\sum_{t=1}^{n} Z_{t-1}^2} \xrightarrow{D} \frac{\frac{1}{2}\left\{\left[\sigma_a\psi(1)W(1)\right]^2 - \sigma_X^2\right\}}{\left[\sigma_a\psi(1)\right]^2 \int_0^1 \left[W(x)\right]^2 dx}. \tag{9.4.16}$$

Now,

$$\frac{\frac{1}{2}\left\{\left[\sigma_a\psi(1)W(1)\right]^2 - \sigma_X^2\right\}}{\left[\sigma_a\psi(1)\right]^2 \int_0^1 \left[W(x)\right]^2 dx} = \frac{\frac{1}{2}\left\{\left[W(1)\right]^2 - 1\right\}}{\int_0^1 \left[W(x)\right]^2 dx} + \frac{\frac{1}{2}\left\{\left[\sigma_a\psi(1)\right]^2 - \sigma_X^2\right\}}{\left[\sigma_a\psi(1)\right]^2 \int_0^1 \left[W(x)\right]^2 dx}.$$

Note that

$$\frac{\frac{1}{2}\left\{\left[\sigma_a\psi(1)\right]^2 - \sigma_X^2\right\}}{n^{-2}\sum_{t=1}^{n} Z_{t-1}^2} = \frac{\frac{1}{2}n^2\sigma_X^2\left\{\left[\sigma_a\psi(1)\right]^2 - \sigma_X^2\right\}}{\sigma_X^2\sum_{t=1}^{n} Z_{t-1}^2} \xrightarrow{D} \frac{\frac{1}{2}\left\{\left[\sigma_a\psi(1)\right]^2 - \sigma_X^2\right\}}{\left[\sigma_a\psi(1)\right]^2 \int_0^1 \left[W(x)\right]^2 dx},$$

which follows from (9.4.14). Further note that $\sigma_X^2/\left[\sum_{t=1}^{n} Z_{t-1}^2\right]$ is the variance of the estimator $\hat{\phi}$, $\hat{\sigma}_{\hat{\phi}}^2$, in the OLS regression in (9.4.3). Hence, following Phillips and Perron (1988), we can use the following modified statistic,

$$n(\hat{\phi} - 1) - \frac{1}{2}\frac{n^2\hat{\sigma}_{\hat{\phi}}^2}{\hat{\sigma}_X^2}\{\hat{\gamma}(1) - \hat{\sigma}_X^2\} \xrightarrow{D} \frac{\frac{1}{2}\left\{\left[W(1)\right]^2 - 1\right\}}{\int_0^1 \left[W(x)\right]^2 dx}, \tag{9.4.17}$$

where $\hat{\sigma}_X^2$ and $\hat{\gamma}(1)$ are consistent estimates of σ_X^2 and $\gamma(1)$, in our test. Here we note from (2.6.8) and (2.6.9) that $\gamma(B) = \sigma_a^2\psi(B)\psi(B^{-1})$ is the autocovariance generating function. Thus,

$$\gamma(1) = \left[\sigma_a\psi(1)\right]^2 = \sum_{k=-\infty}^{\infty} \gamma_k, \tag{9.4.18}$$

where γ_k is the kth autocovariance of the stationary process X_t. With the modified statistic (9.4.17), we can use the same Table F of the appendix for the significance test.

In practice, we can use residual mean square error from the OLS regression fitting in (9.4.3) for $\hat{\sigma}_X^2$, and

$$\hat{\gamma}(1) = \hat{\gamma}_0 + 2\sum_{k=1}^{m}\hat{\gamma}_k, \tag{9.4.19}$$

where the $\hat{\gamma}_k$'s are the sample autocovariances of the OLS regression residuals. The value of m is chosen such that $|\hat{\gamma}_k|$ is negligible for $k > m$.

Similarly, the t statistic from the OLS regression fitting in (9.4.3) can also be modified so that the same Table G of the appendix can be used for the significance test. For more detailed discussion on the procedure of Phillips and Perron, we refer readers to their 1988 paper and to Hamilton (1994).

Alternatively, for any given accuracy, because any stationary process can be approximated by an autoregressive model, Dickey and Fuller (1979) suggest the use of a higher-order AR model to solve the problem.

Suppose that the series is generated from the AR(p) model

$$\Phi_p(B)Z_t = a_t, \tag{9.4.20}$$

where the a_t is the Gaussian $N(0, \sigma_a^2)$ white noise process with $E(a_t^4) < \infty$ and $\Phi_p(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$, which may contain a unit root.

To test for a unit root, we assume that $\Phi_p(B) = \varphi_{p-1}(B)(1 - B)$, where $\varphi_{p-1}(B) = (1 - \varphi_1 B - \cdots - \varphi_{p-1} B^{p-1})$ has roots lying outside the unit circle. This assumption implies that

$$\varphi_{p-1}(B)(1 - B)Z_t = (Z_t - Z_{t-1}) - \sum_{j=1}^{p-1}\varphi_j(Z_{t-j} - Z_{t-j-1}) = a_t. \tag{9.4.21}$$

Thus, testing for a unit root is equivalent to testing $\phi = 1$ in the model

$$Z_t = \phi Z_{t-1} + \sum_{j=1}^{p-1}\varphi_j \Delta Z_{t-j} + a_t, \tag{9.4.22}$$

where $\Delta Z_{t-j} = (Z_{t-j} - Z_{t-j-1})$. In terms of the matrix form, we can rewrite (9.4.22) as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_n \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} Z_p & \Delta Z_p & \cdots & \Delta Z_2 \\ Z_{p+1} & \Delta Z_{p+1} & \cdots & \Delta Z_3 \\ \vdots & \vdots & & \vdots \\ Z_{n-1} & \Delta Z_{n-1} & \cdots & \Delta Z_{n-p+1} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \phi \\ \varphi_1 \\ \vdots \\ \varphi_{p-1} \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} a_{p+1} \\ a_{p+2} \\ \vdots \\ a_n \end{bmatrix}.$$

The OLS estimator, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Because $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$, it follows that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$. Thus, under the null hypothesis of a unit root, we have

$$\begin{bmatrix} \hat{\phi} - 1 \\ \hat{\varphi}_1 - \varphi_1 \\ \vdots \\ \hat{\varphi}_{p-1} - \varphi_{p-1} \end{bmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

$$= \begin{bmatrix} \sum_{t=p+1}^{n} Z_{t-1}^2 & \sum_{t=p+1}^{n} Z_{t-1}\,\Delta Z_{t-1} & \cdots & \sum_{t=p+1}^{n} Z_{t-1}\,\Delta Z_{t-p+1} \\ \sum_{t=p+1}^{n} Z_{t-1}\,\Delta Z_{t-1} & \sum_{t=p+1}^{n} (\Delta Z_{t-1})^2 & \cdots & \sum_{t=p+1}^{n} \Delta Z_{t-1}\,\Delta Z_{t-p+1} \\ \vdots & \vdots & & \vdots \\ \sum_{t=p+1}^{n} Z_{t-1}\,\Delta Z_{t-p+1} & \sum_{t=p+1}^{n} \Delta Z_{t-1}\,\Delta Z_{t-p+1} & \cdots & \sum_{t=p+1}^{n} (\Delta Z_{t-p+1})^2 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \sum_{t=p+1}^{n} Z_{t-1} a_t \\ \sum_{t=p+1}^{n} \Delta Z_{t-1} a_t \\ \vdots \\ \sum_{t=p+1}^{n} \Delta Z_{t-p+1} a_t \end{bmatrix}. \tag{9.4.23}$$

Let

$$\boldsymbol{T} = \begin{bmatrix} n & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{n} & \ddots & \vdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \vdots & \ddots & \sqrt{n} & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{n} \end{bmatrix}. \tag{9.4.24}$$

We note that

$$\boldsymbol{T}\begin{bmatrix} \hat{\phi} - 1 \\ \hat{\varphi}_1 - \varphi_1 \\ \vdots \\ \hat{\varphi}_{p-1} - \varphi_{p-1} \end{bmatrix} = \left\{\boldsymbol{T}^{-1}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{T}^{-1}\right\}^{-1}\boldsymbol{T}^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}. \tag{9.4.25}$$

Hence,

$$\begin{bmatrix} n(\hat{\phi}-1) \\ \sqrt{n}(\hat{\varphi}_1-\varphi_1) \\ \vdots \\ \sqrt{n}(\hat{\varphi}_{p-1}-\varphi_{p-1}) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\sum_{t=p+1}^{n} Z_{t-1}^2}{n^2} & \frac{\sum_{t=p+1}^{n} Z_{t-1}\Delta Z_{t-1}}{n^{3/2}} & \cdots & \frac{\sum_{t=p+1}^{n} Z_{t-1}\Delta Z_{t-p+1}}{n^{3/2}} \\ \frac{\sum_{t=p+1}^{n} Z_{t-1}\Delta Z_{t-1}}{n^{3/2}} & \frac{\sum_{t=p+1}^{n} (\Delta Z_{t-1})^2}{n} & \cdots & \frac{\sum_{t=p+1}^{n} \Delta Z_{t-1}\Delta Z_{t-p+1}}{n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\sum_{t=p+1}^{n} Z_{t-1}\Delta Z_{t-p+1}}{n^{3/2}} & \frac{\sum_{t=p+1}^{n} \Delta Z_{t-1}\Delta Z_{t-p+1}}{n} & \cdots & \frac{\sum_{t=p+1}^{n} (\Delta Z_{t-p+1})^2}{n} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \frac{\sum_{t=p+1}^{n} Z_{t-1}a_t}{n} \\ \frac{\sum_{t=p+1}^{n} \Delta Z_{t-1}a_t}{\sqrt{n}} \\ \vdots \\ \frac{\sum_{t=p+1}^{n} \Delta Z_{t-p+1}a_t}{\sqrt{n}} \end{bmatrix} \tag{9.4.26}$$

From (9.4.21), it is clear that under the null hypothesis we have

$$(1-B)Z_t = \Delta Z_t = X_t = \psi(B)a_t, \tag{9.4.27}$$

where $\psi(B) = 1/\varphi_{p-1}(B)$ is assumed to satisfy stationarity and the conditions given in (9.4.2). Thus, from (9.4.14) and (9.4.15), we have

$$\frac{\sum_{t=1}^{n} Z_{t-1}^2}{n^2} \xrightarrow{D} \left[\sigma_a \psi(1)\right]^2 \int_0^1 \left[W(x)\right]^2 dx \tag{9.4.28}$$

and

$$\frac{\sum_{t=p+1}^{n} Z_{t-1}\Delta Z_{t-j}}{n^{3/2}} = \frac{n^{-1}\sum_{t=p+1}^{n} Z_{t-1}X_{t-j}}{\sqrt{n}} \xrightarrow{P} 0, \tag{9.4.29}$$

for $j = 1, \ldots, (p-1)$. Furthermore, from (9.4.7), we have

$$\begin{aligned}\frac{\sum_{t=p+1}^{n} Z_{t-1}a_t}{n} &= \frac{1}{n}\sum_{t=p+1}^{n}\left[\psi(1)(a_1 + \cdots + a_{t-1}) + Y_{t-1} - Y_0 + Z_0\right]a_t \\ &= \psi(1)\frac{1}{n}\sum_{t=p+1}^{n}(a_1 + \cdots + a_{t-1})a_t + \frac{1}{n}\sum_{t=p+1}^{n}(Y_{t-1} - Y_0 + Z_0)a_t \\ &\xrightarrow{D} \psi(1)\frac{1}{2}\sigma_a^2\left\{\left[W(1)\right]^2 - 1\right\}, \end{aligned} \tag{9.4.30}$$

where we use (9.2.11) and that $\sum_{t+p+1}^{n}(Y_{t-1} - Y_0 + Z_0)a_t/n \xrightarrow{P} 0$.

Applying (9.4.27), (9.4.28), (9.4.29), and (9.4.30) to (9.4.26), we obtain the following interesting results:

1. Asymptotically, the estimators of ϕ and φ's are independent as shown in (9.4.29).
2. The φ_j's are parameters on the stationary regressors $(\Delta Z_{t-1}, \ldots, \Delta Z_{t-p+1})$. The limiting distribution of their OLS estimators $\hat{\varphi}_j$'s obtained from (9.4.22) is the same as the standard asymptotic distribution for the OLS estimators obtained by regressing the differenced series ΔZ_t on $\Delta Z_{t-1}, \ldots, \Delta Z_{t-p+1}$. Hence, they can be tested using the standard t statistics provided in the regression output.
3. The parameter ϕ relates to the unit root test, and asymptotically by (9.4.28) and (9.4.30).

$$n(\hat{\phi} - 1) \xrightarrow{D} \frac{\psi(1)\frac{1}{2}\sigma_a^2\left\{\left[W(1)\right]^2 - 1\right\}}{\left[\sigma_a\psi(1)\right]^2\int_0^1\left[W(x)\right]^2 dx} = \frac{\frac{1}{2}\left\{\left[W(1)\right]^2 - 1\right\}}{\psi(1)\int_0^1\left[W(x)\right]^2 dx},$$

or, equivalently,

$$n(\hat{\phi} - 1)\psi(1) \xrightarrow{D} \frac{\frac{1}{2}\left\{\left[W(1)\right]^2 - 1\right\}}{\int_0^1\left[W(x)\right]^2 dx}. \tag{9.4.31}$$

Comparing (9.4.31) with (9.3.4), we see that Table F(a) can be used when applying the statistic $n(\hat{\phi} - 1)\psi(1)$.

Similarly, following (9.3.7) and from (9.4.26), (9.4.28), and (9.4.30), we see that the "t statistic" in this case is

$$T = \frac{(\hat{\phi} - 1)}{S_{\hat{\phi}}} = \frac{n(\hat{\phi} - 1)\left[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2}}{\left[\hat{\sigma}_a^2\right]^{1/2}} = \frac{n^{-1}\sum_{t=1}^{n} Z_{t-1}a_t}{\left[n^{-2}\sum_{t=1}^{n} Z_{t-1}^2\right]^{1/2}\left[\hat{\sigma}_a^2\right]^{1/2}}$$

$$\xrightarrow{D} \frac{\psi(1)\frac{1}{2}\sigma_a^2\{[W(1)]^2 - 1\}}{\left\{[\sigma_a\psi(1)]^2\int_0^1[W(x)]^2\,dx\right\}^{1/2}[\sigma_a^2]^{1/2}} = \frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int_0^1[W(x)]^2dx\right\}^{1/2}}. \tag{9.4.32}$$

As a result, the Table G(a) can be used in this general case. Similar results extend to the use of Tables F(b) and (c) and G(b) and (c), for the model with a constant term and the model with a linear time trend, respectively. Again we often use $(n-p)(\hat{\phi}-1)\psi(1)$ instead of $n(\hat{\phi}-1)\psi(1)$ if no initial values are assumed in Equation (9.4.22) and the regression is computed for $(n-p)$ observations, i.e., for Z_t, $t=(p+1),\ldots,n$. The limiting distributions of $(n-p)(\hat{\phi}-1)\psi(1)$ and $n(\hat{\phi}-1)\psi(1)$ are exactly the same. The tests such as (9.4.31) and (9.4.32) resulting from the use of a higher-order AR model are often referred to as the augmented Dickey–Fuller tests in the literature.

EXAMPLE 9.4 As an illustration, let us consider the 54 yearly advertising expenditures of the Lydia Pinkham data listed as Series W12 in the appendix. Based on the ACFs and PACFs of the original series and its differencing, we can show that a reasonable model for the series is an AR(2) after differencing. More rigorously, we can assume an AR(3) model and test for a unit root. Thus, we fit (9.4.22) for $p=3$ with a constant, i.e.,

$$Z_t = \alpha + \phi Z_{t-1} + \varphi_1\Delta Z_{t-1} + \varphi_2\Delta Z_{t-2} + a_t \tag{9.4.33}$$

where $\Delta Z_t = (Z_t - Z_{t-1})$. Without using assumed initial values, we will fit the model with $t=4,\ldots,54$. The OLS regression equation becomes

$$\begin{aligned}\hat{Z}_t = {} & 139 + .856Z_{t-1} + .141\Delta Z_{t-1} - .326\Delta Z_{t-2},\\ & (87.97)(.08699)\quad (.1353)\qquad (.1369)\end{aligned}$$

where the values in the parentheses are the standard errors. To compute the test statistic, we note that under the null hypothesis (9.4.33) becomes

$$(1-B)Z_t = \mu + X_t,$$

where $\mu = \alpha/(1-\varphi_1-\varphi_2)$ and $X_t = \psi(B)a_t = [1/\varphi_{p-1}(B)]a_t = [1/(1-\varphi_1B-\varphi_2B^2)]a_t$. Thus,

$$\psi(B) = \frac{1}{[\varphi_{p-1}(B)]} = \frac{1}{[1-.141B+.326B^2]}$$

and

$$(n-p)(\hat{\phi}-1)\psi(1) = (54-3)(.856-1)\frac{1}{(1-.141+.326)} = -6.1975,$$

which is not less than the 5% critical value -13.3 for $n = 50$ given in Table F(b). Thus, the null hypothesis, $H_0 : \phi = 1$, cannot be rejected, and we conclude that the process for yearly advertising expenditures can be approximated by an AR(3) model with a unit root.

The t statistic in this case becomes

$$T = \frac{\hat{\phi} - 1}{S_{\hat{\phi}}} = \frac{.856 - 1}{.08669} = -1.65536,$$

which is not less than the 5% critical value -2.93 for $n = 50$ given in Table G(b). Again the null hypothesis, $H_0 : \phi = 1$, cannot be rejected.

Some remarks are in order.

1. It is clear that Equation (9.4.22) can also be written as

$$\Delta Z_t = \alpha Z_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{t-j} + a_t, \tag{9.4.34}$$

where $\alpha = \phi - 1$. Thus, in testing a unit root, we can also equivalently run the regression (9.4.34) and test $H_0 : \alpha = 0$ against $H_1 : \alpha < 0$. For the AR(1) model, Equation (9.4.34) becomes

$$\Delta Z_t = \alpha Z_{t-1} + a_t. \tag{9.4.35}$$

2. If the null hypothesis is not rejected, we conclude that the model contains a unit root. This conclusion implies, however, only that the series Z_t is integrated of order d greater than or equal to 1. To find out the order of required differencing, we can repeat the above test on the series ΔZ_t, $\Delta^2 Z_t$, and so on until an order of integration is reached.
3. A nonstationary series may not be homogeneous and no differencing of any order will transform it into a stationary series. In such a case, some other transformations such as those discussed in Chapter 4 may be necessary. In practice, however, for most homogeneous nonstationary series the required order of differencing is rarely greater than 2.

It is clear that with the general structure introduced in this section, the test can be used for any general mixed ARMA model. For more references, see Said and Dickey (1985) or Phillips and Perron (1988), among others.

9.5 Testing for a Unit Root in Seasonal Time Series Models

As discussed in Chapter 8, a seasonal time series can also be nonstationary and requires a seasonal differencing to remove its nonstationarity. Recall that if Z_t is a seasonal time series with a seasonal period s, then its seasonal differencing is $(1 - B^s)Z_t = Z_t - Z_{t-s}$. In such

a case, we say that the associated time series model contains a seasonal unit root. A formal testing procedure is introduced in this section.

9.5.1 Testing the Simple Zero Mean Seasonal Model

Consider the seasonal time series model $(1 - \Phi B^s)Z_t = a_t$ or, equivalently,

$$Z_t = \Phi Z_{t-s} + a_t, \tag{9.5.1}$$

where $Z_{1-s}, Z_{2-s}, \ldots, Z_0$ are initial conditions and the a_t are i.i.d. random variables with mean 0 and constant variance σ_a^2. The OLS estimator of Φ is given by

$$\hat{\Phi} = \frac{\sum_{t=1}^{n} Z_{t-s}Z_t}{\sum_{t=1}^{n} Z_{t-s}^2}, \tag{9.5.2}$$

which is also the maximum likelihood estimator when a_t is Gaussian. The Studentized statistic for testing the null hypothesis $H_0 : \Phi = 1$ is

$$T = \frac{\hat{\Phi} - 1}{S_{\hat{\Phi}}}, \tag{9.5.3}$$

where

$$S_{\hat{\Phi}} = \sqrt{\frac{\hat{\sigma}_a^2}{\sum_{t=1}^{n} Z_{t-s}^2}} \tag{9.5.4}$$

and

$$\hat{\sigma}_a^2 = \frac{\sum_{t=1}^{n} (Z_t - \hat{\Phi} Z_{t-s})^2}{(n-1)}. \tag{9.5.5}$$

The sampling distributions of $n(\hat{\Phi} - 1)$ and T, under the null hypothesis $H_0 : \Phi = 1$, are discussed by Dickey, Hasza, and Fuller (1984), and their percentiles for various sample sizes and for $s = 2, 4$, and 12 are reproduced in Tables H and I.

9.5.2 Testing the General Multiplicative Zero Mean Seasonal Model

For the more general multiplicative seasonal model, consider

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Phi B^s)Z_t = a_t, \tag{9.5.6}$$

where the roots of $(1 - \phi_1 B - \cdots - \phi_p B^p) = 0$ lie outside of the unit circle. We note that the a_t in (9.5.6) is a nonlinear function of $(\Phi, \boldsymbol{\phi})$ where $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$. The linear

approximation of this function $a_t(\Phi, \boldsymbol{\phi})$, evaluated at $(\hat{\Phi}, \hat{\boldsymbol{\phi}})$, using Taylor's expansion gives

$$a_t(\Phi, \boldsymbol{\phi}) = a_t(\hat{\Phi}, \hat{\boldsymbol{\phi}}) - (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p) Z_{t-s}(\Phi - \hat{\Phi}) - \sum_{i=1}^{p} (Z_{t-i} - \hat{\Phi} Z_{t-s-i})(\phi_i - \hat{\phi}_i) + R_t,$$

where R_t is the Taylor series remainder. Equivalently,

$$\begin{aligned} a_t(\hat{\Phi}, \hat{\boldsymbol{\phi}}) &= (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p) Z_{t-s}(\Phi - \hat{\Phi}) \\ &\quad + \sum_{i=1}^{p} (Z_{t-i} - \hat{\Phi} Z_{t-s-i})(\phi_i - \hat{\phi}_i) + e_t \\ &= (\Phi - \hat{\Phi})(1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p) Z_{t-s} \\ &\quad + \sum_{i=1}^{p} (\phi_i - \hat{\phi}_i)(Z_{t-i} - \hat{\Phi} Z_{t-s-i}) + e_t, \end{aligned} \tag{9.5.7}$$

where $e_t = a_t(\Phi, \boldsymbol{\phi}) - R_t$.

To test the null hypothesis $H_0 : \Phi = 1$, Equations (9.5.6) and (9.5.7) suggest the following two-step regression procedure:

1. Regress $\Delta^s Z_t$ on $\Delta^s Z_{t-1}, \ldots, \Delta^s Z_{t-p}$ to obtain an initial estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$, where $\Delta^s Z_t \equiv (1 - B^s) Z_t$. This estimator can be easily seen as consistent because under the null hypothesis $H_0 : \Phi = 1$, Equation (9.5.6) becomes

$$(1 - \phi_1 B - \cdots - \phi_p B^p) \Delta^s Z_t = a_t, \tag{9.5.8}$$

   which is a stationary AR(p) model for the seasonally differenced series $\Delta^s Z_t$.

2. Compute $a_t(1, \hat{\boldsymbol{\phi}})$ from (9.5.8) and use Equation (9.5.7) to regress $a_t(1, \hat{\boldsymbol{\phi}})$ on $\left[(1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p) Z_{t-s}, \Delta^s Z_{t-1}, \Delta^s Z_{t-2}, \ldots, \Delta^s Z_{t-p}\right]$ to obtain the OLS estimator of $(\Phi - 1) \equiv \alpha$ and $(\phi_1 - \hat{\phi}_1, \cdots, \phi_p - \hat{\phi}_p)$. The estimator of $\alpha = (\Phi - 1)$ can be used to test the null hypothesis $H_0 : \Phi = 1$. Under the null hypothesis, we have $\alpha = 0$. If α is much less than 0, then it implies that $\Phi < 1$ and the process is stationary. Hence, we will reject $H_0 : \Phi = 1$ if the estimator of $\alpha = (\Phi - 1)$ is significantly negative.

Dickey, Hasza, and Fuller (1984) show that under $H_0 : \Phi = 1$, the percentiles of Tables H and I for $n(\hat{\Phi} - 1)$ and T in (9.5.1) are also applicable for the general multiplicative seasonal model in (9.5.6).

The models in (9.5.1) and (9.5.6) imply that $E(Z_t) = 0$. The extension to the seasonal model with a nonzero mean is straightforward from the results of Section 9.3. For most practical purposes a purely seasonal model without any regular ARMA component is rare. Most of the time the question of a zero or nonzero constant and a zero or nonzero mean arises when one needs to decide the proper regular differencing discussed in Sections 9.3 and 9.4. In other words, in testing a unit root with a nonzero constant or a nonzero mean, we usually

apply the results given in Sections 9.3 and 9.4. Thus, we will not present the extension of the purely seasonal model with a nonzero constant here. Instead, we refer the interested reader to Dickey, Hasza, and Fuller (1984).

In actual data analysis, when there are no good initial values of $Z_{1-s}, Z_{2-s}, \ldots$, and Z_0 available, the estimator $\hat{\Phi}$ of Φ and its associated statistics presented earlier should be modified accordingly. Thus,

$$\hat{\Phi} = \frac{\sum_{t=s+1}^{n} Z_{t-s}Z_t}{\sum_{t=s+1}^{n} Z_{t-s}^2}, \tag{9.5.9}$$

$$S_{\hat{\Phi}} = \sqrt{\hat{\sigma}_a^2 / \sum_{t=s+1}^{n} Z_{t-s}^2}, \tag{9.5.10}$$

and

$$\hat{\sigma}_a^2 = \frac{\sum_{t=s+1}^{n} (Z_t - \hat{\Phi} Z_{t-s})^2}{(n - s - 1)}. \tag{9.5.11}$$

In testing the hypothesis $H_0 : \Phi = 1$, we will use $(n - s)(\hat{\Phi} - 1)$ instead of $n(\hat{\Phi} - 1)$, although the limiting distributions of the two statistics are exactly the same.

EXAMPLE 9.5 Consider the series, Z_t, which is the mean adjusted data set of Series W10 of the 32 quarterly U.S. beer productions from the first quarter of 1975 to the fourth quarter of 1982, which we studied in Example 8.4. That is, $Z_t = Y_t - \overline{Y}$, where Y_t is quarterly U.S. beer productions and $\overline{Y}$ is the mean of the series. With no knowledge of the underlying model, we can begin with the test of whether the model contains a unit root at the seasonal period, $s = 4$. Thus, we fit the regression model

$$Z_t = \Phi Z_{t-4} + a_t, \tag{9.5.12}$$

for $t = 5, \ldots, 32$, without assuming any initial values for Z_t with $t < 1$. The OLS regression becomes

$$\hat{Z}_t = \underset{(.0785)}{.9014} Z_{t-4},$$

where the value in the parenthesis is the standard error. The value of the test statistic becomes $(n - s)(\hat{\Phi} - 1) = (32 - 4)(.9014 - 1) = -2.7608$, which is not less than the 5% critical value -8.67 for $n = 40$ given in Table H. Thus, the null hypothesis, $H_0 : \Phi = 1$, cannot be rejected, and we conclude that the model for the series of quarterly beer productions contains a unit root at the seasonal period of 4.

The t statistic in this case becomes

$$T = \frac{\hat{\Phi} - 1}{S_{\hat{\Phi}}} = \frac{.9014 - 1}{.0785} = -1.2561,$$

which is not less than the 5% critical value -1.87 for $n = 40$ given in Table I. Thus, again the null hypothesis, $H_0 : \Phi = 1$, cannot be rejected.

The above test result suggests that a seasonal difference, $(1 - B^4)Z_t$, is called for. To see whether the second-order seasonal difference is required we can let $W_t = (1 - B^4)Z_t$, for $t = 5, \ldots, 32$, with a total of 28 W_t's, and test the hypothesis, $H_0 : \Phi = 1$, in the regression model

$$W_t = \Phi W_{t-4} + a_t. \tag{9.5.13}$$

The OLS regression results are

$$\hat{W}_t = \underset{(.2097)}{.0628} W_{t-4}.$$

The value of the test statistic becomes $(n - s)(\hat{\Phi} - 1) = (28 - 4)(.0628 - 1) = -22.4928$, which is much less than the 5% critical value -8.67 for $s = 4$ and $n = 40$ given in Table H. Thus, the null hypothesis, $H_0 : \Phi = 1$, should be rejected, and we conclude that the model for the first seasonal differences of quarterly beer productions no longer contains a unit root at the seasonal period. The t statistic in this case becomes

$$T = \frac{\hat{\Phi} - 1}{S_{\hat{\Phi}}} = \frac{.0628 - 1}{.2097} = -4.469,$$

which is also much less than the 5% critical value -1.87 for $s = 4$ and $n = 40$ given in Table I, and we should reject the null hypothesis, $H_0 : \Phi = 1$.

The above results imply that the series requires one seasonal differencing of period 4. To see whether the series also needs a regular differencing, we can apply the results in Section 9.3.1 and test the hypothesis, $H_0 : \phi = 1$, in the regression model

$$W_t = \phi W_{t-1} + a_t, \tag{9.5.14}$$

for $t = 1, \ldots, 28$, where for convenience we reindex the 28 available differences starting from time 1. From the OLS regression results

$$\hat{W}_t = \underset{(.1848)}{.336} W_{t-1}.$$

we have $(n - 1)(\hat{\phi} - 1) = (28 - 1)(.336 - 1) = -17.928$, which is much less than the 5% critical value -7.3 for $n = 25$ or -7.7 for $n = 50$ given in Table F(a). Thus, the null hypothesis, $H_0 : \phi = 1$, should be rejected, and we conclude that the model for the first seasonal differences of quarterly beer production no longer contains unit roots. The t statistic in this case becomes

$$T = \frac{\hat{\phi} - 1}{S_{\hat{\phi}}} = \frac{.336 - 1}{.1848} = -3.593,$$

which is also much less than the 5% critical value -1.95 given in Table G(a); hence, we reach the same conclusion. The implication is that we can identify a stationary model for the quarterly differences, $(1 - B^4)Z_t$.

EXERCISES

9.1 Let $Z_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$ or, equivalently, $Z_t = \mu + \psi(B)a_t$, where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, $\psi_0 = 1$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

(a) Assume that the a_t are white noise with mean 0 and variance σ_a^2. Show that $\sum_{j=-\infty}^{\infty} \gamma_j = \sigma_a^2[\psi(1)]^2$, where γ_j is the jth autocovariance of the process Z_t.

(b) Assume that $\sum_{j=0}^{\infty} \psi_j \neq 0$, and the a_t are i.i.d. $(0, \sigma_a^2)$ random variables. Prove that $\sqrt{n}(\bar{Z} - \mu) \xrightarrow{D} N\left(0, \sigma_a^2[\psi(1)]^2\right)$.

(c) Show that if Z_t is an AR(1) process, $\dot{Z}_t = \phi \dot{Z}_{t-1} + a_t$, where $\dot{Z}_t = Z_t - \mu$ and $|\phi| < 1$, then $\sqrt{n}(\bar{Z} - \mu) \xrightarrow{D} N\left(0, \sigma_a^2(1 - \phi^2)^{-1}\right)$.

9.2 Derive the limiting distribution of $n(\hat{\phi}_t - 1)$ for the model given in (9.3.17).

9.3 Derive the limiting distribution of T_t for the model given in (9.3.17).

9.4 (a) Test for a unit root on Series W7.

(b) Test for a unit root on Series W16.

9.5 (a) Find the limiting distribution of $n(\hat{\phi}_t - 1)$ in testing a unit root for the model

$$(1 - \phi B)Z_t = (1 - \theta B)a_t,$$

where $-1 < \theta < 1$.

(b) Test for a unit root on Series W6.

9.6 Derive the limiting distribution of $n(\hat{\Phi} - 1)$ for the model given in (9.5.1).

9.7 Derive the limiting distribution of T for the model given in (9.5.1).

9.8 Perform unit root tests on the series of U.S. personal consumption of gasoline and oil given in Exercise 8.7 and determine the required differencing to make the series stationary.

9.9 Perform unit root tests on Series W14 and determine the required differencing to make the series stationary.

10

Intervention Analysis and Outlier Detection

Time series are frequently affected by certain external events such as holidays, strikes, sales promotions, and other policy changes. We call these external events *interventions*. In this chapter, we introduce the technique, called *intervention analysis*, to evaluate the effect of these external events. Intervention analysis has been successfully used to study the impact of air pollution control and economic policies (Box and Tiao, 1975), the impact of the Arab oil embargo (Montgomery and Weatherby, 1980), the impact of the New York blackout (Izenman and Zabell, 1981), and many other events. We first discuss the analysis when the timing of the interventions is known. The method is then generalized to study the impact of the events when the timing of interventions is unknown and hence leads to the general time series outlier analysis.

10.1 Intervention Models

Given that a known intervention occurs at time T, is there any evidence of a change in the time series (such as the increase of the mean level) and, if so, by how much? One may initially think that the traditional two-sample t-test could be used to analyze this problem in terms of comparing the pre-intervention data with the post-intervention data. The t-test, however, assumes both normality and independence. Even though the t-test is known to be robust with respect to the normality assumption, it is extremely sensitive to the violation of the independence assumption as shown by Box and Tiao (1965), who developed the intervention analysis to study a time series structural change due to external events (Box and Tiao, 1975).

There are two common types of intervention variables. One represents an intervention occurring at time T that remains in effect thereafter. That is, the intervention is a step function,

$$S_t^{(T)} = \begin{cases} 0, & t < T, \\ 1, & t \geq T. \end{cases} \tag{10.1.1}$$

The other one represents an intervention taking place at only one time period. Thus, it is a pulse function,

$$P_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T. \end{cases} \tag{10.1.2}$$

Note that the pulse function can be produced by differencing the step function $S_t^{(T)}$. That is, $P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1 - B)S_t^{(T)}$. Therefore, an intervention model can be represented equally well with the step function or with the pulse function. The use of a specific form is usually based on the convenience of interpretation.

There are many possible responses to the step and pulse interventions. We illustrate some commonly encountered ones.

1. A fixed unknown impact of an intervention is felt b periods after the intervention. Thus, depending on the type of intervention, the impact is

$$\omega B^b S_t^{(T)} \tag{10.1.3}$$

or

$$\omega B^b P_t^{(T)}. \tag{10.1.4}$$

2. An impact of an intervention is felt b periods after the intervention, but the response is gradual. For a step input, we have

$$\frac{\omega B^b}{(1 - \delta B)} S_t^{(T)}, \tag{10.1.5}$$

and for a pulse input,

$$\frac{\omega B^b}{(1 - \delta B)} P_t^{(T)}, \tag{10.1.6}$$

where $0 \leq \delta \leq 1$. For $\delta = 0$, (10.1.5) and (10.1.6) reduce to (10.1.3) and (10.1.4), respectively. If $\delta = 1$, the impact increases linearly without bound. For most cases, we have $0 < \delta < 1$, and the response is gradual.

For illustrations, we plot the above interventions with $b = 1$ and $0 < \delta < 1$ in Figure 10.1.

Note that various responses can be produced by different combinations of step and pulse inputs. For example, we may have the response

$$\frac{\omega_0 B}{(1 - \delta B)} P_t^{(T)} + \omega_1 B S_t^{(T)} \tag{10.1.7}$$

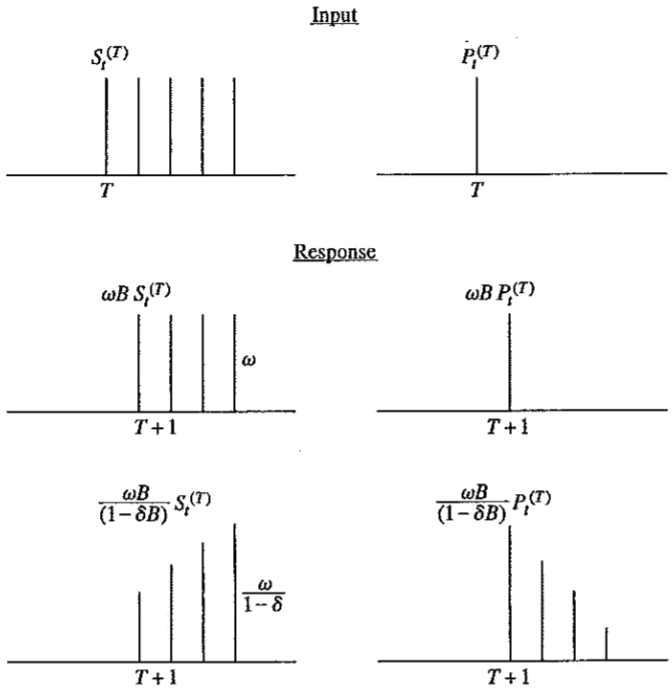


FIGURE 10.1 Responses to step and pulse inputs.

as represented in Figure 10.2. As mentioned earlier, however, because $P_t^{(T)} = (1 - B)S_t^{(T)}$, response (10.1.7) can also be written as

$$\left[\frac{\omega_0 B}{(1 - \delta B)} + \frac{\omega_1 B}{(1 - B)}\right] P_t^{(T)}. \tag{10.1.8}$$

This model is useful to represent the phenomenon in which an intervention produces a response that tapers off gradually but leaves a permanent residue effect in the system. The impact of an intervention such as advertising on sales can be represented as shown in Figure 10.2(a), and the effect of a price or a tax increase on imports may be represented in Figure 10.2(b).

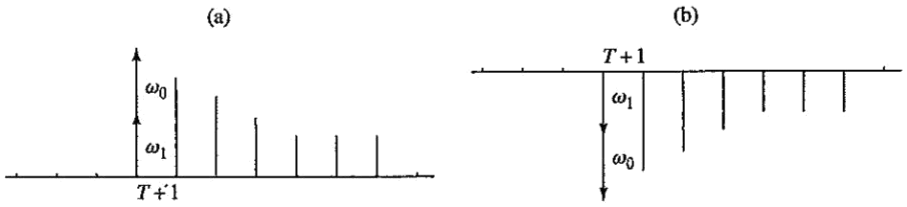


FIGURE 10.2 Response to combined inputs, $[\omega_0 B/(1 - \delta B)]P_t^{(T)} + \omega_1 B S_t^{(T)}$. (a) $\omega_0 > 0$ and $\omega_1 > 0$. (b) $\omega_0 < 0$ and $\omega_1 < 0$.

More generally, a response may be represented as a rational function

$$\frac{\omega(B)B^b}{\delta(B)} \tag{10.1.9}$$

where $\omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$ are polynomials in B, b is the time delay for the intervention effect, and the weights ω_j's in the polynomial $\omega(B)$ often represent the expected initial effects of the intervention. The polynomial $\delta(B)$, on the other hand, measures the behavior of the permanent effect of the intervention. The roots of $\delta(B) = 0$ are assumed to be on or outside the unit circle. The unit root represents an impact that increases linearly, and the root outside the unit circle represents a phenomenon that has a gradual response.

For multiple intervention inputs, we have the following general class of models:

$$Z_t = \theta_0 + \sum_{j=1}^{k} \frac{\omega_j(B)B^{b_j}}{\delta_j(B)} I_{jt} + \frac{\theta(B)}{\psi(B)} a_t, \tag{10.1.10}$$

where I_{jt}, $j = 1, 2, \ldots, k$ are intervention variables. These intervention variables can be either step or pulse functions. More generally, they can be proper indicator variables, as shown later in Examples 10.2 and 10.4. The form $\omega_j(B)B^{b_j}/\delta_j(B)$ for the jth intervention is postulated based on the expected form of the response given knowledge of the intervention. The main purpose of intervention models is to measure the effect of interventions. Thus, with respect to the intervention variables I_{jt}, the time series free of intervention is called the *noise series* and is denoted by N_t, and its model is hence known as the *noise model*. The noise model $\left[\theta(B)/\Psi(B)\right]a_t$ is usually identified using the univariate model identification procedure based on the time series Z_t before the date of intervention, i.e., $\{Z_t : t < T\}$. If diagnostic checking of the model reveals no model inadequacy, then we can make appropriate inferences about the intervention. Otherwise, appropriate modifications must be made to the model and estimation and diagnostic checking repeated. For a nonstationary process, the model in (10.1.10) normally does not contain a constant term θ_0.

10.2 Examples of Intervention Analysis

EXAMPLE 10.1 A utility company provides electricity, light and power, gas and gas transmission, telephone and telegraph services, and water. Its business is regulated by federal and state authorities and normally is given an exclusive right to perform its services in a specific area, free from any competition. As the result, its revenue-generating ability is relatively stable, and its stock does not fluctuate in value as much as most common stocks. It generally pays a high rate of dividend in addition to offering moderate growth potential. Therefore, its stock becomes an important component in many investment portfolios.

Duke Energy Corporation is an utility company headquartered in Charlotte, North Carolina. It contains many business units, including Franchised Electric, Natural Gas Transmission, Field

Services, Duke Energy North America, International Energy, and Other Operations. Through these units, it engages in businesses not only in North Carolina but also other areas inside and outside the United States. The company's shares are traded in the New York Stock Exchange as DUK. Like all other companies, it issued statements from time to time about its business and financial conditions.

On July 9, 2002, Schiffrin & Barroway, LLP, a law firm located outside Philadelphia, Pennsylvania, that specializes in representing shareholders and consumers in many class-action litigations throughout the United States, filed a complaint in the United States District Court for the Southern District of New York against Duke Energy Corporation alleging that the company misled investors about its business and financial conditions and, consequently, about the true value of DUK stock between July 22, 1999, and May 17, 2002. The law firm tried to recover damages for the company's breaches of its fiduciary duties and violations of the federal securities laws on behalf of all investors who bought Duke Energy Corporation stock during the period. Following the announcement, the stock of Duke Energy Corporation fell \$1.25 and closed at \$29.06 per share that day.

In this example, we perform an intervention analysis to assess the effect of this class-action securities litigation announcement on Duke Energy Corporation's stock prices. Figure 10.3 shows Series W11, the daily closing stock price of Duke Energy Corporation between January 3, 2002, and August 31, 2002. There are 166 observations in total, and the closing price, \$29.06, of July 9, 2002, corresponds to the 129th observation.

Because the announcement was made late in the day and July 10, 2002, was the first trading day after the announcement, the period between January 3, 2002, and July 9, 2002, is regarded as the noise series containing no major intervention that would affect the stock

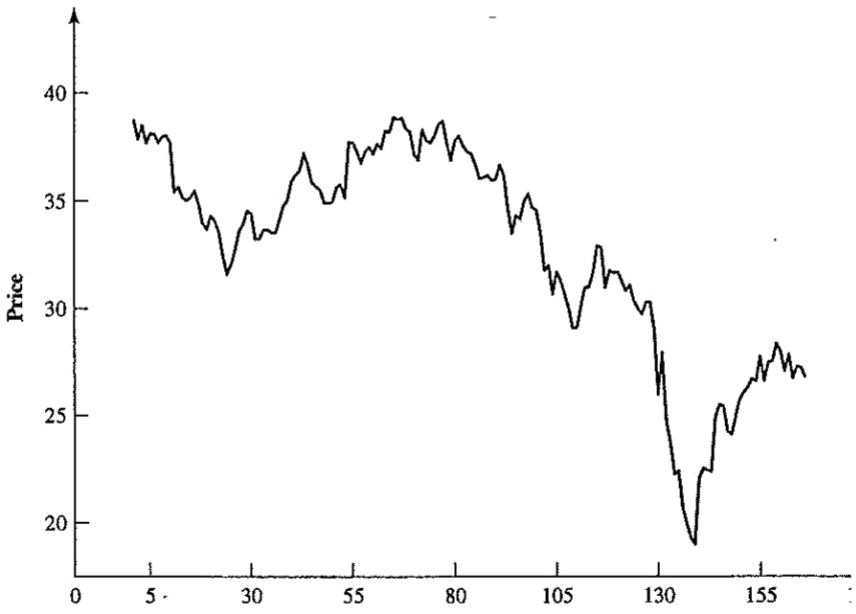


FIGURE 10.3 Daily stock price of Duke Energy from January 3, 2002, to August 31, 2002.

TABLE 10.1 Sample ACF and PACF of Duke Energy Corporation stock prices

			(a) $\{N_t\}$		$(\bar{N} = 34.94, S_N = 2.712, n = 129)$					
k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.94	.89	.84	.79	.74	.69	.66	.62	.58	.53
St.E.	.09	.15	.18	.21	.23	.25	.27	.28	.29	.30
$\hat{\phi}_k$	.94	.03	−.01	−.01	−.03	.01	.11	−.09	−.09	.02
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09
			(b) $\{W_t = (1 - B)N_t\}$		$(\bar{W} = -.0757, S_W = .6942, n = 128)$					
k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.05	−.07	−.04	.00	−.03	−.15	.10	.13	.03	−.09
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09
$\hat{\phi}_k$	.05	−.07	−.03	.00	−.04	−.15	.11	.09	.02	−.08
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

price. The sample ACF and sample PACF of the original and the differenced noise series as shown in Table 10.1 suggest the random walk model

$$(1 - B)N_t = a_t. \tag{10.2.1}$$

By assuming that the effect of the class-action litigation against the company is to cause an immediate level change in stock prices, one could propose the response function

$$\omega_0 I_t, \tag{10.2.2}$$

where ω_0 represents the impact of the litigation announcement and

$$I_t = \begin{cases} 0, & t < 130 \text{ (July 10, 2002)}, \\ 1, & t \geq 130 \text{ (July 10, 2002)}. \end{cases}$$

The intervention model becomes

$$Z_t = \omega_0 I_t + \frac{a_t}{(1 - B)}. \tag{10.2.3}$$

The estimation results are

Parameter	Estimate	St. Error
ω_0	−3.06	.8019

The sample ACF for the residuals show no signs of model inadequacy, and the estimate of the intervention parameter is highly significant. Thus, there is evidence that the announcement of the litigation against Duke Energy Corporation did cause a downward effect on its stock prices.

EXAMPLE 10.2 Los Angeles has been known to be plagued by a special air pollution problem. The problem comes from substances produced by chemical reactions in sunlight among some primary pollutants such as oxides of nitrogen and reactive hydrocarbons. The products of these chemical reactions are responsible for the notorious Los Angeles smog, which causes such health hazards as eye irritation and lung damage. One measured product that is indicative of the degree of this photochemical pollution is ozone, often denoted by O_3. Figure 10.4 shows the monthly average of hourly readings of O_3 in parts per hundred million (pphm) in downtown Los Angeles from 1955 to 1972.

To ease the air pollution problem, different methods were instituted, including the diversion of traffic in early 1960 by the opening of the Golden State Freeway and the inception of a new law (Rule 63) that reduced the allowable proportion of reactive hydrocarbons in the gasoline sold locally. Also, after 1966 special regulations were implemented to require engine design changes in new cars in order to reduce the production of O_3. It was through the study of the effect of these events on the pollution problem that Box and Tiao (1975) introduced the intervention analysis.

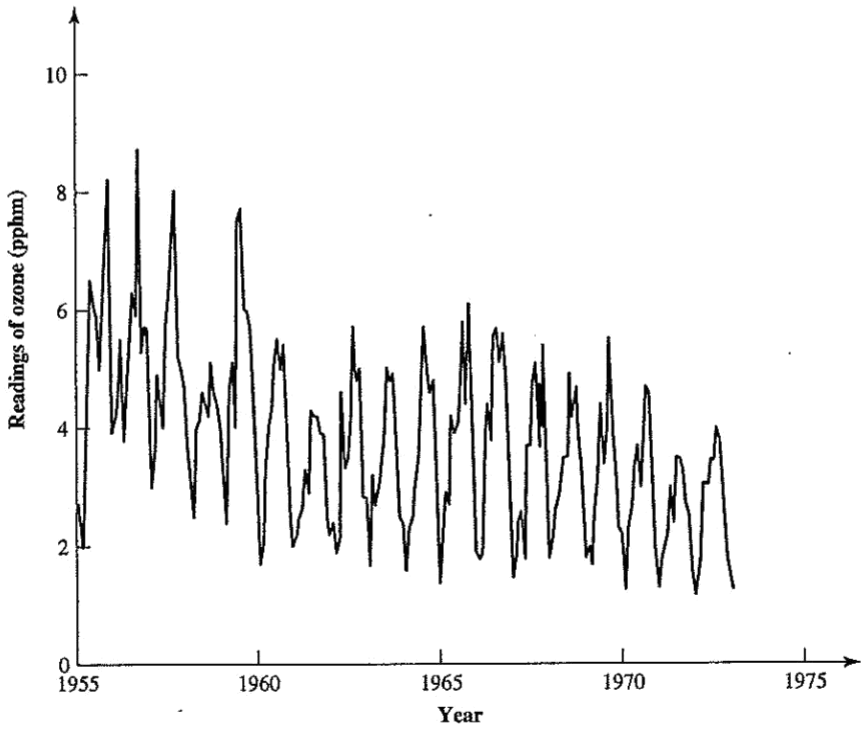


FIGURE 10.4 Monthly average of hourly readings of O_3 (pphm) in downtown Los Angeles, 1955–1972.

The period from 1955 to 1960 is assumed to be free of intervention effects and is used to estimate the noise model for N_t. The sample ACF within this period suggests nonstationary and highly seasonal behavior. The ACF of the seasonally differenced series $(1 - B^{12})N_t$ has significant spikes only at lags 1 and 12, which implies the following noise model:

$$(1 - B^{12})N_t = (1 - \theta B)(1 - \Theta B^{12})a_t. \tag{10.2.4}$$

Box and Tiao (1975) suggest that the opening of the Golden State Freeway and the implementation of Rule 63 in 1960 represent an intervention I_1, which might be expected to produce a step change in the O_3 level at the beginning of 1960. Intervention I_2 would be represented by the regulations implemented in 1966 requiring engine changes in new cars. The effect of I_2 would be most accurately measured by the proportion of new cars having specified engine changes in the car population over time. Unfortunately, no such data are available. One might, however, represent the effect of I_2 as an annual trend reflecting the effect of the increased proportion of new design vehicles in the population. Because of the differences in the intensity of sunlight and other meteorological conditions between the summer months and the winter months, the effect of I_2 would be different in these two seasons. Thus, Box and Tiao (1975) proposed the model

$$Z_t = \omega_1 I_{1t} + \frac{\omega_2}{1 - B^{12}} I_{2t} + \frac{\omega_3}{1 - B^{12}} I_{3t} + \frac{(1 - \theta B)(1 - \Theta B^{12})}{(1 - B^{12})} a_t, \tag{10.2.5}$$

where

$$I_{1t} = \begin{cases} 0, & t < \text{January 1960}, \\ 1, & t \geq \text{January 1960}, \end{cases}$$

$$I_{2t} = \begin{cases} 1, & \text{"summer" months June–October beginning in 1966}, \\ 0, & \text{otherwise}, \end{cases}$$

$$I_{3t} = \begin{cases} 1, & \text{"winter" months November–May beginning in 1966}, \\ 0, & \text{otherwise}. \end{cases}$$

Estimation results of the above model are

Parameter	Estimate	St. Error
ω_1	−1.09	.13
ω_2	−.25	.07
ω_3	−.07	.06
θ	−.24	.03
Θ	.55	.04

Residuals of $\hat{a}_t$ show no obvious inadequacies in the model. Thus, there is evidence that (1) intervention I_1 reduces the level of O_3, and (2) associated with intervention I_2, there is a reduction of O_3 level in the summer months but not in the winter months.

EXAMPLE 10.3 The Arab oil embargo in November 1973 greatly affected the supply of petroleum products to the United States. The prices of petroleum products rocketed, and the cries for energy conservation were loud and clear everywhere in the United States. It is generally believed that the rate of growth of energy consumption following this event was lower for the post-embargo years. To test the validity of this belief, Montgomery and Weatherby (1980) applied an intervention model to the natural logarithm of monthly electricity consumption from January 1951 to April 1977 using a total of 316 observations.

Although the embargo started in November 1973, Montgomery and Weatherby (1980) assumed that the effect of this embargo was not felt until December 1973. Thus, they used 275 months of data from January 1951 to November 1973 to model the noise series and obtained the following ARIMA(0, 1, 2) $\times$ $(0, 1, 1)_{12}$ model:

$$(1 - B)(1 - B^{12})\ln N_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^{12})a_t. \tag{10.2.6}$$

By assuming that the effect of the embargo is to cause a gradual change in consumption, one could propose the intervention model

$$\ln Z_t = \frac{\omega_0}{(1 - \delta_1 B)} I_t + \frac{(1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^{12})}{(1 - B)(1 - B^{12})} a_t, \tag{10.2.7}$$

where ω_0 represents the initial impact of the oil embargo and

$$I_t = \begin{cases} 0, & t \leq 275, \\ 1, & t > 275. \end{cases}$$

The estimates and the associated standard errors are

Parameter	Estimate	St. Error
ω_0	$-.07$	.03
δ_1	.18	.36
θ_1	.40	.06
θ_2	.27	.06
Θ	.64	.05

Because the estimate of δ_1 is not statistically significant, Montgomery and Weatherby dropped the parameter δ_1 in the following modified model:

$$\ln Z_t = \omega_0 I_t + \frac{(1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^{12})}{(1 - B)(1 - B^{12})} a_t, \tag{10.2.8}$$

where I_t was defined as in (10.2.7). The estimation results are

Parameter	Estimate	St. Error
ω_0	−.07	.02
θ_1	.40	.06
θ_2	.28	.06
Θ	.64	.05

The estimates are all statistically significant. The residual ACF does not exhibit any model inadequacy. Thus, the intervention model in (10.2.8) is satisfactory. The result implies that the embargo induced a permanent level change on electricity consumption. Because the model was built using the natural logarithm of electricity consumption, the estimate of the intervention effect in terms of the original megawatt hours (MWH) metric is $e^{\hat{\omega}_0} = e^{-.07} = .93$. Therefore, the post-intervention level of electricity consumption was 93% of the pre-intervention level, or, equivalently, the effect of the Arab oil embargo reduced electricity consumption by 7%.

EXAMPLE 10.4 At exactly 5:27 P.M., November 9, 1965, most of New York City was plunged into darkness due to a massive power failure. The blackout was long, and most of the city remained dark during the night. On Wednesday, August 10, 1966, the *New York Times* carried a front-page article with the headline "Births Up 9 Months After the Blackout." Numerous articles were published afterward in newspapers and magazines both inside and outside the United States alleging a sharp increase in the city's birthrate. A number of medical and demographic articles then appeared with contradictory statements regarding the blackout effect. Using a total of 313 weekly births in New York City from 1961 to 1966, which is plotted in Figure 10.5, Izenman and Zabell (1981) applied the intervention analysis to the phenomenon.

Because the blackout, which occurred on November 9, 1965, falls in the middle of the 254th week, the first 254 weekly birth totals are used to model the noise series with the following process:

$$(1 - B)(1 - B^{52})N_t = (1 - \theta B)(1 - \Theta B^{52})a_t. \tag{10.2.9}$$

Furthermore, the obstetrical and gynecological studies show that the mode of the gestational interval, which is defined as the time from onset of the last menstrual period (LMP) to birth, occurs at 40 or 41 weeks after the LMP. Izenman and Zabell (1981) suggested the intervention form

$$\omega_0 I_t, \tag{10.2.10}$$

where ω_0 is used to represent the effect of the blackout and

$$I_t = \begin{cases} 1, & t = 292, 293, 294, 295, \\ 0, & \text{otherwise.} \end{cases} \tag{10.2.11}$$

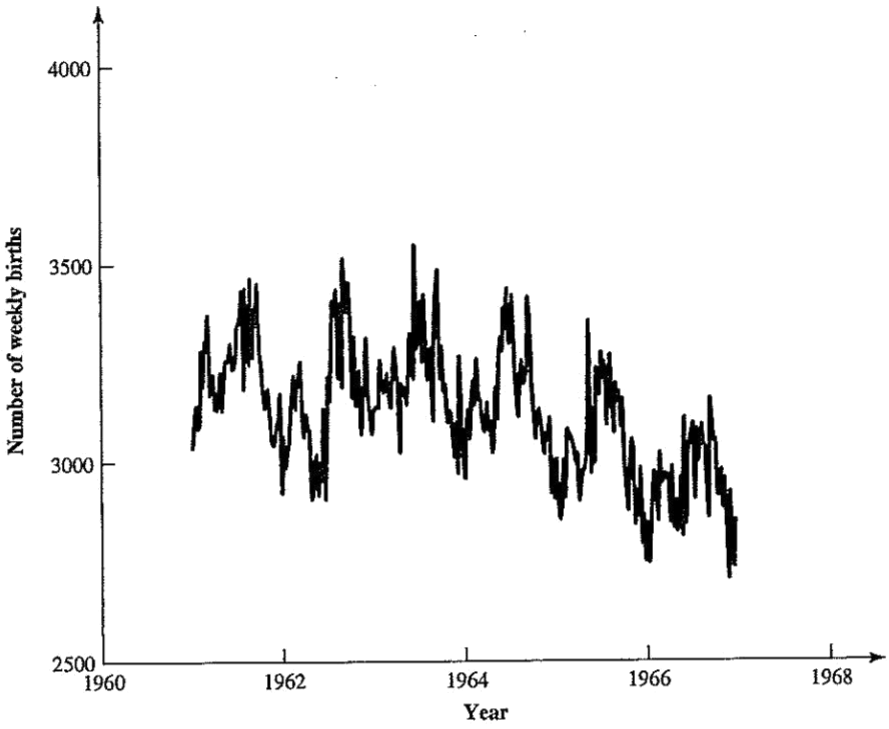


FIGURE 10.5 Total weekly births in New York City, 1961–1966.

The previous intervention variable I_t was introduced to take the value 1 for weeks 38 through 41 after the blackout to give the model the maximum possible chance of detecting the effect. Thus, we have the intervention model

$$Z_t = \omega_0 I_t + \frac{(1 - \theta B)(1 - \Theta B^{52})}{(1 - B)(1 - B^{52})} a_t, \tag{10.2.12}$$

where I_t is defined as in (10.2.11).

The estimates and associated standard errors are

Parameter	Estimate	St. Error
ω_0	28.63	47.36
θ	.74	.04
Θ	.82	.02

The residual ACF shows no signs of model inadequacy. Because the parameter estimate of ω_0 is not statistically significant, we conclude that intervention analysis of the New York City birth data does not detect a significant increase in births that can be ascribed to the blackout. Here, we note that the model estimated in (10.2.12) is slightly different from the model given in Izenman and Zabell (1981).

10.3 Time Series Outliers

Time series observations are sometimes influenced by interruptive events, such as strikes, outbreaks of war, sudden political or economic crises, unexpected heat or cold waves, or even unnoticed errors of typing and recording. The consequences of these interruptive events create spurious observations that are inconsistent with the rest of the series. Such observations are usually referred to as *outliers*. When the timing and causes of interruptions are known, their effects can be accounted for by using the intervention model discussed in Sections 10.1 and 10.2. In practice, however, the timing of interruptive events is sometimes unknown. Because outliers are known to wreak havoc in data analysis, making the resultant inference unreliable or even invalid, it is important to have procedures that will detect and remove such outliers effects. The detection of time series outliers was first studied by Fox (1972), where two statistical models, additive and innovational, were introduced. Other references on this topic include Abraham and Box (1979), Martin (1980), Chang and Tiao (1983), Hillmer, Bell, and Tiao (1983), Tsay (1986), Chang, Tiao, and Chen (1988).

10.3.1 Additive and Innovational Outliers

With no loss of generality, consider a zero mean stationary process. Let Z_t be the observed series and X_t be the outlier-free series. Assume that $\{X_t\}$ follows a general ARMA(p, q) model

$$\phi(B)X_t = \theta(B)a_t, \tag{10.3.1}$$

where $\phi(B) = 1-\phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$ are stationary and invertible operators sharing no common factors and $\{a_t\}$ is a sequence of white noise, identically and independently distributed as $N(0, \sigma_a^2)$. An additive outlier (AO) model is defined as

$$Z_t = \begin{cases} X_t, & t \neq T \\ X_t + \omega, & t = T \end{cases} \tag{10.3.2a}$$

$$= X_t + \omega I_t^{(T)} \tag{10.3.2b}$$

$$= \frac{\theta(B)}{\phi(B)}a_t + \omega I_t^{(T)}, \tag{10.3.2c}$$

where

$$I_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T, \end{cases}$$

is the indicator variable representing the presence or absence of an outlier at time T. An innovational outlier (IO) model is defined as

$$Z_t = X_t + \frac{\theta(B)}{\phi(B)}\omega I_t^{(T)} \tag{10.3.3a}$$

$$= \frac{\theta(B)}{\phi(B)}\left(a_t + \omega I_t^{(T)}\right). \tag{10.3.3b}$$

Hence, an additive outlier affects only the Tth observation, Z_T, whereas an innovational outlier affects all observations $Z_T, Z_{T+1}, \ldots$, beyond time T through the memory of the system described by $\theta(B)/\phi(B)$.

More generally, a time series might contain several, say k, outliers of different types, and we have the following general outlier model:

$$Z_t = \sum_{j=1}^{k} \omega_j \nu_j(B) I_t^{(T)} + X_t, \tag{10.3.4}$$

where $X_t = (\theta(B)/\phi(B))\, a_t$, $\nu_j(B) = 1$ for an AO, and $\nu_j(B) = \theta(B)/\phi(B)$ for an IO at time $t = T_j$.

10.3.2 Estimation of the Outlier Effect When the Timing of the Outlier Is Known

To motivate the procedure for detecting AO and IO, we consider a simpler case when T and all parameters in (10.3.1) are known. Letting

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = (1 - \pi_1 B - \pi_2 B^2 - \cdots) \tag{10.3.5}$$

and defining

$$e_t = \pi(B) Z_t, \tag{10.3.6}$$

we have from (10.3.2c) and (10.3.3b) that

$$\text{AO:} \quad e_t = \omega \pi(B) I_t^{(T)} + a_t \tag{10.3.7}$$

and

$$\text{IO:} \quad e_t = \omega I_t^{(T)} + a_t. \tag{10.3.8}$$

For n available observations, the AO model in (10.3.7) can be written as

$$\begin{bmatrix} e_1 \\ \vdots \\ e_{T-1} \\ e_T \\ e_{T+1} \\ e_{T+2} \\ \vdots \\ e_n \end{bmatrix} = \omega \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\pi_1 \\ -\pi_2 \\ \vdots \\ -\pi_{n-T} \end{bmatrix} + \begin{bmatrix} a_1 \\ \vdots \\ a_{T-1} \\ a_T \\ a_{T+1} \\ a_{T+2} \\ \vdots \\ a_n \end{bmatrix}. \tag{10.3.9}$$

Let $\hat{\omega}_{AT}$ be the least square estimator of ω for the AO model. Because $\{a_t\}$ is white noise, from the least squares theory, we have that

$$\begin{aligned} \text{AO:} \quad \hat{\omega}_{AT} &= \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=0}^{n-T} \pi_j^2} \\ &= \frac{\pi^*(F) e_T}{\tau^2} \end{aligned} \tag{10.3.10}$$

where $\pi^*(F) = (1 - \pi_1 F - \pi_2 F^2 - \cdots - \pi_{n-T} F^{n-T})$, F is the forward shift operator such that $Fe_t = e_{t+1}$, and $\tau^2 = \sum_{j=0}^{n-T} \pi_j^2$. The variance of the estimator is

$$\begin{aligned} \text{Var}(\hat{\omega}_{AT}) &= \text{Var}\left(\frac{\pi^*(F) e_T}{\tau^2}\right) \\ &= \frac{1}{\tau^4} \text{Var}\left[\pi^*(F) a_T\right] \\ &= \frac{\sigma_a^2}{\tau^2}. \end{aligned} \tag{10.3.11}$$

Similarly, letting $\hat{\omega}_{IT}$ be the least squares estimator of ω for the IO model, we have

$$\text{IO:} \quad \hat{\omega}_{IT} = e_T \tag{10.3.12}$$

and

$$\begin{aligned} \text{Var}(\hat{\omega}_{IT}) &= \text{Var}(e_T) = \text{Var}(\omega I_t^{(T)} + a_T) \\ &= \sigma_a^2. \end{aligned} \tag{10.3.13}$$

Thus, the best estimate of the effect of an IO at time T is the residual e_T, whereas the best estimate of the effect of an AO is a linear combination of e_T, e_{T+1}, . . . , and e_n with weight depending on the structure of the time series process. It is easily seen that $\text{Var}(\hat{\omega}_{AT}) \leq \text{Var}(\hat{\omega}_{IT}) = \sigma_a^2$, and in some cases, $\text{Var}(\hat{\omega}_{AT})$ can be much smaller than σ_a^2.

Various tests can be performed for the hypotheses

H_0: Z_T is neither an AO nor an IO

H_1: Z_T is an AO

H_2: Z_T is an IO.

The likelihood ratio test statistics for AO and IO are

$$H_1 \text{ vs. } H_0: \quad \lambda_{1,T} = \frac{\tau\hat{\omega}_{AT}}{\sigma_a} \tag{10.3.14}$$

and

$$H_2 \text{ vs. } H_0: \quad \lambda_{2,T} = \frac{\hat{\omega}_{IT}}{\sigma_a}. \tag{10.3.15}$$

Under the null hypothesis H_0, both $\lambda_{1,T}$ and $\lambda_{2,T}$ are distributed as $N(0, 1)$.

10.3.3 Detection of Outliers Using an Iterative Procedure

If T is unknown but the time series parameters are known, we can proceed to calculate $\lambda_{1,t}$ and $\lambda_{2,t}$ for each $t = 1, 2, \ldots, n$, and then make decisions based on the above sampling results. In practice, however, the time series parameters ϕ_j, θ_j, π_j, and σ_a^2 are usually unknown and have to be estimated. It is known that existence of outliers makes the estimates of the parameters seriously biased. In particular, as expected, σ_a^2 will tend to be overestimated. Chang and Tiao (1983) proposed an iterative detecting procedure to handle the situation when an unknown number of AO or IO may exist.

Step 1. Model the series $\{Z_t\}$ by assuming that there are no outliers. Compute the residuals from the estimated model, i.e.,

$$\begin{aligned} \hat{e}_t &= \hat{\pi}(B)Z_t \\ &= \frac{\hat{\phi}(B)}{\hat{\theta}(B)}Z_t, \end{aligned} \tag{10.3.16}$$

where $\hat{\phi}(B) = \left(1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p\right)$ and $\hat{\theta}(B) = \left(1 - \hat{\theta}_1 B - \cdots - \hat{\theta}_q B^q\right)$. Let

$$\hat{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2$$

be the initial estimate of σ_a^2.

Step 2. Calculate $\hat{\lambda}_{1,t}$ and $\hat{\lambda}_{2,t}$ for $t = 1, 2, \ldots, n$, using the estimated model. Define

$$\hat{\lambda}_T = \max_t \max_i \{|\hat{\lambda}_{i,t}|\}, \tag{10.3.17}$$

where T denotes the time when the maximum occurs. If $\hat{\lambda}_T = |\hat{\lambda}_{1,T}| > C$, where C is a predetermined positive constant usually taken to be some value between 3 and 4, then there is an AO at time T with its effect estimated by $\hat{\omega}_{AT}$. One can modify the data using (10.3.2b) as

$$\widetilde{Z}_t = Z_t - \hat{\omega}_{AT} I_t^{(T)} \tag{10.3.18}$$

and define the new residuals using (10.3.7) as

$$\widetilde{e}_t = \hat{e}_t - \hat{\omega}_{AT} \hat{\pi}(B) I_t^{(T)}. \tag{10.3.19}$$

If $\hat{\lambda}_T = |\hat{\lambda}_{2,T}| > C$, then there is an IO at time T with its effect being $\hat{\omega}_{IT}$. This IO effect can be removed by modifying the data using (10.3.3a), i.e.,

$$\widetilde{Z}_t = Z_t - \frac{\hat{\theta}(B)}{\hat{\phi}(B)} \hat{\omega}_{IT} I_t^{(T)}, \tag{10.3.20}$$

and defining the new residuals using (10.3.8) as

$$\widetilde{e}_t = \hat{e}_t - \hat{\omega}_{IT} I_t^{(T)}. \tag{10.3.21}$$

A new estimate $\widetilde{\sigma}_a^2$ is then computed from the modified residuals.

Step 3. Recompute $\hat{\lambda}_{1,t}$ and $\hat{\lambda}_{2,t}$ based on the modified residuals and $\widetilde{\sigma}_a^2$, and repeat step 2 until all outliers are identified. The initial estimates in $\pi(B)$ remain unchanged.

Step 4. Suppose that step 3 terminated and k outliers have been tentatively identified at times $T_1, T_2, \ldots,$ and T_k. Treat these times as if they are known, and estimate the outlier parameters $\omega_1, \omega_2, \ldots, \omega_k$ and the time series parameters simultaneously using the model

$$Z_t = \sum_{j=1}^{k} \omega_j \nu_j(B) I_j^{(T_j)} + \frac{\theta(B)}{\phi(B)} a_t, \tag{10.3.22}$$

where $\nu_j(B) = 1$ for an AO and $\nu_j(B) = \theta(B)/\phi(B)$ for an IO at $t = T_j$. This result leads to the new residuals

$$\hat{e}_t^{(1)} = \hat{\pi}^{(1)}(B)\left[Z_t - \sum_{j=1}^{k} \hat{\omega}_j \hat{\nu}_j(B) I_t^{(T_j)}\right]. \tag{10.3.23}$$

A revised estimate of σ_a^2 can then be calculated.

Steps 2 through 4 are repeated until all outliers are identified and their impacts simultaneously estimated. Thus, we have the fitted outlier model

$$Z_t = \sum_{j=1}^{k} \hat{\omega}_j \hat{\nu}_j(B) I_t^{(T_j)} + \frac{\hat{\theta}(B)}{\hat{\phi}(B)} a_t, \tag{10.3.24}$$

where $\hat{\omega}_j$, $\hat{\phi}(B) = (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p)$ and $\hat{\theta}(B) = (1 - \hat{\theta}_1 B - \cdots - \hat{\theta}_q B^q)$ are obtained in the final iteration.

10.4 Examples of Outlier Analysis

The above outlier detection procedure can be easily implemented in any existing intervention analysis software or linear regression package. Both time series software SCA (1992) and AUTOBOX have implemented the procedure and make the analysis easier. A computer program based on a modification of the above procedure was also written by Bell (1983). We illustrate the following examples by using AUTOBOX and a standard regression package.

EXAMPLE 10.5 The outlier detection procedure discussed in Section 10.3 has been applied using software AUTOBOX to the U.S. quarterly series of beer production between 1975 and 1982, which was fitted earlier by a seasonal ARIMA(0, 0, 0) × (0, 1, 1)$_4$ model in Chapter 8. The result indicates that there are no outliers evident with a significance level .05 in the series. To check the efficiency of the proposed procedure, we artificially contaminated the series by replacing the original observation $Z_{12} = 36.54$ with a new observation $Z_{12} = 56.54$, which could be due to a typing error. We then apply the procedure to this outlier contaminated series, which is plotted in Figure 10.6.

The analysis produces the following result:

Detected Outliers

Iteration	Time	Type	Magnitude ($\hat{\omega}$)
1	12	AO	16.26
2	27	IO	−2.31

Thus, the procedure correctly identifies the AO at $t = 12$ in the first iteration. Although Z_{27} has also been detected as an IO in the second iteration, the effect is much smaller.

EXAMPLE 10.6 Series W1 was analyzed in Sections 6.2 and 7.6 resulting in an AR(1) model

$$(1 - .43B)Z_t = 1.04 + a_t \tag{10.4.1}$$

with $\hat{\sigma}_a^2 = .21$. Inspection of the residuals from the fitted model indicates the possible existence of a number of outliers. The series is the daily average number of defects per truck

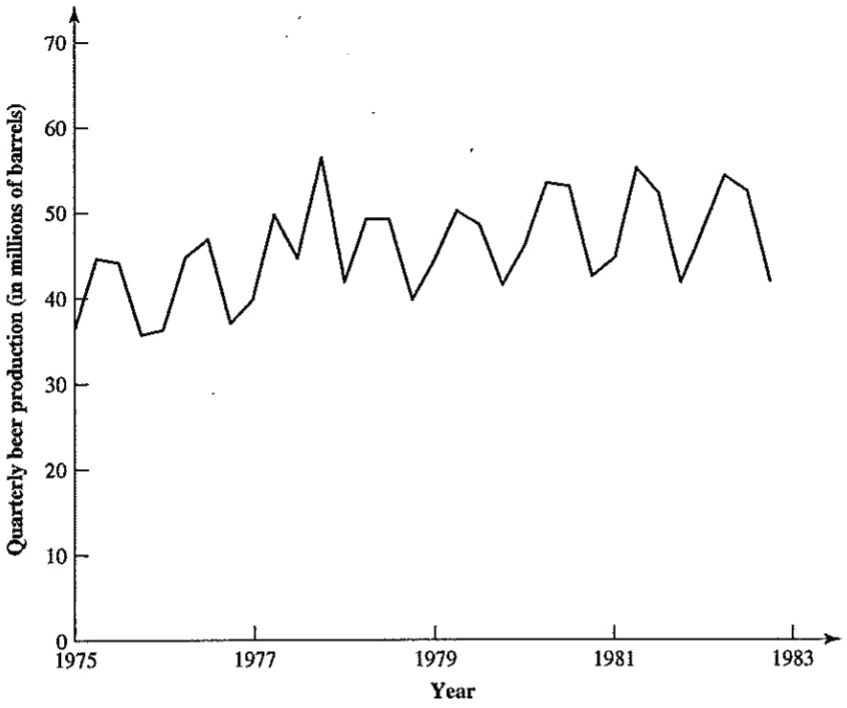


FIGURE 10.6 Outlier contaminated U.S. quarterly series of beer production between 1975 and 1982.

found in the final inspection at the end of the assembly line of a truck manufacturing plant. To ensure the maintenance of quality, the detection of outliers is always an important task in quality control. In this example, we apply the previous outlier detection procedure to the data set and obtain the following results:

Detected Outliers

Iteration	Time	Type
1	36	AO
2	9	IO
3	7	AO
4	4	IO

Thus, we consider the following outlier model:

$$Z_t = \theta_0 + \omega_1 I_t^{(36)} + \omega_2 \frac{1}{(1-\phi B)} I_t^{(9)} + \omega_3 I_t^{(7)} + \omega_4 \frac{1}{(1-\phi B)} I_t^{(4)} + \frac{1}{(1-\phi B)} a_t$$

$$= \theta_0 + \omega_1 I_t^{(36)} + \omega_3 I_t^{(7)} + \frac{1}{(1-\phi B)}\left(\omega_2 I_t^{(9)} + \omega_4 I_t^{(4)} + a_t\right) \qquad (10.4.2)$$

Simultaneous estimation of the previous parameters yields

$$Z_t = \underset{(.11)}{1.14} + \underset{(.11)}{1.39}I_t^{(36)} + \underset{(.37)}{.99}I_t^{(7)} + \frac{1}{\underset{(.11)}{(1 - .28B)}}\left(-\underset{(.19)}{.61}I_t^{(9)} + \underset{(.31)}{.66}I_t^{(4)} + a_t\right) \tag{10.4.3}$$

and $\hat{\sigma}_a^2 = .11$. Comparing (10.4.3) with (10.4.1), we see that a reduction of about 100% in the estimated variance $\hat{\sigma}_a^2$ of the a_t's occurs, from .21 to .11, when the effects of the four outliers are taken into account. Furthermore, the change in the estimate of the autoregressive parameter is also substantial, decreasing from .43 to a much smaller value of .28. For a production process under proper quality control, one would expect the series of defects to be random white noise. That would occur if a smaller critical value C were chosen and a few more additional outliers were identified.

10.5 Model Identification in the Presence of Outliers

As indicated earlier, the intervention analysis is used when the timing and causes of interruptions are known. Because the timing and causes of interruptions are sometimes unknown, in addition to the additive outlier (AO) and innovational outlier (IO), the following level shift (LS) and temporary change (TC) can also be incorporated in the iterative outlier detection procedure discussed in Section 10.3:

$$\text{LS: } Z_t = X_t + \frac{1}{(1 - B)}\omega_L I_t^{(T)} \tag{10.5.1}$$

and

$$\text{TC: } Z_t = X_t + \frac{1}{(1 - \delta B)}\omega_C I_t^{(T)}, \tag{10.5.2}$$

where we recall that X_t is the underlying outlier-free process. For empirical examples that contain possible LS and TC situations, we refer readers to Chen and Liu (1991, 1993).

Other than the model based iterative detection procedure, many other time series outlier detection procedures have been introduced in the literature. For example, because the underlying outlier-free model is often unknown and outliers can distort model identification as shown in the following, Lee and Wei (1995) proposed a model-independent outlier detection procedure. Other detection procedures include Tsay (1988), Ledolter (1990), and Ljung (1993).

After outliers are identified, one can adjust data using (10.3.18) or (10.3.20) and then pursue the analysis based on the adjusted data. A more fruitful approach, however, is to search for the causes of the identified outliers and further fine-tune the fitted model in (10.3.24). This process is appropriate not only for parameter estimation but also for model checking and forecasting. In searching for the causes of an outlier, one may find the nature of the disturbance.

For example, some outliers may turn out to be important intervention variables due to some policy changes with which the analyst was unfamiliar and hence were overlooked during the preliminary stage of data analysis. Thus, instead of using the adjusted data by removing the effects of outliers, the analyst should incorporate the information into the model by introducing proper intervention variables and response functions as discussed in Sections 10.1 and 10.2. This explicit form of a combined intervention-outlier model is usually more useful in forecasting and control than the univariate fitted model based on the outlier-adjusted data.

EXAMPLE 10.7 As an example let us consider Series W14, the monthly airline passengers in the United States from January 1995 to March 2002 plotted in Figure 10.7.

Without looking at the plot and blindly applying the outlier detection method introduced in the chapter to the series, we obtain the following result:

Detected Outliers

Iteration	Time	Type
1	81	TC
2	82	TC

If we use a significance level less than .01, then the only outlier found is the observation at time 81 that corresponds to September 2001, the month of the World Trade Center tragedy in New York City, which is clearly an intervention event. The outlier procedure not only detects the event but also suggests the form of the intervention.

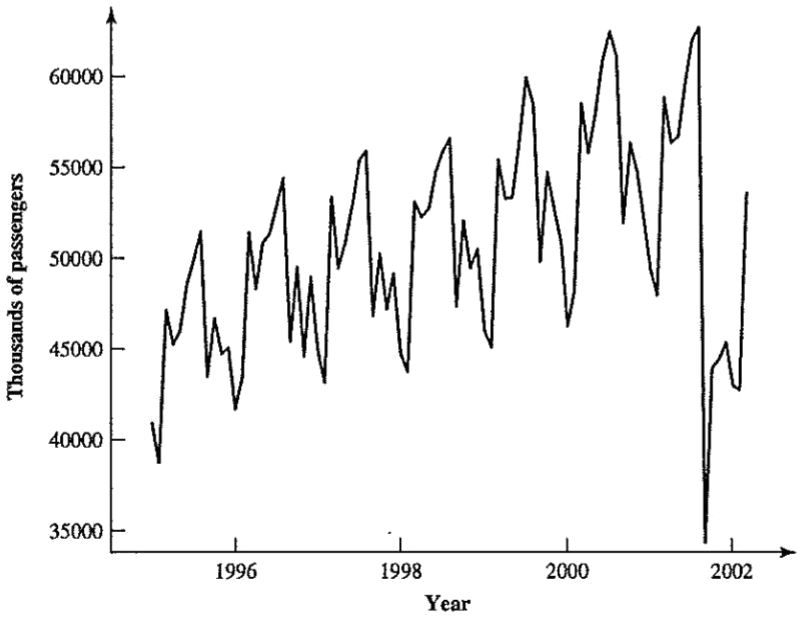


FIGURE 10.7 The monthly airline passengers in the United States from January 1995 to March 2002 (Series W14).

The standard time series modeling on the subseries from January 1995 to August 2001 suggests the $ARIMA(2, 0, 0) \times (0, 1, 0)_{12}$ seasonal model:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})Z_t = a_t. \tag{10.5.3}$$

Thus, we will combine model (10.5.3) and the information about the observation at time 81 in the intervention model

$$Z_t = \frac{\omega}{(1 - \delta B)} I_t + \frac{1}{(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})} a_t, \tag{10.5.4}$$

where

$$I_t = \begin{cases} 0, & t < 81 \text{ (Sept. 2001)} \\ 1, & t \geq 81 \text{ (Sept. 2001).} \end{cases}$$

The estimation results are

Parameter	Estimate	St. Error
ω	−18,973.5	1299.3
δ	.76	.06
ϕ_1	.62	.1
ϕ_2	.21	.1

The effect of the September 11, 2001, tragedy on the airline industry is clearly devastating.

The iterative outlier detection procedure is based on the assumption that the underlying model for the outlier-free series is either known or can be identified. In practice, however, the underlying model is usually unknown and has to be identified through the standard analysis of sample statistics, such as sample ACF, sample PACF, sample IACF, and ESACF. Thus, it should be emphasized that the detection procedure of Section 10.3 works well only when the outlier effect is moderate and will not overshadow and obscure the underlying pattern contained in the sample statistics of an outlier-free series. In more serious cases, outlier contamination may make model identification impossible. For example, the blowfly data exhibit a very clear autocorrelation pattern of AR(1) process, as shown in Table 6.6. The estimate of the AR(1) parameter ϕ_1 was found to be .73 in Table 7.3. The data are listed as Series W3 in the appendix. Now, suppose, due to a typing error, that the value 22221 was used for Z_{20} instead of 2221. Figure 10.8 shows the series contaminated by this outlier at Z_{20}. Its sample ACF, sample PACF, and ESACF are computed and shown in Table 10.2. The results all indicate a white noise phenomenon. The underlying AR(1) characteristics have been completely washed out by this single outlier. A white noise series is itself an outlier in empirical time series. Thus, when encountering a white noise phenomenon in studying sample statistics such as sample ACF, one should first examine whether there is an obvious outlier in the data. Such an outlier can be easily detected by plotting the series, which is always the first aid for any data and outlier analysis.

To reduce the adverse effect of outliers on the parameter estimation, a number of robust estimation methods have been introduced in the literature. In terms of estimation of the ACF, Chan and Wei (1992) introduced the following *α-trimmed* sample autocorrelation function (TACF).

Let $Z_{(1)} \leq Z_{(2)}, \ldots, \leq Z_{(n)}$ be the ordered observations of the given time series $Z_1, Z_2, \ldots, Z_n$. The *α-trimmed* sample autocorrelation function is defined by

$$\hat{\rho}_k^{(\alpha)} = \frac{\hat{\gamma}_k^{(\alpha)}}{\hat{\gamma}_0^{(\alpha)}}, \tag{10.5.5}$$

where

$$\hat{\gamma}_k^{(\alpha)} = \frac{1}{\sum_{t=k+1}^{n} L_{t-k}^{(\alpha)} L_t^{(\alpha)}} \left\{ \sum_{t=k+1}^{n} \left(Z_{t-k} - \bar{Z}^{(\alpha)}\right)\left(Z_t - \bar{Z}^{(\alpha)}\right) L_{t-k}^{(\alpha)} L_t^{(\alpha)} \right\}, \tag{10.5.6}$$

$$\bar{Z}^{(\alpha)} = \left(\frac{\sum_{t=1}^{n} Z_t L_t^{(\alpha)}}{\sum_{t=1}^{n} L_t^{(\alpha)}} \right), \tag{10.5.7}$$

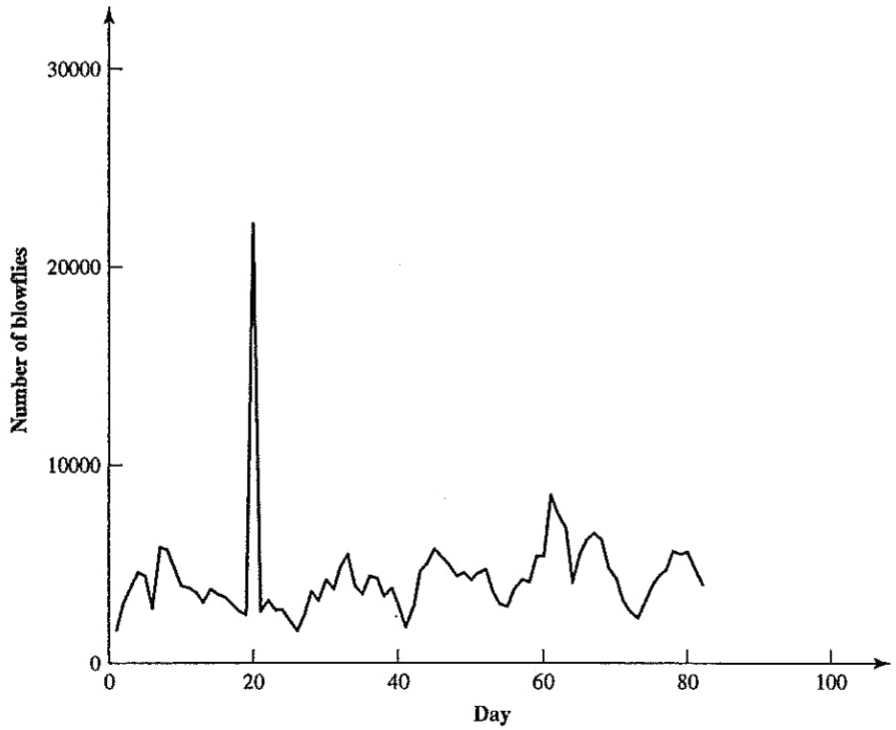


FIGURE 10.8 Outlier contaminated blowfly data.

TABLE 10.2 Sample ACF, sample PACF, and ESACF for an outlier contaminated AR(1) series.

(a) $\hat{\rho}_k$

1–10	.09	.05	−.03	−.05	−.09	−.12	−.14	−.02	−.06	−.01
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.12	.12
11–20	.02	.06	.10	.04	.03	.07	.06	.00	−.00	−.08
St.E.	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12

(b) $\hat{\phi}_{kk}$

1–10	.09	.04	−.04	−.04	−.08	−.11	−.12	−.04	−.06	−.02
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11
11–20	−.01	.02	.06	−.01	.01	.06	.06	.01	.03	−.04
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11

(c) ESACF

AR \ MA	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0
1	X	0	0	0	0	0	0	0	0	0	0
2	X	X	0	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0	0	0

$$L_t^{(\alpha)} = \begin{cases} 0, & \text{if } Z_t \le Z_{(g)} \text{ or } Z_t \ge Z_{(n-g+1)}, \\ 1, & \text{otherwise,} \end{cases} \tag{10.5.8}$$

and g is the integer part of $[\alpha n]$ and $0 \le \alpha \le .05$. In other words, the TACF $\hat{\rho}_k^{(\alpha)}$ in (10.5.5) trimmed α% extreme values from the computation of the sample ACF.

Because $\{L_t^{(\alpha)}\}$ is a deterministic sequence of real numbers (0 or 1), the asymptotic theory for time series containing amplitude modulated observations can be applied to the TACF. Following Dunsmuir and Robinson (1981), under the white noise model it can be easily shown that the $\sqrt{n}\hat{\rho}_k^{(\alpha)}$ for $k \ge 1$ are asymptotically independent normal random variates with asymptotic variances $1/\nu(k)$, where

$$\nu(k) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=k+1}^{n} L_{t-k}^{(\alpha)} L_t^{(\alpha)}. \tag{10.5.9}$$

Thus, we can approximate the standard error of the TACF by

$$S_{\hat{\rho}_k^{(\alpha)}} = \frac{1}{\sqrt{n\hat{\nu}(k)}}, \tag{10.5.10}$$

where

$$\hat{\nu}(k) = \frac{1}{n}\sum_{t=k+1}^{n} L_{t-k}^{(\alpha)} L_t^{(\alpha)}. \tag{10.5.11}$$

Note that $\hat{\nu}(k) < 1$ when $\alpha > 0$. Thus, $S_{\hat{\rho}_k^{(\alpha)}} > S_{\hat{\rho}_k}$ for the white noise model, which is expected because the information in the trimmed series is always less than that in the original series under the outlier-free situation.

The choice of α is very important for the TACF. If we choose a large value for α, the TACF is no longer an efficient estimator because $\hat{\nu}(k)$ becomes smaller as α increases. On the other hand, if we select a very small value for α, the purpose of the estimator (to trim out possible outliers in the series) might not be fully fulfilled. In practice, we recommend using $\alpha = 1$ to 2% for general series, $\alpha = 3$ to 5% for medium contaminated series, and $\alpha = 6$ to 10% for heavily contaminated series.

Other robust estimations of ACF include the jackknife estimator proposed by Quenouilli (1949) and the robust sample partial autocorrelation and autocorrelation functions by Masarotto (1987). Chan and Wei (1992) show that in estimating the ACF, the simple TACF in (10.5.5) not only compares favorably with other robust estimators in terms of the root mean squared error, but it also performs well in the presence of both additive and innovational outliers.

When there are outliers in the series, all the proposed robust ACF estimations outperform the standard sample ACF. Because, in practice, we may not know whether outliers exist, for proper model identification the use of robust estimates of autocorrelations is recommended.

There are many studies on time series outliers. For other useful discussions on robust methods and model identification in the presence of outliers, we refer readers to Tsay (1986), Abraham and Chuang (1989), Chen and Liu (1991, 1993), Wei and Wei (1998), and Li (2003).

In closing the chapter, we note that intervention analysis is a useful technique to study the impacts of intervention and outliers that normally cause the level shift of a time series. On occasion, however, the shift occurs in the variance of a series. Different techniques are needed to study this problem. Because of the length of the chapter, we will not discuss this interesting problem and refer readers to Wichern, Miller, and Hsu (1976) and Abraham and Wei (1984).

EXERCISES

10.1 Consider the following response functions of interventions:

(1) $\omega_0 p_t^{(T)}$

(2) $\dfrac{\omega_0}{(1-\delta B)} p_t^{(T)}$

(3) $\dfrac{\omega_0}{(1-B)} p_t^{(T)}$

(4) $\left[\dfrac{\omega_0}{(1-\delta B)} + \dfrac{\omega_1}{(1-B)}\right] p_t^{(T)}$

(5) $\left[\omega_0 + \frac{\omega_1 B}{(1 - \delta B)}\right] p_t^{(T)}$

(6) $\left[\omega_0 + \frac{\omega_1}{(1 - \delta B)} + \frac{\omega_2}{(1 - B)}\right] p_t^{(T)}$

d(7) $\frac{\omega_0}{(1 - \delta B)(1 - B)} p_t^{(T)}$

(a) Plot the above response functions.

(b) Discuss possible applications of the various interventions.

10.2 Find a time series that was affected by some external events. Carry out intervention analysis and submit a written report of your analysis.

10.3 Perform and report an iterative outlier analysis for the following time series (read across):

.561	.664	.441	.635	1.083	.961	.057
1.349	1.100	.544	−.132	−1.567	−1.277	−1.192
−1.346	1.401	.037	−.272	−.591	−.542	−.574
−.742	−1.416	.549	−1.446	1.883	1.050	1.134
1.947	−1.839	.803	.321	.470	−.279	1.913
−.785	.236	.147	−.690	.667	−.270	.221
−.633	−.245	−1.705	−1.648	−.723	−1.316	−.642
−.510	−.065	−.553	−1.058	−14.960	−.764	−.556
−.079	.047	−.203	.244	−.407	−.438	−1.616
−.231	−.371	−1.643	.203	−.338	−.830	−1.749
−1.025	−2.218	.360	−1.332	.199	−.034	.621
2.008	−.154	.308				

10.4 Find a time series of your interest, which was likely contaminated by some outliers or interventions.

(a) Fit a best ARIMA model for the series.

(b) Detect all possible outlier and intervention variables, and build an outlier-intervention model for the series.

(c) Compare and discuss your findings in parts (a) and (b).

10.5 (a) Find a time series model for the data given in Exercise 10.3 using the sample ACF, $\hat{\rho}_k$.

(b) Find a time series model for the same data in part (a) using the 5% trimmed sample autocorrelation function, $\hat{\rho}_k^{(\alpha)}$.

(c) Compare and comment your results in parts (a) and (b).

11
Fourier Analysis

The time series approach presented in earlier chapters, which uses functions such as autocorrelations and partial autocorrelations to study the evolution of a time series through parametric models, is known as *time domain analysis*. An alternative approach, which tries to describe the fluctuation of time series in terms of sinusoidal behavior at various frequencies, is known as *frequency domain analysis*. In this chapter, we introduce some basic concepts of Fourier analysis, which are fundamental for the frequency domain analysis.

11.1 General Concepts

Often it is more convenient to represent a function by a set of elementary functions called a *basis* such that all functions under study can be written as linear combinations of the elementary functions in the basis. One very useful set of these elementary functions involves the sines and cosines or complex exponentials. In this chapter, we study how to construct an arbitrary function using these sinusoids. The study of this subject is often referred to as *Fourier analysis* due to the 18th-century French mathematician J. B. J. Fourier, who claimed in 1807 that any periodic function could be represented as a series of harmonically related sinusoids. Later he also obtained a representation for nonperiodic functions as weighted integrals of sinusoids that are not all harmonically related.

The development of the techniques of Fourier analysis has a long history involving many mathematicians, including E. Euler, D. Bernoulli, P. S. Laplace, J. L. Lagrange, and P. L. Dirichlet, in addition to Fourier. Although the original work of Fourier analysis was solely concerned with phenomena in continuous time, such as the physical motion of vibrating string and heat propagation and diffusion, the essential ideas of Fourier analysis carry over to phenomena in discrete time. For our purposes, we concentrate on Fourier analysis of discrete time functions or sequences. Specifically, we establish the orthogonality property of sine-cosine functions and the complex exponentials in Section 11.2. Using this property, we develop the Fourier series representations for finite sequences and periodic sequences in Sections 11.3 and 11.4. The Fourier transform of an arbitrary nonperiodic sequence is discussed in Section 11.5. The Fourier representation of a continuous function is presented in Section 11.6. The fast Fourier transform used in the actual computation of the Fourier representation is presented in Section 11.7. With the proper understanding of these techniques, we are then ready for the frequency domain approach of time series analysis in Chapters 12 and 13.

11.2 Orthogonal Functions

Let $\phi_k(t)$ and $\phi_j(t)$ be complex valued functions defined on the domain D, which is a subset of the real line. Discrete-time functions $\phi_k(t)$ and $\phi_j(t)$ defined on a discrete set are said to be orthogonal if

$$\sum_{t\in D}\phi_k(t)\phi_j^*(t)\begin{cases}=0, & k\neq j,\\ \neq 0, & k=j,\end{cases} \tag{11.2.1}$$

where $\phi_j^*(t)$ denotes the complex conjugate of $\phi_j(t)$. Continuous time functions $\phi_k(t)$ and $\phi_j(t)$ defined on an interval of the real line are orthogonal if

$$\int_D \phi_k(t)\phi_j^*(t)\,dt\begin{cases}=0, & k\neq j,\\ \neq 0, & k=j.\end{cases} \tag{11.2.2}$$

In this book, we concentrate on the discrete-time functions, which are also called *sequences*. There are many kinds of orthogonal functions. In this section, we introduce two equivalent systems of orthogonal functions that are useful in time series analysis.

Suppose that the trigonometric sine and cosine functions $\sin(2\pi kt/n)$ and $\cos(2\pi kt/n)$ are defined on a finite number of n points, i.e., for $t = 1, 2, \ldots, n$. For $k = 0, 1, 2, \ldots, [n/2]$, where $[x]$ is the greatest integer less than or equal to x, the system

$$\left\{\sin\left(\frac{2\pi kt}{n}\right), \cos\left(\frac{2\pi kt}{n}\right): k = 0, 1, \ldots, \left[\frac{n}{2}\right]\right\} \tag{11.2.3}$$

contains exactly n nonzero functions, i.e., none of which is identically zero, which follows from the sine function being identically zero for $k = 0$ and for $k = [n/2]$ if n is even. More specifically, $[n/2] = n/2$ when n is even, and the system contains $\cos(2\pi 0t/n) = 1$, $\sin(2\pi kt/n)$ and $\cos(2\pi kt/n)$ for $k = 1, 2, \ldots, n/2 - 1$, and $\cos(2\pi(n/2)t/n) = \cos(\pi t) = (-1)^t$. If n is odd, then $[n/2] = (n - 1)/2$ and the system consists of $\cos(2\pi 0t/n) = 1$, $\sin(2\pi kt/n)$ and $\cos(2\pi kt/n)$ for $k = 1, 2, \ldots, (n - 1)/2$. In either case, it has exactly n functions in the system.

Next, we show that the system (11.2.3) is actually a collection of orthogonal functions. To do so, we use the Euler relation

$$e^{i\omega} = \cos\omega + i\sin\omega \tag{11.2.4}$$

and the identities

$$\sin\omega = \frac{e^{i\omega} - e^{-i\omega}}{2i}, \tag{11.2.5}$$

$$\cos\omega = \frac{e^{i\omega} + e^{-i\omega}}{2}. \tag{11.2.6}$$

Then

$$\begin{aligned}\sum_{t=1}^{n} e^{i\omega t} &= e^{i\omega}\left(\frac{1-e^{i\omega n}}{1-e^{i\omega}}\right) = e^{i\omega}\left(\frac{e^{i\omega n}-1}{e^{i\omega}-1}\right)\\ &= e^{i\omega}\left[\frac{e^{i\omega n/2}(e^{i\omega n/2}-e^{-i\omega n/2})/2i}{e^{i\omega/2}(e^{i\omega/2}-e^{-i\omega/2})/2i}\right]\\ &= e^{i\omega(n+1)/2}\frac{\sin(\omega n/2)}{\sin(\omega/2)}\\ &= \cos\left(\frac{\omega(n+1)}{2}\right)\frac{\sin(\omega n/2)}{\sin(\omega/2)} + i\sin\left(\frac{\omega(n+1)}{2}\right)\frac{\sin(\omega n/2)}{\sin(\omega/2)}. \qquad (11.2.7)\end{aligned}$$

But

$$\sum_{t=1}^{n} e^{i\omega t} = \sum_{t=1}^{n}\cos\omega t + i\sum_{t=1}^{n}\sin\omega t.$$

Hence, we have

$$\sum_{t=1}^{n}\cos\omega t = \cos\left(\frac{\omega(n+1)}{2}\right)\frac{\sin(\omega n/2)}{\sin(\omega/2)}, \qquad (11.2.8)$$

$$\sum_{t=1}^{n}\sin\omega t = \sin\left(\frac{\omega(n+1)}{2}\right)\frac{\sin(\omega n/2)}{\sin(\omega/2)}. \qquad (11.2.9)$$

Let $\omega = 2\pi k/n$. Because

$$\frac{\sin(\omega n/2)}{\sin(\omega/2)} = \frac{\sin(\pi k)}{\sin(\pi k/n)} = \begin{cases} n, & k = 0,\\ 0, & k \neq 0,\end{cases}$$

and $\cos\omega = 1$ and $\sin\omega = 0$ for $\omega = 0$, Equations (11.2.8) and (11.2.9) imply that

$$\sum_{t=1}^{n}\cos\left(\frac{2\pi kt}{n}\right) = \begin{cases} n, & k = 0,\\ 0, & k \neq 0;\end{cases} \qquad (11.2.10)$$

$$\sum_{t=1}^{n}\sin\left(\frac{2\pi kt}{n}\right) = 0, \qquad k = 0, 1, \ldots, [n/2]. \qquad (11.2.11)$$

Next, using the trigonometric identities

$$\cos\omega\cos\lambda = \tfrac{1}{2}\{\cos(\omega+\lambda)+\cos(\omega-\lambda)\}, \tag{11.2.12a}$$

$$\sin\omega\sin\lambda = \tfrac{1}{2}\{\cos(\omega-\lambda)-\cos(\omega+\lambda)\}, \tag{11.2.12b}$$

$$\sin\omega\cos\lambda = \tfrac{1}{2}\{\sin(\omega+\lambda)+\sin(\omega-\lambda)\}, \tag{11.2.12c}$$

and from Equations (11.2.10) and (11.2.11), we obtain

$$\sum_{t=1}^{n}\cos\left(\frac{2\pi kt}{n}\right)\cos\left(\frac{2\pi jt}{n}\right) = \begin{cases} n, & k=j=0 \text{ or } n/2 \ (n \text{ even}), \\ n/2, & k=j\neq 0 \text{ or } n/2 \ (n \text{ even}), \\ 0, & k\neq j; \end{cases} \tag{11.2.13}$$

$$\sum_{t=1}^{n}\sin\left(\frac{2\pi kt}{n}\right)\sin\left(\frac{2\pi jt}{n}\right) = \begin{cases} 0, & k=j=0 \text{ or } n/2 \ (n \text{ even}), \\ n/2, & k=j\neq 0 \text{ or } n/2 \ (n \text{ even}), \\ 0, & k\neq j; \end{cases} \tag{11.2.14}$$

$$\sum_{t=1}^{n}\sin\left(\frac{2\pi kt}{n}\right)\cos\left(\frac{2\pi jt}{n}\right) = 0 \text{ for all } k \text{ and } j. \tag{11.2.15}$$

This result shows that the system (11.2.3) is a set of orthogonal functions.

From Equations (11.2.5) and (11.2.6), it is clear that the system of the trigonometric functions in (11.2.3) can also be represented in complex form. This form is much simpler and compact and is very useful in certain applications. In terms of this representation, the corresponding system contains the following complex exponentials:

$$\left\{ e^{i2\pi kt/n} : -\frac{n}{2}+1 \le k \le \frac{n}{2} \text{ if } n \text{ is even and } -\frac{(n-1)}{2} \le k \le \frac{(n-1)}{2} \right\}, \tag{11.2.16}$$
$$\text{if } n \text{ is odd}$$

which again consist of exactly n functions. To show the orthogonal property of these complex exponentials, we note that

$$\sum_{t=1}^{n} e^{i2\pi kt/n} = \begin{cases} n, & k=0, \\ 0, & k\neq 0, \end{cases} \tag{11.2.17}$$

where we use that

$$\sum_{t=1}^{n} e^{i2\pi kt/n} = e^{i2\pi k/n}\left[\frac{1-\left(e^{i2\pi k/n}\right)^n}{1-e^{i2\pi k/n}}\right]$$

and $(e^{i2\pi k/n})^n = e^{i2\pi k} = 1$. Applying Equation (11.2.17) we immediately obtain the following result:

$$\sum_{t=1}^{n} e^{i2\pi kt/n} e^{-i2\pi jt/n} = \begin{cases} n, & k = j, \\ 0, & k \neq j. \end{cases} \tag{11.2.18}$$

That is, the system (11.2.16) is orthogonal.

11.3 Fourier Representation of Finite Sequences

Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of n numbers. This sequence can be regarded as a set of coordinates of a point in an n-dimensional space. In vector analysis, we often construct a set of vectors called a *basis* so that every vector in the space can be written as a linear combination of the elements of the basis. For a given n-dimensional space, it is known that any set of n orthogonal vectors forms a basis. Thus, the given sequence of n numbers, $\{Z_t\}$, can be written as a linear combination of the orthogonal trigonometric functions given in the system (11.2.3). That is,

$$Z_t = \sum_{k=0}^{[n/2]} \left[a_k \cos\left(\frac{2\pi kt}{n}\right) + b_k \sin\left(\frac{2\pi kt}{n}\right) \right], \qquad t = 1, 2, \ldots, n. \tag{11.3.1}$$

Equation (11.3.1) is called the Fourier series of the sequence Z_t. The a_k and b_k are called Fourier coefficients. Using the orthogonal property of the trigonometric functions given in (11.2.13) to (11.2.15), we obtain the a_k and b_k by multiplying $\cos(2\pi kt/n)$ and $\sin(2\pi kt/n)$ on both sides of (11.3.1), respectively, and then summing over $t = 1, 2, \ldots, n$. To avoid possible confusion, the reader can replace the index k by j in Equation (11.3.1) in this operation. Thus, we have

$$a_k = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=1}^{n} Z_t \cos\left(\frac{2\pi kt}{n}\right), & k = 0, \text{ and } k = n/2 \text{ if } n \text{ is even}, \\ \dfrac{2}{n} \displaystyle\sum_{t=1}^{n} Z_t \cos\left(\frac{2\pi kt}{n}\right), & k = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right], \end{cases}$$

$$b_k = \frac{2}{n} \sum_{t=1}^{n} Z_t \sin\left(\frac{2\pi kt}{n}\right), \qquad k = 1, 2, \ldots, \left[\frac{n-1}{2}\right]. \tag{11.3.2}$$

Let $\omega_k = 2\pi k/n$, $k = 0, 1, \ldots, [n/2]$. These frequencies are called the Fourier frequencies. Using the system of complex exponentials given in (11.2.16), we can also write the Fourier series of Z_t as

$$Z_t = \begin{cases} \displaystyle\sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is odd}, \\ \displaystyle\sum_{k=-n/2+1}^{n/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is even}, \end{cases} \tag{11.3.3}$$

where the Fourier coefficients c_k are given by

$$c_k = \frac{1}{n}\sum_{t=1}^{n} Z_t e^{-i\omega_k t}. \tag{11.3.4}$$

Equations (11.3.1) and (11.3.3) imply that

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t)$$

$$= \begin{cases} \sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is odd,} \\ \sum_{k=-(n/2)+1}^{n/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is even.} \end{cases}$$

Thus, from the relationships (11.2.5) and (11.2.6), the Fourier coefficients a_k, b_k, and c_k are easily seen to be related as

$$\begin{cases} c_0 = a_0;\ c_{n/2} = a_{n/2} \text{ (even } n), \\ c_k = \dfrac{a_k - ib_k}{2}, \\ c_{-k} = c_k^* = \dfrac{a_k + ib_k}{2}. \end{cases} \tag{11.3.5}$$

The coefficient $c_0 = a_0 = \sum_{t=1}^{n} Z_t/n$ is often referred to as the d.c. value, i.e., the constant average value of the sequence.

The above material shows that any finite sequence can be written as a linear combination of the sine-cosine sequences or the complex exponentials.

11.4 Fourier Representation of Periodic Sequences

A general function $f(t)$ is said to be periodic with period P if there exists a positive constant P such that

$$f(t + P) = f(t), \tag{11.4.1}$$

for all t. It is obvious that a function that is periodic with period P is also periodic with periods $2P, 3P, \ldots$. The smallest positive value of P for which (11.4.1) holds is called the fundamental period or simply the period of the function.

Suppose that the sequence (or the discrete time function) Z_t is periodic with period n, where n is a positive integer. That is,

$$Z_{t+n} = Z_t, \tag{11.4.2}$$

for all integer t. A fundamental phenomenon of a periodic function is that the function is uniquely determined by its pattern in a range of one period. Outside that range the function is just a repetition of the pattern in that range. Thus, a periodic sequence with period n is uniquely defined by its values at $t = 1, 2, \ldots,$ and n.

Following the result of the Fourier series of a finite sequence in Section 11.3, we can write Z_t for $t = 1, 2, \ldots, n$ as the linear combination of the orthogonal sine and cosine functions as follows:

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t) \tag{11.4.3}$$

$$a_k = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=1}^{n} Z_t \cos \omega_k t, & k = 0, \text{ and } k = \dfrac{n}{2} \text{ if } n \text{ is even}, \\ \dfrac{2}{n} \displaystyle\sum_{t=1}^{n} Z_t \cos \omega_k t, & k = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right], \end{cases} \tag{11.4.4}$$

$$b_k = \frac{2}{n} \sum_{t=1}^{n} Z_t \sin \omega_k t, \qquad k = 1, 2, \ldots, \left[\frac{n-1}{2}\right] \tag{11.4.5}$$

with $\omega_k = 2\pi k/n$. Similarly, we can also write Z_t for $t = 1, 2, \ldots, n$ as the linear combination of complex exponentials

$$Z_t = \begin{cases} \displaystyle\sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is odd}, \\ \displaystyle\sum_{k=-(n/2)+1}^{n/2} c_k e^{i\omega_k t}, & \text{if } n \text{ is even}, \end{cases} \tag{11.4.6}$$

$$c_k = \frac{1}{n} \sum_{t=1}^{n} Z_t e^{-i\omega_k t}. \tag{11.4.7}$$

The Fourier coefficients a_k and b_k in (11.4.4) and (11.4.5) and c_k in (11.4.7) are related by the equations in (11.3.5).

It is easy to see that the sine-cosine functions $\sin \omega_k t$ and $\cos \omega_k t$ in (11.4.3) and the complex exponentials $e^{i\omega_k t}$ in (11.4.6) are periodic with period n, which leads to the desired consequence that

$$Z_{t+jn} = Z_t,$$

for all integers t and j. In other words, the Equations (11.4.3) and (11.4.6) are valid for all integers t.

We have demonstrated how to represent an arbitrary periodic sequence of period n as a linear combination of n trigonometric sequences or n complex exponentials. The smallest positive value of n for which the Fourier series representation in (11.4.3) and (11.4.6) holds is called the fundamental period, and the corresponding value $2\pi/n$ is called the fundamental frequency. The terms for $k = +1$ and $k = -1$ in the above representation both have the same fundamental period equal to n (and hence the same fundamental frequency $2\pi/n$) and are collectively referred to as the first harmonic components. More generally, the terms for $k = +j$ and $k = -j$ both have the frequency $j(2\pi/n)$ and are referred to as the jth harmonic components. Thus, all the terms in the Fourier series representation have frequencies that are multiples of the same fundamental frequency, $2\pi/n$, and hence are harmonically related.

For a given periodic sequence Z_t of period n, the energy associated with the sequence in one period is defined as

$$\sum_{t=1}^{n} Z_t^2. \tag{11.4.8}$$

Now, multiplying by Z_t on the both sides of (11.4.3), summing from $t = 1$ to $t = n$, and using the relation (11.4.4) and (11.4.5), we have

$$\begin{aligned}\sum_{t=1}^{n} Z_t^2 &= \sum_{k=0}^{[n/2]}\left[a_k \sum_{t=1}^{n} Z_t \cos \omega_k t + b_k \sum_{t=1}^{n} Z_t \sin \omega_k t\right] \\ &= \begin{cases} na_0^2 + \dfrac{n}{2}\displaystyle\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2), & \text{if } n \text{ is odd,} \\ na_0^2 + \dfrac{n}{2}\displaystyle\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2) + na_{n/2}^2, & \text{if } n \text{ is even.} \end{cases}\end{aligned} \tag{11.4.9}$$

Equation (11.4.9) is known as Parseval's relation for Fourier series. Equivalently, using Equations (11.4.6) and (11.4.7), we have the corresponding form of Parseval's relation as

$$\sum_{t=1}^{n} Z_t^2 = \begin{cases} n \displaystyle\sum_{k=-(n-1)/2}^{(n-1)/2} |c_k|^2, & \text{if } n \text{ is odd,} \\ n \displaystyle\sum_{k=-n/2+1}^{n/2} |c_k|^2, & \text{if } n \text{ is even,} \end{cases} \tag{11.4.10}$$

where $|c_k|^2 = c_k c_k^*$.

Equations (11.4.9) and (11.4.10) imply that the total energy for a periodic sequence over the whole time horizon $t = 0, +1, +2, \ldots$ is infinite. Hence, we consider its energy per unit time, which is called the power of the sequence. This equation is given by

$$\text{Power} = \begin{cases} a_0^2 + \frac{1}{2}\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2), & \text{if } n \text{ is odd,} \\ a_0^2 + \frac{1}{2}\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2) + a_{n/2}^2, & \text{if } n \text{ is even,} \end{cases} \tag{11.4.11}$$

or, equivalently,

$$\text{Power} = \begin{cases} \sum_{k=-(n-1)/2}^{(n-1)/2} |c_k|^2, & \text{if } n \text{ is odd,} \\ \sum_{k=-n/2+1}^{n/2} |c_k|^2, & \text{if } n \text{ is even.} \end{cases} \tag{11.4.12}$$

As noted above, the jth harmonic components include the terms for both $k = +j$ and $k = -j$ as they correspond to the same frequency $j(2\pi/n)$. Hence, we can interpret the quantity

$$\begin{cases} f_0 = c_0^2 = a_0^2; f_{n/2} = |c_{n/2}|^2 \text{ (even } n), \\ f_k = |c_{-k}|^2 + |c_k|^2 = 2|c_k|^2 = \frac{1}{2}(a_k^2 + b_k^2), \end{cases} \tag{11.4.13}$$

from the term in the Fourier series of Z_t at the kth frequency $\omega_k = 2\pi k/n$ as the contribution to the total power. The quantity f_k plotted as a function of ω_k shown in Figure 11.1 is called the power spectrum and describes how the total power is distributed over the various frequency components of the sequence Z_t.

EXAMPLE 11.1 Let $Z_1 = 1$, $Z_2 = 2$, $Z_3 = 3$, and $Z_{t+3j} = Z_t$, $t = 1, 2, 3$, $j = \pm 1, \pm 2, \pm 3, \ldots$. Clearly, the function is periodic, with period 3 as shown in Figure 11.2.

Now, the period $n = 3$, which is odd, and $[n/2] = 1$. The Fourier representation using (11.4.3) becomes

$$Z_t = a_0 + a_1 \cos\left(\frac{2\pi t}{3}\right) + b_1 \sin\left(\frac{2\pi t}{3}\right),$$

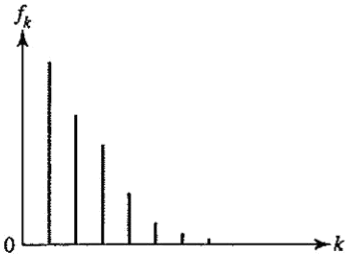


FIGURE 11.1 Power spectrum.

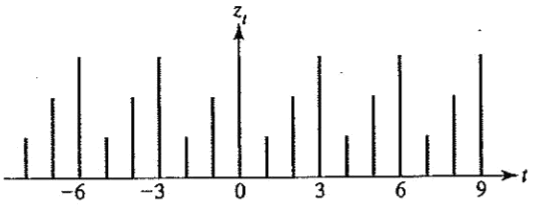


FIGURE 11.2 Periodic sequence in Example 11.1.

where the Fourier coefficients are obtained through (11.4.4) as

$$a_0 = \frac{1}{n}\sum Z_t = \bar{Z} = \frac{1}{3}(1 + 2 + 3) = 2,$$

$$a_1 = \frac{2}{3}\left[1\cos\left(\frac{2\pi}{3}\right) + 2\cos\left(\frac{4\pi}{3}\right) + 3\cos\left(\frac{6\pi}{3}\right)\right] = 1,$$

$$b_1 = \frac{2}{3}\left[1\sin\left(\frac{2\pi}{3}\right) + 2\sin\left(\frac{4\pi}{3}\right) + 3\sin\left(\frac{6\pi}{3}\right)\right] = -.5773503.$$

That is,

$$Z_t = 2 + \cos\left(\frac{2\pi t}{3}\right) - .5773503\sin\left(\frac{2\pi t}{3}\right), \qquad t = 1, 2, 3, 4, 5, \ldots.$$

Similarly, in terms of (11.4.6), we can also represent Z_t as

$$Z_t = c_{-1}e^{-i2\pi t/3} + c_0 + c_1 e^{i2\pi t/3}, \qquad t = 1, 2, 3, 4, \ldots,$$

where the coefficients are calculated using (11.4.7):

$$c_0 = \tfrac{1}{3}[1 + 2 + 3] = 2,$$

$$c_{-1} = \tfrac{1}{3}[e^{i2\pi/3} + 2e^{i4\pi/3} + 3e^{i6\pi/3}] = \tfrac{1}{2}(1 - .5773503i),$$

$$c_1 = \tfrac{1}{3}[e^{-i2\pi/3} + 2e^{-i4\pi/3} + 3e^{-i6\pi/3}] = \tfrac{1}{2}(1 + .5773503i).$$

That is,

$$Z_t = \tfrac{1}{2}(1 - .5773503i)e^{-i2\pi t/3} + 2 + \tfrac{1}{2}(1 + .5773503i)e^{i2\pi t/3}, \qquad t = 1, 2, 3, 4, \ldots.$$

The coefficients a_k, b_k, and c_k are clearly related as shown in (11.3.5).

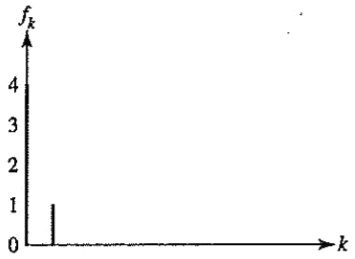


FIGURE 11.3 Power spectrum of Example 11.1.

Using (11.4.13), the power spectrum of the sequence is given by

$$f_k = \begin{cases} 2^2 = 4, & k = 0, \\ \frac{1}{2}[(1)^2 + (-.5773503)^2] = \frac{2}{3}, & k = 1, \\ 0, & \text{otherwise.} \end{cases}$$

The plot is given in Figure 11.3.

11.5 Fourier Representation of Nonperiodic Sequences: The Discrete-Time Fourier Transform

Consider a general nonperiodic sequence or discrete-time function, Z_t, of finite duration such that $Z_t = 0$ for $|t| > M$. Let $n = (2M + 1)$ and define a new function

$$Y_{t+jn} = Z_t, \quad -\frac{n-1}{2} \le t \le \frac{n-1}{2}, \qquad j = 0, \pm 1, \pm 2, \ldots, \tag{11.5.1}$$

which is clearly periodic with period n. Thus, using (11.4.6), we can express (11.5.1) in a Fourier series of the form

$$Y_t = \sum_{k=-(n-1)/2}^{(n-1)/2} c_k e^{i2\pi kt/n}, \tag{11.5.2}$$

$$c_k = \frac{1}{n} \sum_{t=-(n-1)/2}^{(n-1)/2} Y_t e^{-i2\pi kt/n}. \tag{11.5.3}$$

In the interval $-(n-1)/2 \le t \le (n-1)/2$, however, $Y_t = Z_t$. Therefore,

$$\begin{aligned} c_k &= \frac{1}{n} \sum_{t=-(n-1)/2}^{(n-1)/2} Z_t e^{-i2\pi kt/n} \\ &= \frac{1}{n} \sum_{t=-\infty}^{\infty} Z_t e^{-i2\pi kt/n}, \end{aligned} \tag{11.5.4}$$

where we have used that $Z_t = 0$ for $t < -(n-1)/2$ or $t > (n-1)/2$. Let

$$f(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t}. \tag{11.5.5}$$

We have

$$c_k = \frac{2\pi}{n} f(k\Delta\omega), \tag{11.5.6}$$

where $\Delta\omega = 2\pi/n$ is the sample spacing of frequencies. Combining (11.5.2) and (11.5.6) yields

$$Y_t = \sum_{k=-(n-1)/2}^{(n-1)/2} \frac{2\pi}{n} f(k\Delta\omega) e^{ik\Delta\omega t}. \tag{11.5.7}$$

Because $\Delta\omega = 2\pi/n$, (11.5.7) can be written as

$$Y_t = \sum_{k=-(n-1)/2}^{(n-1)/2} f(k\Delta\omega) e^{ik\Delta\omega t} \Delta\omega. \tag{11.5.8}$$

Now each term in the summation of (11.5.8) represents the area of a rectangle of height $f(k\Delta\omega)e^{ik\Delta\omega t}$ and width $\Delta\omega$. As $n \to \infty$, we have $Y_t \to Z_t$ and $\Delta\omega \to 0$. Thus, the limit of the summation becomes an integral. Furthermore, because the summation is carried out over n consecutive values of width $\Delta\omega = 2\pi/n$, the total interval of integration will always have a width of 2π. Therefore,

$$\begin{aligned} Z_t &= \lim_{n\to\infty} Y_t \\ &= \lim_{\Delta\omega\to 0} \sum_{k=-\infty}^{\infty} f(k\Delta\omega) e^{ik\Delta\omega t} \Delta\omega \\ &= \int_{2\pi} f(\omega) e^{i\omega t}\, d\omega. \end{aligned} \tag{11.5.9}$$

Because $f(\omega)e^{i\omega t}$, as a function of ω, is periodic with period 2π, the interval of integration can be taken as any interval of length 2π. Specifically, we can consider $-\pi \le \omega \le \pi$. Thus, we have the following:

$$Z_t = \int_{-\pi}^{\pi} f(\omega)e^{i\omega t}\,d\omega, \qquad t = 0, \pm 1, \pm 2, \ldots \tag{11.5.10}$$

and

$$f(\omega) = \frac{1}{2\pi}\sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t}, \qquad -\pi \le \omega \le \pi. \tag{11.5.11}$$

The function $f(\omega)$ in (11.5.11) is usually referred to as the (discrete time) Fourier transform of Z_t, and Z_t in (11.5.10) is often referred to as the (discrete time) inverse Fourier transform of $f(\omega)$. They form a Fourier transform pair.

From the above discussion, we see that we can also define the Fourier transform $f(\omega)$ as

$$f(\omega) = \sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t} \tag{11.5.12}$$

or

$$f(\omega) = \frac{1}{\sqrt{2\pi}}\sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t} \tag{11.5.13}$$

instead of the given (11.5.5). These modified definitions of $f(\omega)$ lead to the following Fourier transform pairs:

$$\begin{cases} Z_t = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(\omega)e^{i\omega t}\,d\omega \\ f(\omega) = \displaystyle\sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t} \end{cases} \tag{11.5.14}$$

and

$$\begin{cases} Z_t = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\pi}^{\pi} f(\omega)e^{i\omega t}\,d\omega \\ f(\omega) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t}, \end{cases} \tag{11.5.15}$$

respectively, which may be found in other books. In this book, we use the Fourier transform pairs as given in (11.5.10) and (11.5.11).

The derivation of the integral in (11.5.10) as the limiting form of a sum implies that the inverse Fourier transform given in (11.5.10) represents the sequence Z_t as a linear combination of complex sinusoids infinitesimally close in frequency with amplitudes $|f(\omega)|(d\omega)$. Thus, the quantity $|f(\omega)|$ is often referred to as the spectrum or amplitude spectrum of the sequence, as it provides us with the information on how Z_t is composed of complex sinusoids at different frequencies.

The energy associated with the sequence is given by the following Parseval's relation:

$$\begin{aligned}\sum_{t=-\infty}^{\infty} |Z_t|^2 &= \sum_{t=-\infty}^{\infty} Z_t \int_{-\pi}^{\pi} f^*(\omega) e^{-i\omega t}\, d\omega \\ &= \int_{-\pi}^{\pi} f^*(\omega) \sum_{t=-\infty}^{\infty} Z_t e^{-i\omega t}\, d\omega \\ &= 2\pi \int_{-\pi}^{\pi} f^*(\omega) f(\omega)\, d\omega \\ &= 2\pi \int_{-\pi}^{\pi} |f(\omega)|^2\, d\omega. \end{aligned} \tag{11.5.16}$$

Parseval's relation thus relates energy in the time domain to energy in the frequency domain. In other words, the energy may be determined either by computing the energy per unit time $|Z_t|^2$ and summing over all time or by computing the energy per unit frequency $2\pi|f(\omega)|^2$ and integrating over all frequencies. Hence, $g(\omega) = 2\pi|f(\omega)|^2$, as a function of ω, is also referred to as the energy spectrum or the energy spectral density function.

In the above construction, Z_t was assumed to be of arbitrary finite duration. Equations (11.5.10) and (11.5.11) remain valid for a general nonperiodic sequence of infinite duration. In this case, however, we must consider the question of convergence of the infinite summation in (11.5.11). A condition on Z_t which guarantees the convergence of this sum is that the sequence $\{Z_t\}$ is absolutely summable. That is,

$$\sum_{t=-\infty}^{\infty} |Z_t| < \infty. \tag{11.5.17}$$

In fact, the above theory holds also when Z_t is square summable, i.e.,

$$\sum_{t=-\infty}^{\infty} Z_t^2 < \infty. \tag{11.5.18}$$

The proof of the result using a weak condition depends on an alternative presentation of the Fourier transform and is omitted. For our purposes, it suffices to note that the condition in Equation (11.5.17) implies the condition in Equation (11.5.18), but the converse is not true.

Important differences between the frequency domain properties of periodic and nonperiodic sequences exist:

1. The spectrum frequencies of periodic sequences are harmonically related and form a finite discrete set, whereas those of nonperiodic sequences form a continuum of frequencies.
2. The energy over whole time horizon $t = 0, \pm 1, \pm 2, \ldots$ for periodic sequences is infinite. Hence, we study their properties in terms of the power spectrum over a finite set of harmonically related frequencies. The resulting spectra are hence sometimes referred to as line spectra. The energy over the whole time horizon for nonperiodic sequences is finite as guaranteed by the condition in Equation (11.5.17). Hence, we describe their properties in terms of the energy spectrum over a continuum of frequencies.

EXAMPLE 11.2 Consider the sequence

$$Z_t = \left(\frac{1}{2}\right)^{|t|}, \qquad t = 0, \pm 1, \pm 2, \ldots$$

which is graphed in Figure 11.4.

Now

$$\sum_{t=-\infty}^{\infty} |Z_t| = \sum_{t=-\infty}^{\infty} \left(\frac{1}{2}\right)^{|t|} = 1 + 2\sum_{t=1}^{\infty}\left(\frac{1}{2}\right)^{t}$$
$$= 1 + \sum_{t=0}^{\infty}\left(\frac{1}{2}\right)^{t} = 1 + \frac{1}{1-\frac{1}{2}} = 1 + 2 = 3 < \infty.$$

Hence, we can write

$$f(\omega) = \frac{1}{2\pi}\sum_{t=-\infty}^{\infty}\left(\frac{1}{2}\right)^{|t|} e^{-i\omega t}$$
$$= \frac{1}{2\pi}\left\{\sum_{t=-\infty}^{-1}\left[\left(\frac{1}{2}\right)^{-t} e^{-i\omega t}\right] + 1 + \sum_{t=1}^{\infty}\left[\left(\frac{1}{2}\right)^{t} e^{-i\omega t}\right]\right\}$$
$$= \frac{1}{2\pi}\left[\sum_{t=1}^{\infty}\left(\frac{1}{2}e^{i\omega}\right)^{t} + 1 + \sum_{t=1}^{\infty}\left(\frac{1}{2}e^{-i\omega}\right)^{t}\right]$$

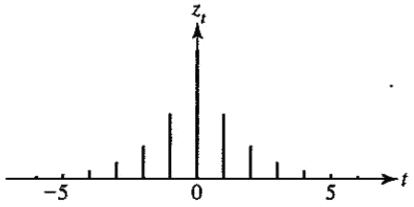


FIGURE 11.4 Nonperiodic sequence in Example 11.2.

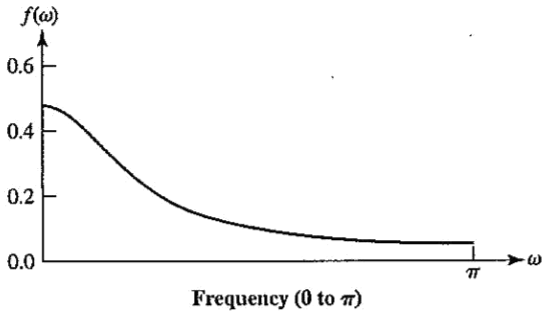


FIGURE 11.5 The spectrum of the sequence in Example 11.2.

$$= \frac{1}{2\pi}\left[\frac{1}{2}e^{i\omega}\sum_{t=0}^{\infty}\left(\frac{1}{2}e^{i\omega}\right)^t + 1 + \frac{1}{2}e^{-i\omega}\sum_{t=0}^{\infty}\left(\frac{1}{2}e^{-i\omega}\right)^t\right]$$

$$= \frac{1}{2\pi}\left[\frac{1}{2}e^{i\omega}\frac{1}{1-\frac{1}{2}e^{i\omega}} + 1 + \frac{1}{2}e^{-i\omega}\frac{1}{1-\frac{1}{2}e^{-i\omega}}\right]$$

$$= \frac{1}{2\pi}\left[1 + \frac{e^{i\omega}}{2-e^{i\omega}} + \frac{e^{-i\omega}}{2-e^{-i\omega}}\right]$$

$$= \frac{1}{2\pi}\left[1 + \frac{2(e^{i\omega}+e^{-i\omega})-2}{5-2(e^{i\omega}+e^{-i\omega})}\right]$$

$$= \frac{1}{2\pi}\left[1 + \frac{4\cos\omega - 2}{5-4\cos\omega}\right] = \frac{3}{2\pi(5-4\cos\omega)}, \qquad -\pi \le \omega \le \pi.$$

The function $f(\omega)$ is graphed in Figure 11.5. It shows that the spectrum is a positive and symmetric function dominated by low frequencies.

EXAMPLE 11.3 Let the sequence be

$$Z_t = \left(-\frac{1}{2}\right)^{|t|}, \qquad t = 0, \pm 1, \pm 2, \ldots$$

as shown in Figure 11.6.

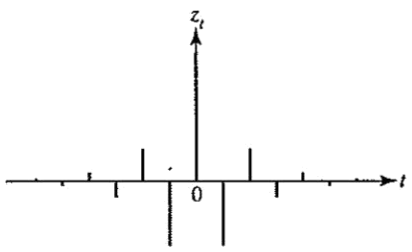


FIGURE 11.6 Symmetric oscillating sequences in Example 11.3.

Following the same arguments as in Example 11.2, we obtain

$$
\begin{aligned}
f(\omega) &= \frac{1}{2\pi}\sum_{t=-\infty}^{\infty}\left(-\frac{1}{2}\right)^{|t|}e^{-i\omega t} \\
&= \frac{1}{2\pi}\left[\sum_{t=1}^{\infty}\left(-\frac{1}{2}e^{i\omega}\right)^{t} + 1 + \sum_{t=1}^{\infty}\left(-\frac{1}{2}e^{-i\omega}\right)^{t}\right] \\
&= \frac{1}{2\pi}\left[\left(-\frac{1}{2}e^{i\omega}\right)\frac{1}{1+\frac{1}{2}e^{i\omega}} + 1 + \left(-\frac{1}{2}e^{-i\omega}\right)\frac{1}{1+\frac{1}{2}e^{-i\omega}}\right] \\
&= \frac{1}{2\pi}\left[1 - \frac{e^{i\omega}}{2+e^{i\omega}} - \frac{e^{-i\omega}}{2+e^{-i\omega}}\right] \\
&= \frac{1}{2\pi}\left[1 - \frac{4\cos\omega + 2}{5 + 4\cos\omega}\right] = \frac{3}{2\pi(5+4\cos\omega)}, \qquad -\pi \le \omega \le \pi.
\end{aligned}
$$

The spectrum $f(\omega)$ as shown in Figure 11.7 is again positive and symmetric. Now however, it is dominated by high frequencies. This phenomenon is typical for a sequence that tends to oscillate rapidly.

EXAMPLE 11.4 In this example, we find the corresponding sequence Z_t for the spectrum

$$
f(\omega) = \frac{1+\cos\omega}{2\pi}, \qquad -\pi \le \omega \le \pi.
$$

For a given spectrum, we can calculate the corresponding sequence by using inverse Fourier transform given in (11.5.10). Thus,

$$
\begin{aligned}
Z_t &= \int_{-\pi}^{\pi} f(\omega)e^{i\omega t}\,d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}(1+\cos\omega)e^{i\omega t}\,d\omega
\end{aligned}
$$

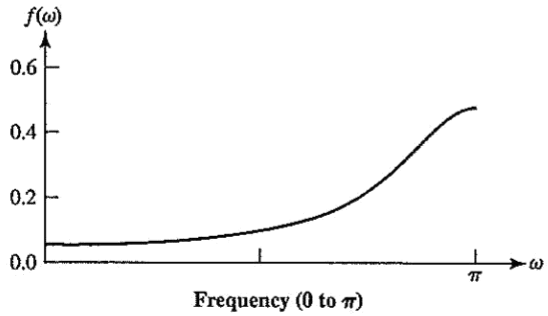


FIGURE 11.7 The spectrum of the oscillating sequence in Example 11.3.

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t} + \left(\frac{e^{i\omega} + e^{-i\omega}}{2}\right)e^{i\omega t}\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}\, d\omega + \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[e^{i\omega(t+1)} + e^{i\omega(t-1)}\right] d\omega$$

$$= \begin{cases} 1, & t = 0, \\ \frac{1}{2}, & t = -1, 1, \\ 0, & \text{otherwise.} \end{cases}$$

11.6 Fourier Representation of Continuous-Time Functions

11.6.1 Fourier Representation of Periodic Functions

We now consider a function $f(t)$ that is defined for all real values t. First, we assume that $f(t)$ is periodic with the fundamental period P, i.e.,

$$f(t) = f(t + jP) \text{ for all real values } t \text{ and integers } j. \tag{11.6.1}$$

This periodic function is completely characterized by its behavior in any interval of length P. Thus, we only need to consider the function for t in the interval $[-P/2, P/2]$. Now recall that the fundamental period P and the fundamental frequency ω_0 are related as follows:

$$\omega_0 = \frac{2\pi}{P}. \tag{11.6.2}$$

For this fundamental frequency, consider the set of harmonically related complex exponentials

$$\{e^{ik\omega_0 t}: \quad k = 0, \pm 1, \pm 2, \ldots\}. \tag{11.6.3}$$

Clearly,

$$e^{ik\omega_0 t} = e^{ik\omega_0(t+2\pi/\omega_0)} = e^{i(k\omega_0 t + 2\pi k)} = e^{ik\omega_0 t}. \tag{11.6.4}$$

Hence, the complex exponentials in (11.6.3) are periodic with the same period $P = 2\pi/\omega_0$ as the function $f(t)$. Also,

$$\int_{-P/2}^{P/2} e^{ik\omega_0 t} e^{-ij\omega_0 t}\, dt = \int_{-P/2}^{P/2} e^{i(k-j)\omega_0 t}\, dt$$

$$= \begin{cases} P, & \text{if } k = j, \\ 0, & \text{if } k \neq j. \end{cases} \tag{11.6.5}$$

Equation (11.6.5) follows because if $k = j$, then the integral becomes

$$\int_{-P/2}^{P/2} e^{i(k-j)\omega_0 t}\,dt = \int_{-P/2}^{P/2} dt = P;$$

if $k \neq j$, then the integral is

$$\begin{aligned}\int_{-P/2}^{P/2} e^{i(k-j)\omega_0 t}\,dt &= \left[\frac{1}{i(k-j)\omega_0} e^{i(k-j)\omega_0 t}\right]_{-P/2}^{P/2} \\ &= \frac{1}{i(k-j)\omega_0}\left[e^{i(k-j)\pi} - e^{-i(k-j)\pi}\right] = 0.\end{aligned}$$

Thus, the system in (11.6.3) is a periodic system with period P that is orthogonal on any interval of width P. Hence, we might represent the function $f(t)$ as the following Fourier series:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t}. \tag{11.6.6}$$

If the representation is valid, i.e., if the series in (11.6.6) converges uniformly to $f(t)$, then term-by-term integration is permissible and the Fourier coefficient can be easily shown to be

$$c_k = \frac{\int_{-P/2}^{P/2} f(t)e^{-ik\omega_0 t}\,dt}{\int_{-P/2}^{P/2} e^{ik\omega_0 t}e^{-ik\omega_0 t}\,dt} = \frac{1}{P}\int_{-P/2}^{P/2} f(t)e^{-ik\omega_0 t}\,dt. \tag{11.6.7}$$

For the representation in (11.6.6) to be valid, we need to show that the series uniformly converges to the function $f(t)$. There are a number of necessary and sufficient conditions for the convergence of the series. For our purposes, it suffices to note the following sufficient conditions proposed by Dirichlet.

Dirichlet Conditions for Periodic Functions If $f(t)$ is a bounded periodic function with period P and has at most a finite number of maxima and minima in one period and a finite number of discontinuities, then the Fourier series converges to $f(t)$ at every continuity point of $f(t)$ and converges to the average of the right-hand and left-hand limits of $f(t)$ at each point where $f(t)$ is discontinuous.

The Parseval's relation for (11.6.6) is easily seen to be

$$\int_{-P/2}^{P/2} |f(t)|^2\,dt = P\sum_{k=-\infty}^{\infty} |c_k|^2. \tag{11.6.8}$$

11.6.2 Fourier Representation of Nonperiodic Functions: The Continuous-Time Fourier Transform

For a general nonperiodic function $f(t)$, which is defined on all real value t but of finite duration, i.e., $f(t) = 0$ for $|t| > P/2$, we define a new function that is identical to $f(t)$ in the interval $[-P/2, P/2]$, but is periodic outside this interval. That is,

$$\begin{cases} g(t) = f(t), & -P/2 \le t \le P/2, \\ g(t + jP) = g(t), & j = \pm 1, \pm 2, \ldots. \end{cases} \tag{11.6.9}$$

Clearly, $g(t)$ is periodic with period P. Thus, we can write

$$g(t) = \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t} \tag{11.6.10}$$

and

$$c_k = \frac{1}{P}\int_{-P/2}^{P/2} g(t)e^{-ik\omega_0 t}\,dt, \tag{11.6.11}$$

where $\omega_0 = 2\pi/P$. Now $g(t) = f(t)$ in the interval $[-P/2, P/2]$. Hence, in this interval, we have

$$\begin{aligned} c_k &= \frac{\omega_0}{2\pi}\int_{-P/2}^{P/2} f(t)e^{-ik\omega_0 t}\,dt \\ &= \frac{\omega_0}{2\pi}\int_{-P/2}^{P/2} f(u)e^{-ik\omega_0 u}\,du \end{aligned}$$

and

$$\begin{aligned} g(t) &= \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t} \\ &= \sum_{k=-\infty}^{\infty}\left[\frac{\omega_0}{2\pi}\int_{-P/2}^{P/2} f(u)e^{-ik\omega_0 u}\,du\right]e^{ik\omega_0 t} \\ &= \sum_{k=-\infty}^{\infty}\frac{\omega_0}{2\pi}\int_{-P/2}^{P/2} f(u)e^{ik\omega_0(t-u)}\,du \\ &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} H(k\omega_0)\omega_0, \end{aligned} \tag{11.6.12}$$

where

$$H(k\omega_0) = \int_{-P/2}^{P/2} f(u)e^{ik\omega_0(t-u)}\,du.$$

Now each term in the summation of (11.6.12) represents the area of a rectangle of height $H(k\omega_0)$ and width ω_0. As $P \longrightarrow \infty$, $g(t) \longrightarrow f(t)$, $\omega_0 \longrightarrow 0$, and the limit of the summation becomes an integral. Thus,

$$f(t) = \lim_{P\to\infty} g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)\,d\omega$$

and

$$H(\omega) = \lim_{P\to\infty} H(k\omega_0) = \int_{-\infty}^{\infty} f(u)e^{i\omega(t-u)}\,du,$$

where we note that in the limit the quantity $k\omega_0$ becomes a continuous variable, which we denote by ω. That is,

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)e^{i\omega(t-u)}\,du\,d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)e^{-i\omega u}\,du\,e^{i\omega t}\,d\omega \\ &= \int_{-\infty}^{\infty} \mathcal{F}(\omega)e^{i\omega t}\,d\omega \end{aligned} \tag{11.6.13}$$

and

$$\begin{aligned} \mathcal{F}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(u)e^{-i\omega u}\,du \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt. \end{aligned} \tag{11.6.14}$$

The function $\mathcal{F}(\omega)$ in (11.6.14) is called the Fourier transform or Fourier integral of $f(t)$, and $f(t)$ in (11.6.13) is called the inverse Fourier transform or inverse Fourier integral

of $\mathcal{F}(\omega)$. They form a Fourier transform (integral) pair. Similarly, by the same arguments as in Section 11.5, the following alternative forms of the Fourier transform pairs,

$$\begin{cases} f(t) &= \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} \mathcal{F}(\omega)e^{i\omega t}\,d\omega \\ \mathcal{F}(\omega) &= \displaystyle\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt \end{cases} \tag{11.6.15}$$

and

$$\begin{cases} f(t) &= \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} \mathcal{F}(\omega)e^{i\omega t}\,d\omega \\ \mathcal{F}(\omega) &= \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt, \end{cases} \tag{11.6.16}$$

may also be found in other books.

The Parseval's relation for the Fourier transform pair in (11.6.13) and (11.6.14) can be easily shown to be

$$\int_{-\infty}^{\infty} |\mathcal{F}(\omega)|^2\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(t)|^2\,dt. \tag{11.6.17}$$

In the above construction, $f(t)$ is assumed to be of finite duration. Although the results are valid for some functions of infinite duration, not all functions of infinite durations have a Fourier transform. The following Dirichlet conditions provide a useful set of sufficient conditions of the existence of $\mathcal{F}(\omega)$.

Dirichlet Conditions for Nonperiodic Functions

1. $f(t)$ is absolutely integrable, i.e., $\int_{-\infty}^{\infty} |f(t)|\,dt < \infty$.
2. $f(t)$ has only a finite number of maxima and minima and a finite number of discontinuities in any finite interval.

11.7 The Fast Fourier Transform

Because, in actual time series analysis, we are given a sequence of n values $Z_1, Z_2, \ldots, Z_n$, we use (11.3.4) to calculate the following Fourier coefficients, also known as Fourier transforms:

$$c_k = \frac{1}{n}\sum_{t=1}^{n} Z_t e^{-i\omega_k t}, \qquad k = -\frac{n}{2} + 1, \ldots, 0, \ldots, \frac{n}{2}, \tag{11.7.1}$$

where we assume that n is even and recall that $\omega_k = 2\pi k/n$. Rewriting (11.7.1) in the form

$$\begin{aligned} c_k &= \frac{1}{n}\sum_{t=1}^{n} Z_t e^{-i\omega_k t}, \\ &= \frac{1}{n}\left[Z_1 e^{-i\omega_k} + Z_2 e^{-i\omega_k 2} + \cdots + Z_n e^{-i\omega_k n}\right], \end{aligned} \tag{11.7.2}$$

we see that for each k, the computation of c_k requires approximately n complex multiplications and additions if we ignore that for some values of t and k, $e^{-i\omega_k t}$ is equal to ± 1 and does not, strictly speaking, require a complex multiplication. Hence, the direct computation of the whole set $\{c_k : k = -n/2 + 1, -n/2 + 2, \ldots, 0, \ldots, n/2\}$ using (11.7.1) requires approximately n^2 complex multiplications and additions. When n is large, the required number of operations is clearly substantial. Additionally, there is a corresponding storage problem. For example, if $n = 450$, then the direct computation requires 202,500 complex multiplications and additions and storage locations. For large values of n, this computation may become formidable if not impossible. Hence, an efficient algorithm, known as the fast Fourier transform (FFT), was developed to compute these Fourier transforms. Major contributions to the method include Good (1958) and Cooley and Tukey (1965).

To illustrate the algorithm, we consider an even value of n, and we let $X_t = Z_{2t}$ represent the even-indexed values of $\{Z_t\}$ and $Y_t = Z_{2t-1}$ represent the odd-indexed values of $\{Z_t\}$. Furthermore, to simplify the notations we let

$$\lambda_n = e^{-i2\pi/n}. \tag{11.7.3}$$

Then, the n-point Fourier transforms in (11.7.1) become

$$\begin{aligned} c_k &= \frac{1}{n}\sum_{t=1}^{n} Z_t e^{-i\omega_k t} \\ &= \frac{1}{n}\sum_{t=1}^{n} Z_t \lambda_n^{kt} \\ &= \frac{1}{n}\sum_{t=1}^{n/2} Z_{2t} e^{-i2\pi k 2t/n} + \frac{1}{n}\sum_{t=1}^{n/2} Z_{2t-1} e^{-i2\pi k(2t-1)/n} \\ &= \frac{1}{n}\sum_{t=1}^{n/2} X_t e^{-i2\pi kt/(n/2)} + \frac{1}{n} e^{i2\pi k/n}\sum_{t=1}^{n/2} Y_t e^{-i2\pi kt/(n/2)} \\ &= \frac{1}{2} f_k + \frac{1}{2}\lambda_n^{-k} g_k, \end{aligned} \tag{11.7.4}$$

where

$$f_k = \frac{1}{(n/2)}\sum_{t=1}^{n/2} X_t \lambda_{n/2}^{kt}, \tag{11.7.5}$$

$$g_k = \frac{1}{(n/2)}\sum_{t=1}^{n/2} Y_t \lambda_{n/2}^{kt}. \tag{11.7.6}$$

Note that f_k and g_k in (11.7.5) and (11.7.6) are clearly the $(n/2)$-point Fourier transforms of X_t and Y_t, respectively. Hence, Equation (11.7.4) implies that the n-point Fourier transforms can be calculated in terms of the linear combination of the simpler Fourier transforms of length $n/2$. Because

$$\begin{aligned}\lambda_{n/2}^{k} &= e^{-i2\pi k/(n/2)} = e^{-i2\pi k/(n/2)} e^{i2\pi} \\ &= e^{-i2\pi(k-n/2)/(n/2)} = \lambda_{n/2}^{k-n/2},\end{aligned} \tag{11.7.7}$$

we have

$$f_k = f_{k-n/2}, \tag{11.7.8}$$

$$g_k = g_{k-n/2}. \tag{11.7.9}$$

Thus, in Equations (11.7.5) and (11.7.6), we only need to calculate f_k and g_k for $k = 1, 2, \ldots, n/2$. The direct computation of these f_k and g_k through (11.7.5) and (11.7.6) requires approximately $2(n/2)^2$ complex multiplications and additions. Then the linear combination of these two $(n/2)$-point Fourier transforms through (11.7.4) requires n complex multiplications corresponding to multiplying g_k by λ_n^{-k} and also n complex additions corresponding to adding the product $\frac{1}{2}\lambda_n^{-k} g_k$ to $\frac{1}{2} f_k$. Consequently, the total number of complex multiplications and additions required in the computation of Equation (11.7.4) is

$$n + 2\left(\frac{n}{2}\right)^2. \tag{11.7.10}$$

If $n/2$ is also even, then by repeating the same process, each of the $(n/2)$-point Fourier transforms f_k and g_k in (11.7.5) and (11.7.6) can be computed via two $(n/4)$-point Fourier transforms. To find the number of required operations to compute each of the $(n/2)$-point transforms via the $(n/4)$-point transforms, we simply follow the same logic as described in the previous paragraph and note that it is equal to $n/2 + 2(n/4)^2$. Hence, the required total number of operations for the overall computation leading to c_k's is

$$n + 2\left[\frac{n}{2} + 2\left(\frac{n}{4}\right)^2\right] = n + n + 2^2\left(\frac{n}{2^2}\right)^2. \tag{11.7.11}$$

If $n = 2^r$ with r being a prime number, then we can continue this process for r iterations and end with a set of 2-point Fourier transforms. The total number of required complex multiplications and additions for the overall computation is seen to be approximately equal to

$$n + n + \cdots + 2^r\left(\frac{n}{2^r}\right)^2 \simeq n + n + \cdots + n = nr = n \log_2 n. \tag{11.7.12}$$

Clearly, for a large value of n, the reduction in the required operations from n^2 to $n \log_2 n$ is very substantial. For example, if $n = 2048 = 2^{11}$, then the direct computation using (11.7.1) requires approximately 4,194,304 operations, whereas the fast Fourier transform requires only 22,528 operations. The reduction is more than 99%. Furthermore, because $\lambda_2 = e^{-i2\pi/2} = -1$, the 2-point Fourier transforms in the final iteration are actually obtained only by simple addition and subtraction.

In essence, the FFT is an iterative algorithm that uses the properties of trigonometric functions and complex exponentials in reducing the computation of n-point Fourier transforms to simpler transforms. In this section, we illustrated the algorithm when $n = 2^r$. In practice, the actual length of a sequence is, of course, rarely of the form 2^r. One can, however, always add zeros to the sequence (known as padding) until it reaches the required form. It is clear that for a given frequency ω_k these appended zeros will not affect the numerical values of the Fourier transforms. After we add zeros to increase the length of a sequence from n to m (say), however, then the transforms will be calculated at the frequencies $\omega_k = 2\pi k/m$ rather than at $\omega_k = 2\pi k/n$. To attenuate potential undesirable effects, some authors have recommended that the unpadded portion of a modified sequence be "tapered" to smooth the transition from nonzero to zero values (see, e.g., Tukey [1967], Godfrey [1974], and Brillinger [1975]).

Although n is not necessarily of the form $n = 2^r$, we know from number theory that an integer can always be written as the product of prime numbers, i.e., we can always write

$$n = r_1 \cdot r_2 \cdot \cdots \cdot r_m,$$

where $r_1, r_2, \ldots, r_m$ are prime numbers not necessarily equal to 2. It is clear that the FFT for this form of n can be similarly developed. Because most available computer software for fast Fourier transforms is based on the form $n = 2^r$ introduced in this section, we will not go into the method with other forms of n. We leave as an exercise the simple case when n is the product of two integers, say $n = n_1 n_2$. For other cases, we refer readers to Bloomfield (2000).

EXERCISES

11.1 Prove the identity

$$2\sum_{t=1}^{n} \cos \omega t = \frac{\sin\left[(n + 1/2)\omega\right]}{\sin(\omega/2)} - 1.$$

11.2 Consider the following finite sequence:

t	1	2	3	4	5	6	7	8	9	10	11	12
Z_t	−4	−23	−18	−6	1	4	10	−2	−16	−10	2	8

Find the Fourier representation of the sequence.

11.3 Find the Fourier representation for each of the following sequences:

(a) $Z_t = \begin{cases} 1, & \text{if } 1 \le t \le m, \\ 0, & \text{if } m < t \le n, \ (n > m + 2). \end{cases}$

(b) $Z_t = n - t$, for $1 \le t \le n$.

11.4 Consider $Z_t = \sum_{k=0}^{p-1} X_k \exp(-i(2\pi kt)/p)$. Show that Z_t is periodic with period p.

11.5 Let $Z_1 = 4$, $Z_2 = 3$, $Z_3 = 2$, $Z_4 = 1$, and $Z_{t+4j} = Z_t$, for $t = 1, 2, 3, 4$ and $j = \pm 1, \pm 2, \ldots$.

(a) Find the Fourier representation of Z_t.

(b) Find the Fourier representation of Z_t in terms of complex exponentials.

(c) Find and plot the power spectrum of Z_t.

11.6 Consider the following sequence:

$$Z_t = \frac{16}{25}\left(\frac{1}{2}\right)^{|t|} + \frac{9}{25}\left(-\frac{1}{3}\right)^{|t|}, t = 0, \pm 1, \pm 2, \ldots.$$

(a) Is the sequence absolutely summable?

(b) Find the Fourier transform of Z_t.

(c) Plot and discuss the spectrum of the sequence.

11.7 Find the corresponding sequence Z_t for each of the following spectra:

(a) $f(\omega) = \dfrac{1 - \cos\omega + \cos 2\omega}{2\pi}, \quad -\pi \le \omega \le \pi.$

(b) $f(\omega) = \dfrac{1}{2\pi}, \quad -\pi \le \omega \le \pi.$

(c) $f(\omega) = \dfrac{1 + \cos 4\omega}{2\pi}, \quad -\pi \le \omega \le \pi.$

11.8 For absolutely integrable functions $g(x)$ and $h(x)$ defined on the real line, the function

$$f(x) = \int_{-\infty}^{\infty} g(x - y)h(y)\, dy$$

is called the convolution of $g(x)$ and $h(x)$. Let X_t and Y_t be absolutely summable sequences, and define $Z_t = X_t Y_t$. Show that the spectrum of Z_t is the convolution of the spectrums for X_t and Y_t.

11.9 Find the Fourier representation for each of the following functions:

(a) $f(t) = 1$, $-\pi \le t \le \pi$, and $f(t) = 0$ otherwise.

(b) $f(t) = \cos \phi t$, where ϕ is not an integer.

(c) $f(t) = \sin \phi t$, where ϕ is not an integer.

11.10 Show the Parseval's relation given in Equation (11.6.17).

11.11 A generalized function $\delta(t)$ represents a sequence of functions $\delta_n(t)$ for $n = 1, 2, \ldots$ such that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(t)\delta_n(t - t_0)dt = f(t_0)$$

i.e.,

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_0)dt = f(t_0)$$

provided that $f(t)$ is continuous at $t = t_0$. The function $\delta(t)$ is also called Dirac's delta function.

(a) Show that the sequence $s_n(t) = (n/\pi)^{1/2} e^{-nt^2}$, $-\infty < t < \infty$, yields a Dirac's delta function.

(b) Find the Fourier transform pair of a delta function.

11.12 Discuss the fast Fourier transform when the number of observations n is the product of two integers n_1 and n_2, i.e., $n = n_1 n_2$.

12

Spectral Theory of Stationary Processes

With the proper background of the Fourier transform discussed in Chapter 11, we are now ready to study the frequency domain approach of time series analysis. In fact, the spectrum of a stationary process is simply the Fourier transform of the absolutely summable autocovariance function of the process. More generally, a stationary process can always be represented by a spectral distribution function. After developing the general spectral theory, we examine the spectrum of some common processes, such as the spectrum of ARMA models. Related questions of the effect of linear filters and aliasing are also addressed.

12.1 The Spectrum

12.1.1 The Spectrum and Its Properties

Let Z_t be a real-valued stationary process with absolutely summable autocovariance sequence γ_k. Then, from the results in Section 11.5, the Fourier transform of γ_k exists and by Equation (11.5.11) equals

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k} \tag{12.1.1}$$

$$= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k \cos \omega k \tag{12.1.2}$$

$$= \frac{1}{2\pi}\gamma_0 + \frac{1}{\pi}\sum_{k=1}^{\infty}\gamma_k \cos \omega k, \qquad -\pi \leq \omega \leq \pi, \tag{12.1.3}$$

where we have used the properties that $\gamma_k = \gamma_{-k}$, $\sin 0 = 0$, $\sin \omega(-k) = -\sin \omega k$, and $\cos \omega(-k) = \cos \omega k$. The sequence γ_k can be recovered from $f(\omega)$ through the inverse Fourier transform

$$\gamma_k = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k}\, d\omega. \tag{12.1.4}$$

The function $f(\omega)$ has the following important properties.

1. $f(\omega)$ is a continuous real-valued nonnegative function, i.e., $|f(\omega)| = f(\omega)$. Thus, from the discussion following Equation (11.5.15) in Section 11.5, $f(\omega)$ is also referred to as the spectrum of the autocovariance sequence γ_k or the corresponding stationary process Z_t. It is obvious that $f(\omega)$ is a continuous real-valued function. To show it is nonnegative, we note that γ_k, as an autocovariance function, is positive semidefinite, i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \gamma_{(k_i - k_j)} \geq 0 \tag{12.1.5}$$

for any real numbers c_i, c_j and any integers k_i and k_j. In particular, let $c_i = \cos \omega i$, $c_j = \cos \omega j$, $k_i = i$, and $k_j = j$. We have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{(i-j)} \cos \omega i \cos \omega j \geq 0.$$

Likewise,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{(i-j)} \sin \omega i \sin \omega j \geq 0.$$

Thus,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{(i-j)} &[\cos \omega i \cos \omega j + \sin \omega i \sin \omega j] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{(i-j)} \cos \omega(i-j) \\ &= \sum_{k=-(n-1)}^{n-1} (n - |k|)\gamma_k \cos \omega k \\ &= n \sum_{k=-(n-1)}^{n-1} \left[1 - \frac{|k|}{n}\right] \gamma_k \cos \omega k \geq 0, \end{aligned} \tag{12.1.6}$$

where $k = (i - j)$. Therefore,

$$\sum_{k=-(n-1)}^{n-1} \left[1 - \frac{|k|}{n}\right] \gamma_k \cos \omega k \geq 0. \tag{12.1.7}$$

Now, that γ_k is absolutely summable implies that for any given $\epsilon > 0$, we can choose an $N > 0$ such that

$$\sum_{k=-\infty}^{-(N+1)} |\gamma_k| + \sum_{k=(N+1)}^{\infty} |\gamma_k| < \epsilon.$$

Thus, for $n > N$, we have

$$\begin{aligned}\sum_{k=-n}^{n} \frac{|k|}{n}|\gamma_k| &< \sum_{k=-N}^{N} \frac{|k|}{n}|\gamma_k| + \sum_{k=-\infty}^{-(N+1)}|\gamma_k| + \sum_{k=(N+1)}^{\infty}|\gamma_k| \\ &< \sum_{k=-N}^{N} \frac{|k|}{n}|\gamma_k| + \epsilon.\end{aligned}$$

Clearly, for any fixed N,

$$\lim_{n\to\infty} \sum_{k=-N}^{N} \frac{|k|}{n}|\gamma_k| = 0.$$

Because ϵ is arbitrary, it follows that

$$\lim_{n\to\infty} \sum_{k=-n}^{n} \frac{|k|}{n}|\gamma_k| = 0.$$

Hence,

$$\lim_{n\to\infty} \sum_{k=-n}^{n} \frac{|k|}{n}\gamma_k \cos \omega k = 0 \tag{12.1.8}$$

because $|\gamma_k \cos \omega k| \leq |\gamma_k|$. Therefore, by (12.1.7) and (12.1.8),

$$\begin{aligned}f(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \cos \omega k \\ &= \frac{1}{2\pi} \lim_{n\to\infty} \sum_{k=-(n-1)}^{n-1} \left[1 - \frac{|k|}{n}\right] \gamma_k \cos \omega k \geq 0.\end{aligned} \tag{12.1.9}$$

2. $f(\omega) = f(\omega + 2\pi)$ and hence $f(\omega)$ is periodic with period 2π. Furthermore, because $f(\omega) = f(-\omega)$, i.e., $f(\omega)$ is a symmetric even function, its graph is normally presented only for $0 \leq \omega \leq \pi$.

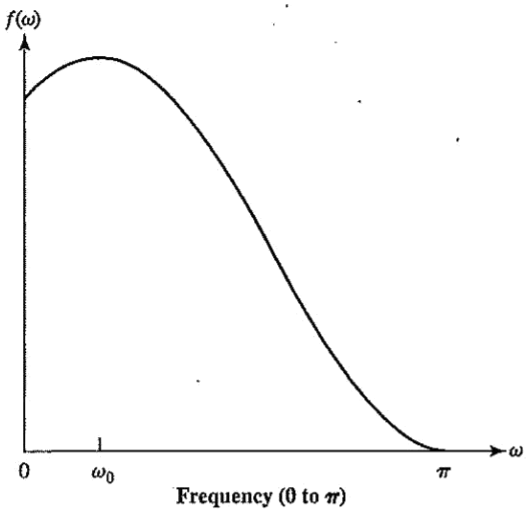


FIGURE 12.1 An example of a spectrum.

3. From (12.1.4), we have

$$\text{Var}(Z_t) = \gamma_0 = \int_{-\pi}^{\pi} f(\omega)\, d\omega, \tag{12.1.10}$$

which shows that the spectrum $f(\omega)$ may be interpreted as the decomposition of the variance of a process. The term $f(\omega)\, d\omega$ is the contribution to the variance attributable to the component of the process with frequencies in the interval $(\omega, \omega + d\omega)$. A peak in the spectrum indicates an important contribution to the variance from the components at frequencies in the corresponding interval. For example, Figure 12.1 shows that the components near frequency ω_0 are of greatest importance and the high-frequency components near π are of little importance.

4. Equations (12.1.1) and (12.1.4) imply that the spectrum $f(\omega)$ and the autocovariance sequence γ_k form a Fourier transform pair, with one being uniquely determined from the other. Hence, the time domain approach and the frequency domain approach are theoretically equivalent. The reason to consider both approaches is that there are some occasions when one approach is preferable to the other for presentation or interpretation.

12.1.2 The Spectral Representation of Autocovariance Functions: The Spectral Distribution Function

Note that the spectral representation of γ_k given in (12.1.1) and (12.1.4) holds only for an absolutely summable autocovariance function. More generally, for a given autocovariance function γ_k, we can always have its spectral representation in terms of the following Fourier–Stieltjes integral:

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k}\, dF(\omega), \tag{12.1.11}$$

where $F(\omega)$ is known as the spectral distribution function. Equation (12.1.11) is usually referred to as the spectral representation of the autocovariance function γ_k. Like any statistical distribution function, the spectral distribution function is a nondecreasing function that can be partitioned into three components: (1) a step function consisting of a countable number of finite jumps, (2) an absolutely continuous function, and (3) a "singular" function. The third component is insignificant and is ignored in most applications. Thus, we may write the spectral distribution function as

$$F(\omega) \simeq F_s(\omega) + F_c(\omega), \tag{12.1.12}$$

where $F_s(\omega)$ is the step function and $F_c(\omega)$ is the absolutely continuous component. Equation (12.1.4) shows that for a process with absolutely summable autocovariance function, $F(\omega) = F_c(\omega)$ and $dF(\omega) = f(\omega)\, d\omega$.

To illustrate a step spectral distribution function, consider the general linear cyclical model

$$Z_t = \sum_{i=1}^{M} A_i \sin(\omega_i t + \Theta_i), \tag{12.1.13}$$

where the A_i are constants and the Θ_i are independent uniformly distributed random variables on the interval $[-\pi, \pi]$. The ω_i are distinct frequencies contained in the interval $[-\pi, \pi]$. Then

$$\begin{aligned} E(Z_t) &= \sum_{i=1}^{M} A_i E[\sin(\omega_i t + \Theta_i)] \\ &= \sum_{i=1}^{M} \frac{A_i}{2\pi} \int_{-\pi}^{\pi} \sin(\omega_i t + \Theta_i)\, d\Theta_i = 0, \end{aligned} \tag{12.1.14}$$

and

$$\begin{aligned} \gamma_k &= E(Z_t Z_{t+k}) \\ &= \sum_{i=1}^{M} A_i^2 E\{\sin(\omega_i t + \Theta_i)\sin[\omega_i(t + k) + \Theta_i)]\} \\ &= \sum_{i=1}^{M} A_i^2 E\left\{\frac{1}{2}[\cos \omega_i k - \cos(\omega_i(2t + k) + 2\Theta_i)]\right\} \\ &= \frac{1}{2}\sum_{i=1}^{M} A_i^2 \cos \omega_i k, \qquad k = 0, \pm 1, \pm 2, \ldots, \end{aligned} \tag{12.1.15}$$

where we note that

$$E\{\cos[\omega_i(2t + k) + 2\Theta_i]\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} cos[\omega_i(2t + k) + 2\Theta_i] d\Theta_i = 0.$$

Clearly, the autocovariance function γ_k in (12.1.15) is not absolutely summable. Hence, we cannot represent it in the form of (12.1.4). We can, however, always represent it in the form of (12.1.11). To do that we make the following important observations:

1. Z_t is the sum of M independent components.
2. ω_i can be anywhere in the interval $[-\pi, \pi]$, and γ_k in (12.1.11) is also the autocovariance function for the process given in (12.1.13) at frequencies $-\omega_i$.
3. $$\gamma_0 = \frac{1}{2}\sum_{i=1}^{M} A_i^2, \tag{12.1.16}$$

and hence the variance $A_i^2/2$ of the ith component is its contribution to the total variance of Z_t at both frequencies $-\omega_i$ and ω_i. In other words, one half of the variance of the ith component is associated with the frequency $-\omega_i$, whereas the other half is associated with the frequency ω_i. Thus, Equation (12.1.15) can be rewritten as

$$\gamma_k = \sum_{i=-M,\, i\neq 0}^{M} c_i \cos \omega_i k, \tag{12.1.17}$$

where

$$c_i = c_{-i} = \begin{cases} A_i^2/4, & \omega_i \neq 0, \\ A_i^2/2, & \omega_i = 0, \end{cases}$$

or, equivalently,

$$\begin{aligned} \gamma_k &= \int_{-\pi}^{\pi} \cos \omega k \, dF(\omega) \\ &= \int_{-\pi}^{\pi} e^{i\omega k} \, dF(\omega), \end{aligned} \tag{12.1.18}$$

where $F(\omega)$ is a monotonically nondecreasing step function with steps of size $A_i^2/4$ at $\omega = \pm\omega_i \neq 0$ and a step of size $A_i^2/2$ for $\omega = \omega_i = 0$ for $i = 1, \ldots, M$.

Figure 12.2 illustrates the spectral distribution function for a process with $M = 2$, $A_1 = 1$, $A_2 = 2$, and nonzero frequencies ω_1 and ω_2. The corresponding spectrum representing the variance contributions at different frequencies is a set of discrete lines and is shown in Figure 12.3.

Because (12.1.17) is clearly periodic, this result is certainly expected. As shown in Section 11.5, the spectrum for a periodic sequence is a discrete or line spectrum.

Although the spectral distribution function $F(\omega)$ is a nonnegative and nondecreasing function, it does not quite have the properties of a probability distribution function because $\int_{-\pi}^{\pi} dF(\omega) = \gamma_0$, which is not necessarily equal to 1. If we define

$$G(\omega) = \frac{F(\omega)}{\gamma_0}, \tag{12.1.19}$$

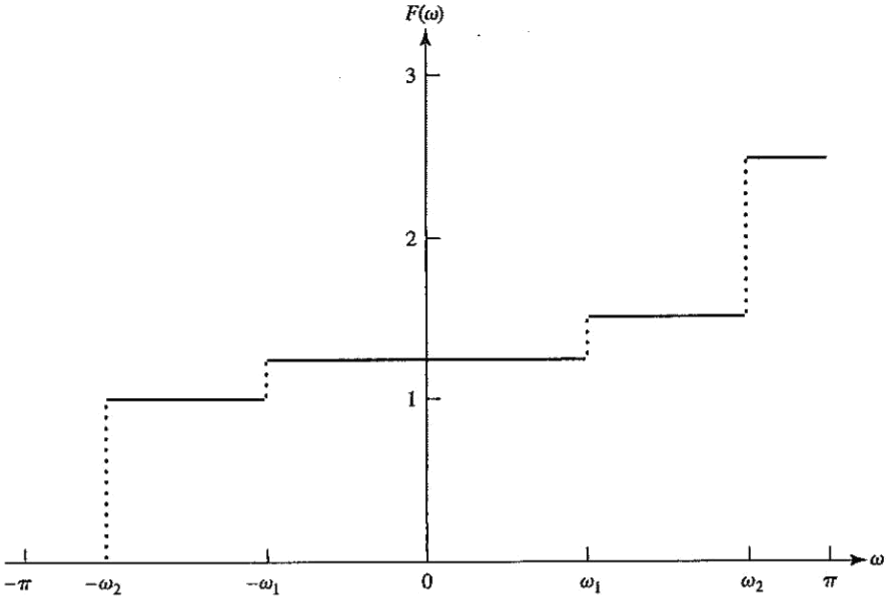


FIGURE 12.2 Spectral distribution function for $Z_t = \sum_{i=1}^{2} A_i \sin(\omega_i t + \Theta_i), A_1 = 1, A_2 = 2.$

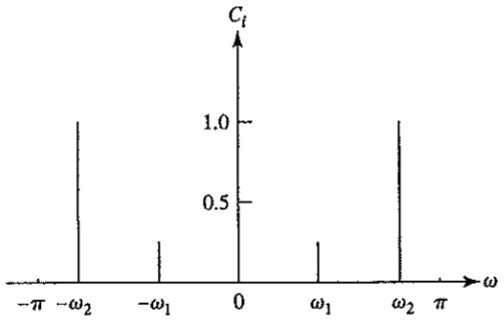


FIGURE 12.3 Line spectrum for $Z_t = \sum_{i=1}^{2} A_i \sin(\omega_i t + \Theta_i), A_1 = 1, A_2 = 2.$

however, then $G(\omega) \geq 0$ and $\int_{-\pi}^{\pi} dG(\omega) = 1$. When $dF(\omega) = f(\omega)\, d\omega$, we have

$$p(\omega)\, d\omega = dG(\omega) = \frac{f(\omega)}{\gamma_0}\, d\omega. \tag{12.1.20}$$

Hence, from (12.1.1) and (12.1.4), we get the following corresponding Fourier transform pair:

$$p(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\rho_k e^{-i\omega k} \tag{12.1.21}$$

$$= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\rho_k \cos \omega k \tag{12.1.22}$$

$$= \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\rho_k \cos \omega k, \qquad -\pi \leq \omega \leq \pi \tag{12.1.23}$$

and

$$\rho_k = \int_{-\pi}^{\pi} p(\omega)e^{i\omega k}\, d\omega, \qquad k = 0, \pm 1, \pm 2, \ldots. \tag{12.1.24}$$

The function $p(\omega)$ has the properties of a probability density function over the range $[-\pi, \pi]$ and is often referred to as the spectral density function, although the same term has also been used to refer to the spectrum $f(\omega)$ in the literature.

12.1.3 Wold's Decomposition of a Stationary Process

Note that the process Z_t in (12.1.13) is clearly stationary because $E(Z_t) = 0$ and $E(Z_t Z_{t+k}) = \frac{1}{2}\sum_{i=1}^{M} A_i^2 \cos \omega_i k = \gamma_k$. There is, however, a very important difference between this process and the processes discussed in Chapters 2 through 10. The linear cyclical model given in (12.1.13) is deterministic in the sense that it can be predicted without error from its past record. A process such as the ARIMA model that does not have this property is termed nondeterministic. Obviously, a process may contain both deterministic and nondeterministic components. For example, the model

$$Z_t = A \sin\left(\frac{2\pi t}{12} + \lambda\right) + \left[\frac{\theta(B)}{\phi(B)}\right]a_t, \tag{12.1.25}$$

where A is a constant, λ is a uniformly distributed random variable in $[-\pi, \pi]$, and $[\theta(B)/\phi(B)]$ a_t is a stationary and invertible ARMA model, contains both deterministic and nondeterministic components. In fact, Wold (1938) showed that any covariance stationary process can be represented as

$$Z_t = Z_t^{(d)} + Z_t^{(n)}, \tag{12.1.26}$$

where $Z_t^{(d)}$ is a purely deterministic component and $Z_t^{(n)}$ is a purely nondeterministic component. This representation is known as Wold's decomposition of a stationary process. This decomposition is similar to the decomposition of spectral distribution function given in (12.1.12) with $Z_t^{(d)}$ having a step spectral distribution function and $Z_t^{(n)}$ having an absolutely continuous spectral distribution function.

12.1.4 The Spectral Representation of Stationary Processes

Using exactly the same argument that an autocovariance sequence γ_k can always be represented in the form given in (12.1.11), a given time series realization Z_t can be written as the following Fourier–Stieltjes integral:

$$Z_t = \int_{-\pi}^{\pi} e^{i\omega t}\, dU(\omega), \tag{12.1.27}$$

where $U(\omega)$ plays a similar role to $F(\omega)$ in Equation (12.1.11) for a single realization. Because every realization can be represented in this form, however, we certainly should not expect the function $U(\omega)$ to be the same for all realizations. Instead, the function $U(\omega)$ should, in general, vary from realization to realization. In other words, corresponding to each realization of the process Z_t there will be a realization of $U(\omega)$. If (12.1.27) is used to represent the process Z_t of all possible realizations, then $U(\omega)$ is itself a complex valued stochastic process for each ω. The integral on the right-hand side of (12.1.27) thus becomes a stochastic integral, and the equality is defined in the mean-square sense, i.e.,

$$E\left[\left|Z_t - \int_{-\pi}^{\pi} e^{i\omega t}\, dU(\omega)\right|^2\right] = 0. \tag{12.1.28}$$

The relation (12.1.27) is called the spectral representation of the stationary process Z_t.

To study the properties of $U(\omega)$, we assume, with no loss of generality, that Z_t is a zero-mean stationary process. Hence,

$$E[dU(\omega)] = 0, \qquad \text{for all } \omega. \tag{12.1.29}$$

Next, we note that for a zero-mean complex valued stationary process Z_t, the autocovariance function γ_k is defined as

$$\gamma_k = E(Z_t Z_{t+k}^*), \tag{12.1.30}$$

where Z_{t+k}^* denotes the complex conjugate of Z_{t+k}. This general definition is necessary so that the variance of Z_t will be real. Clearly, when Z_t is a real valued process, we have $Z_{t+k}^* = Z_{t+k}$ and hence $\gamma_k = E(Z_t Z_{t+k})$. Taking the complex conjugate of both sides of (12.1.27), we obtain

$$Z_t^* = \int_{-\pi}^{\pi} e^{-i\omega t}\, dU^*(\omega) = \int_{-\pi}^{\pi} e^{i\omega t}\, dU^*(-\omega). \tag{12.1.31}$$

Hence, Z_t is real-valued if and only if for all ω

$$dU^*(-\omega) = dU(\omega). \tag{12.1.32}$$

Now, consider

$$\begin{aligned}\gamma_k &= E(Z_t Z_{t+k}^*) \\ &= E\left[\int_{-\pi}^{\pi} e^{i\omega t}\, dU(\omega) \int_{-\pi}^{\pi} e^{-i\omega(t+k)}\, dU^*(\omega)\right] \\ &= E\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i\omega t} e^{-i\lambda(t+k)}\, dU(\omega)\, dU^*(\lambda)\right] \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(\omega-\lambda)t} e^{-i\lambda k} E[dU(\omega)\, dU^*(\lambda)], \end{aligned} \tag{12.1.33}$$

which should be independent of t as Z_t is assumed to be stationary. This result implies that the contribution to the double integral in (12.1.33) must be zero when $\omega \neq \lambda$, i.e.,

$$E[dU(\omega)\, dU^*(\lambda)] = 0, \qquad \text{for all } \omega \neq \lambda. \tag{12.1.34}$$

In other words, $U(\omega)$ is an orthogonal process. Letting $\lambda = \omega$ in (12.1.33), we get

$$\begin{aligned}\gamma_k &= \int_{-\pi}^{\pi} e^{-i\omega k} E\{|dU(\omega)^2|\} = \int_{-\pi}^{\pi} e^{i\omega k} E\{|dU(\omega)|^2\} \\ &= \int_{-\pi}^{\pi} e^{i\omega k}\, dF(\omega), \end{aligned} \tag{12.1.35}$$

where

$$dF(\omega) = E[|dU(\omega)|^2] = E\, dU(\omega)\, dU^*(\omega), \qquad \text{for } -\pi \leq \omega \leq \pi, \tag{12.1.36}$$

and $F(\omega)$ is the spectral distribution function of Z_t. If Z_t is purely nondeterministic so that $F(\omega) = F_c(\omega)$, then we have

$$f(\omega)\, d\omega = dF(\omega) = E[|dU(\omega)|^2], \qquad \text{for } -\pi \leq \omega \leq \pi, \tag{12.1.37}$$

and

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} f(\omega) d\omega. \tag{12.1.38}$$

The above result shows that a stationary process can always be represented as the limit of the sum of complex exponentials or, equivalently, sine and cosine functions with uncorrelated random coefficients. This representation is also known as Cramer's representation. The derivation leading from (12.1.27) to (12.1.38) provides a more explicit physical interpretation of the spectrum discussed earlier in (12.1.10). That is, $f(\omega)\, d\omega$ is the contribution to the variance of Z_t attributable to the component $Z_t(\omega)$ of Z_t having frequencies in the range $(\omega, \omega + d\omega)$.

12.2 The Spectrum of Some Common Processes

12.2.1 The Spectrum and the Autocovariance Generating Function

Recall from Section 2.6 that, for a given sequence of autocovariances γ_k, $k = 0, \pm 1, \pm 2, \ldots$, the autocovariance generating function is defined as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k, \tag{12.2.1}$$

where the variance of the process γ_0 is the coefficient of $B^0 = 1$ and the autocovariance of lag k, γ_k, is the coefficient of both B^k and B^{-k}. Now, we know that if the given autocovariance sequence γ_k is absolutely summable, then the spectrum or the spectral density exists and equals

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k}. \tag{12.2.2}$$

Comparing Equations (12.2.1) and (12.2.2), we note that for a process with an absolutely summable autocovariance sequence, its spectrum and autocovariance generating function are related as follows:

$$f(\omega) = \frac{1}{2\pi} \gamma(e^{-i\omega}). \tag{12.2.3}$$

This result is very useful in deriving the spectrum for many commonly used processes.

12.2.2 The Spectrum of ARMA Models

Any nondeterministic stationary process Z_t can be written as

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B) a_t \tag{12.2.4}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, $\psi_0 = 1$. Without loss of generality, we assume that $E(Z_t) = 0$. From (2.6.9), the autocovariance generating function of this linear process is given by

$$\gamma(B) = \sigma_a^2 \psi(B) \psi(B^{-1}). \tag{12.2.5}$$

Now, a general stationary ARMA(p, q) model

$$\phi_p(B) Z_t = \theta_q(B) a_t \tag{12.2.6}$$

with $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ sharing no common factors can be written as

$$Z_t = \psi(B) a_t$$

with $\psi(B) = \theta_q(B)/\phi_p(B)$. Hence, the autocovariance generating function of the ARMA(p, q) model becomes

$$\gamma(B) = \sigma_a^2 \frac{\theta_q(B)\theta_q(B^{-1})}{\phi_p(B)\phi_p(B^{-1})}. \tag{12.2.7}$$

When the model is stationary, the roots of $\phi_p(B) = 0$ lie outside of the unit circle, guaranteeing the absolute summability of the autocovariance function. Consequently, the spectrum of a stationary ARMA(p, q) model is given by

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi}\gamma(e^{-i\omega}) \\ &= \frac{\sigma_a^2}{2\pi}\frac{\theta_q(e^{-i\omega})\theta_q(e^{i\omega})}{\phi_p(e^{-i\omega})\phi_p(e^{i\omega})} \qquad (12.2.8a) \\ &= \frac{\sigma_a^2}{2\pi}\left|\frac{\theta_q(e^{-i\omega})}{\phi_p(e^{-i\omega})}\right|^2, \qquad (12.2.8b) \end{aligned}$$

which is also known as a rational spectrum.

If the model is invertible, i.e., the roots of $\theta_q(B) = 0$ lie outside of the unit circle, $\theta_q(e^{-i\omega})\theta_q(e^{i\omega})$ will not vanish. Hence, the inverse of $f(\omega)$ exists and is given by

$$\begin{aligned} f^{-1}(\omega) &= \frac{2\pi}{\sigma_a^2}\frac{\phi_p(e^{-i\omega})\phi_p(e^{i\omega})}{\theta_q(e^{-i\omega})\theta_q(e^{i\omega})} \qquad (12.2.9a) \\ &= \frac{2\pi}{\sigma_a^2}\left|\frac{\phi_p(e^{-i\omega})}{\theta_q(e^{-i\omega})}\right|^2, \qquad (12.2.9b) \end{aligned}$$

which can be easily seen to be the spectrum of an ARMA(q, p) process. Applying the inverse Fourier transform of $f^{-1}(\omega)$, we get

$$\gamma_k^{(I)} = \int_{-\pi}^{\pi} f^{-1}(\omega)e^{i\omega k}d\omega, \tag{12.2.10}$$

which is known as the inverse autocovariance function. Naturally,

$$\rho_k^{(I)} = \frac{\gamma_k^{(I)}}{\gamma_0^{(I)}} = \frac{1}{\gamma_0^{(I)}}\int_{-\pi}^{\pi} f^{-1}(\omega)e^{i\omega k}d\omega \tag{12.2.11}$$

is called the inverse autocorrelation function, a concept due to Cleveland (1972). As discussed in Section 6.3, the inverse autocorrelation function can be used as an identification tool that has similar characteristics as the partial autocorrelation function. Equations (12.2.9) and (12.2.11) provide an alternative method to calculate the sample inverse autocorrelation function.

The Spectrum of a White Noise Process A white noise process

$$Z_t = a_t \tag{12.2.12}$$

is a series of uncorrelated random variables with autocovariance function

$$\gamma_k = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{12.2.13}$$

and autocovariance generating function $\gamma(B) = \sigma_a^2$. Hence, by (12.1.1) or equivalently by (12.2.3), its spectrum is

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k} \\ &= \frac{\sigma_a^2}{2\pi}, \qquad -\pi \leq \omega \leq \pi, \end{aligned} \tag{12.2.14}$$

which is a horizontal straight line as shown in Figure 12.4. This result implies that the contributions to the variance at all frequencies are equal. It is well known that white light is produced when the light spectrum at all frequencies is the same. In fact, the white noise process is named because of its similarity to white light in terms of the spectrum. In model building, the diagnostic checking for white noise can be performed by examining whether the spectrum of residuals is relatively flat.

The Spectrum of an AR(1) Process The autocovariance generating function of a stationary AR(1) process, $(1 - \phi B)Z_t = a_t$, is given by

$$\gamma(B) = \sigma_a^2 \frac{1}{(1 - \phi B)(1 - \phi B^{-1})}. \tag{12.2.15}$$

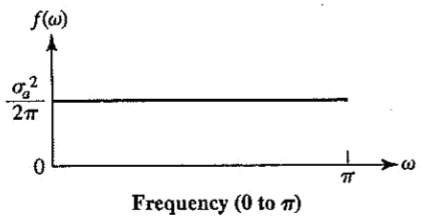


FIGURE 12.4 The spectrum of a white noise process.

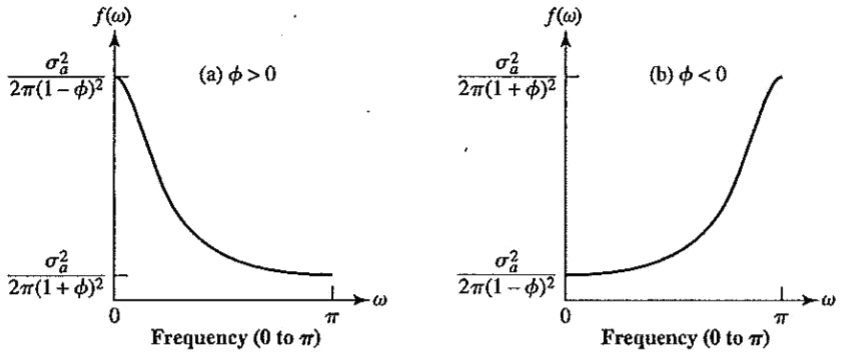


FIGURE 12.5 Spectrum for the AR(1) process.

Thus, by (12.2.8a) the spectrum is

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \frac{1}{(1 - \phi e^{-i\omega})(1 - \phi e^{i\omega})}$$
$$= \frac{\sigma_a^2}{2\pi} \frac{1}{(1 + \phi^2 - 2\phi \cos \omega)}. \tag{12.2.16}$$

The shape of the spectrum depends on the sign of ϕ. As shown in Figure 12.5, when $\phi > 0$ and hence the series is positively autocorrelated, the spectrum is dominated by low-frequency (long-period) components. If $\phi < 0$, then the series is negatively autocorrelated and the spectrum is dominated by high-frequency (short-period) components. In other words, a low-frequency dominating spectrum indicates a relatively smooth series, whereas a high-frequency dominating spectrum implies a more ragged series. We remind the reader that these phenomena were, in fact, illustrated earlier in Examples 11.2 and 11.3.

As ϕ approaches 1, the AR(1) process in the limit becomes a random walk model; hence,

$$f(\omega) = \frac{\sigma_a^2}{4\pi} \frac{1}{(1 - \cos \omega)} \tag{12.2.17}$$

can be taken to be the spectrum of a random walk model, although, strictly speaking, a random walk does not have a spectrum because its autocovariance sequence is not absolutely summable. The function (12.2.17) however, is a well-behaved nonnegative even function other than a possible problem at $\omega = 0$ where the peak of the function becomes infinite in height. The implication of this phenomenon in data analysis is that a sample spectrum with strong peak near zero frequencies indicates a possible need for differencing.

The Spectrum of an MA(1) Process The autocovariance generating function of a moving average process of order 1, $Z_t = (1 - \theta B)a_t$, is

$$\gamma(B) = \sigma_a^2(1 - \theta B)(1 - \theta B^{-1}), \tag{12.2.18}$$

and its spectrum is

$$f(\omega) = \frac{\sigma_a^2}{2\pi}(1 - \theta e^{-i\omega})(1 - \theta e^{i\omega})$$
$$= \frac{\sigma_a^2}{\pi}(1 + \theta^2 - 2\theta \cos \omega). \tag{12.2.19}$$

Again, the shape of $f(\omega)$ depends on the sign of θ.

Recall that

$$\rho_k = \begin{cases} 1, & k = 0, \\ \dfrac{-\theta}{1 + \theta^2}, & k = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, when θ is positive, the series is negatively autocorrelated and hence relatively ragged, which is represented by a spectrum with higher values at the higher frequencies. On the other hand, when θ is negative, the series is positively autocorrelated and hence relatively smooth. The corresponding phenomenon in terms of the spectrum is its dominating values at the lower frequencies. Both cases are shown in Figure 12.6. As expected, the autocorrelation function in the time domain and the spectrum in the frequency domain contain exactly the same information about the process.

12.2.3 The Spectrum of the Sum of Two Independent Processes

Consider the process Z_t, which is the sum of two independent stationary processes X_t and Y_t,

$$Z_t = X_t + Y_t. \tag{12.2.20}$$

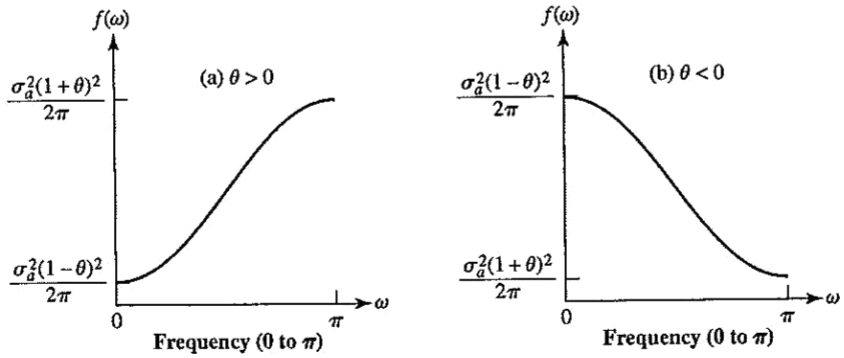


FIGURE 12.6 Spectrum for the MA(1) process.

Let $\gamma_Z(B)$, $\gamma_X(B)$, and $\gamma_Y(B)$ be the autocovariance generating functions for the Z_t, X_t, and Y_t processes. Then,

$$\begin{aligned}\gamma_Z(B) &= \sum_{k=-\infty}^{\infty} \gamma_Z(k)B^k \\ &= \sum_{k=-\infty}^{\infty} \mathrm{Cov}(Z_t, Z_{t+k})B^k \\ &= \sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_t + Y_t, X_{t+k} + Y_{t+k})B^k \\ &= \sum_{k=-\infty}^{\infty} [\mathrm{Cov}(X_t, X_{t+k}) + \mathrm{Cov}(Y_t, Y_{t+k})]B^k \\ &= \sum_{k=-\infty}^{\infty} \gamma_X(k)B^k + \sum_{k=-\infty}^{\infty} \gamma_Y(k)B^k \\ &= \gamma_X(B) + \gamma_Y(B). \end{aligned} \tag{12.2.21}$$

Hence, it follows that

$$f_Z(\omega) = f_X(\omega) + f_Y(\omega). \tag{12.2.22}$$

That is, the spectrum of the sum of two independent processes is the sum of the two independent spectra.

12.2.4 The Spectrum of Seasonal Models

From the previous results in Section 11.4 and Section 12.1.2, we know that a deterministic seasonal component with seasonal period s will have a discrete (line) spectrum at the harmonic (Fourier) frequencies $\omega_k = 2\pi k/s$, $k = \pm 1, \pm 2, \ldots, \pm[s/2]$. For example, the model

$$Z_t = \alpha_0 + \sum_{k=1}^{6}\left[\alpha_k \cos\left(\frac{2\pi kt}{12}\right) + \beta_k \sin\left(\frac{2\pi kt}{12}\right)\right] + a_t, \tag{12.2.23}$$

where the a_t represent a white noise error process, can be used to fit a monthly series with a cyclical trend. The spectrum of Z_t in (12.2.23), from the result in Section 12.2.3, will be the sum of two independent spectra. One represents the deterministic component in (12.2.23), and the other represents the error process a_t. Because the spectrum of the white noise error process is a horizontal flat line, the resulting spectrum will have spikes at the harmonic seasonal frequencies $\omega_k = 2\pi k/12$ for $k = \pm 1, \pm 2, \ldots, \pm 6$.

Now consider the following seasonal ARMA model with seasonal period 12:

$$(1 - \Phi B^{12})Z_t = a_t. \tag{12.2.24}$$

The autocovariance generating function is

$$\gamma(B) = \sigma_a^2 \frac{1}{(1 - \Phi B^{12})(1 - \Phi B^{-12})}. \tag{12.2.25}$$

If the process is stationary, then its spectrum exists and equals

$$\begin{aligned} f(\omega) &= \frac{\sigma_a^2}{2\pi} \frac{1}{(1 - \Phi e^{-i12\omega})(1 - \Phi e^{i12\omega})} \\ &= \frac{\sigma_a^2}{2\pi} \frac{1}{1 + \Phi^2 - 2\Phi \cos(12\omega)}. \end{aligned} \tag{12.2.26}$$

Note that $\cos(12\omega) = -1$ at $\omega = \pi(2k - 1)/12$ for $k = 1, 2, 3, 4, 5, 6$, and $\cos(12\omega) = 1$ for $\omega = 0$ and $\omega = 2\pi k/12$, $k = 1, 2, 3, 4, 5, 6$. Thus, for $\Phi > 0$, other than a peak at $\omega = 0$, the spectrum will also exhibit peaks at seasonal harmonic frequencies $2\pi k/12$ for $k = 1, 2, 3, 4, 5$, and 6 and troughs at frequencies $\omega = \pi(2k - 1)/12$ for $k = 1, 2, 3, 4, 5$, and 6 as shown in Figure 12.7. The spectrum for $\Phi < 0$ can be derived similarly.

Consider a multiplicative seasonal ARMA model having both seasonal and nonseasonal ARMA components,

$$(1 - \Phi B^{12})(1 - \phi B)Z_t = a_t. \tag{12.2.27}$$

From (12.2.16) and (12.2.26), the spectrum can easily be seen to be

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \frac{1}{(1 + \Phi^2 - 2\Phi \cos 12\omega)(1 + \phi^2 - 2\phi \cos \omega)}. \tag{12.2.28}$$

The general shape of the spectrum for $\Phi > 0$ and $\phi > 0$ is shown in Figure 12.8.

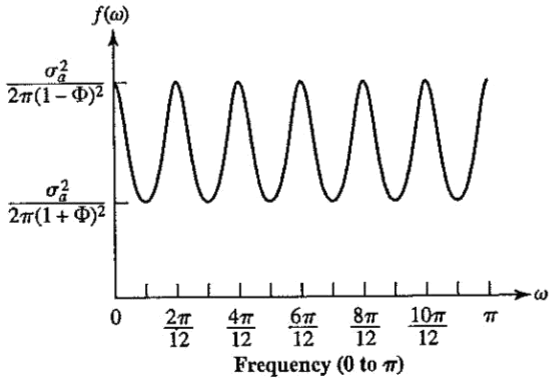


FIGURE 12.7 Spectrum for $(1 - \Phi B^{12})Z_t = a_t$.

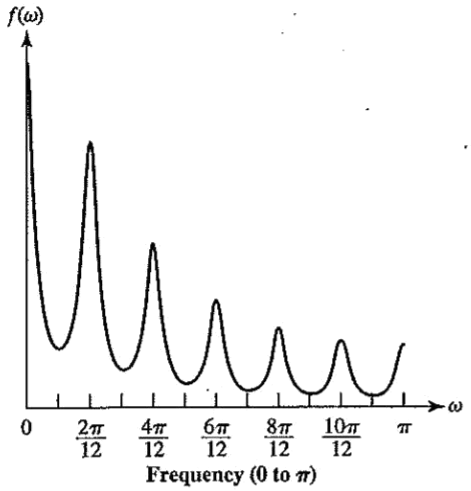


FIGURE 12.8 Spectrum for $(1 - \Phi B^{12})(1 - \phi B)Z_t = a_t$.

The sample spectrum discussed in the next chapter does not behave as regularly as those in Figures 12.7 or 12.8. If the data indeed contain a seasonal phenomenon, though, then the spectrum should exhibit peaks or bumps at the harmonically related seasonal frequencies.

Again, we remind the reader about the difference between the spectra for the Equations (12.2.23) and (12.2.24). Although both are purely seasonal models, the cyclical trend term in (12.2.23) is deterministic and hence has a discrete (line) spectrum, whereas the seasonal phenomenon in (12.2.24) is nondeterministic with a continuous spectrum. This difference implies that, in data analysis, a sharp spike in the sample spectrum may indicate a possible deterministic cyclical component, whereas broad peaks, unless due to the smoothing effect to be discussed in the next chapter, often imply a nondeterministic seasonal component similar to the multiplicative seasonal ARIMA process. In either case, it is generally impossible to determine the exact form or the order of the underlying model from a spectrum. In terms of the ARIMA model building, the autocorrelations and partial autocorrelations are, in general, more useful. In any event, for model building the autocorrelation and the spectrum should complement rather than compete with each other.

12.3 The Spectrum of Linear Filters

12.3.1 The Filter Function

Let Z_t be a stationary process with an absolutely summable autocovariance sequence $\gamma_Z(k)$ and corresponding spectrum

$$f_Z(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_Z(l) e^{-i\omega l}.$$

Consider the following linear filter of Z_t given by

$$Y_t = \sum_{j=-\infty}^{\infty} \alpha_j Z_{t-j} = \sum_{j=-\infty}^{\infty} \alpha_j B^j Z_t = \alpha(B) Z_t, \tag{12.3.1}$$

where $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$ and $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$.

Without loss of generality, assume that $E(Z_t) = 0$ and hence $E(Y_t) = 0$. We have

$$\begin{aligned} E(Y_t Y_{t+k}) &= E\left[\left(\sum_{h=-\infty}^{\infty} \alpha_h Z_{t-h}\right)\left(\sum_{j=-\infty}^{\infty} \alpha_j Z_{t+k-j}\right)\right] \\ &= E\left[\sum_{h=-\infty}^{\infty} \alpha_h \sum_{j=-\infty}^{\infty} \alpha_j Z_{t-h} Z_{t+k-j}\right] \\ &= \sum_{h=-\infty}^{\infty} \alpha_h \sum_{j=-\infty}^{\infty} \alpha_j \gamma_Z(k - j + h) \\ &= \sum_{h=-\infty}^{\infty} \alpha_h \sum_{l=-\infty}^{\infty} \alpha_{h+k-l} \gamma_Z(l) = \gamma_Y(k), \end{aligned} \tag{12.3.2}$$

where we let $l = k - j + h$.

Now, $\gamma_Y(k)$ is absolutely summable, because

$$\begin{aligned} \sum_{k=-\infty}^{\infty} |\gamma_Y(k)| &\le \sum_{k=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} |\alpha_h| \sum_{l=-\infty}^{\infty} |\alpha_{h+k-l}| |\gamma_Z(l)| \\ &= \sum_{h=-\infty}^{\infty} |\alpha_h| \sum_{l=-\infty}^{\infty} |\gamma_Z(l)| \sum_{k=-\infty}^{\infty} |\alpha_{h+k-l}| \\ &= \sum_{h=-\infty}^{\infty} |\alpha_h| \sum_{l=-\infty}^{\infty} |\gamma_Z(l)| \sum_{k=-\infty}^{\infty} |\alpha_k| < \infty. \end{aligned}$$

Hence, we have

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-i\omega k} \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\omega k} \sum_{h=-\infty}^{\infty} \alpha_h \sum_{l=-\infty}^{\infty} \alpha_{h+k-l} \gamma_Z(l) \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \alpha_h \alpha_{h+k-l} \gamma_Z(l) e^{i\omega h} e^{-i\omega l} e^{-i\omega(h+k-l)} \\ &= \sum_{h=-\infty}^{\infty} \alpha_h e^{i\omega h} \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_Z(l) e^{-i\omega l} \sum_{k=-\infty}^{\infty} \alpha_{h+k-l} e^{-i\omega(h+k-l)} \\ &= \sum_{h=-\infty}^{\infty} \alpha_h e^{i\omega h} \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_Z(l) e^{-i\omega l} \sum_{t=-\infty}^{\infty} \alpha_t e^{-i\omega t} \\ &= |\alpha(e^{i\omega})|^2 f_Z(\omega), \end{aligned} \tag{12.3.3}$$

where

$$|\alpha(e^{i\omega})|^2 = \alpha(e^{i\omega})[\alpha(e^{i\omega})]^* = \alpha(e^{i\omega})\alpha(e^{-i\omega}) \tag{12.3.4}$$

and which is often known as the filter function or the transfer function. This filter function can be used to measure the effect of applying a filter on a series. In many studies, one important objective is to design a good filter so that the output series or signals satisfy certain desired properties.

12.3.2 Effect of Moving Average

Consider the simple moving average

$$Y_t = \frac{1}{m}\sum_{j=0}^{m-1} Z_{t-j} = \alpha(B)Z_t, \qquad m \geq 2, \tag{12.3.5}$$

where $\alpha(B) = (\sum_{j=0}^{m-1} B^j)/m$. Then, from (12.3.4),

$$\begin{aligned} |\alpha(e^{i\omega})|^2 &= \frac{1}{m^2}\left(\sum_{j=0}^{m-1} e^{i\omega j}\right)\left(\sum_{j=0}^{m-1} e^{-i\omega j}\right) \\ &= \frac{1}{m^2}\left(\frac{1 - e^{im\omega}}{1 - e^{i\omega}}\right)\left(\frac{1 - e^{-im\omega}}{1 - e^{-i\omega}}\right) \\ &= \frac{1}{m^2}\frac{2 - 2\cos m\omega}{2 - 2\cos \omega} \\ &= \frac{1}{m^2}\frac{1 - \cos m\omega}{1 - \cos \omega}. \end{aligned} \tag{12.3.6}$$

Note that $\cos m\omega = -1$ when $m\omega = (2k - 1)\pi$, i.e., $\omega = (2k - 1)\pi/m$ for $k = 1, \ldots, [(m + 1)/2]$, and $\cos m\omega = 1$ when $m\omega = 2k\pi$ or $\omega = 2k\pi/m$ for $k = 1, \ldots, [m/2]$. Also,

$$\lim_{\omega \to 0} |\alpha(e^{i\omega})|^2 = \lim_{\omega \to 0} \frac{1}{m^2}\left(\frac{1 - \cos m\omega}{1 - \cos \omega}\right) = 1.$$

Hence, for $m \geq 2$, the filter function has its absolute maximum at $\omega = 0$ and relative maxima at frequencies $\omega = (2k - 1)\pi/m$ for $k = 1, \ldots, [(m + 1)/2]$. The function vanishes at frequencies $\omega = 2k\pi/m$ for $k = 1, \ldots, [m/2]$. Because $(1 - \cos \omega)$ is an increasing function between 0 and π, the overall shape of the filter function looks like that shown in Figure 12.9. When m is large, the first peak around the zero frequency becomes dominating and the relative maxima at other frequencies become negligible, which implies that Y_t will mainly contain

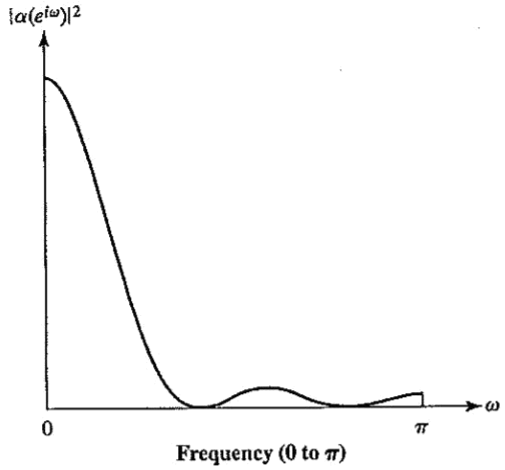


FIGURE 12.9 Filter function for $|\alpha(e^{i\omega})|^2 = 1/m^2[(1 - \cos m\omega)/(1 - \cos \omega)]$.

only the low frequency component of the original series Z_t. Because a monotonic trend has an infinite period that corresponds to a zero frequency, the moving average has often been used to estimate the monotonic trend component of a series. A filter like this one, which preserves the low-frequency component and attenuates or eliminates the high-frequency component of a series, is called a low pass filter.

12.3.3 Effect of Differencing

Next, we consider the differencing operator

$$\begin{aligned} W_t &= (1 - B)Z_t \\ &= \alpha(B)Z_t, \end{aligned} \tag{12.3.7}$$

where $\alpha(B) = (1 - B)$. From (12.3.4), we have

$$\begin{aligned} |\alpha(e^{i\omega})|^2 &= (1 - e^{i\omega})(1 - e^{-i\omega}) \\ &= 2(1 - \cos \omega). \end{aligned} \tag{12.3.8}$$

Clearly, $|\alpha(e^{i\omega})|^2$ is 0 at $\omega = 0$. It increases between 0 and π and attains its maximum at $\omega = \pi$, as shown in Figure 12.10. Thus, the differencing filter preserves the higher-frequency component and essentially eliminates the low-frequency component of a series. Such a filter is called a high pass filter and is often employed to remove the trend of a series. The seasonal differencing has an effect similar to a high pass filter and is left as an exercise.

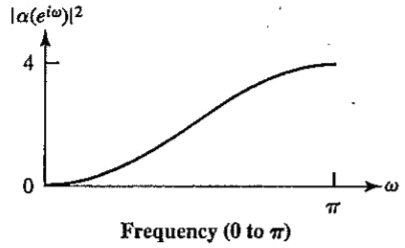


FIGURE 12.10 Filter function for differencing $|\alpha(e^{i\omega})|^2 = 2(1 - \cos\omega)$.

12.4 Aliasing

Consider the following cosine curves:

$$\cos\left[(\omega \pm 2\pi j/\Delta t)t\right], \quad \text{for } j = 0, \pm 1, \pm 2, \ldots \quad \text{and} \quad -\infty < t < \infty. \tag{12.4.1}$$

For any given ω, the curves $\cos(\omega t)$ and $\cos[(\omega \pm 2\pi j/\Delta t)t]$ for $j \neq 0$ are clearly different, as shown in Figure 12.11 for $\cos(\omega t)$ and $\cos[(\omega + 2\pi/\Delta t)t]$ with $\omega = \pi/2$ and $\Delta t = 1$. Now, suppose that we study the curve $\cos(\omega t)$ as shown in the dotted line and observe it only at the time points $k\,\Delta t$ for $k = 0, \pm 1, \pm 2, \ldots$, where Δt is the sampling interval. These curves become indistinguishable at the time points $t = k\,\Delta t$ for $k = 0, \pm 1, \pm 2, \ldots$ because

$$\cos\left[\left(\frac{\omega \pm 2\pi j}{\Delta t}\right)k\,\Delta t\right] = \cos(\omega k\,\Delta t \pm 2\pi jk) = \cos(\omega k\,\Delta t). \tag{12.4.2}$$

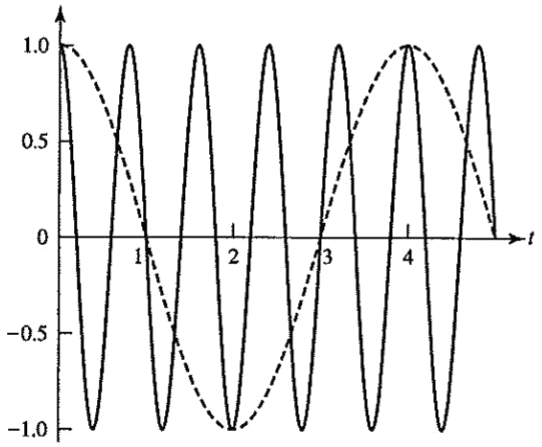


FIGURE 12.11 The aliasing.

The same problem occurs for the sine and exponential functions. Thus, in Fourier representation, if we observe a process Z_t at $t = k\,\Delta t$, for $k = 0, \pm 1, \pm 2, \ldots$, then the components of Z_t with frequencies $\omega \pm 2\pi j/\Delta t$, for $j = \pm 1, \pm 2, \ldots$ will all appear to have frequency ω. These frequencies are said to be *aliases* of ω. The phenomenon, in which a high frequency component takes on the identity of a lower frequency, is called *aliasing*.

For a given sample interval Δt, if we observe a process at time points $k\,\Delta t$ for $k = 0, \pm 1, \pm 2, \ldots$, the fastest observable oscillation goes above the mean at one time point and below the mean at the next time point. The period of this fastest oscillation is $2\Delta t$. Thus, the highest frequency used in Fourier transform is $2\pi/2\Delta t = \pi/\Delta t$, which is also known as the Nyquist frequency with the given sampling interval Δt. With this Nyquist frequency $\pi/\Delta t$, if $\omega < \pi/\Delta t$, then, as discussed above, the components with frequencies $\omega \pm 2\pi j/\Delta t$, for $j = 0, \pm 1, \pm 2, \ldots$ are all confounded and are aliases of one another, and the sum of all their contributions will appear at frequency ω in $F(\omega)$ or $f(\omega)$, making the estimation of contributions of different components difficult or impossible. To avoid this confounding or aliasing problem, we thus should choose the sampling interval so that the frequencies under study are all less than the Nyquist frequency. For example, if a phenomenon is a linear combination of sine-cosine components with frequencies $\omega_1, \omega_2, \ldots$, and ω_m, then we should choose Δt so that $|\omega_j| < (\pi/\Delta t)$ for $j = 1, 2, \ldots, m$. Otherwise, if $\pi/\Delta t < \omega_3$, say, then ω_3 will be aliased with some other frequency in the interval $[0, \pi/\Delta t]$ and hence make the estimation of the component at ω_3 impossible. In data analysis, we often take $\Delta t = 1$.

EXERCISES

12.1 Consider the process

$$Z_t = A_1 \cos\frac{\pi}{2}t + A_2 \sin\frac{\pi}{2}t, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

where A_1 and A_2 are independent random variables, each with mean 0 and variance π.

(a) Find the autocovariance function γ_k for Z_t.

(b) Is γ_k absolutely summable? Find the spectral distribution function or the spectrum of Z_t.

12.2 Consider the process

$$Z_t = A \sin\left[\frac{\pi}{2}t + \lambda\right] + a_t, \qquad \text{for } t = 0, \pm 1, \pm 2, \ldots,$$

where A is a constant, λ is a uniformly distributed random variable on $[-\pi, \pi]$, and the a_t's are i.i.d. random variables with mean 0 and variance 2π. Also, λ and a_t are independent for all t.

(a) Find the autocovariance function γ_k for Z_t.

(b) Is γ_k absolutely summable? Find the spectral distribution function or the spectrum of Z_t.

12.3 A spectral distribution function $G(\omega)$ can be decomposed in the form

$$G(\omega) = pG^a(\omega) + qG^d(\omega),$$

where p and q are nonnegative, $p + q = 1$, $G^a(\omega)$ is an absolutely continuous spectral distribution function, and $G^d(\omega)$ is a discrete or step spectral distribution function. Express the following spectral distribution function in the above decomposition form:

$$G(\omega) = \begin{cases} 0, & -\pi \le \omega < 0, \\ \omega^2 + .2, & 0 \le \omega < .5, \\ \omega, & .5 \le \omega < 1, \\ 1, & 1 \le \omega \le \pi. \end{cases}$$

12.4 Given the following spectral distribution of Z_t,

$$F_Z(\omega) = \begin{cases} .5\pi + .5\omega, & -\pi \le \omega < -\dfrac{\pi}{2}, \\ 4.5\pi + .5\omega, & -\dfrac{\pi}{2} \le \omega < \dfrac{\pi}{2}, \\ 8.5\pi + .5\omega, & \dfrac{\pi}{2} \le \omega < \pi. \end{cases}$$

Find the autocovariance function γ_k for Z_t.

12.5 Find and discuss the shape of the spectra for AR(1) processes, $(1 - \phi B)Z_t = a_t$, with the following values of ϕ:

(a) $\phi = .6$,

(b) $\phi = .99$,

(c) $\phi = -.6$.

12.6 Find and discuss the spectrum for each of the following processes:

(a) $(1 - .7B + .12B^2)Z_t = a_t$,

(b) $Z_t = (1 - .7B - .18B^2)a_t$,

(c) $(1 - .5B)Z_t = (1 + .6B)a_t$,

(d) $Z_t = (1 - .4B)(1 - .8B^4)a_t$, where the a_t's are i.i.d. $N(0, 1)$.

12.7 Find and discuss the shape of the spectrum for

$$(1 - \Phi B^4)Z_t = a_t,$$

where the a_t is a Gaussian white noise process with mean 0 and variance 1.

12.8 Let X_t and Y_t be independent stationary processes with the spectra $f_X(\omega)$ and $f_Y(\omega)$, respectively. Assume that Y_t follows the AR(2) model, $(1 - .6B - .2B^2)Y_t = a_t$, and X_t and a_t are independent white noise processes, each with mean 0 and variance σ^2. If $Z_t = X_t + Y_t$, then find and discuss the shape of the spectrum of Z_t.

12.9 Consider the following linear filter of Z_t:

$$Y_t = \frac{1}{4}Z_t + \frac{1}{2}Z_{t-1} + \frac{1}{4}Z_{t-2},$$

where Z_t is a stationary process with an absolutely summable autocovariance function. Find and discuss the filter function defined in (12.3.4).

12.10 Differencing operators have been used widely to reduce a nonstationary process to a stationary process. Discuss the effect through the filter function for each of the following differencing operators:

(a) $W_t = (1 - B)^2 Z_t$,

(b) $W_t = (1 - B^4)Z_t$,

(c) $W_t = (1 - B)(1 - B^{12})Z_t$.

13

Estimation of the Spectrum

In this chapter, we discuss estimation of the spectrum. We begin with periodogram analysis, which is a useful technique to search for hidden periodicities. The smoothing of the sample spectrum and related concepts such as the lag and spectral windows are discussed. Empirical examples are presented to illustrate the procedures and results.

13.1 Periodogram Analysis

13.1.1 The Periodogram

Given a time series of n observations, we can use the results in Section 11.3 to represent the n observations in the following Fourier representation:

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t), \tag{13.1.1}$$

where $\omega_k = 2\pi k/n$, $k = 0, 1, \ldots, [n/2]$, are Fourier frequencies, and

$$a_k = \begin{cases} \dfrac{1}{n}\sum_{t=1}^{n} Z_t \cos \omega_k t, & k = 0 \quad \text{and} \quad k = \dfrac{n}{2} \quad \text{if } n \text{ is even}, \\ \dfrac{2}{n}\sum_{t=1}^{n} Z_t \cos \omega_k t, & k = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right], \end{cases}$$

and

$$b_k = \frac{2}{n}\sum_{t=1}^{n} Z_t \sin \omega_k t, \qquad k = 1, 2, \ldots, \left[\frac{n-1}{2}\right],$$

are Fourier coefficients. It is interesting to note the close relationship between the above Fourier representation and regression analysis. In fact, the Fourier coefficients are essentially the least squares estimates of the coefficients in fitting the following regression model:

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t) + e_t, \tag{13.1.2}$$

where the ω_k are the same Fourier frequencies. The fitting will be perfect, but we can use the notion from regression analysis to regard the Parseval's relation

$$\sum_{t=1}^{n} Z_t^2 = \begin{cases} na_0^2 + \dfrac{n}{2}\displaystyle\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2), & \text{if } n \text{ is odd,} \\ na_0^2 + \dfrac{n}{2}\displaystyle\sum_{k=1}^{[(n-1)/2]} (a_k^2 + b_k^2) + na_{n/2}^2, & \text{if } n \text{ is even,} \end{cases} \tag{13.1.3}$$

as an analysis of variance presented in Table 13.1.

The quantity $I(\omega_k)$ defined by

$$I(\omega_k) = \begin{cases} na_0^2, & k = 0, \\ \dfrac{n}{2}(a_k^2 + b_k^2), & k = 1, \ldots, [(n-1)/2], \\ na_{n/2}^2, & k = \dfrac{n}{2} \text{ when } n \text{ is even,} \end{cases} \tag{13.1.4}$$

is called the periodogram. It was introduced by Schuster (1898) to search for a periodic component in a series.

13.1.2 Sampling Properties of the Periodogram

Assume that $Z_1, Z_2, \ldots,$ and Z_n are i.i.d. $N(0, \sigma^2)$. Then

$$E(a_k) = \frac{2}{n}\sum_{t=1}^{n} E(Z_t)\cos \omega_k t = 0$$

TABLE 13.1 Analysis of variance table for periodogram analysis.

Source	Degrees of freedom	Sum of squares
Frequency $\omega_0 = 0$ (Mean)	1	na_0^2
Frequency $\omega_1 = 2\pi/n$	2	$\frac{n}{2}(a_1^2 + b_1^2)$
Frequency $\omega_2 = 4\pi/n$	2	$\frac{n}{2}(a_2^2 + b_2^2)$
⋮	⋮	⋮
Frequency $\omega_{[(n-1)/2]} = [(n-1)/2]2\pi/n$	2	$\frac{n}{2}(a_{[(n-1)/2]}^2 + b_{[(n-1)/2]}^2)$
Frequency $\omega_{n/2} = \pi$ (exists only for even n)	1	$na_{n/2}^2$
Total	n	$\sum_{t=1}^{n} Z_t^2$

and

$$\begin{aligned}\text{Var}(a_k) &= \frac{4}{n^2}\sum_{t=1}^{n}\sigma^2(\cos\omega_k t)^2 \\ &= \frac{4\sigma^2}{n^2}\sum_{t=1}^{n}(\cos\omega_k t)^2 \\ &= \frac{4\sigma^2}{n^2}\cdot\frac{n}{2} = \frac{2\sigma^2}{n},\end{aligned}$$

where we use the fact that a_k and a_j are independent for $k \neq j$ because of (11.2.13). Hence, the a_k for $k = 1, 2, \ldots, [(n-1)/2]$, are i.i.d. $N(0, 2\sigma^2/n)$, and the $na_k^2/2\sigma^2, k = 1, 2, \ldots, [(n-1)/2]$, are i.i.d. chi-squares with one degree of freedom. Similarly, the $nb_k^2/2\sigma^2, k = 1, 2, \ldots, [(n-1)/2]$, are i.i.d. chi-squares with one degree of freedom. Furthermore, $na_k^2/2\sigma^2$ and $nb_j^2/2\sigma^2$ are independent for $k = 1, \ldots, [(n-1)/2]$ and $j = 1, 2, \ldots, [(n-1)/2]$, because, a_k and b_j are normal and by the orthogonality property of the cosine and sine system, we have

$$\begin{aligned}\text{Cov}(a_k, b_j) &= \frac{4}{n^2}E\left(\sum_{t=1}^{n} Z_t\cos\omega_k t\cdot\sum_{u=1}^{n} Z_u\sin\omega_j u\right) \\ &= \frac{4}{n^2}\sum_{t=1}^{n}[E(Z_t^2)\cos\omega_k t\cdot\sin\omega_j t] \\ &= \frac{4\sigma^2}{n^2}\sum_{t=1}^{n}[\cos\omega_k t\cdot\sin\omega_j t] \\ &= 0 \qquad \text{for any } k \text{ and } j. \end{aligned} \tag{13.1.5}$$

It follows that the periodogram ordinates

$$\frac{I(\omega_k)}{\sigma^2} = \frac{n}{2\sigma^2}(a_k^2 + b_k^2) \tag{13.1.6}$$

for $k = 1, 2, \ldots, [(n-1)/2]$ are i.i.d. chi-squares with two degrees of freedom. Clearly, by the same arguments, each of $I(0)/\sigma^2$ and $I(\pi)/\sigma^2$ (for even n) follows the chi-square distribution with one degree of freedom. With the adjustment for $I(\pi)$ in mind, we assume, without loss of generality, that the sample size n is odd in the following discussion.

Suppose that a time series of n observations contains a periodic component and can be represented by

$$Z_t = a_0 + a_k\cos\omega_k t + b_k\sin\omega_k t + e_t, \tag{13.1.7}$$

where $\omega_k = 2\pi k/n$, $k \neq 0$, and the e_t are i.i.d. $N(0, \sigma^2)$. To test the hypothesis in (13.1.7), it is equivalent to test in Table 13.1

$$H_0: a_k = b_k = 0 \qquad \text{vs.} \qquad H_1: a_k \neq 0 \qquad \text{or} \qquad b_k \neq 0$$

using the test statistic

$$F = \frac{[n(a_k^2 + b_k^2)/2]/2}{\left[\sum_{\substack{j=1 \\ j \neq k}}^{[n/2]} n(a_j^2 + b_j^2)/2\right]/(n-3)}$$

$$= \frac{(n-3)(a_k^2 + b_k^2)}{2\sum_{\substack{j=1 \\ j \neq k}}^{[n/2]}(a_j^2 + b_j^2)}, \tag{13.1.8}$$

which follows the F-distribution $F(2, n-3)$ with 2 and $(n-3)$ degrees of freedom. Neither the numerator nor the denominator in (13.1.8) includes the term a_0, corresponding to the zero frequency. In fact, because the periodogram at frequency zero reflects the sample mean and not the periodicity of the series, it is usually removed from consideration in the analysis. More generally, we can test whether a series contains multiple m periodic components by postulating the model

$$Z_t = a_0 + \sum_{i=1}^{m}(a_{k_i} \cos \omega_{k_i} t + b_{k_i} \sin \omega_{k_i} t) + e_t, \tag{13.1.9}$$

where the e_t are i.i.d. $N(0, \sigma^2)$, $\omega_{k_i} = 2\pi k_i/n$, and the set $I = \{k_i : i = 1, \ldots, m\}$ is a subset of $\{k : k = 1, 2, \ldots, [n/2]\}$. The corresponding test statistic from Table 13.1

$$F = \frac{(n-2m-1)\sum_{i=1}^{m}(a_{k_i}^2 + b_{k_i}^2)}{2m\sum_{j \notin I}(a_j^2 + b_j^2)} \tag{13.1.10}$$

follows the F-distribution $F(2m, n-2m-1)$ with $2m$ and $(n-2m-1)$ degrees of freedom.

13.1.3 Tests for Hidden Periodic Components

In practice, even if we believe that a time series contains a periodic component, the underlying frequency is often unknown. For example, we might test

$$H_0: \alpha = \beta = 0 \quad \text{vs.} \quad H_1: \alpha \neq 0 \quad \text{or} \quad \beta \neq 0$$

for the model

$$Z_t = \mu + \alpha \cos \omega t + \beta \sin \omega t + e_t, \tag{13.1.11}$$

where $\{e_t\}$ is a Gaussian white noise series of i.i.d. $N(0, \sigma^2)$ and the frequency ω is unknown. Because the frequency ω is not known, the F-distribution and the test statistic as discussed in

the previous section are not directly applicable. The periodogram analysis, however, can still be used. In fact, the original purpose of the periodogram was to search for "hidden periodicities." If the time series indeed contains a single periodic component at frequency ω, it is hoped that the periodogram $I(\omega_k)$ at the Fourier frequency ω_k closest to ω will be the maximum. Thus, we can search out the maximum periodogram ordinate and test whether this ordinate can be reasonably considered as the maximum in a random sample of $[n/2]$ i.i.d. random variables, each being a multiple of a chi-square distribution with two degrees of freedom. In this case, a natural test statistic will be

$$I^{(1)}(\omega_{(1)}) = \max\{I(\omega_k)\}, \tag{13.1.12}$$

where we use $\omega_{(1)}$ to indicate the Fourier frequency with the maximum periodogram ordinate.

Now, under the null hypothesis H_0, the $\{Z_t\}$ are Gaussian white noise $N(0, \sigma^2)$. Hence, the periodogram ordinates $I(\omega_k)/\sigma^2$, $k = 1, 2, \ldots, [n/2]$, are i.i.d. chi-square random variables with two degrees of freedom, which have the probability density function

$$p(x) = \frac{1}{2}e^{-(x/2)}, \qquad 0 \le x < \infty. \tag{13.1.13}$$

Thus, for any $g \ge 0$, we have

$$\begin{aligned} P\left[\frac{I^{(1)}(\omega_{(1)})}{\sigma^2} > g\right] &= 1 - P\left[\frac{I^{(1)}(\omega_{(1)})}{\sigma^2} \le g\right] \\ &= 1 - P\left\{\frac{I(\omega_k)}{\sigma^2} \le g, \qquad k = 1, 2, \ldots, \left[\frac{n}{2}\right]\right\} \\ &= 1 - \left\{\int_0^g \frac{1}{2}e^{-(x/2)}dx\right\}^{[n/2]} \\ &= 1 - (1 - e^{-(g/2)})^{[n/2]}. \end{aligned} \tag{13.1.14}$$

If σ^2 were known, then we could use (13.1.14) to derive an exact test for the maximum ordinate.

In practice, however, σ^2 is often unknown and has to be estimated. To derive an unbiased estimator of σ^2, we note that under the null hypothesis

$$E\left[\sum_{k=1}^{[n/2]} I(\omega_k)\right] = \left[\frac{n}{2}\right]2\sigma^2.$$

Thus,

$$\hat{\sigma}^2 = \frac{1}{2[n/2]}\sum_{k=1}^{[n/2]} I(\omega_k) \tag{13.1.15}$$

is an unbiased estimator of σ^2, and by replacing σ^2 by $\hat{\sigma}^2$ it leads to the test statistics

$$V = \frac{I^{(1)}(\omega_{(1)})}{2[n/2]\sum_{k=1}^{[n/2]} I(\omega_k)}. \tag{13.1.16}$$

Using that $I(\omega_k)$, $k = 1, 2, \ldots, [n/2]$ are independent, we have

$$\text{Var}(\hat{\sigma}^2) \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty. \tag{13.1.17}$$

It follows that $\hat{\sigma}^2$ is a consistent estimator of σ^2. Thus, for large samples, the distribution of V can be approximated by the same distribution as $I^{(1)}(\omega_{(1)})/\sigma^2$, i.e., for any $g \geq 0$,

$$P(V > g) \simeq 1 - \left(1 - e^{-(g/2)}\right)^{[n/2]}. \tag{13.1.18}$$

An exact test for $\max\{I(\omega_k)\}$ was derived by Fisher (1929), based on the following statistic:

$$T = \frac{I^{(1)}(\omega_{(1)})}{\sum_{k=1}^{[n/2]} I(\omega_k)}. \tag{13.1.19}$$

Under the null hypothesis of the Gaussian white noise process $N(0, \sigma^2)$ for Z_t, Fisher (1929) showed that

$$P(T > g) = \sum_{j=1}^{m} (-1)^{(j-1)} \binom{N}{j} (1 - jg)^{N-1}, \tag{13.1.20}$$

where $N = [n/2]$, $g > 0$, and m is the largest integer less than $1/g$. Thus, for any given significance level α, we can use Equation (13.1.20) to find the critical value g_α such that

$$P(T > g_\alpha) = \alpha.$$

If the T value calculated from the series is larger than g_α, then we reject the null hypothesis and conclude that the series Z_t contains a periodic component. This test procedure is known as Fisher's test.

The critical values of T for the significance level $\alpha = .05$ as given by Fisher (1929) is shown in Table 13.2.

The third column in Table 13.2 is an approximation obtained by using only the first term in (13.1.20), i.e.,

$$P(T > g) \simeq N(1 - g)^{N-1}. \tag{13.1.21}$$

TABLE 13.2 · The critical values for $\alpha = .05$ for the ratio of the largest periodogram ordinate to the sum.

N^*	g_α (by exact formula)	g_α (by first term only)
5	.68377	.68377
10	.44495	.44495
15	.33462	.33463
20	.27040	.27046
25	.22805	.22813
30	.19784	.19794
35	.17513	.17525
40	.15738	.15752
45	.14310	.14324
50	.13135	.13149

*$N = (n - 1)/2$ if n is odd and $N = (n/2 - 1)$ if n is even.

The approximation is very close to the exact result. Hence, for most practical purposes, we can use Equation (13.1.21) to derive the critical value g_α for the test.

Under the null hypothesis H_0 in (13.1.11), a significant value of $I^{(1)}(\omega_{(1)})$ leads to the rejection of H_0, which implies that there exists a periodic component in the series at some frequency ω. This ω is not necessarily equal to $\omega_{(1)}$, because $\omega_{(1)}$ is chosen only from the Fourier frequencies and not from all the possible frequencies between 0 and π. Hartley (1949), however, has shown that the unknown ω with the maximum periodic component can be safely estimated by $\omega_{(1)}$ and that the probability that $|\omega - \omega_{(1)}| > 2\pi/n$ is less than the significance level of the test.

Let $I^{(2)}(\omega_{(2)})$ be the second largest periodogram ordinate at Fourier frequency $\omega_{(2)}$. Whittle (1952) suggested extending Fisher's test for this second largest ordinate based on the test statistic

$$T_2 = \frac{I^{(2)}(\omega_{(2)})}{\left\{\sum_{k=1}^{[n/2]} I(\omega_k)\right\} - I^{(1)}(\omega_{(1)})}, \tag{13.1.22}$$

where the distribution in (13.1.20) is taken as the distribution of T_2 with N being replaced by $(N - 1)$. The procedure can be continued until an insignificant result is obtained. It leads to an estimate of m, the number of periodic components present in the series.

Alternatively, we can use standard regression analysis and construct the usual regression test. For example, to examine whether a monthly time series is seasonal with period 12, we might postulate the model

$$Z_t = a_0 + \sum_{k=1}^{6} [a_k \cos(2\pi k/12) + b_k \sin(2\pi k/12)] + e_t, \tag{13.1.23}$$

where the e_t are i.i.d. $N(0, \sigma^2)$. The test for seasonality is then performed by the F ratio of the mean square for the six seasonal frequencies $2\pi k/12$, for $k = 1, 2, 3, 4, 5, 6$, to the residual mean square.

Before illustrating the procedure with an empirical series, we should note that we could fit a general sine-cosine model for Z_t

$$Z_t = \mu + \sum_{j=1}^{m} (\alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t) + e_t, \tag{13.1.24}$$

where the e_t are i.i.d. $N(0, \sigma^2)$. Unless the λ_j are Fourier frequencies, however, the OLS estimators at different frequencies λ_i and λ_j are not independent because the sine-cosine and complex exponential systems are complete and orthogonal only over the Fourier frequencies.

EXAMPLE 13.1 In this example, we perform the periodogram analysis of the logarithms of the annual lynx pelt sales of the Hudson's Bay Company between 1857 and 1911 ($n = 55$), which we fitted an ARMA model in Chapters 6 and 7. The results are given in Table 13.3 and plotted in Figure 13.1 in terms of both frequency and period. The periodogram is clearly domi-

TABLE 13.3 Periodogram analysis of the lynx pelt sales.

k	Freq. (ω_k)	Period (P)	$I(\omega_k)$	F
1	.11424	55.0000	.8853	.5598
2	.22848	27.5000	.0489	.0303
3	.34272	18.3333	2.0284	1.4327
4	.45696	13.7500	.9876	.6260
5	.57120	11.0000	.9593	.6077
6	.68544	9.1667	31.5768	78.7436
7	.79968	7.8571	1.7398	1.1235
8	.91392	6.8750	.6567	.4130
9	1.02816	6.1111	.4326	.2706
10	1.14240	5.5000	.3162	.1972
11	1.25664	5.0000	.3212	.2004
12	1.37088	4.5833	.4072	.2545
13	1.48512	4.2308	.2437	.1517
14	1.59936	3.9286	.0121	.0075
15	1.71360	3.6667	.2094	.1303
16	1.82784	3.4375	.2823	.1759
17	1.94208	3.2353	.0929	.0576
18	2.05632	3.0556	.0483	.0299
19	2.17055	2.8947	.1396	.0867
20	2.28479	2.7500	.0596	.0369
21	2.39903	2.6190	.0370	.0229
22	2.51327	2.5000	.0071	.0044
23	2.62751	2.3913	.3390	.2115
24	2.74175	2.2917	.0640	.0397
25	2.85599	2.2000	.0222	.0137
26	2.97023	2.1154	.0140	.0087
27	3.08447	2.0370	.0720	.0446

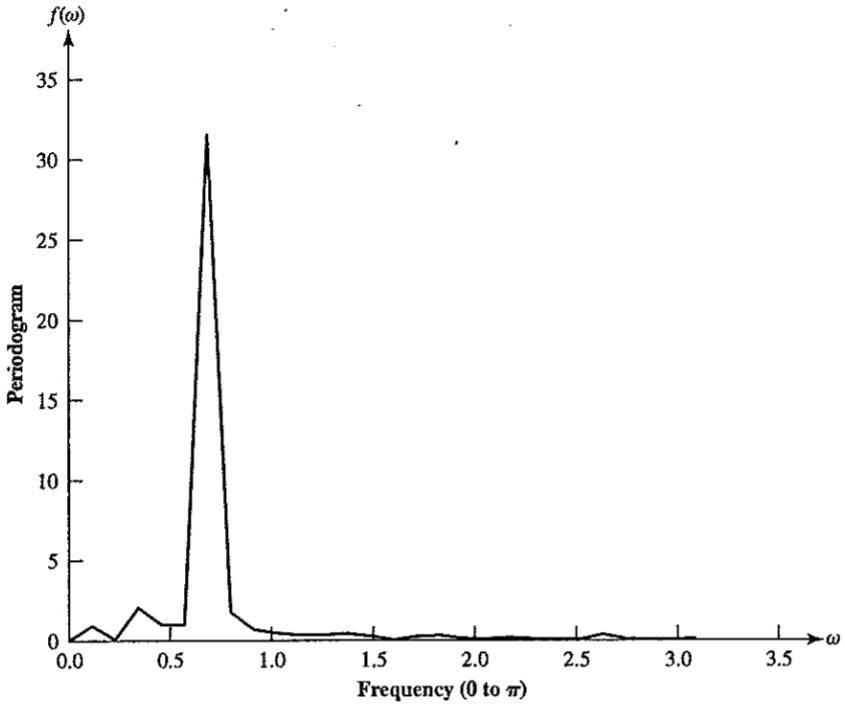


FIGURE 13.1 Periodogram of the logarithms of the annual Canadian lynx pelt sales.

nated by a very large peak at frequency $\omega_6 = 2\pi(6)/55 = .68544$. This frequency corresponds to a period of $P = 2\pi/\omega = 2\pi/\omega_6 = 9.1667$ years. It indicates that the data exhibit an approximate nine-year cycle. Also included in Table 13.3 are the values of the F statistics given in (13.1.8) to test the significance of the periodogram at each Fourier frequency. For significance level $\alpha = .05$, $F_{.05}(2, 52) = 3.19533$, and the periodogram is significant only at frequency $\omega_6 = .68544$.

For Fisher's exact test for the maximum periodogram, we have

$$T = \frac{I^{(1)}(\omega_{(1)})}{\sum_{k=1}^{[n/2]} I(\omega_k)} = \frac{I^{(1)}(\omega_6)}{\sum_{k=1}^{27} I(\omega_k)}$$
$$= \frac{31.5768}{42.003} = .7518.$$

From Table 13.2, $g_{.05} = .22805$ for $N = 25$, and $g_{.05} = .19784$ for $N = 30$. More precisely for $N = 27$, we can use the first term approximation given in (13.1.21). For $\alpha = .05$, we have

$$N(1 - g)^{N-1} = 27(1 - g)^{26} = .05,$$

which gives $g = .21493$. Because $T = .7518 >> .21493$, the result is highly significant and we conclude that the series contains a periodic component at the frequency $\omega_6 = .68544$.

13.2 The Sample Spectrum

In this section, we investigate the estimation of the spectrum of time series with absolutely summable autocovariances. From (12.1.3), the spectrum is given by

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k} \tag{13.2.1a}$$

$$= \frac{1}{2\pi}\left(\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos \omega k\right), \qquad -\pi \leq \omega \leq \pi. \tag{13.2.1b}$$

Based on sample data, it is natural to estimate $f(\omega)$ by replacing the theoretical autocovariances γ_k by the sample autocovariances $\hat{\gamma}_k$. For a given time series of n observations, however, we can only calculate $\hat{\gamma}_k$ for $k = 0, 1, \ldots, (n-1)$. Hence, we estimate $f(\omega)$ by

$$\hat{f}(\omega) = \frac{1}{2\pi}\sum_{k=-(n-1)}^{(n-1)}\hat{\gamma}_k e^{-i\omega k} \tag{13.2.2a}$$

$$= \frac{1}{2\pi}\left(\hat{\gamma}_0 + 2\sum_{k=1}^{n-1}\hat{\gamma}_k \cos \omega k\right) \tag{13.2.2b}$$

and call it the sample spectrum. Because $\hat{\gamma}_k$ is asymptotically unbiased, as discussed in Section 2.5.2, we have

$$\lim_{n\to\infty} E(\hat{f}(\omega)) = f(\omega). \tag{13.2.3}$$

Thus, $\hat{f}(\omega)$ is asymptotically unbiased and looks very promising as a potential estimator for $f(\omega)$.

To further examine the properties of the sample spectrum, let us consider $\hat{f}(\omega_k)$ at the Fourier frequencies $\omega_k = 2\pi k/n$, $k = 1, \ldots, [n/2]$. At these Fourier frequencies, the sample spectrum and the periodogram are closely related. To see, we note that

$$\begin{aligned} I(\omega_k) &= \frac{n}{2}(a_k^2 + b_k^2) \\ &= \frac{n}{2}(a_k - ib_k)(a_k + ib_k) \\ &= \frac{n}{2}\left[\frac{2}{n}\sum_{t=1}^{n} Z_t(\cos \omega_k t - i \sin \omega_k t)\right]\left[\frac{2}{n}\sum_{t=1}^{n} Z_t(\cos \omega_k t + i \sin \omega_k t)\right] \\ &= \frac{2}{n}\left[\sum_{t=1}^{n} Z_t e^{-i\omega_k t}\right]\left[\sum_{t=1}^{n} Z_t e^{i\omega_k t}\right] \end{aligned}$$

$$= \frac{2}{n}\left[\sum_{t=1}^{n}(Z_t - \bar{Z})e^{-i\omega_k t}\right]\left[\sum_{t=1}^{n}(Z_t - \bar{Z})e^{i\omega_k t}\right]$$

$$= \frac{2}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}(Z_t - \bar{Z})(Z_s - \bar{Z})e^{-i\omega_k(t-s)}, \tag{13.2.4}$$

where we use that $\sum_{t=1}^{n} e^{i\omega_k t} = \sum_{t=1}^{n} e^{-i\omega_k t} = 0$.

Because

$$\hat{\gamma}_j = \frac{1}{n}\sum_{t=1}^{n-j}(Z_t - \bar{Z})(Z_{t+j} - \bar{Z})$$

and letting $j = t - s$ in (13.2.4), we get

$$I(\omega_k) = 2\sum_{j=-(n-1)}^{n-1}\hat{\gamma}_j e^{-i\omega_k j} \tag{13.2.5a}$$

$$= 2\left\{\hat{\gamma}_0 + 2\sum_{j=1}^{n-1}\hat{\gamma}_j \cos\omega_k j\right\}. \tag{13.2.5b}$$

Hence, from (13.2.2b), we have

$$\hat{f}(\omega_k) = \frac{1}{4\pi}I(\omega_k), \qquad k = 1, 2, \ldots, [n/2], \tag{13.2.6}$$

where we note that $\hat{f}(\omega_{n/2}) = I(\omega_{n/2})/2\pi = na_{n/2}^2/2\pi$ if n is even.

It follows that if Z_t is a Gaussian white noise series with mean 0 and constant variance σ^2, then $\hat{f}(\omega_k)$, for $k = 1, 2, \ldots, (n-1)/2$, are distributed independently and identically as $(\sigma^2/4\pi)\chi^2(2) = (\sigma^2/2\pi)\chi^2(2)/2$, and we denote it as

$$\hat{f}(\omega_k) \sim \frac{\sigma^2}{2\pi}\frac{\chi^2(2)}{2}, \tag{13.2.7}$$

where $\chi^2(2)$ is the chi-square distribution with two degrees of freedom. Here, we note that $\sigma^2/2\pi$ in (13.2.7) is the spectrum of Z_t. In general, following the same arguments, it can be shown that if Z_t is a Gaussian process with the spectrum $f(\omega)$, then

$$\hat{f}(\omega_k) \sim f(\omega_k)\frac{\chi^2(2)}{2}. \tag{13.2.8}$$

Now,

$$E(\hat{f}(\omega_k)) = E\left[f(\omega_k)\frac{\chi^2(2)}{2}\right] = f(\omega_k) \tag{13.2.9}$$

and

$$\mathrm{Var}(\hat{f}(\omega_k)) = \mathrm{Var}\left[f(\omega_k)\frac{\chi^2(2)}{2} \right] = [f(\omega_k)]^2, \tag{13.2.10}$$

which does not depend on the sample size n. Thus, although the sample spectrum calculated at the Fourier frequencies is unbiased, it is an unsatisfactory estimator because it is not consistent as the variance of $\hat{f}(\omega_k)$ does not tend to zero as the sample size n tends to infinity. Furthermore, by (13.2.6) and that the periodogram ordinates are independent, as shown in (13.1.6), we have

$$\mathrm{Cov}[\hat{f}(\omega_k), \hat{f}(\omega_j)] = 0 \tag{13.2.11}$$

for any two different Fourier frequencies ω_k and ω_j. More generally, even for other distinct non-Fourier frequencies ω and λ, as the sample size n increases and the mesh of the Fourier

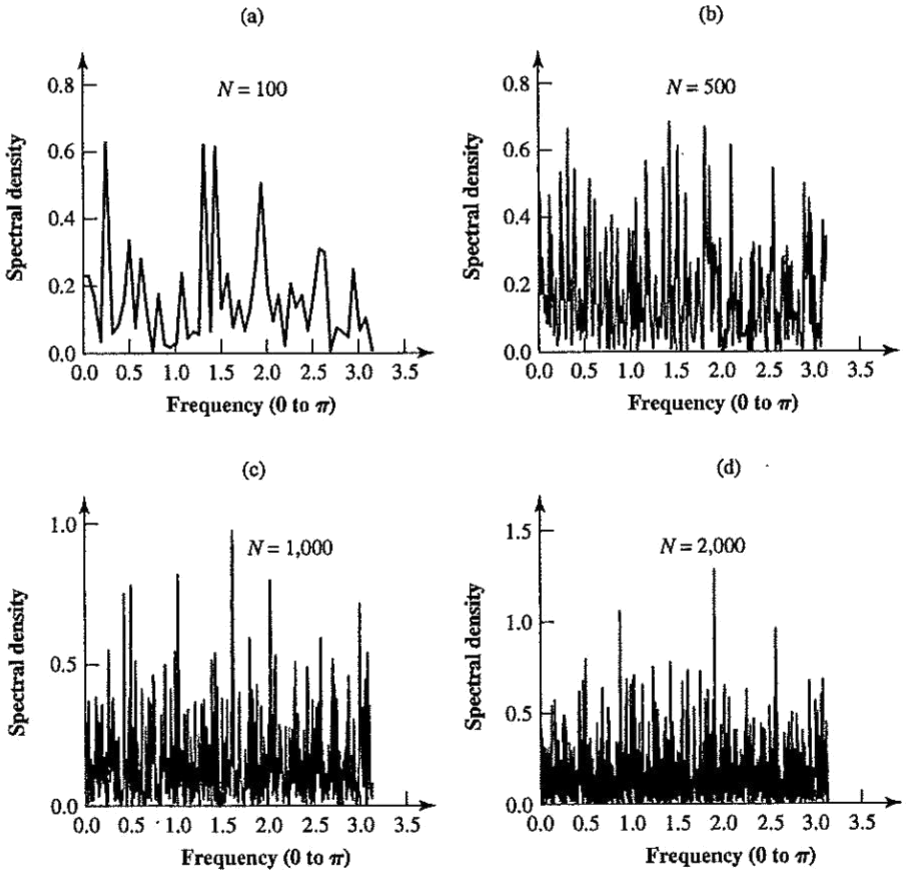


FIGURE 13.2 Sample spectrum of a white noise process.

frequencies becomes finer and finer, the covariance between $\hat{f}(\omega)$ and $\hat{f}(\lambda)$ still tends to zero as $n \to \infty$. As a result, $\hat{f}(\omega)$ is unstable and tends to be very jagged, as illustrated in Figure 13.2. Note that the pattern persists regardless of the size of the sample. The effort of correcting these unsatisfactory properties leads to the smoothing of the periodogram and the sample spectrum.

13.3 The Smoothed Spectrum

13.3.1 Smoothing in the Frequency Domain: The Spectral Window

A natural way to reduce the variance of the sample spectrum is to smooth the sample spectrum locally in the neighborhood of the target frequency. In other words, the spectral estimator is the smoothed spectrum obtained from the following weighted average of m values to the right and left of a target frequency ω_k, i.e.,

$$\hat{f}_W(\omega_k) = \sum_{j=-m_n}^{m_n} W_n(\omega_j)\hat{f}(\omega_k - \omega_j), \tag{13.3.1}$$

where the $\omega_k = 2\pi k/n$, n being the sample size, $k = 0, \pm 1, \ldots, \pm[n/2]$, are Fourier frequencies; m_n indicates that m is a function of n, which is often chosen such that $m_n \to \infty$ but $(m_n/n) \to 0$ as $n \to \infty$; and $W_n(\omega_j)$ is the weighting function having properties that

$$\sum_{j=-m_n}^{m_n} W_n(\omega_j) = 1, \tag{13.3.2}$$

$$W_n(\omega_j) = W_n(-\omega_j), \tag{13.3.3}$$

and

$$\lim_{n\to\infty} \sum_{j=-m_n}^{m_n} W_n^2(\omega_j) = 0. \tag{13.3.4}$$

The weighting function $W_n(\omega_j)$ is called the spectral window because only some of the spectral ordinates are utilized and shown in the smoothing summation. If $f(\omega)$ is flat and approximately constant within the window, then we have

$$\begin{aligned} E[\hat{f}_W(\omega_k)] &= \sum_{j=-m_n}^{m_n} W_n(\omega_j)E[\hat{f}(\omega_k - \omega_j)] \\ &\simeq f(\omega_k)\sum_{j=-m_n}^{m_n} W_n(\omega_j) = f(\omega_k). \end{aligned} \tag{13.3.5}$$

Also, by (13.2.10),

$$\begin{aligned}\operatorname{Var}[\hat{f}_W(\omega_k)] &\simeq \sum_{j=-m_n}^{m_n} \mathcal{W}_n^2(\omega_j)[f(\omega_k)]^2 \\ &\simeq [f(\omega_k)]^2 \sum_{j=-m_n}^{m_n} \mathcal{W}_n^2(\omega_j), \end{aligned} \tag{13.3.6}$$

where we also use that the sample spectrum is independent at Fourier frequencies.

The property (13.3.4) of the spectral window implies that the variance of the smoothed spectrum decreases as m_n increases. The value of m_n represents the number of frequencies used in the smoothing. This value is directly related to the width of the spectral window, also known as the bandwidth of the window. As the bandwidth increases, more spectral ordinates are averaged; hence, the resulting estimator becomes smoother, more stable, and has smaller variance. Unless $f(\omega)$ is really flat, however, the bias will also increase as the bandwidth increases because more and more spectral ordinates are used in the smoothing. We are thus forced to compromise between variance reduction and bias, a common dilemma with statistical estimators.

More generally, from (13.2.2a), the sample spectrum is defined at any frequency ω between $-\pi$ to π, and not only at the Fourier frequencies. Thus, we can write the general smoothed spectrum in terms of the following integral form:

$$\hat{f}_{\mathcal{W}}(\omega) = \int_{-\pi}^{\pi} \mathcal{W}_n(\lambda)\hat{f}(\omega - \lambda)\, d\lambda, \tag{13.3.7}$$

$$= \int_{-\pi}^{\pi} \mathcal{W}_n(\omega - \lambda)\hat{f}(\lambda)\, d\lambda, \tag{13.3.8}$$

where $\mathcal{W}_n(\lambda)$ is the spectral window satisfying the conditions

$$\int_{-\pi}^{\pi} \mathcal{W}_n(\lambda)\, d\lambda = 1, \tag{13.3.9}$$

$$\mathcal{W}_n(\lambda) = \mathcal{W}_n(-\lambda), \tag{13.3.10}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \int_{-\pi}^{\pi} \mathcal{W}_n^2(\lambda)\, d\lambda = 0. \tag{13.3.11}$$

The spectral window is also known as the kernel in the literature.

If the spectrum is approximately constant within the bandwidth of the spectral window, the condition (13.3.9) implies that

$$E(\hat{f}_w(\omega)) \cong f(\omega). \tag{13.3.12}$$

For the variance, we note that (13.3.8) can be approximated by the sum

$$\hat{f}_w(\omega) \simeq \frac{2\pi}{n} \sum_{k=-[n/2]}^{[n/2]} W_n(\omega - \omega_k)\hat{f}(\omega_k), \tag{13.3.13}$$

where $\omega_k = 2\pi k/n$. Thus,

$$\begin{aligned} \operatorname{Var}(\hat{f}_w(\omega)) &\simeq \left(\frac{2\pi}{n}\right)^2 f^2(\omega) \sum_{k=-[n/2]}^{[n/2]} W_n^2(\omega - \omega_k) \\ &\simeq \frac{2\pi}{n} f^2(\omega) \sum_{k=-[n/2]}^{[n/2]} W_n^2(\omega - \omega_k)\frac{2\pi}{n} \\ &\simeq \frac{2\pi}{n} f^2(\omega) \int_{-\pi}^{\pi} W_n^2(\lambda)\, d\lambda. \end{aligned} \tag{13.3.14}$$

By condition (13.3.11), $\operatorname{Var}(\hat{f}_w(\omega)) \to 0$ as $n \to \infty$; hence, $\hat{f}_w(\omega)$ is a consistent estimator of $f(\omega)$.

Clearly, spectral ordinates at two different frequencies ω and λ are in general correlated because they may contain some common terms in the smoothing. Because the sample spectral ordinates at different Fourier frequencies are independent, however, the covariance between smoothed spectral ordinates is easily seen to be proportional to the amount of overlap between the spectral windows centered at ω and λ. The covariance will be large if the spectral windows overlap considerably; the covariance will be small if the spectral windows overlap only slightly.

A few other remarks are in order. In the actual computation of the smoothed spectrum, the discrete form of (13.3.1) is used at the Fourier frequencies. Thus, the smoothing is essentially applied to the periodogram. Because the periodogram is periodic with period 2π, when the window covers frequencies that fail to lie entirely in the range between $-\pi$ and π, we can extend the periodogram and hence the sample spectrum using the periodic property. Equivalently, we can fold the weights back into the interval $-\pi$ to π. As the periodogram is also symmetric about frequency zero, calculation is only necessary for the frequency range between zero and π. Also, because the periodogram at frequency zero reflects the sample mean of Z_t and not the spectrum, it is not included in the smoothing, and the value at ω_1 is used in its place.

13.3.2 Smoothing in the Time Domain: The Lag Window

Note that the spectrum $f(\omega)$ is the Fourier transform of the autocovariance function γ_k. Hence, from (13.2.2a), an alternative to spectrum smoothing is to apply a weighting function $W(k)$ to the sample autocovariances, i.e.,

$$\hat{f}_W(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} W(k)\hat{\gamma}_k e^{-i\omega k}. \tag{13.3.15}$$

Because the sample autocovariance function $\hat{\gamma}_k$ is symmetric and because $\hat{\gamma}_k$ is less reliable for larger k, the weighting function $W(k)$ should also be chosen to be symmetric, with its weights inversely proportional to the magnitude of k. Thus, we have

$$\hat{f}_W(\omega) = \frac{1}{2\pi} \sum_{k=-M}^{M} W_n(k)\hat{\gamma}_k e^{-i\omega k}, \tag{13.3.16}$$

where the weighting function $W_n(k)$ is chosen to be an absolutely summable sequence

$$W_n(k) = W\left(\frac{k}{M}\right), \tag{13.3.17}$$

which is often derived from a bounded even continuous function $W(x)$ satisfying

$$\begin{aligned} |W(x)| &\le 1, \\ W(0) &= 1, \\ W(x) &= W(-x), \\ W(x) &= 0, \qquad |x| > 1. \end{aligned} \tag{13.3.18}$$

The value of M is the truncation point that depends on the sample size n. This weighting function $W_n(k)$ for the autocovariances is called the lag window.

The lag window and the spectral window are closely related because the autocovariance is the inverse Fourier transform of the spectrum. From (13.2.2a), the inverse Fourier transform of $\hat{f}(\lambda)$ is

$$\hat{\gamma}_k = \int_{-\pi}^{\pi} \hat{f}(\lambda) e^{i\lambda k}\, d\lambda \qquad \text{for } k = 0, \pm 1, \ldots, \pm(n-1). \tag{13.3.19}$$

Thus,

$$\begin{aligned} \hat{f}_W(\omega) &= \frac{1}{2\pi} \sum_{k=-M}^{M} W_n(k)\hat{\gamma}_k e^{-i\omega k} \\ &= \frac{1}{2\pi} \sum_{k=-M}^{M} W_n(k) \int_{-\pi}^{\pi} \hat{f}(\lambda) e^{i\lambda k} e^{-i\omega k}\, d\lambda \end{aligned}$$

$$= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-M}^{M} W_n(k) e^{-i(\omega-\lambda)k} \hat{f}(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} \mathcal{W}_n(\omega - \lambda) \hat{f}(\lambda) d\lambda$$

$$= \int_{-\pi}^{\pi} \mathcal{W}_n(\lambda) \hat{f}(\omega - \lambda) d\lambda, \tag{13.3.20}$$

where

$$\mathcal{W}_n(\omega) = \frac{1}{2\pi} \sum_{k=-M}^{M} W_n(k) e^{-i\omega k} \tag{13.3.21}$$

is the spectral window. It is clear from (13.3.21) that the spectral window is the Fourier transform of the lag window and that the lag window is the inverse Fourier transform of the spectral window, i.e.,

$$W_n(k) = \int_{-\pi}^{\pi} \mathcal{W}_n(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \ldots, \pm M. \tag{13.3.22}$$

Hence, the lag window and the spectral window form a Fourier transform pair, with one being uniquely determined by the other. Both the terms *lag window* and *spectral window* were introduced by Blackman and Tukey (1959). The *weighting function* was the standard term used in the earlier literature.

The variance expression given in (13.3.14) can also be written in terms of a lag window. By Parseval's relation from (11.5.16),

$$\sum_{k=-M}^{M} W_n^2(k) = 2\pi \int_{-\pi}^{\pi} \mathcal{W}_n^2(\omega) d\omega; \tag{13.3.23}$$

hence,

$$\text{Var}[\hat{f}_w(\omega)] \simeq \frac{1}{n} f^2(\omega) \sum_{k=-M}^{M} W_n^2(k). \tag{13.3.24}$$

Because the lag window is often derived from a bounded even continuous function as given in (13.3.18), however, we have

$$\sum_{k=-M}^{M} W_n^2(k) = \sum_{k=-M}^{M} W^2\left(\frac{k}{M}\right)$$

$$= M \sum_{k=-M}^{M} W^2\left(\frac{k}{M}\right) \cdot \frac{1}{M}$$

$$\simeq M \int_{-1}^{1} W^2(x) dx. \tag{13.3.25}$$

Thus,

$$\text{Var}[\hat{f}_w(\omega)] \simeq \frac{M}{n} f^2(\omega) \int_{-1}^{1} W^2(x)\,dx. \tag{13.3.26}$$

This result implies that the variance of the smoothed spectrum for a given lag window is related to the ratio M/n, which is the proportion of sample autocovariances for which the lag weights are nonzero.

13.3.3 Some Commonly Used Windows

With the above general background, we now introduce some lag and spectral windows commonly used in time series analysis.

The Rectangular Window The so-called rectangular or truncated lag window

$$W_n^R(k) = \begin{cases} 1, & |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.27}$$

where M is the truncation point less than $(n-1)$, is derived from the continuous function

$$W(x) = \begin{cases} 1, & |x| \le 1, \\ 0, & |x| > 1. \end{cases} \tag{13.3.28}$$

From (13.3.21), the corresponding spectral window is given by

$$\begin{aligned} \mathcal{W}_n^R(\omega) &= \frac{1}{2\pi} \sum_{k=-M}^{M} W_n^R(k) e^{-i\omega k} \\ &= \frac{1}{2\pi} \sum_{k=-M}^{M} e^{-i\omega k} \\ &= \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{M} \cos \omega k \right) \\ &= \frac{1}{2\pi} \left[1 + \frac{2 \cos(\omega(M+1)/2) \sin(\omega M/2)}{\sin(\omega/2)} \right], \qquad \text{(using (11.2.8))}, \\ &= \frac{1}{2\pi} \frac{\sin(\omega/2) + [\sin(\omega(M+1/2)) - \sin(\omega/2)]}{\sin(\omega/2)}, \qquad \text{(using (11.2.12c))}, \\ &= \frac{1}{2\pi} \frac{\sin(\omega(M+1/2))}{\sin(\omega/2)}. \end{aligned} \tag{13.3.29}$$

The rectangular lag and spectral windows are both shown in Figure 13.3.

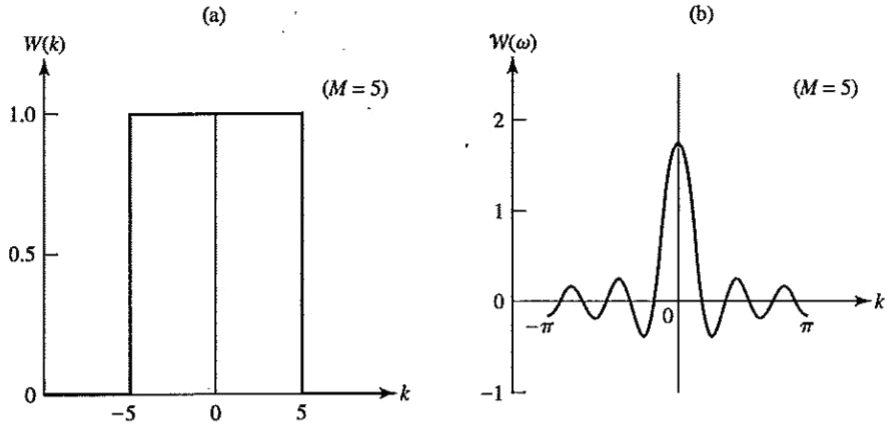


FIGURE 13.3 The rectangular lag and spectral windows. (a) Rectangular lag window. (b) Rectangular spectral window.

Note that the spectral window as shown in Figure 13.3 has a main lobe at $\omega = 0$ with a height of $(2M + 1)/2\pi$, zeros at $\omega = \pm 2j\pi/(2M + 1)$, and side peaks (lobes) of decreasing magnitude approximately at $\omega = \pm(4j + 1)\pi/(2M + 1)$ for $j = 1, 2, \ldots$. As a result, the smoothed spectrum estimator using this window may lead to negative values for some frequencies ω. Because the spectrum is known to be a nonnegative function, this result is certainly not desirable.

For a given spectral window $\mathcal{W}_n(\omega)$ that attains its maximum at $\omega = 0$, one commonly used definition for the bandwidth of the window is the distance between the half-power points on the main lobe, as shown in Figure 13.4. That is,

$$\text{Bandwidth} = 2\omega_I, \tag{13.3.30}$$

where ω_I is such that $\mathcal{W}_n(\pm\omega_I) = \frac{1}{2}\mathcal{W}_n(0)$.

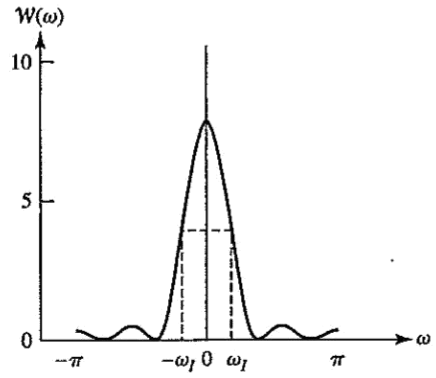


FIGURE 13.4 Bandwidth of the spectral window.

If ω_{II} is the first zero of $\mathcal{W}_n(\omega)$, the value of ω_I can be approximated by $\frac{1}{2}\omega_{II}$; hence, the bandwidth is approximately equal to ω_{II}. In terms of the above rectangular spectral window $\mathcal{W}_n^R(\omega)$, the bandwidth is approximately equal to $2\pi/(2M + 1)$. In general, the bandwidth of the spectral window is inversely related to the truncation point M used in the lag window. Thus, as M increases, the bandwidth decreases; hence, the variance of the smoothed spectrum increases although the bias decreases, as discussed in Section 13.3.1. On the other hand, as M decreases, the bandwidth increases and the variance decreases while the bias increases.

Bartlett's Window Bartlett (1950) proposed the lag window

$$W_n^B(k) = \begin{cases} 1 - |k|/M, & |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.31}$$

based on the triangular function

$$W(x) = \begin{cases} 1 - |x|, & |x| \le 1, \\ 0, & |x| > 1. \end{cases} \tag{13.3.32}$$

Hence, the window is also known as the triangular window. The corresponding spectral window is given by

$$\begin{aligned}
\mathcal{W}_n^B(\omega) &= \frac{1}{2\pi}\sum_{k=-M}^{M}\left(1 - \frac{|k|}{M}\right)e^{-i\omega k} \\
&= \frac{1}{2\pi M}\sum_{k=-M}^{M}\big(M - |k|\big)e^{-i\omega k} \\
&= \frac{1}{2\pi M}\sum_{j=0}^{M-1}\sum_{k=-j}^{j} e^{-i\omega k} \\
&= \frac{1}{2\pi M}\sum_{j=0}^{M-1}\frac{\sin(\omega(j + 1/2))}{\sin(\omega/2)} \\
&= \frac{1}{2\pi M\sin(\omega/2)}\left[\sin(\omega/2) + \sum_{j=1}^{M-1}\sin(\omega(j + 1/2))\right] \\
&= \frac{1}{2\pi M\sin(\omega/2)}\left\{\sin(\omega/2) + \sum_{j=1}^{M-1}[\sin(\omega j)\cos(\omega/2) + \cos(\omega j)\sin(\omega/2)]\right\} \\
&= \frac{1}{2\pi M\sin(\omega/2)}\left\{\sin(\omega/2) + \cos(\omega/2)\sum_{j=1}^{M-1}\sin(\omega j) + \sin(\omega/2)\sum_{j=1}^{M-1}\cos(\omega j)\right\} \\
&= \frac{1}{2\pi M\sin(\omega/2)}\left\{\sin(\omega/2) + \frac{\cos(\omega/2)\sin(\omega M/2)\sin(\omega(M-1)/2)}{\sin(\omega/2)}\right. \\
&\quad \left. + \frac{\sin(\omega/2)\cos(\omega M/2)\sin(\omega(M-1)/2)}{\sin(\omega/2)}\right\}, \qquad \text{by (11.2.9) and (11.2.8)}
\end{aligned}$$

$$= \frac{1}{2\pi M \sin(\omega/2)} \left\{ \sin(\omega/2) + \frac{\sin(\omega(M-1)/2)}{\sin(\omega/2)} [\cos(\omega/2)\sin(\omega M/2) + \sin(\omega/2)\cos(\omega M/2)] \right\}$$

$$= \frac{1}{2\pi M \sin(\omega/2)} \left\{ \sin(\omega/2) + \frac{\sin(\omega(M-1)/2)}{\sin(\omega/2)} \sin(\omega(M+1)/2) \right\}$$

$$= \frac{1}{2\pi M [\sin(\omega/2)]^2} \{ [\sin(\omega/2)]^2 + \sin(\omega(M-1)/2)\sin(\omega(M+1)/2) \}$$

$$= \frac{1}{2\pi M [\sin(\omega/2)]^2} \left\{ \frac{1}{2}(1-\cos\omega) + \frac{1}{2}(\cos\omega - \cos\omega M) \right\}, \quad \text{by (11.2.12b)}$$

$$= \frac{1}{2\pi M [\sin(\omega/2)]^2} \left\{ \frac{1}{2}(1-\cos\omega M) \right\}$$

$$= \frac{1}{2\pi M} \left[\frac{\sin(\omega M/2)}{\sin(\omega/2)} \right]^2. \tag{13.3.33}$$

Bartlett's triangular lag window and its corresponding spectral window are shown in Figure 13.5. Because the spectral window $\mathcal{W}_n^B(\omega)$ is nonnegative, the Bartlett's spectrum estimator is always nonnegative. Furthermore, a direct comparison of Equations (13.3.29) and (13.3.33) shows that the side lobes of the Bartlett window are smaller than the side lobes of the rectangular window. The effect of large side lobes is to allow $\hat{f}(\lambda)$ to make large contributions in the smoothing at frequencies distant from ω. Hence, the resulting spectrum

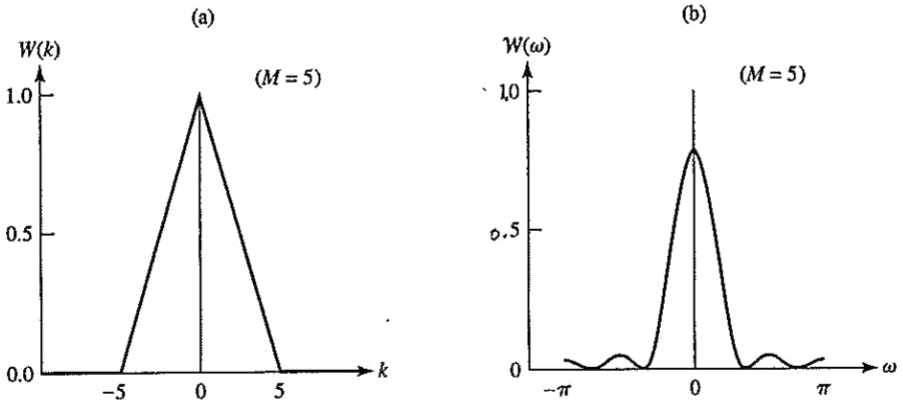


FIGURE 13.5 Bartlett's lag and spectral windows. (a) Bartlett's triangular lag window. (b) Bartlett's spectral window.

estimate $\hat{f}_W(\omega)$ may reflect significant spectral components at other frequencies different from ω. This phenomenon is known as leakage. Because lobes are produced by the Fourier transform of sharp corners, the general principle of the selection of lag windows is to avoid functions with sharp corners.

The Blackman–Tukey Window Blackman and Tukey (1959) suggested the lag window

$$W_n^T(k) = \begin{cases} 1 - 2a + 2a\cos(\pi k/M), & |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.34}$$

based on the continuous weighting function

$$W(x) = \begin{cases} 1 - 2a + 2a\cos \pi x, & |x| \le 1, \\ 0, & |x| > 1. \end{cases} \tag{13.3.35}$$

Again, M is the truncation point for the sample autocovariance function and the constant a is chosen in the range $0 < a \le .25$ so that $W_n^T(k) \ge 0$ for all k. The corresponding spectral window can be derived as follows:

$$\begin{aligned}
\mathcal{W}_n^T(\omega) &= \frac{1}{2\pi}\sum_{k=-M}^{M} W_n^T(k)e^{-i\omega k} \\
&= \frac{1}{2\pi}\sum_{k=-M}^{M}[1 - 2a + 2a\cos(\pi k/M)]e^{-i\omega k} \\
&= \frac{1}{2\pi}\sum_{k=-M}^{M}[1 - 2a + a(e^{i\pi k/M} + e^{-i\pi k/M})]e^{-i\omega k} \\
&= a\frac{1}{2\pi}\sum_{k=-M}^{M} e^{-i(\omega-\pi/M)k} + (1 - 2a)\frac{1}{2\pi}\sum_{k=-M}^{M} e^{-i\omega k} + a\frac{1}{2\pi}\sum_{k=-M}^{M} e^{-i(\omega+\pi/M)k} \\
&= \frac{a}{2\pi}\frac{\sin[(\omega - \pi/M)(M + 1/2)]}{\sin[(\omega - \pi/M)/2]} + \frac{(1 - 2a)}{2\pi}\frac{\sin[\omega(M + 1/2)]}{\sin(\omega/2)} \\
&\quad + \frac{a}{2\pi}\frac{\sin[(\omega + \pi/M)(M + 1/2)]}{\sin[(\omega + \pi/M)/2]},
\end{aligned}$$

where we have used (13.3.29). Thus, the Blackman–Tukey spectral window is a weighted linear combination of the spectral windows for the rectangular lag function at the frequencies $(\omega - \pi/M)$, ω, and $(\omega + \pi/M)$, i.e.,

$$\mathcal{W}_n^T(\omega) = a\mathcal{W}_n^R\left(\omega - \frac{\pi}{M}\right) + (1 - 2a)\mathcal{W}_n^R(\omega) + a\mathcal{W}_n^R\left(\omega + \frac{\pi}{M}\right), \tag{13.3.36}$$

where $\mathcal{W}_n^R(\omega)$ is given by (13.3.29). As a result, the Blackman–Tukey spectrum estimator may also be negative at some frequencies ω.

Blackman and Tukey (1959) called the window in (13.3.34) with $a = .23$ as *hamming* after R.W. Hamming, one of Tukey's coauthors who had worked on the subject. The resulting window is

$$W_n(k) = \begin{cases} .54 + .46\cos(\pi k/M), & |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.37}$$

which is also known as the Tukey–Hamming window in the literature. The window (13.3.34) with $a = .25$ is called as *hanning* after the Austrian meteorologist Julius von Hann, who was not directly related with the subject. The window is given by

$$W_n(k) = \begin{cases} 1/2[1 + \cos(\pi k/M)], & |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.38}$$

which is also known as the Tukey–Hanning window or the Tukey window in the literature.

The Parzen Window Parzen (1961b) suggested the lag window

$$W_n^P(k) = \begin{cases} 1 - 6(k/M)^2 + 6(|k|/M)^3, & |k| \le M/2, \\ 2(1 - |k|/M)^3, & M/2 < |k| \le M, \\ 0, & |k| > M, \end{cases} \tag{13.3.39}$$

which is based on the continuous weighting function

$$W(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \le 1/2, \\ 2(1 - |x|)^3, & 1/2 < |x| \le 1, \\ 0, & |x| > 1. \end{cases} \tag{13.3.40}$$

The corresponding spectral window for an even value of M is given by

$$\begin{aligned} \mathcal{W}_n^P(\omega) &= \frac{1}{2\pi}\sum_{k=-M}^{M} W_n^P(k)\cos\omega k \\ &= \frac{1}{2\pi}\left\{\sum_{k=-M/2}^{M/2}[1 - 6(k/M)^2 + 6(|k|/M)^3]\cos\omega k \right. \\ &\qquad \left. + 2\sum_{M/2<|k|\le M}(1 - |k|/M)^3\cos\omega k\right\} \\ &= \frac{3}{8\pi M^3}\left[\frac{\sin(\omega M/4)}{1/2\sin(\omega/2)}\right]^4\{1 - 2/3[\sin(\omega/2)]^2\}, \end{aligned} \tag{13.3.41}$$

where we refer readers to a more detailed derivation in Parzen (1963). When M is large, (13.3.41) is approximated by

$$\mathcal{W}_n^P(\omega) \simeq \frac{3}{8\pi M^3}\left[\frac{\sin(\omega M/4)}{1/2\sin(\omega/2)}\right]^4. \tag{13.3.42}$$

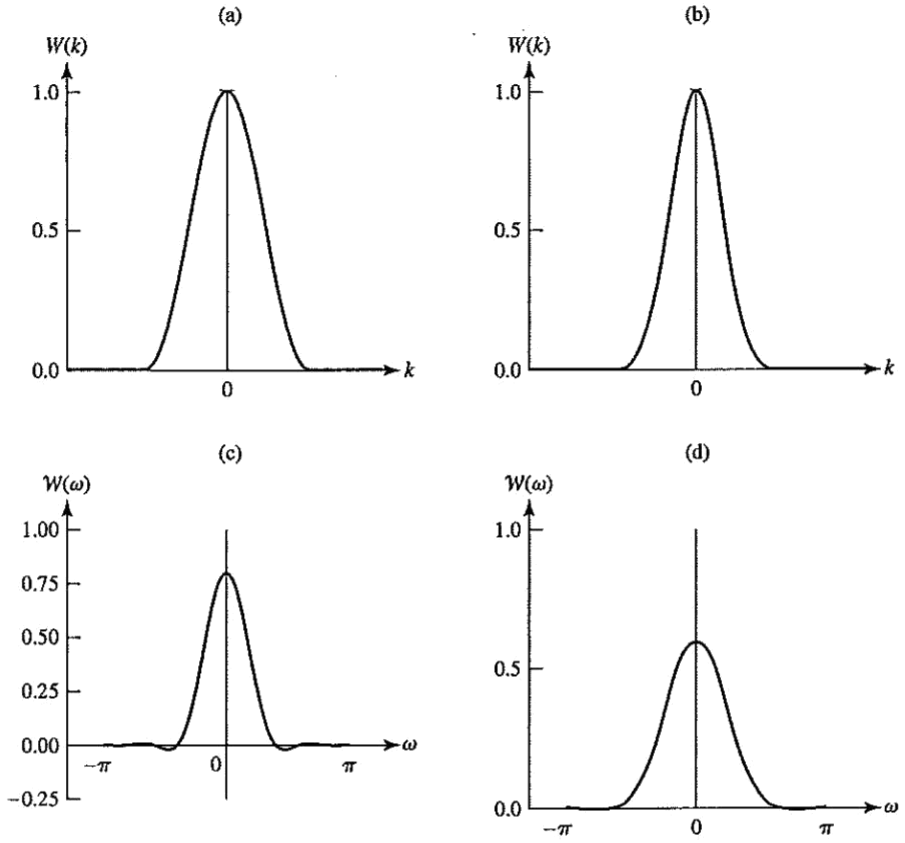


FIGURE 13.6 The Tukey–Hanning and the Parzen windows. (a) Tukey–Hanning lag window ($M = 5$). (b) Parzen lag window ($M = 5$). (c) Tukey–Hanning spectral window ($M = 5$). (d) Parzen spectral window ($M = 5$).

Both the Tukey–Hanning window and the Parzen window are plotted in Figure 13.6. As can be seen from Figure 13.6(b) and (d), for the same truncation point M, the bandwidth of the Tukey–Hanning spectral window is narrower than the bandwidth of the Parzen spectral window. Hence, the Tukey–Hanning estimator will have a smaller bias than the Parzen estimator. Because $\mathcal{W}_n^P(\omega)$ is always nonnegative, however, the Parzen window will produce a nonnegative and smoother spectrum estimator.

The above windows are some commonly used ones, particularly in terms of the commercially available software. Many other windows are introduced in the literature. Interested readers are referred to an excellent book by Priestley (1981, Chapter 6) for more details.

The quality of a smoothed spectrum is determined by the shape of the window (i.e., the form of the weighting function) and the bandwidth of the window (or, equivalently, the truncation point). The spectrum estimates produced for the same window shape and different bandwidths vary. Thus, in spectrum smoothing, we are concerned not only about the design

of spectral windows with desirable shapes (or "window carpentry" as John W. Tukey called it), but also about the bandwidth of the windows. The latter concern often is a more crucial and difficult problem in time series analysis because for a given window shape there is no single criterion for choosing the optimal bandwidth. Additionally, wider bandwidths produce smoother spectrums with smaller variance of the estimators; on the other hand, narrower bandwidths lead to smaller bias and smudging and hence to higher resolution of the resulting spectrum estimates. A compromise is needed between high stability and high resolution. To ease the difficulty, the following steps are often suggested. First, choose a spectral window with an acceptable and desirable shape. Then initially calculate the spectral estimates using a large bandwidth and then recalculate the estimates by using gradually smaller bandwidths until the required stability and resolution are achieved. This procedure is often referred to as "window closing."

Because the bandwidth of a spectral window is inversely related to the truncation point M used in the lag window, Jenkins and Watts (1968, Chapter 7) suggest choosing three values of M, say, M_1, M_2, and M_3 with $M_3 = 4M_1$. The sequences of spectral estimates are then computed and examined with these M values to see whether an intermediate value between M_1 and M_3, or a value smaller than M_1, or a value larger than M_3 should be chosen. The spectral bandwidth can also be determined by choosing a truncation point M such that $\hat{\gamma}(k)$ for $k > M$ are negligible. This choice of M has been found practical and was used successfully in a number of cases. Alternatively, a truncation point M can also be chosen as a function of n, the number of observations in a series. For a moderate size n, M can be chosen to be $M = [n/10]$.

In principle, if we wish to be able to distinguish several peaks of $f(\omega)$ at frequencies λ_1, λ_2, $\lambda_3, \ldots$, the bandwidth of the spectral window should not exceed the minimum interval between these adjacent peaks. Otherwise, the smoothed spectrum will blur these peaks. In fact, a commonly chosen bandwidth is equal to $\min_i|\lambda_{i+1} - \lambda_i|$.

EXAMPLE 13.2 Figure 13.7 shows the sample spectrum of the lynx pelt sales data over the interval $(0, \pi)$. Also shown are the smoothed sample spectrum functions using the Bartlett window with truncation points $M = 2$, 5, and 10. The spectrum for $M = 10$ clearly is improper. To compare different estimates, Figure 13.8 shows the spectrum estimates using various windows with the same truncation point $M = 5$. The smoothed spectrum given by the rectangular window is clearly undesirable. The estimates using the Bartlett, Parzen, and Tukey–Hanning windows all lead to the periodic component at the frequency $\omega = .68544$, but the rectangular window does not lead to this periodic component.

13.3.4 Approximate Confidence Intervals for Spectral Ordinates

Recall that for a given sample $Z_1, Z_2, \ldots, Z_n$ from a process with the spectrum $f(\omega)$, the sample spectral ordinates $\hat{f}(\omega_k)$ at Fourier frequencies $\omega_k = 2\pi k/n$, with $\omega_k \neq 0$ and π, are independent and identically distributed as

$$\hat{f}(\omega_k) \sim f(\omega_k)\frac{\chi^2(2)}{2}. \tag{13.3.43}$$

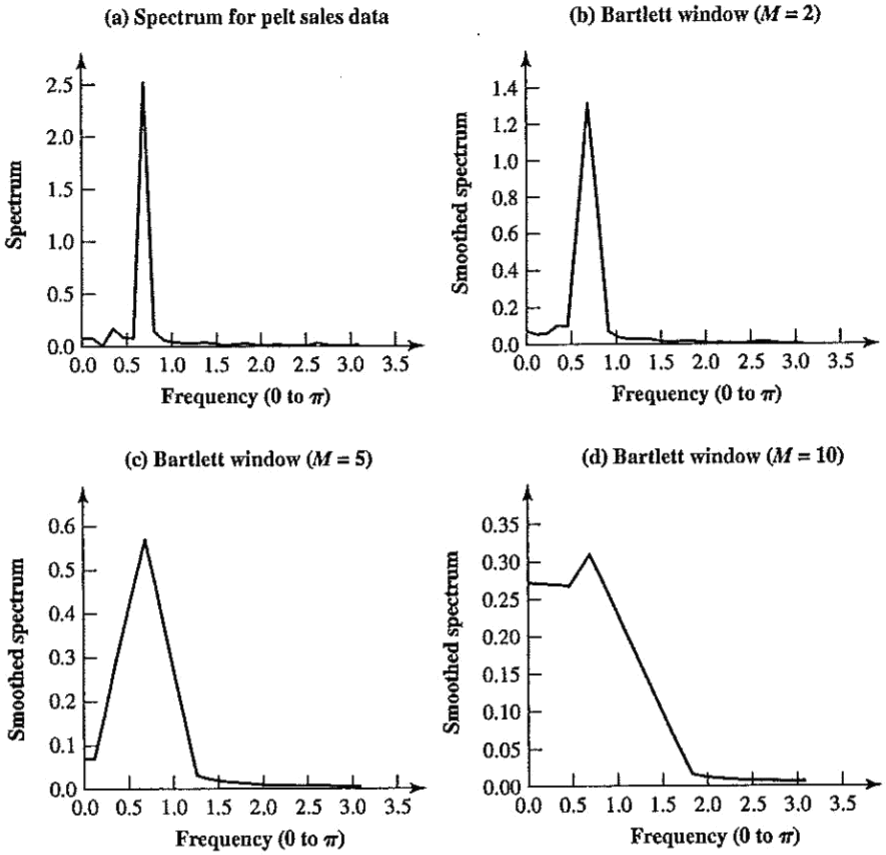


FIGURE 13.7 Sample spectrum of the logarithms of the Canadian lynx pelt sales and Bartlett estimates for different values of M.

Thus, if we smooth the sample spectrum with the simple $(2m + 1)$ term moving average, i.e.,

$$\hat{f}_W(\omega_k) = \frac{1}{2m + 1} \sum_{j=-m}^{m} \hat{f}(\omega_k - \omega_j), \tag{13.3.44}$$

the smoothed spectral ordinate $\hat{f}_W(\omega_k)$ will be distributed as

$$\hat{f}_W(\omega_k) \sim f(\omega_k) \frac{\chi^2(\mathrm{DF})}{\mathrm{DF}}, \tag{13.3.45}$$

where the degrees of freedom $\mathrm{DF} = (4m + 2)$ is simply the sum of the degrees of freedom of $(2m + 1)$ $\chi^2(2)$ random variables. In fact, this property is the first smoothing method introduced by Daniell (1946), one of the pioneers of spectrum estimation. The estimator in

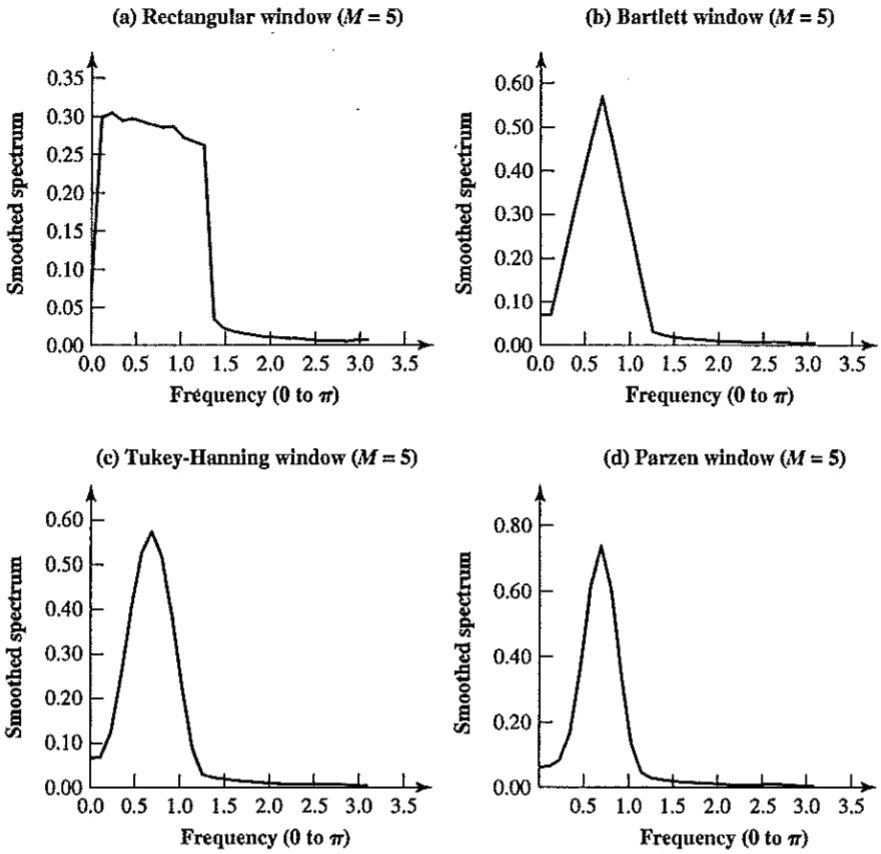


FIGURE 13.8 Smoothed sample spectrum of the logarithms of the Canadian lynx pelt sales using different windows.

(13.3.44) is also known as Daniell's estimator. However, this nice additive property no longer holds for $\hat{f}_W(\omega)$ if ω is not a Fourier frequency or if the sample spectrum is smoothed by a spectral window with nonuniform weights. Thus, we can only approximate a general smoothed spectrum $\hat{f}_W(\omega)$ by

$$\hat{f}_W(\omega) \sim c\chi^2(\nu), \tag{13.3.46}$$

where c and ν are chosen such that

$$E[\hat{f}_W(\omega)] = E(c\chi^2(\nu)) = c\nu \tag{13.3.47}$$

and

$$\text{Var}[\hat{f}_W(\omega)] = \text{Var}(c\chi^2(\nu)) = 2c^2\nu. \tag{13.3.48}$$

Now, for any given spectral window $\mathcal{W}_n(\omega)$, from (13.3.12) and (13.3.14), we have

$$E[\hat{f}_w(\omega)] \simeq f(\omega)$$

and

$$\text{Var}[\hat{f}_w(\omega)] \simeq \frac{2\pi}{n}[f(\omega)]^2 \int_{-\pi}^{\pi} \mathcal{W}_n^2(\omega)\, d\omega.$$

Hence,

$$c\nu = f(\omega)$$

and

$$2c^2\nu = \frac{2\pi}{n} f^2(\omega) \int_{-\pi}^{\pi} \mathcal{W}_n^2(\omega)\, d\omega,$$

giving

$$\nu = \frac{n}{\pi \int_{-\pi}^{\pi} \mathcal{W}_n^2(\omega)\, d\omega}, \tag{13.3.49}$$

$$c = \frac{f(\omega)}{\nu}. \tag{13.3.50}$$

It follows that

$$\hat{f}_{\mathcal{W}}(\omega) \sim f(\omega) \frac{\chi^2(\nu)}{\nu}, \tag{13.3.51}$$

where ν is often referred to as equivalent degrees of freedom for a smoothed spectrum. Using (13.3.26) or (13.3.23) and (13.3.25), we can also write the equivalent degrees of freedom in terms of the lag window

$$\nu = \frac{2n}{M \int_{-1}^{1} W^2(x)\, dx}, \tag{13.3.52}$$

where $W(x)$ is the continuous weighting function used in the associated lag window. The equivalent degrees of freedom ν for some commonly used windows are given in Table 13.4.

TABLE 13.4 Equivalent degrees of freedom for various windows.

Window	ν
Rectangular	n/M
Bartlett	$3n/M$
Tukey–Hamming	$2.516n/M$
Tukey–Hanning	$8n/3M$
Parzen	$3.709n/M$

By (13.3.51), we have

$$P\left\{f(\omega)\frac{\chi^2_{1-\alpha/2}(\nu)}{\nu} \le \hat{f}_w(\omega) \le f(\omega)\frac{\chi^2_{\alpha/2}(\nu)}{\nu}\right\} = (1-\alpha),$$

where $\chi^2_\alpha(\nu)$ is the upper α% point of the chi-square distribution with ν degrees of freedom. Hence, a $(1-\alpha)100\%$ confidence interval for $f(\omega)$ is given by

$$\frac{\nu\hat{f}_w(\omega)}{\chi^2_{\alpha/2}(\nu)} \le f(\omega) \le \frac{\nu\hat{f}_w(\omega)}{\chi^2_{1-\alpha/2}(\nu)}, \tag{13.3.53}$$

where ν is the equivalent degrees of freedom calculated by (13.3.49) or (13.3.52).

Recall that the asymptotic mean and variance of $\hat{f}_w(\omega)$ are proportional to $f(\omega)$ and $f^2(\omega)$, respectively. Hence, from the discussion in Section 4.3.2, the logarithmic transformation of the spectrum estimator $\ln\hat{f}_w(\omega)$ is often suggested. From (13.3.53), the $(1-\alpha)100\%$ confidence interval for $\ln f(\omega)$ is given by

$$\ln\hat{f}_w(\omega) + \ln\left[\frac{\nu}{\chi^2_{\alpha/2}(\nu)}\right] \le \ln f(\omega) \le \ln\hat{f}_w(\omega) + \ln\left[\frac{\nu}{\chi^2_{1-\alpha/2}(\nu)}\right]. \tag{13.3.54}$$

Note that the width of the interval for $f(\omega)$ as given in (13.3.53) is proportional to $\hat{f}_w(\omega)$ and hence varies with frequency. The width of the confidence interval for $\ln f(\omega)$ as given in (13.3.54), however, will be the same for any frequency ω.

EXAMPLE 13.3 For illustration, we calculate the 95% confidence interval for the spectrum of the Canadian lynx pelt sales discussed in earlier examples using Parzen's window with $M = 5$. For $n = 55$ and $M = 5$, we have, from Table 13.4, $\nu = 3.709(55/5) = 40.79 \simeq 41$. Because $\chi^2_{.975}(41) = 25.22$ and $\chi^2_{.025}(41) = 60.56$, the 95% confidence interval for $f(\omega)$, from (13.3.53), becomes

$$.68\,\hat{f}_w(\omega) \le f(\omega) \le 1.63\,\hat{f}_w(\omega),$$

where $\hat{f}_w(\omega)$ is the estimated spectrum using Parzen's window with $M = 5$ given in Figure 13.8. Note that the maximum of $\hat{f}_w(\omega) = .74$ occurs at $\omega = .69$. Thus, the 95% confidence interval for $f(\omega)$ at $\omega = .69$ is given by

$$.5 \leq f(\omega = .69) \leq 1.2.$$

The confidence intervals for $f(\omega)$ at other frequencies ω can be calculated similarly and are shown in Figure 13.9.

13.4 ARMA Spectral Estimation

As discussed in Chapters 2 and 3, any stationary process can be approximated by an AR(p) model,

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\dot{Z}_t = a_t \tag{13.4.1}$$

for some p. Let $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and $\hat{\sigma}_a^2$ be the estimates of $\phi_1, \ldots, \phi_p$, and σ_a^2. A reasonable alternative to spectrum estimation is to substitute these parameter estimates in the theoretical expression for the spectrum of the AR(p) model discussed in Section 12.2.2, i.e.,

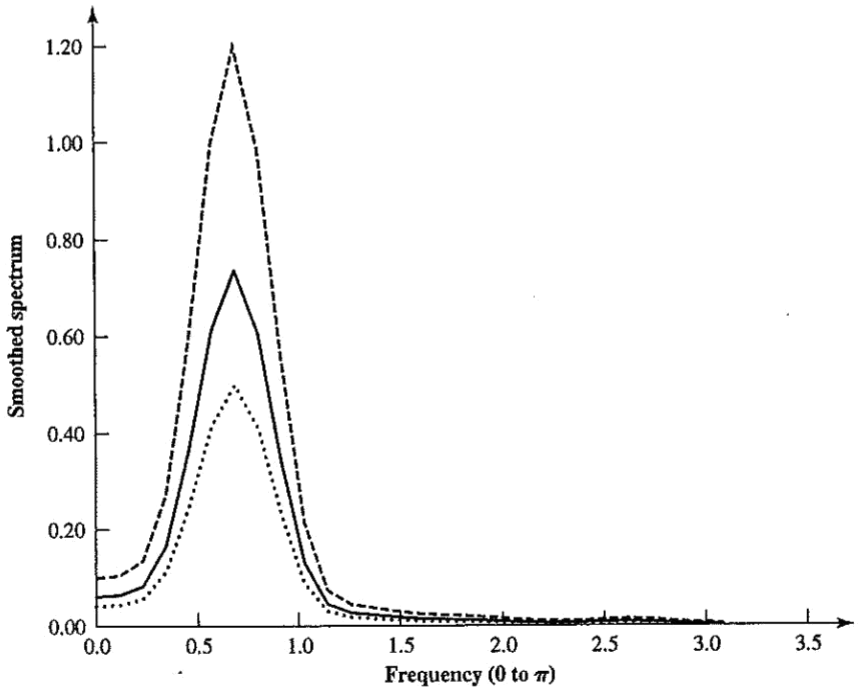


FIGURE 13.9 The 95% confidence intervals for the spectrum of the logarithms of the Canadian lynx pelt sales using Parzen's window.

$$\hat{f}_A(\omega) = \frac{\hat{\sigma}_a^2}{2\pi} \frac{1}{\hat{\phi}_p(e^{-i\omega})\hat{\phi}_p(e^{i\omega})}, \tag{13.4.2}$$

where $\hat{\phi}_p(e^{-i\omega}) = (1 - \hat{\phi}_1 e^{-i\omega} - \cdots - \hat{\phi}_p e^{-ip\omega})$. This method of spectrum estimation through an AR approximation was suggested by Akaike (1969) and Parzen (1974) and is often known as autoregressive spectral estimation.

For large n, Parzen (1974) showed that

$$\text{Var}[\hat{f}_A(\omega)] \simeq \frac{2pf^2(\omega)}{n}. \tag{13.4.3}$$

Thus, to control the variance, the order of p chosen to approximate the process should not be too large. On the other hand, p should not be too small because an inadequate order of p leads to a poor approximation and hence may increase the bias of the estimated spectrum. Consequently, similar to the choice of bandwidth, truncation point, and different windows, the determination of the order p in the autoregressive approximation becomes a crucial step in the AR spectral estimation. Parzen (1977) suggested the use of CAT criterion in choosing the optimal order p. If the AR order p is correctly determined, then asymptotic simultaneous confidence bands for autoregressive spectra can be constructed using the method by Newton and Pagano (1984).

More generally, we can approximate the unknown process by an ARMA(p, q) model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(Z_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \tag{13.4.4}$$

for some p and q. Let $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q$, and $\hat{\sigma}_a^2$ be the consistent estimates of $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, and σ_a^2. The spectrum for the underlying process can also be estimated by

$$\hat{f}_A(\omega) = \frac{\hat{\sigma}_a^2}{2\pi} \frac{\hat{\theta}_q(e^{-i\omega})\hat{\theta}_q(e^{i\omega})}{\hat{\phi}_p(e^{-i\omega})\hat{\phi}_p(e^{i\omega})}, \tag{13.4.5}$$

where

$$\hat{\phi}_p(e^{-i\omega}) = (1 - \hat{\phi}_1 e^{-i\omega} - \cdots - \hat{\phi}_p e^{-ip\omega})$$

and

$$\hat{\theta}_q(e^{-i\omega}) = (1 - \hat{\theta}_1 e^{-i\omega} - \cdots - \hat{\theta}_q e^{-iq\omega}).$$

Equation (13.4.5) is known as ARMA spectral estimation. Similar to the AR spectral estimation, the quality of the ARMA spectral estimation relies on the proper choices of p and q in the approximation. As discussed in Section 7.7, the choice of the orders p and q can be determined using criteria such as Akaike's AIC.

EXAMPLE 13.4 Recall that we fitted the logarithms of Canadian lynx pelt sales with ARMA models in Chapter 7 and found that both the AR(3) model

$$(1 - .97B + .12B^2 + .5B^3)X_t = a_t \tag{13.4.6}$$

with $\hat{\sigma}_a^2 = .124$ and the ARMA(2,1) model

$$(1 - 1.55B + .94B^2)X_t = (1 - .59B)a_t \tag{13.4.7}$$

with $\hat{\sigma}_a^2 = .116$, where $X_t = (\ln Z_t - .98)$, were adequate. Thus, the spectrum of the logarithms of Canadian lynx pelt sales can also be estimated by

$$\hat{f}_A(\omega) = \frac{.124}{2\pi} \frac{1}{(1 - .97e^{-i\omega} + .12e^{-i2\omega} + .5e^{-i3\omega})(1 - .97e^{i\omega} + .12e^{i2\omega} + .5e^{i3\omega})} \tag{13.4.8}$$

or

$$\hat{f}_A(\omega) = \frac{.116}{2\pi} \frac{(1 - .59e^{-i\omega})(1 - .59e^{i\omega})}{(1 - 1.55e^{-i\omega} + .94e^{-i2\omega})(1 - 1.55e^{i\omega} + .94e^{i2\omega})}. \tag{13.4.9}$$

Both the spectra (13.4.8) and (13.4.9) are plotted in Figure 13.10. They are extremely similar, and both have the sharp peak at $\omega \simeq .68544$ yielding a period of 9.17 years. After carefully examining Figures 13.7, 13.8, and 13.10, one should find that the ARMA spectral estimation is comparable with other estimates for this data set.

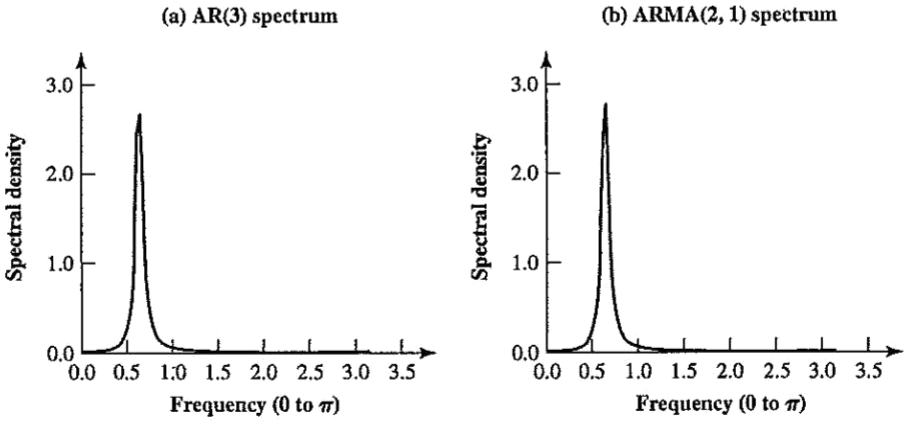


FIGURE 13.10 ARMA spectral estimation for the logarithms of Canadian lynx pelt sales.

EXERCISES

13.1 Consider the following data set (read across):

1.033	−.947	−.804	.053	−.424	1.157	−.123
−.899	−1.512	−.875	−1.348	2.079	1.042	.564
2.876	2.480	1.462	.364	.147	.467	.454
.472	1.882	−2.541	−.288	−.176	.564	−.027
.052	−1.373	−2.723	−.247	−.212	−.135	−1.245
.260	.300	−1.034	−1.666	−.351	.489	.192
3.425	.744	.878	.090	1.068	.564	−.745
−1.096	−.039	−1.451	−1.454	−.329	−.855	.232
−2.710	.552	−1.388	.455	1.192	.615	1.282
.834	−.948	1.756	.787	.032	.298	−.248
.606	−1.551	−.369	−.252	−1.542	−.689	1.575
1.152	−1.527	−1.383	1.132	.136	−2.463	−.093
−.297	−.727	2.328	−.878	−.074	−1.405	2.761
1.523	−.954	.056	−.030	1.000	−1.635	.433
−.454	.089					

(a) Compute the periodogram.

(b) Perform Fisher's test for the maximum periodogram ordinate.

13.2 Consider each of the series W2 and W3 discussed in Chapters 6 and 7.

(a) Compute the periodogram.

(b) Perform Fisher's test for the maximum periodogram ordinate.

(c) Perform Whittle's test for the second largest periodogram ordinate.

13.3 Compute and discuss the smoothed spectrum for the data set given in Exercise 13.1 using Bartlett's window with truncation points $M = 5$, 10, and 15.

13.4 (a) Compute the smoothed spectrum for the data set given in Exercise 13.1 using the Tukey–Hanning window with $M = 10$.

(b) Find the 95% confidence interval for the spectrum of the unknown underlying generating process for the data set using the results obtained in part (a).

13.5 (a) Compute the smoothed spectrum for the Series W2 using Parzen's window with a proper choice of M.

(b) Find the 95% confidence interval for the spectrum of the unknown underlying generating process for Series W2 using the result obtained in part (a).

13.6 Consider the first differences of Series W4 discussed in Chapters 6 and 7.

(a) Compute the periodogram.

(b) Compute and plot a smoothed spectrum using a proper spectral window.

(c) Find and plot an ARMA spectral estimate.

(d) Comment your results obtained in parts (b) and (c).

13.7 Consider the proper differenced sequence of Series W9 discussed in Chapter 8. Repeat the analysis requested in Exercise 13.6.

13.8 Find and discuss the ARMA spectral estimates for Series W2 and W3.

14

Transfer Function Models

In earlier chapters, we were concerned with the time-domain and the frequency-domain analysis of a univariate time series. In this chapter, we consider transfer function models where an output series is related to one or more input series. For example, sales may be related to advertising expenditures or daily electricity consumption series may be related to certain weather variable series such as daily maximum outdoor temperature, and relative humidity. We study the single-output linear system with both a single-input and a multiple-input linear system. After studying the basic characteristics of these transfer function models, we discuss the identification, estimation, and diagnostic checking of these models. Forecasting using transfer function models is also discussed. Detailed step-by-step model building is illustrated with an empirical example. The frequency domain approach of bivariate processes is also presented.

14.1 Single-Input Transfer Function Models

14.1.1 General Concepts

Assume that x_t and y_t are properly transformed series so that they are both stationary. In a single-input, single-output linear system, the output series y_t and the input series x_t are related through a linear filter as

$$y_t = \nu(B)x_t + n_t, \tag{14.1.1}$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$ is referred to as the transfer function of filter by Box, Jenkins, and Reinsel (1994) and n_t is the noise series of the system that is independent of the input series x_t. Recall from Section 12.3 that the term *transfer function* is also used to describe the frequency response function in the literature. Box, Jenkins, and Reinsel call Equation (14.1.1) the transfer function model. For an obvious reason, when x_t and n_t are assumed to follow some ARMA models, Equation (14.1.1) is also known as the ARMAX model.

The coefficients ν_j in the transfer function model (14.1.1) are often called the impulse response weights. As a function of j, ν_j is also called the impulse response function. The transfer function model is said to be stable if the sequence of these impulse response weights is absolutely summable, i.e., if $\sum|\nu_j| < \infty$. Thus, in a stable system, a bounded input

always produces a bounded output. The transfer function model is said to be causal if $\nu_j = 0$ for $j < 0$. Thus, in a causal model, the system does not respond to input series until they have been actually applied to the system. In other words, the present output is affected only by the system's current and past input values. A causal model is also called a realizable model as it appears that all real physical systems are causal. In practice, we often consider only the stable and causal model

$$\begin{aligned} y_t &= \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \cdots + n_t \\ &= \nu(B)x_t + n_t, \end{aligned} \tag{14.1.2}$$

where $\nu(B) = \sum_{j=0}^{\infty} \nu_j B^j$, $\sum_{j=0}^{\infty} |\nu_j| < \infty$, and x_t and n_t are independent. The system is presented in Figure 14.1.

The purposes of transfer function modeling are to identify and estimate the transfer function $\nu(B)$ and the noise model for n_t based on the available information of the input series x_t and the output series y_t. The difficulties are that the information on x_t and y_t is finite and the transfer function $\nu(B)$ in (14.1.2) may contain an infinite number of coefficients. To alleviate these difficulties, we represent the transfer function $\nu(B)$ in the rational form

$$\nu(B) = \frac{\omega(B)B^b}{\delta(B)}, \tag{14.1.3}$$

where $\omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$, $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$, and b is a delay parameter representing the actual time lag that elapses before the impulse of the input variable produces an effect on the output variable. For a stable system, we assume that the roots of $\delta(B) = 0$ are outside the unit circle.

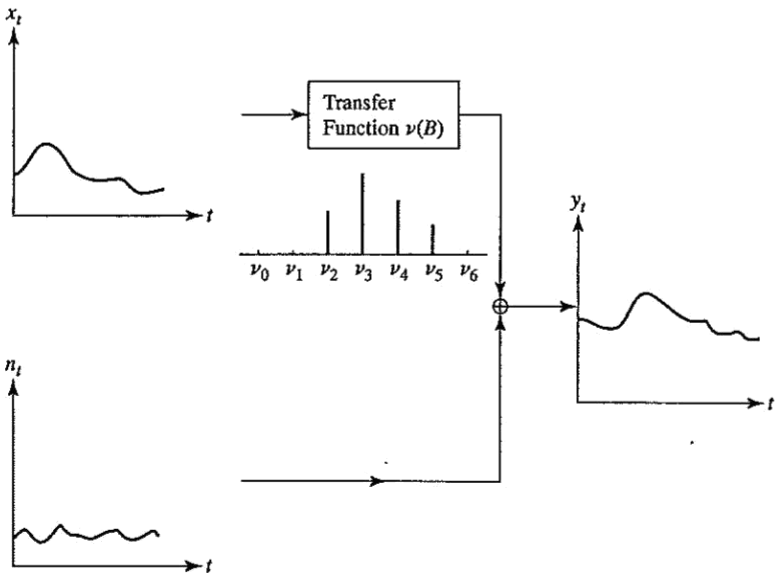


FIGURE 14.1 Dynamic transfer function system.

The orders of s, r, and b and their relationships to the impulse response weights ν_j can be found by equating the coefficients of B^j in both sides of the equation

$$\delta(B)\nu(B) = \omega(B)B^b,$$

or

$$[1 - \delta_1 B - \cdots - \delta_r B^r][\nu_0 + \nu_1 B + \nu_2 B^2 + \cdots] = [\omega_0 - \omega_1 B - \cdots - \omega_s B^s]B^b. \tag{14.1.4}$$

Thus, we have that

$$\begin{aligned}
&\nu_j = 0, && j < b,\\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_r\nu_{j-r} + \omega_0, && j = b,\\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_r\nu_{j-r} - \omega_{j-b}, && j = b+1, b+2, \ldots, b+s,\\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_r\nu_{j-r}, && j > b+s.
\end{aligned}$$

The r impulse response weights $\nu_{b+s}, \nu_{b+s-1}, \ldots, \nu_{b+s-r+1}$ serve as starting values for the difference equation

$$\delta(B)\nu_j = 0, \qquad j > b + s. \tag{14.1.5}$$

In other words, the impulse response weights for the system (14.1.3) consist of the following:

1. b zero weights $\nu_0, \nu_1, \ldots, \nu_{b-1}$
2. $s - r + 1$ weights $\nu_b, \nu_{b+1}, \ldots, \nu_{b+s-r}$ that do not follow a fixed pattern
3. r starting impulse response weights $\nu_{b+s-r+1}, \nu_{b+s-r+2}, \ldots,$ and ν_{b+s}
4. ν_j, for $j > b + s$, that follows the pattern given in (14.1.5)

In summary, b is determined by $\nu_j = 0$ for $j < b$ and $\nu_b \neq 0$. The value r is determined by the pattern of the impulse response weights in a similar manner to the identification of the order p for a univariate ARIMA model through the pattern of the autocorrelation function. For a given value of b, if $r = 0$, then the value of s can be easily found using that $\nu_j = 0$ for $j > b + s$; if $r \neq 0$, then the value of s is found by checking where the pattern of decay for impulse response weight begins.

14.1.2 Some Typical Impulse Response Functions

In practice, the values of r and s in the system (14.1.3) rarely exceed 2. Some typical transfer functions are illustrated as follows:

Type 1. $r = 0$. In this case, the transfer function contains only a finite number of impulse response weights starting with $\nu_b = \omega_0$ and ending at $\nu_{b+s} = -\omega_s$ as shown in Table 14.1.

Type 2. $r = 1$. In this case, the impulse response weights exhibit an exponential decay starting from ν_b if $s = 0$, from ν_{b+1} if $s = 1$, and from ν_{b+2} if $s = 2$ as shown in Table 14.2.

TABLE 14.1 Transfer function for $r = 0$.

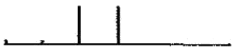
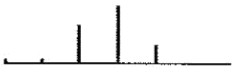
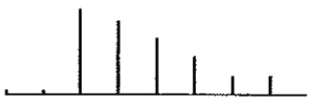
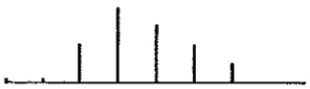
(b,r,s)	Transfer function	Typical impulse weights
(2, 0, 0)	$\nu(B)x_t = \omega_0 x_{t-2}$	
(2, 0, 1)	$\nu(B)x_t = (\omega_0 - \omega_1 B)x_{t-2}$	
(2, 0, 2)	$\nu(B)x_t = (\omega_0 - \omega_1 B - \omega_2 B^2)x_{t-2}$	

TABLE 14.2 Transfer function for $r = 1$.

(b,r,s)	Transfer function	Typical impulse weights
(2, 1, 0)	$\nu(B)x_t = \dfrac{\omega_0}{(1 - \delta_1 B)}x_{t-2}$	
(2, 1, 1)	$\nu(B)x_t = \dfrac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)}x_{t-2}$	
(2, 1, 2)	$\nu(B)x_t = \dfrac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B)}x_{t-2}$	

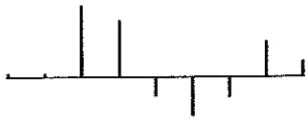
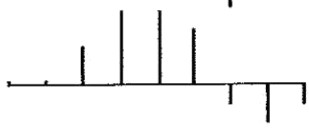
Type 3. $r = 2$. In this case, the impulse response weights exhibit either a damped exponential or a damped sine wave depending on the nature of the roots of the polynomial $\delta(B) = (1 - \delta_1 B - \delta_2 B^2) = 0$. They follow an exponential decay if the roots are real, i.e., if $\delta_1^2 + 4\delta_2 \geq 0$; they follow a damped sine wave if the roots are complex, i.e., if $\delta_1^2 + 4\delta_2 < 0$. The value of s is determined similarly as described earlier. Table 14.3 illustrates some examples of impulse response weights exhibiting a damped sine wave.

14.2 The Cross-Correlation Function and Transfer Function Models

14.2.1 The Cross-Correlation Function (CCF)

The cross-correlation function is a useful measure of strength and direction of correlation between two random variables. Given two stochastic processes x_t and y_t for $t = 0, \pm 1, \pm 2, \ldots$, we say that x_t and y_t are jointly stationary if x_t and y_t are both univariate stationary processes and the cross-covariance function between x_t and y_s, $\text{Cov}(x_t, y_s)$, is a function

TABLE 14.3 Transfer function for $r = 2$.

(b,r,s)	Transfer function	Typical impulse weights
(2, 2, 0)	$\nu(B)x_t = \dfrac{\omega_0}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
(2, 2, 1)	$\nu(B)x_t = \dfrac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
(2, 2, 2)	$\nu(B)x_t = \dfrac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	

of the time difference $(s - t)$ only. In such cases, we have the following cross-covariance function between x_t and y_t:

$$\gamma_{xy}(k) = E[(x_t - \mu_x)][(y_{t+k} - \mu_y)] \tag{14.2.1}$$

for $k = 0, \pm 1, \pm 2, \ldots$ where $\mu_x = E(x_t)$ and $\mu_y = E(y_t)$. Upon standardization, we have the following cross-correlation function (CCF):

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \tag{14.2.2}$$

for $k = 0, \pm 1, \pm 2, \ldots$, where σ_x and σ_y are the standard deviations of x_t and y_t. It is important to note that the cross-covariance function $\gamma_{xy}(k)$ and cross-correlation functions $\rho_{xy}(k)$ are generalizations of autocovariances and autocorrelation functions because $\gamma_{xx}(k) = \gamma_x(k)$ and $\rho_{xx}(k) = \rho_x(k)$. Unlike the autocorrelation function $\rho_x(k)$, however, which is symmetric around the origin, i.e., $\rho_x(k) = \rho_x(-k)$, the cross-correlation function is not symmetric, i.e., $\rho_{xy}(k) \neq \rho_{xy}(-k)$. Instead, because $\gamma_{xy}(k) = E(x_t - \mu_x)(y_{t+k} - \mu_y) = E(y_{t+k} - \mu_y)(x_t - \mu_x) = \gamma_{yx}(-k)$, we have $\rho_{xy}(k) = \rho_{yx}(-k)$. Thus, the CCF measures not only the strength of an association but also its direction. To see the full picture of the relationship between the series x_t and the series y_t, it is important to examine the CCF, $\rho_{xy}(k)$, for both positive lags $k > 0$ and negative lags $k < 0$. The graph of the CCF is sometimes called the cross-correlogram.

EXAMPLE 14.1 Consider the simple AR(1) process

$$(1 - \phi B)Z_t = a_t, \tag{14.2.3}$$

where $|\phi| < 1$ and a_t is a white noise series with zero mean and constant variance σ_a^2. For time $t + k$, we can rewrite Equation (14.2.3) as

$$Z_{t+k} = \frac{1}{(1 - \phi B)} a_{t+k}$$
$$= a_{t+k} + \phi a_{t+k-1} + \phi^2 a_{t+k-2} + \cdots. \tag{14.2.4}$$

The cross-covariance function between a_t and Z_t is

$$\gamma_{az}(k) = E(a_t Z_{t+k}) = \begin{cases} \phi^k \sigma_a^2, & k \geq 0, \\ 0, & k < 0. \end{cases}$$

Hence, the CCF becomes

$$\rho_{az}(k) = \begin{cases} \phi^k \sqrt{1 - \phi^2}, & k \geq 0, \\ 0, & k < 0, \end{cases}$$

where we recall that $\sigma_z^2 = \text{Var}(Z_t) = \sigma_a^2/(1 - \phi^2)$. The graph of this CCF for $\phi = .5$ is shown in Figure 14.2.

Because a univariate ARMA model $\phi_p(B)Z_t = \theta_q(B)a_t$ can be written as

$$Z_t = \frac{\theta_q(B)}{\phi_p(B)} a_t,$$

the above example shows that the univariate ARMA model can be regarded as a special case of the transfer function model without a noise term. In this formulation, the univariate time series Z_t is the output, the white noise series a_t is the input, and the transfer function $\nu(B) = \theta_q(B)/\phi_p(B)$.

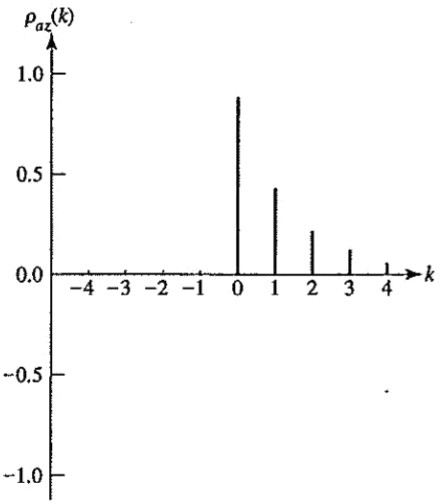


FIGURE 14.2 CCF between a_t and Z_t for $(1 - \phi B)Z_t = a_t$ with $\phi = .5$.

14.2.2 The Relationship between the Cross-Correlation Function and the Transfer Function

For time $t + k$, the transfer function model (14.1.2) can be written as

$$y_{t+k} = \nu_0 x_{t+k} + \nu_1 x_{t+k-1} + \nu_2 x_{t+k-2} + \cdots + n_{t+k}. \tag{14.2.5}$$

Without loss of generality we assume that $\mu_x = 0$ and $\mu_y = 0$. Multiplying by x_t on both sides of Equation (14.2.5) and taking expectations, we have

$$\gamma_{xy}(k) = \nu_0 \gamma_{xx}(k) + \nu_1 \gamma_{xx}(k-1) + \nu_2 \gamma_{xx}(k-2) + \cdots, \tag{14.2.6}$$

where we recall that $\gamma_{xn}(k) = 0$ for all k. Hence,

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y}[\nu_0 \rho_x(k) + \nu_1 \rho_x(k-1) + \nu_2 \rho_x(k-2) + \cdots]. \tag{14.2.7}$$

The relationship between the CCF, $\rho_{xy}(k)$ and the impulse response function ν_j is clearly contaminated by the autocorrelation structure of the input series x_t. Even if $r = 0$ in (14.1.3) and hence the transfer function $\nu(B)$ contains only a finite number of impulse response weights, the use of Equation (14.2.7) to form a system of equations to solve for ν_j as a function of $\rho_{xy}(k)$ and $\rho_x(k)$ is difficult. The variance and covariance of the sample estimate of $\rho_{xy}(k)$ are clearly contaminated by the autocorrelation structure of the input series x_t as shown in Equation (14.2.7), making the identification of both $\rho_{xy}(k)$ and ν_k difficult.

If the input series is white noise, i.e., $\rho_x(k) = 0$ for $k \neq 0$, however, then Equation (14.2.7) reduces to

$$\nu_k = \frac{\sigma_y}{\sigma_x}\rho_{xy}(k). \tag{14.2.8}$$

Thus, the impulse response function ν_k is directly proportional to the cross-correlation function $\rho_{xy}(k)$.

Some remarks are in order.

1. The CCF, $\rho_{xy}(k)$, is defined only when x_t and y_t are jointly bivariate stationary processes. To achieve the required stationarity, some differencing operators and variance stabilization transformations may be necessary. Thus, unless mentioned otherwise, the processes x_t and y_t are assumed to be jointly stationary in our discussions.
2. In the general transfer function model

$$y_t = \nu(B)x_t + n_t, \tag{14.2.9}$$

we assume that the input series x_t follows an ARMA process

$$\phi_x(B)x_t = \theta_x(B)\alpha_t,$$

where α_t is white noise. The series α_t,

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} x_t, \tag{14.2.10}$$

is often called the prewhitened input series. Applying the same prewhitening transformation to the output series y_t, we obtain a filtered output series,

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)} y_t. \tag{14.2.11}$$

Letting $\epsilon_t = \theta_x^{-1}(B)\phi_x(B)n_t$, the transfer function model (14.2.9) becomes

$$\beta_t = \nu(B)\alpha_t + \epsilon_t. \tag{14.2.12}$$

The impulse response weights ν_j for the transfer function can therefore be found as

$$\nu_k = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\alpha\beta}(k). \tag{14.2.13}$$

This result leads us to the fundamental steps of the identification of the transfer function models to be discussed in the next section.

14.3 Construction of Transfer Function Models

14.3.1 Sample Cross-Correlation Function

For a given set of time series data x_t and y_t, $1 \le t \le n$, the cross-correlation function

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y}, \qquad k = 0, \pm 1, \pm 2, \ldots, \tag{14.3.1}$$

is estimated by the sample cross-correlation function

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{S_x S_y}, \qquad k = 0, \pm 1, \pm 2, \ldots, \tag{14.3.2}$$

where

$$\hat{\gamma}_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), & k \ge 0, \\ \frac{1}{n} \sum_{t=1-k}^{n} (x_t - \bar{x})(y_{t+k} - \bar{y}), & k < 0, \end{cases}$$

$$S_x = \sqrt{\hat{\gamma}_{xx}(0)}, \qquad S_y = \sqrt{\hat{\gamma}_{yy}(0)}, \tag{14.3.3}$$

and $\bar{x}$ and $\bar{y}$ are the sample means of the x_t and y_t series, respectively.

To test whether certain values of the $\rho_{xy}(k)$ are zero, we compare the sample CCF $\hat{\rho}_{xy}(k)$ with their standard errors. Under the normal assumption, Bartlett (1955) derived the approximate variance and covariance between two sample cross-correlations $\hat{\rho}_{xy}(k)$ and $\hat{\rho}_{xy}(k+j)$. The covariance is given by

$$\begin{aligned}\operatorname{Cov}[\hat{\rho}_{xy}(k), \hat{\rho}_{xy}(k+j)] &\simeq (n-k)^{-1} \sum_{i=-\infty}^{\infty} \Big\{ \rho_{xx}(i)\rho_{yy}(i+j) + \rho_{xy}(i+k+j)\rho_{xy}(k-i) \\ &\quad + \rho_{xy}(k)\rho_{xy}(k+j)\Big[\rho_{xy}^2(i) + \frac{1}{2}\rho_{xx}^2(i) + \frac{1}{2}\rho_{yy}^2(i)\Big] \\ &\quad - \rho_{xy}(k)[\rho_{xx}(i)\rho_{xy}(i+k+j) + \rho_{xy}(-i)\rho_{yy}(i+k+j)] \\ &\quad - \rho_{xy}(k+j)[\rho_{xx}(i)\rho_{xy}(i+k) + \rho_{xy}(-i)\rho_{yy}(i+k)]\Big\}. \end{aligned} \tag{14.3.4}$$

Hence,

$$\begin{aligned}\operatorname{Var}[\hat{\rho}_{xy}(k)] \simeq (n-k)^{-1} \sum_{i=-\infty}^{\infty} \Big\{ &\rho_{xx}(i)\rho_{yy}(i) + \rho_{xy}(k+i)\rho_{xy}(k-i) \\ &+ \rho_{xy}^2(k)\Big[\rho_{xy}^2(i) + \frac{1}{2}\rho_{xx}^2(i) + \frac{1}{2}\rho_{yy}^2(i)\Big] \\ &- 2\rho_{xy}(k)[\rho_{xx}(i)\rho_{xy}(i+k) + \rho_{xy}(-i)\rho_{yy}(i+k)]\Big\}. \end{aligned} \tag{14.3.5}$$

Under the hypothesis that the two series x_t and y_t are uncorrelated and the x_t series is white noise, (14.3.4) becomes

$$\operatorname{Cov}[\hat{\rho}_{xy}(k), \hat{\rho}_{xy}(k+j)] \simeq (n-k)^{-1}\rho_{yy}(j). \tag{14.3.6}$$

It follows that

$$\operatorname{Var}[\hat{\rho}_{xy}(k)] \simeq (n-k)^{-1}. \tag{14.3.7}$$

Thus, when the x_t series is white noise, we can test the hypothesis that the two series x_t and y_t are not cross-correlated by comparing the sample CCF $\hat{\rho}_{xy}(k)$ with their approximate standard errors $1/\sqrt{(n-k)}$.

In practice, the x_t series is usually not white noise, and we have to prewhiten it and also filter the output series, as shown in the next section.

14.3.2 Identification of Transfer Function Models

Based on the above discussions, the transfer function $\nu(B)$ is obtained from the following simple steps:

1. Prewhiten the input series,

$$\phi_x(B)x_t = \theta_x(B)\alpha_t, \tag{14.3.8}$$

i.e.,

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)}x_t, \tag{14.3.9}$$

where α_t is a white noise series with mean zero and variance σ_α^2.

2. Calculate the filtered output series. That is, transform the output series y_t using the above prewhitening model to generate the series

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)}y_t. \tag{14.3.10}$$

3. Calculate the sample CCF, $\hat{\rho}_{\alpha\beta}(k)$ between α_t and β_t to estimate ν_k:

$$\hat{\nu}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}\hat{\rho}_{\alpha\beta}(k). \tag{14.3.11}$$

The significance of the CCF and its equivalent $\hat{\nu}_k$ can be tested by comparing it with its standard error $(n-k)^{-1/2}$.

4. Identify b, $\delta(B) = (1 - \delta_1 B - \cdots - \delta_r B^r)$, and $\omega(B) = (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)$ by matching the pattern of $\hat{\nu}_k$ with the known theoretical patterns of the ν_k discussed in Sections 14.1.1 and 14.1.2. Once b, r, and s are chosen, preliminary estimates $\hat{\omega}_j$ and $\hat{\delta}_j$ can be found from their relationships with ν_k, as shown in Equation (14.1.4). Thus, we have a preliminary estimate of the transfer function $\nu(B)$ as

$$\hat{\nu}(B) = \frac{\hat{\omega}(B)}{\hat{\delta}(B)}B^b. \tag{14.3.12}$$

Identification of the Noise Model Once we obtain the preliminary transfer function, we can calculate the estimated noise series

$$\begin{aligned}\hat{n}_t &= y_t - \hat{\nu}(B)x_t \\ &= y_t - \frac{\hat{\omega}(B)}{\hat{\delta}(B)}B^b x_t.\end{aligned} \tag{14.3.13}$$

The appropriate model for the noise can then be identified by examining its sample ACF and PACF or by other univariate time series identification tools, giving

$$\phi(B)n_t = \theta(B)a_t. \tag{14.3.14}$$

Remarks on the Identification of the Transfer Function Model Combining (14.3.12) and (14.3.14), we have the following entertained transfer function model:

$$y_t = \frac{\omega(B)}{\delta(B)}x_{t-b} + \frac{\theta(B)}{\phi(B)}a_t. \tag{14.3.15}$$

Some important remarks are in order in the construction of this transfer function model.

1. In the construction of the model, it is assumed that the variables y_t, x_t, and n_t are all stationary. Thus, for nonstationary time series, some variance-stabilization and differencing transformations should be used first to achieve stationarity.
2. In the above process of identifying the transfer function $\nu(B)$, we prewhiten the input series. The prewhitening model is applied to filter the output series, but not necessarily to whiten it. This method for constructing a causal transfer function model is normal and simple. For constructing a possible noncausal system with feedback phenomenon, where y_t is influenced by x_t and x_t is also influenced by y_t, however, both input and output series should be prewhitened before examining its CCF. This process is often referred to as double prewhitening. For a good reference and example of constructing a noncausal system, see Granger and Newbold (1986). It is, however, generally more advantageous to model a noncausal system using a vector process discussed in Chapter 16.

14.3.3 Estimation of Transfer Function Models

After identifying a tentative transfer function model

$$y_t = \frac{\omega(B)}{\delta(B)}x_{t-b} + \frac{\theta(B)}{\phi(B)}a_t, \tag{14.3.16}$$

we need to estimate the parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_r)'$, $\boldsymbol{\omega} = (\omega_0, \omega_1, \ldots, \omega_s)'$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, and σ_a^2. We can rewrite (14.3.16) as

$$\delta(B)\phi(B)y_t = \phi(B)\omega(B)x_{t-b} + \delta(B)\theta(B)a_t \tag{14.3.17}$$

or, equivalently,

$$c(B)y_t = d(B)x_{t-b} + e(B)a_t, \tag{14.3.18}$$

where

$$\begin{aligned} c(B) = \delta(B)\phi(B) &= (1 - \delta_1 B - \cdots - \delta_r B^r)(1 - \phi_1 B - \cdots - \phi_p B^p) \\ &= (1 - c_1 B - c_2 B^2 - \cdots - c_{p+r}B^{p+r}), \end{aligned}$$

$$d(B) = \phi(B)\omega(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)(\omega_0 - \omega_1 B - \cdots - \omega_s B^s)$$
$$= (d_0 - d_1 B - d_2 B^2 - \cdots - d_{p+s} B^{p+s}),$$

and

$$e(B) = \delta(B)\theta(B) = (1 - \delta_1 B - \cdots - \delta_r B^r)(1 - \theta_1 B - \cdots - \theta_q B^q)$$
$$= (1 - e_1 B - e_2 B^2 - \cdots - e_{r+q} B^{r+q}).$$

Thus,

$$a_t = y_t - c_1 y_{t-1} - \cdots - c_{p+r} y_{t-p-r} - d_0 x_{t-b} + d_1 x_{t-b-1}$$
$$+ \cdots + d_{p+s} x_{t-b-p-s} + e_1 a_{t-1} + \cdots + e_{r+q} a_{t-r-q}, \tag{14.3.19}$$

where c_i, d_j, and e_k are functions of δ_i, ω_j, ϕ_k, and θ_l. Under the assumption that the a_t are $N(0, \sigma_a^2)$ white noise series, we have the following conditional likelihood function:

$$L(\boldsymbol{\delta}, \boldsymbol{\omega}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_a^2 | b, \mathbf{x}, \mathbf{y}, \mathbf{x}_0, \mathbf{y}_0, \mathbf{a}_0) = (2\pi\sigma_a^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right], \tag{14.3.20}$$

where $\mathbf{x}_0$, $\mathbf{y}_0$, $\mathbf{a}_0$ are some proper starting values for computing a_t from (14.3.19) similar to the starting values needed in the estimation of univariate ARIMA models discussed in Section 7.2.

In general, the estimation methods introduced in Chapter 7 can also be used to estimate the parameters $\boldsymbol{\omega}$, $\boldsymbol{\delta}$, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and σ_a^2. For example, by setting the unknown a's equal to their conditional expected values of zero, the nonlinear least squares estimate of these parameters is obtained by minimizing

$$S(\boldsymbol{\delta}, \boldsymbol{\omega}, \boldsymbol{\phi}, \boldsymbol{\theta} | b) = \sum_{t=t_0}^{n} a_t^2, \tag{14.3.21}$$

where $t_0 = \max\{p + r + 1, b + p + s + 1\}$.

Note that so far we have assumed that b is known. For given values of r, s, p, and q, if we need also to estimate b, then (14.3.21) can be optimized for a likely range of values of b. Then b is selected to be the value that gives the overall minimum of the sum of squares.

Because the ordinary least squares method is the most commonly used estimation procedure, in terms of the general transfer function model in (14.2.9), one may be tempted to use it to estimate the parameters. As pointed out by Müller and Wei (1997), however, the OLS estimates in this case are not necessarily consistent. They introduce an iterative regression procedure to produce consistent estimates, and on the basis of these consistent estimates they also propose an alternative method of model specification. We refer readers to Müller and Wei's paper for details.

14.3.4 Diagnostic Checking of Transfer Function Models

After the model has been identified and its parameters estimated, it is necessary to check the model adequacy before we can use it for forecasting, control, and other purposes. In the transfer function model, we assume that the a_t are white noise and are independent of the input series x_t and hence are also independent of the prewhitened input series α_t. Thus, in the diagnostic checking of a transfer function model, we have to examine the residuals $\hat{a}_t$ from the noise model as well as the residuals α_t from the prewhitened input model to see whether the assumptions hold.

1. *Cross-correlation check:* To check whether the noise series a_t and the input series x_t are independent. For an adequate model, the sample CCF, $\hat{\rho}_{\alpha\hat{a}}(k)$, between $\hat{a}_t$ and α_t should show no patterns and lie within their two standard errors $2(n-k)^{-1/2}$. The following portmanteau test can also be used:

$$Q_0 = m(m+2)\sum_{j=0}^{K}(m-j)^{-1}\hat{\rho}^2_{\alpha\hat{a}}(j), \tag{14.3.22}$$

   which approximately follows a χ^2 distribution with $(K+1)-M$ degrees of freedom, where $m = n - t_0 + 1$, which is the number of residuals $\hat{a}_t$ calculated, and M is the number of parameters δ_i and ω_j estimated in the transfer function $\nu(B) = \omega(B)/\delta(B)$. As shown in Pierce (1968), the number of degrees of freedom for Q_0 is independent of the number of parameters estimated in the noise model.
2. *Autocorrelation check:* To check whether the noise model is adequate. For an adequate model, both the sample ACF and PACF of $\hat{a}_t$ should not show any patterns. A portmanteau Q test similar to the one in (7.5.1) can also be used. In other words, we can calculate

$$Q_1 = m(m+2)\sum_{j=1}^{K}(m-j)^{-1}\hat{\rho}^2_{\hat{a}}(j). \tag{14.3.23}$$

   The Q_1 statistic approximately follows a χ^2 distribution with $(K-p-q)$ degrees of freedom depending only on the number of parameters in the noise model.

   In terms of model inadequacy, two situations may arise.

Case I: The transfer function $\nu(B)$ is incorrect. In this case, regardless of whether the noise model is adequate, we have both $\rho_{a\alpha}(k) \neq 0$ and $\rho_a(k) \neq 0$ for some k. To see, suppose that the correct model is

$$y_t = \nu(B)x_t + \psi(B)a_t, \tag{14.3.24}$$

but we incorrectly choose a wrong transfer function $\nu_0(B)$ leading to an error term b_t,

$$y_t = \nu_0(B)x_t + \psi_0(B)b_t. \tag{14.3.25}$$

Then,

$$b_t = \psi_0^{-1}(B)[\nu(B) - \nu_0(B)]x_t + \psi_0^{-1}(B)\psi(B)a_t. \tag{14.3.26}$$

Thus, b_t will not only be cross-correlated with x_t and hence with α_t, but also will be autocorrelated. That is true even when $\psi_0(B) = \psi(B)$, i.e., when the noise model is correct. Hence, the remedy in this case is first to reidentify the transfer function $\nu(B)$ and then to modify the noise model.

Case 2: The transfer function $\nu(B)$ is correct, and only the noise model is inadequate. In this case, $\rho_{\alpha a}(k) = 0$ for all k, but $\rho_a(k) \neq 0$ for some k. The pattern of $\hat{\rho}_{\hat{a}}(k)$ can be used to modify the noise model $[\theta(B)/\phi(B)]a_t$.

In summary, for a transfer function model to be adequate, both $\hat{\rho}_{\alpha\hat{a}}(k)$ and $\hat{\rho}_{\hat{a}}(k)$ should be statistically insignificant and show no patterns. Because the noise model will be contaminated by an incorrect transfer function, in the diagnostic checking of transfer function models, it is prudent to carry out the cross-correlation check first. In essence, this step is similar to the analysis of a variance model with an interaction term where the test of an interaction effect should be performed before the test of the main effects. If the cross-correlation test fails, there is no need to check the autocorrelations for $\hat{a}_t$. One should reidentify the transfer function $\nu(B)$ and repeat the iterative model-building steps until a satisfactory model is obtained.

14.3.5 An Empirical Example

The Data The Lydia Pinkham annual data listed as Series W12 in the appendix and plotted in Figure 14.3 provide an interesting case for studying the dynamic relationship between advertising and sales. Lydia Pinkham's Vegetable Compound was first brewed and prepared

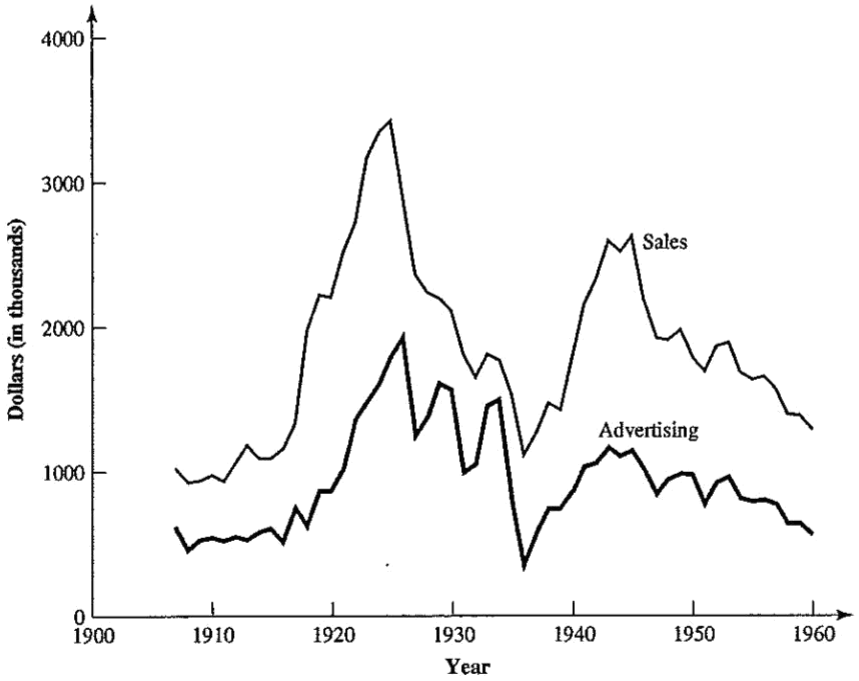


FIGURE 14.3 Lydia Pinkham annual advertising and sales data (1907–1960).

by the lady of the same name as a home remedy for her friends. The compound, an herbal extract in alcoholic solution, was considered to be effective against "women's weakness." Additional medical claims followed with the marketing success of the compound. The first commercial sale of the medicine was in 1873, when sales reached $500,000. A part of the profits was reinvested immediately into high intensity advertising campaigns. Records of a court litigation involving a family argument produced a complete listing of the firm's annual advertising expenditures and sales over the period 1908–1935. Palda (1964) was later able to obtain the data for 1907–1960.

Based on Palda (1964), the product remained essentially unchanged in taste, form, and action through the period in question, although in 1914 the solid content of the liquid compound was increased sevenfold at the prodding of the Internal Revenue Service, which was considering an alcohol tax on the compound. There were no close product substitutes on the market. The company relied exclusively on advertising to increase primary demand for the product; no other marketing instruments, such as a sales force, credit, or quantity discounts, were used. The advertising budget was primarily for newspaper copy, but the company began to experiment with radio and television in the years following World War II. There were some changes in the character of the advertising copy used because of orders from the U.S. Food and Drug Administration in 1925 to cease and desist from continuing to make strong medical claims. Furthermore, the U.S. Federal Trade Commission in 1938 was also examining the medical uses of the compound. Finally, the company was able to demonstrate the possibility that the active herbal ingredients had some medical benefits, and in 1941 the FTC permitted the company to strengthen its advertising copy.

A Transfer Function Model between Advertising and Sales A strong relationship has been demonstrated (Clarke, 1976) between current sales and present and past advertising. Such lagged advertising effects on sales may result from delayed response to the marketing effort, to a holdover of new customer demand, or to an increase in demand from existing customers. In addition, advertising may have a cumulative effect on demand. A transfer function model using sales Y_t as an output series and advertising expenditures X_t as an input series was first suggested by Helmer and Johansson (1977) to model the Lydia Pinkham data. It was reexamined later by Heyse and Wei (1986). For the purpose of forecasting comparisons, they hold back the last 14 observations and use only the first 40 observations (1907–1946) for model building, as adopted in the following.

1. *Prewhitening of the advertising series.* The first step in building a transfer function model is to prewhiten the input series: advertising expenditures. Examination of the sample ACF and PACF, shown in Table 14.4, indicates that the original series X_t is nonstationary but its differences, $x_t = (1 - B)X_t$, are stationary and can be modeled as an AR(2) model. The fitted model is

$$(1 - .07B + .41B^2)x_t = \alpha_t, \tag{14.3.27}$$

where $x_t = (1 - B)X_t$ and the α_t are the white noise with mean 0 and $\hat{\sigma}_\alpha^2 = 52{,}601$. Diagnostic checking does not reveal any inadequacies in the model. Thus, the prewhitening filter for x_t will be used to filter the differenced output series, $y_t = (1 - B)Y_t$, where we note that the original series of sales is also nonstationary.

TABLE 14.4 Sample ACF and PACF of the advertising series X_t.

(a) $\hat{\rho}_k$ and $\hat{\phi}_{kk}$ for X_t

k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$	.82	.60	.53	.50	.31	.11	.02	−.02	−.13	−.27	−.33	−.40
St.E.	.16	.24	.28	.30	.32	.33	.33	.33	.33	.33	.34	.34
$\hat{\phi}_{kk}$	.82	−.19	.31	−.03	−.41	−.01	−.00	−.11	−.01	−.19	−.00	−.32
St.E.	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16

(b) $\hat{\rho}_k$ and $\hat{\phi}_{kk}$ for $x_t = (1 - B)X_t$

k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$	.06	−.40	−.11	.43	.05	−.34	−.10	.17	.10	−.20	−.00	−.05
St.E.	.16	.16	.18	.19	.21	.21	.22	.23	.23	.23	.23	.23
$\hat{\phi}_{kk}$	.06	−.40	−.07	.24	−.10	−.13	.01	−.12	.04	−.05	.10	−.23
St.E.	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16	.16

TABLE 14.5 Sample CCF and impulse response function for the prewhitened input and the filtered output series ($\hat{\sigma}_\alpha = 229.35$ and $\hat{\sigma}_\beta = 287.54$).

Lag(k)	CCF($\hat{\rho}_{\alpha\beta}(k)$)	Impulse response function ($\hat{\nu}_k$)
0	.508	.637
1	.315	.395
2	.109	.136
3	−.094	−.118
4	.004	.005
5	−.257	−.322
6	−.257	−.323
7	−.279	−.349
8	−.140	−.175
9	−.080	−.099
10	−.182	−.228

2. *Identification of the impulse response function and transfer function.* Table 14.5 gives sample cross-correlation, $\hat{\rho}_{\alpha\beta}$, and the impulse response function, $\hat{\nu}_k = (\hat{\sigma}_\beta/\hat{\sigma}_\alpha)\hat{\rho}_{\alpha\beta}(k)$, for the prewhitened input series

$$\alpha_t = (1 - .07B + .41B^2)x_t, \qquad \text{where } \hat{\sigma}_\alpha = 229.35,$$

and the filtered output series

$$\beta_t = (1 - .07B + .41B^2)y_t, \qquad \text{where } \hat{\sigma}_\beta = 287.54. \tag{14.3.28}$$

The pattern of the sample impulse response function $\hat{\nu}_k$ in Table 14.2 suggests a transfer function with $b = 0$, $r = 1$, and $s = 0$, i.e.,

$$\nu(B)x_t = \frac{\omega_0}{(1 - \delta_1 B)}x_t. \tag{14.3.29}$$

If only the first two large impulse weights $\hat{\nu}_0$ and $\hat{\nu}_1$ were retained and the remaining smaller ones were ignored, then, from Table 14.1, a transfer function with $b = 0$, $r = 0$, and $s = 1$ is also a possibility. That is,

$$\nu(B)x_t = (\omega_0 - \omega_1 B)x_t. \tag{14.3.30}$$

3. *Identification of the noise model.* Once a tentative transfer function is obtained, we can obtain preliminary estimates of the parameters in the transfer function by utilizing the relationships between the impulse response weights and the parameters given in Equation (14.1.4). For the transfer function given in (14.3.29), we have

$$(1 - \hat{\delta}_1 B)[\hat{\nu}_0 + \hat{\nu}_1 B + \hat{\nu}_2 B^2 + \cdots] = \hat{\omega}_0.$$

Thus,

$$\hat{\omega}_0 = \hat{\nu}_0 = .637$$

$$\hat{\delta}_1 = \frac{\hat{\nu}_1}{\hat{\nu}_0} = \frac{.395}{.637} = .62.$$

The estimated noise series is then obtained by

$$\hat{n}_t = y_t - \frac{.64}{(1 - .62B)}x_t, \tag{14.3.31}$$

where $y_t = (1 - B)Y_t$ and $x_t = (1 - B)X_t$. Table 14.6 shows the sample ACF and PACF of the estimated noise series. The pattern indicates a white noise model. To be conservative, however, one can try an AR(1) model

$$(1 - \phi_1 B)n_t = a_t \tag{14.3.32}$$

TABLE 14.6 Sample ACF and PACF for the estimated noise series.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.03	−.22	−.20	.22	−.11	−.17	−.11	.41	−.06	−.13
St.E.	.19	.19	.20	.21	.21	.22	.22	.22	.25	.25
$\hat{\phi}_k$	−.03	−.22	−.23	.16	−.20	−.18	.14	.31	−.03	.11
St.E.	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19

and check the result of estimation in the combined transfer function-noise model to see whether $\phi_1 = 0$.

4. *Estimation of the transfer function model.* Combining the above results, the following two tentative transfer function models are obtained.

$$(1 - B)Y_t = \frac{\omega_0}{(1 - \delta_1 B)}(1 - B)X_t + \frac{1}{(1 - \phi_1 B)}a_t \tag{14.3.33}$$

and

$$(1 - B)Y_t = (\omega_0 - \omega_1 B)(1 - B)X_t + \frac{1}{(1 - \phi_1 B)}a_t. \tag{14.3.34}$$

These models were in fact originally suggested by Helmer and Johansson (1977). Estimates of the parameters and their associated standard errors are as follows:

Equation (14.3.33)			Equation (14.3.34)		
Parameter	Estimate	St.E.	Parameter	Estimate	St.E.
ω_0	.52	.16	ω_0	.48	.15
δ_1	.36	.25	ω_1	−.19	.14
ϕ_1	.26	.19	ϕ_1	.28	.19
σ_a^2	45,175		σ_a^2	45,238	

The results indicate that the parameters δ_1, ω_1, and ϕ_1 could be deleted from the models, and one could consider the following model:

$$(1 - B)Y_t = \omega_0(1 - B)X_t + a_t. \tag{14.3.35}$$

The fitted model becomes

$$\underset{(.14)}{(1 - B)Y_t = .6(1 - B)X_t} + a_t \tag{14.3.36}$$

with $\hat{\sigma}_a^2 = 46{,}935$.

5. *Diagnostic checking.* Final adoption of a proposed model requires study of the residuals. Specifically, one has to check whether the a_t are indeed white noise and are independent of the input series X_t and hence independent of the prewhitened input series α_t. The first 11 lags of the residual ACF of $\hat{a}_t$ and the CCF between the prewhitened input α_t and the residuals $\hat{a}_t$ from Equation (14.3.33) are shown in Table 14.7. Inspection of these two functions indicates that both the ACF and CCF are zero at these lags; hence, one might conclude that Equation (14.3.33) is adequate. Similar results hold for the residual ACF and the CCF between the prewhitened input and the residuals from Equations (14.3.34) and (14.3.35).

TABLE 14.7 Sample ACF and CCF for the residuals from Equation (14.3.33).

k	0	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_{\hat{a}}(k)$	1	−.04	.12	.02	−.09	.11	−.01	−.04	−.23	−.21	.03
St.E.	.16	.16	.16	.16	.17	.17	.17	.17	.17	.18	.18
$\hat{\rho}_{\alpha\hat{a}}$	−.01	.02	−.04	−.14	−.10	−.31	−.16	−.17	−.04	−.07	−.07
St.E.	.17	.17	.17	.17	.18	.18	.18	.18	.19	.20	.20

Reexamination of the Transfer Function Model Are the transfer function models in Equations (14.3.33), (14.3.34), and (14.3.35) really adequate for the Lydia Pinkham data? Because in practice one often considers only a realizable system, i.e., a causal transfer function model, the above procedure relied only on nonnegative lags. This approach has been widely used in empirical research. Unfortunately, these procedures are wrong; hence, the resulting conclusions are dubious. Because the ACF is symmetric about zero, one needs consider only its nonnegative lags. As mentioned in Section 14.2, however, the CCF is generally asymmetric and measures both the strength and direction of an association. To see the complete picture of the relationship between two time series x_t and y_t, it is important to examine $\rho_{xy}(k)$ for both positive lags $k > 0$ and negative lags $k < 0$. A causal transfer function model exists between series x_t and y_t only when $\rho_{xy}(k) = 0$ for $k > 0$ or $\rho_{xy}(k) = 0$ for $k < 0$. Series x_t is said to cause series y_t, denoted by $x_t \rightarrow y_t$, if $\rho_{xy}(k) \neq 0$ for some $k > 0$ and $\rho_{xy}(k) = 0$ for all $k < 0$; series y_t is said to cause series x_t, denoted by $x_t \leftarrow y_t$, if $\rho_{xy}(k) \neq 0$ for some $k < 0$ and $\rho_{xy}(k) = 0$ for all $k > 0$. A feedback relationship between x_t and y_t exists, denoted by $x_t \leftrightarrow y_t$, if $\rho_{xy}(k) \neq 0$ for some $k < 0$ as well as for some $k > 0$. Two series x_t and y_t are said to have a contemporaneous relationship if $\rho_{xy}(0) \neq 0$. They are said to be related only contemporaneously if $\rho_{xy}(0) \neq 0$ and $\rho_{xy}(k) = 0$ for all $k \neq 0$.

Now, let us reexamine the Lydia Pinkham data. The cross-correlations between the prewhitened input series α_t and the filtered output series β_t at both negative and nonnegative lags are shown in Table 14.8 and plotted in Figure 14.4. The pattern clearly indicates a contemporaneous relationship, $\hat{\rho}_{\alpha\beta}(0) = .51$, and the presence of a lagged association from sales to advertising, $\hat{\rho}_{\alpha\beta}(-1) = .43$. Because $\rho_{\alpha\beta}(k) \neq 0$ for $k = 0, -1$ and $\rho_{\alpha\beta}(k) = 0$ for $k > 0$ at a 5% significance level, we have sales $(Y_t) \rightarrow$ advertising (X_t) rather than advertising $(X_t) \rightarrow$ sales (Y_t). This phenomenon happens because advertising budgets are often set as a percentage of earlier sales.

TABLE 14.8 Sample CCF between the prewhitened input and the filtered output series.

k	−5	−4	−3	−2	−1	0	1	2	3	4	5
$\hat{\rho}_{\alpha\beta}(k)$	.23	.19	.24	.16	.43	.51	.32	.11	−.09	.00	−.26
St. E.	.18	.18	.18	.17	.17	.17	.17	.17	.18	.18	.18

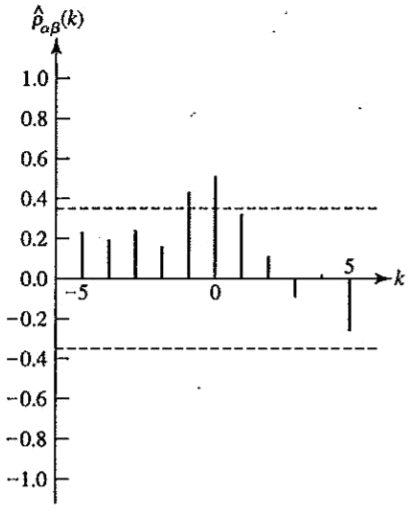


FIGURE 14.4 Sample CCF between the prewhitened input and filtered output series with two standard error limits.

TABLE 14.9 Sample CCF between the prewhitened input series and the residuals.

k	−5	−4	−3	−2	−1	0	1	2	3	4	5
$\hat{\rho}_{\alpha\hat{a}}(k)$	.32	.06	.19	.05	.46	−.01	.02	−.04	−.14	−.10	−.31
St.E.	.18	.18	.18	.17	.17	.17	.18	.18	.18	.18	.18

Equations (14.3.33), (14.3.34), and (14.3.35) are constructed without using the information contained in the CCF at negative lags. As a result, although the residual ACF indicates that the residuals are white noise, the more detailed CCF between the prewhitened input series and the residuals from Equation (14.3.33) as shown in Table 14.9 strongly suggests that the model is inadequate. The significant spike, $\hat{\rho}_{\alpha\hat{a}}(-1)$, at lag $k = -1$ implies that there is a feedback relationship from sales to advertising at one period which is not properly accounted for by transfer function models such as (14.3.33), (14.3.34), or (14.3.35).

14.4 Forecasting Using Transfer Function Models

Once a transfer function model is found to be adequate, it can be used to improve the forecast of the output series Y_t by using the past history of both the output series Y_t and the associated input series X_t. That is particularly true if the input series is a leading indicator.

14.4.1 Minimum Mean Square Error Forecasts for Stationary Input and Output Series

Suppose that Y_t and X_t are stationary and are related in the stable transfer function model

$$Y_t = \frac{\omega(B)}{\delta(B)}B^b X_t + \frac{\theta(B)}{\phi(B)}a_t \tag{14.4.1}$$

and

$$\phi_x(B)X_t = \theta_x(B)\alpha_t, \tag{14.4.2}$$

where $\omega(B)$, $\delta(B)$, $\theta(B)$, $\phi(B)$, $\phi_x(B)$, and $\theta_x(B)$ are finite order of polynomials in B; the roots of $\delta(B) = 0$, $\theta(B) = 0$, $\phi(B) = 0$, $\phi_x(B) = 0$, and $\theta_x(B) = 0$ are all outside of the unit circle; and a_t and α_t are independent zero mean white noise series with variances σ_a^2 and σ_α^2, respectively. Let

$$u(B) = \frac{\omega(B)B^b\theta_x(B)}{\delta(B)\phi_x(B)} = u_0 + u_1B + u_2B^2 + \cdots \tag{14.4.3}$$

and

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = 1 + \psi_1B + \psi_2B^2 + \cdots. \tag{14.4.4}$$

We can rewrite (14.4.1) as

$$\begin{aligned} Y_t &= u(B)\alpha_t + \psi(B)a_t \\ &= \sum_{j=0}^{\infty} u_j\alpha_{t-j} + \sum_{j=0}^{\infty}\psi_j a_{t-j}, \end{aligned} \tag{14.4.5}$$

where $\psi_0 = 1$. Thus, $Y_{t+l} = \sum_{j=0}^{\infty} u_j\alpha_{t+l-j} + \sum_{j=0}^{\infty}\psi_j a_{t+l-j}$. Let $\hat{Y}_t(l) = \sum_{j=0}^{\infty} u_{l+j}^*\alpha_{t-j} + \sum_{j=0}^{\infty}\psi_{l+j}^* a_{t-j}$ be the l-step ahead optimal forecast of Y_{t+l}. Then the forecast error becomes

$$\begin{aligned} Y_{t+l} - \hat{Y}_t(l) = &\sum_{j=0}^{l-1}[u_j\alpha_{t+l-j} + \psi_j a_{t+l-j}] \\ &- \sum_{j=0}^{\infty}[u_{l+j}^* - u_{l+j}]\alpha_{t-j} - \sum_{j=0}^{\infty}(\psi_{l+j}^* - \psi_{l+j})a_{t-j}. \end{aligned} \tag{14.4.6}$$

The mean square of the forecast error $E[Y_{t+l} - \hat{Y}_t(l)]^2$ is given by

$$E[Y_{t+l} - \hat{Y}_t(l)]^2 = \sum_{j=0}^{l-1}(\sigma_a^2 u_j^2 + \sigma_a^2 \psi_j^2) + \sum_{j=0}^{\infty}\sigma_a^2(u_{l+j}^* - u_{l+j})^2 + \sum_{j=0}^{\infty}\sigma_a^2(\psi_{l+j}^* - \psi_{l+j})^2, \quad (14.4.7)$$

which is minimized when $u_{l+j}^* = u_{l+j}$ and $\psi_{l+j}^* = \psi_{l+j}$. In other words, the minimum mean square error forecast $\hat{Y}_t(l)$ of Y_{t+l} at time origin t is given by the conditional expectation of Y_{t+l} at time t. Because $E(Y_{t+l} - \hat{Y}_t(l)) = 0$, the forecast is unbiased and the variance of the forecast is given by

$$V(l) = E(Y_{t+l} - \hat{Y}_t(l))^2 = \sigma_a^2 \sum_{j=0}^{l-1} u_j^2 + \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2. \quad (14.4.8)$$

14.4.2 Minimum Mean Square Error Forecasts for Nonstationary Input and Output Series

By the same argument used in Section 5.2.2 for a general ARIMA model, the minimum mean square error forecast of Y_{t+l} at time t is also given by the conditional expectation of Y_{t+l} at time t even if Y_t and X_t are nonstationary. That is true as long as Y_t and X_t can be reduced to stationary series by some proper transformations and differencing so that they follow some ARIMA(p, d, q) models. Likewise, Equation (14.4.8) holds also for the nonstationary case, and the weights u_j and ψ_j in (14.4.8) can be derived by the same method as Equation (5.2.26) used to compute the ψ_j weights for the ARIMA model. Specifically, suppose that Y_t and X_t are nonstationary but that $(1 - B)^d Y_t$ and $(1 - B)^d X_t$ are stationary and are related in the stable transfer function model

$$(1 - B)^d Y_t = \frac{\omega(B)}{\delta(B)} B^b (1 - B)^d X_t + \frac{\theta(B)}{\phi(B)} a_t \quad (14.4.9)$$

and

$$\phi_x(B)(1 - B)^d X_t = \theta_x(B)\alpha_t, \quad (14.4.10)$$

where $\omega(B)$, $\delta(B)$, $\theta(B)$, $\phi_x(B)$, $\theta_x(B)$, a_t, and α_t are defined as in (14.4.1) and (14.4.2). From (14.4.9), we have

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\phi(B)(1 - B)^d} a_t, \quad (14.4.11)$$

and Y_t is the sum of two independent processes. The required u_j weights come from the first term $(\omega(B)/\delta(B))B^b X_t$, and the ψ_j weights come from the second term $(\theta(B)/\phi(B)(1 - B)^d)a_t$.

To derive the ψ_j weights, we let

$$e_t = \frac{\theta(B)}{\phi(B)(1 - B)^d} a_t \tag{14.4.12}$$

and express e_t in an AR representation that exists because the roots of $\theta(B) = 0$ are outside the unit circle. Thus,

$$\pi^{(a)}(B) e_t = a_t, \tag{14.4.13}$$

where

$$\pi^{(a)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(a)} B^j = \frac{\phi(B)(1 - B)^d}{\theta(B)}. \tag{14.4.14}$$

By the results from (5.2.15) through (5.2.26), the ψ_j weights are calculated recursively from the $\pi_j^{(a)}$ weights in (14.4.14) as follows:

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}^{(a)} \psi_i, \qquad j = 1, \ldots, l - 1, \tag{14.4.15}$$

and $\psi_0 = 1$.

To derive the u_j weights, we first express X_t in (14.4.10) in terms of its AR representation that again exists because the roots of $\theta_x(B) = 0$ are assumed to be outside of the unit circle. Thus,

$$\pi^{(\alpha)}(B) X_t = \alpha_t, \tag{14.4.16}$$

where

$$\pi^{(\alpha)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(\alpha)} B^j = \frac{\phi_x(B)(1 - B)^d}{\theta_x(B)}. \tag{14.4.17}$$

Next, we use the result (5.2.26) to compute the following $\psi_j^{(\alpha)}$ weights:

$$\psi_j^{(\alpha)} = \sum_{i=0}^{j-1} \pi_{j-1}^{(\alpha)} \psi_i^{(\alpha)}, \qquad j = 1, \ldots, l - 1 \tag{14.4.18}$$

and $\psi_0^{(\alpha)} = 1$. The required u_j weights for $j = 0, 1, \ldots, (l - 1)$, are thus equal to the coefficient of B^j in the following expression that exists because $\omega(B)$ is a finite-order polynomial in B and the roots of $\delta(B) = 0$ are outside of the unit circle:

$$\frac{\omega(B)B^b}{\delta(B)}(1 + \psi_1^{(\alpha)} B + \cdots + \psi_{l-1}^{(\alpha)} B^{l-1}). \tag{14.4.19}$$

It can be easily seen that if $b \neq 0$, then $u_j = 0$ for $j < b$; if $b = 0$, then $u_0 \neq 1$ unless $\omega_0 = 1$ in $\omega(B) = (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)$.

For the actual computation of forecasts, we consider the general transfer function model (14.4.9) and rewrite it in the difference equation form

$$c(B)Y_t = d(B)X_{t-b} + e(B)a_t, \tag{14.4.20}$$

where

$$\begin{aligned} c(B) &= \delta(B)\phi(B)(1-B)^d \\ &= 1 - c_1B - c_2B^2 - \cdots - c_{p+r+d}B^{p+r+d}, \\ d(B) &= \omega(B)\phi(B)(1-B)^d \\ &= d_0 - d_1B^1 - d_2B^2 - \cdots - d_{p+s+d}B^{p+s+d}, \end{aligned}$$

and

$$\begin{aligned} e(B) &= \delta(B)\theta(B) \\ &= 1 - e_1B - \cdots - e_{q+r}B^{q+r}. \end{aligned}$$

Thus,

$$\begin{aligned} \hat{Y}_t(l) &= E_t(Y_{t+l}) \\ &= c_1\hat{Y}_t(l-1) + \cdots + c_{p+r+d}\hat{Y}_t(l-p-r-d) \\ &\quad + d_0\hat{X}_t(l-b) - \cdots - d_{p+s+d}\hat{X}_t(l-b-p-s-d) \\ &\quad + \hat{a}_t(l) - e_1\hat{a}_t(l-1) - \cdots - e_{q+r}\hat{a}_t(l-q-r), \end{aligned} \tag{14.4.21}$$

where

$$\begin{aligned} \hat{Y}_t(j) &= \begin{cases} Y_{t+j}, & j \le 0, \\ \hat{Y}_t(j), & j > 0, \end{cases} \\ \hat{X}_t(j) &= \begin{cases} X_{t+j}, & j \le 0, \\ \hat{X}_t(j), & j > 0, \end{cases} \\ \hat{a}_t(j) &= \begin{cases} \hat{a}_{t+j}, & j \le 0, \\ 0, & j > 0, \end{cases} \end{aligned} \tag{14.4.22}$$

and $\hat{a}_t$ is generated either by Equation (14.4.20) or (14.3.19) or as the one-step forecast error $Y_t - \hat{Y}_{t-1}(1)$. It is clear that for $l > (q + r)$, the forecast $\hat{Y}_t(l)$ of Y_{t+l} depends only on the past history of Y and the present and past history of X. The required forecast of the input series can be easily obtained from Equation (14.4.10) as discussed in Section 5.3 for the univariate ARIMA models.

14.4.3 An Example

The numerical computation of the minimum mean square error forecasts for stationary input and output series is simpler than that of nonstationary input and output series. For illustration, we consider the nonstationary case of the Lydia Pinkham example. Although the causal transfer function model has been found inadequate in describing the relationship between the Lydia Pinkham annual advertising budget and sales, the model equations themselves can be used to illustrate the computational aspect of the forecasting procedure described above.

EXAMPLE 14.2 From (14.3.33), we have that

$$(1 - B)Y_t = \frac{.52}{(1 - .36B)}(1 - B)X_t + \frac{1}{(1 - .26B)}a_t \tag{14.4.23}$$

and

$$(1 - .07B + .41B^2)(1 - B)X_t = \alpha_t, \tag{14.4.24}$$

where $\hat{\sigma}_a^2 = 45175$, and $\hat{\sigma}_\alpha^2 = 52601$. To compute the forecast, we rewrite (14.4.23) as

$$(1 - .36B)(1 - .26B)(1 - B)Y_t = (1 - .26B)(.52)(1 - B)X_t + (1 - .36B)a_t.$$

Hence,

$$\begin{aligned} Y_{n+1} = {} & 1.62Y_{n+l-1} - .71Y_{n+l-2} + .09Y_{n+l-3} + .52X_{n+l} \\ & - .66X_{n+l-1} + .14X_{n+l-2} + a_{n+l} - .36a_{n+l-1}. \end{aligned}$$

The one-step ahead forecast from the forecast origin $n = 40$ is

$$\hat{Y}_{40}(1) = 1.62Y_{40} - .71Y_{39} + .09Y_{38} + .52\hat{X}_{40}(1) - .66X_{40} + .14X_{39} - .36a_{40},$$

where $\hat{X}_{40}(1) = X_{41}$ if it is known; otherwise, it equals the one-step ahead forecast of X_{41} determined from its own past through Equation (14.4.24). Now, from the data and model fitting, we have $\hat{a}_{40} = -101.05$, $Y_{38} = 2518$, $Y_{39} = 2637$, $Y_{40} = 2177$, $X_{39} = 1145$, $X_{40} = 1012$, and $X_{41} = 836$. Thus,

$$\begin{aligned} \hat{Y}_{40}(1) = {} & 1.62(2177) - .71(2637) + .09(2518) + .52(836) \\ & - .66(1012) + .14(1145) - .36(-101.05) = 1844.57. \end{aligned}$$

To obtain the forecast variance, we need to compute the weights u_j and ψ_j. Note that Equation (14.4.23) can be written as

$$Y_t = \frac{.52}{(1 - .36B)}X_t + \frac{1}{(1 - .26B)(1 - B)}a_t. \tag{14.4.25}$$

Let

$$e_t = \frac{1}{(1 - .26B)(1 - B)} a_t$$

and

$$\pi^{(a)}(B)e_t = a_t,$$

where

$$\pi^{(a)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(a)} B^j = (1 - .26B)(1 - B)$$
$$= (1 - 1.26B + .26B^2).$$

Thus,

$$\pi_1^{(a)} = 1.26,$$
$$\pi_2^{(a)} = -.26,$$

and

$$\pi_j^{(a)} = 0, \qquad \text{for } j \geq 3.$$

Using (14.4.15), we have the following ψ_j weights:

$$\psi_0 = 1,$$
$$\psi_1 = \pi_1^{(a)} = 1.26,$$
$$\psi_2 = \pi_2^{(a)} + \pi_1^{(a)}\psi_1 = -.26 + (1.26)^2 = 1.33,$$
$$\vdots$$
$$\psi_{l-1} = \sum_{i=0}^{l-2} \pi_{l-1-i}^{(a)} \psi_i.$$

To obtain the u_j weights, we note that

$$\pi^{(\alpha)}(B)X_t = \alpha_t,$$

where

$$\pi^{(\alpha)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(\alpha)} B^j = (1 - .07B + .41B^2)(1 - B)$$
$$= 1 - 1.07B + .48B^2 - .41B^3.$$

Hence,

$$\pi_1^{(\alpha)} = 1.07,$$
$$\pi_2^{(\alpha)} = -.48,$$
$$\pi_3^{(\alpha)} = .41,$$

and

$$\pi_j^{(\alpha)} = 0, \qquad \text{for } j \geq 4.$$

Using (14.4.18), we have

$$\psi_0^{(\alpha)} = 1,$$
$$\psi_1^{(\alpha)} = \pi_1^{(\alpha)} = 1.07,$$
$$\psi_2^{(\alpha)} = \pi_2^{(\alpha)} + \pi_1^{(\alpha)}\psi_1^{(\alpha)} = -.48 + (1.07)^2 = .66,$$
$$\vdots$$

and

$$\psi_{l-1}^{(\alpha)} = \sum_{i=0}^{l-2} \pi_{l-1-i}^{(\alpha)}\psi_i^{(\alpha)}.$$

Thus, by (14.4.19), the u_j weights for $j = 0, 1, \ldots l-1$ are equal to the coefficient of B^j in the following expression:

$$\frac{.52}{(1 - .36B)}(1 + \psi_1^{(\alpha)}B + \psi_2^{(\alpha)}B^2 + \cdots + \psi_{l-1}^{(\alpha)}B^{l-1})$$
$$= .52(1 + .36B + (.36)^2B^2 + (.36)^3B^3 + \cdots)$$
$$\cdot (1 + \psi_1^{(\alpha)}B + \psi_2^{(\alpha)}B^2 + \cdots + \psi_{l-1}^{(\alpha)}B^{l-1}).$$

Thus,

$$u_0 = .52,$$
$$u_1 = .52(.36 + \psi_1^{(\alpha)})$$
$$= .52(.36 + 1.07) = .7436,$$
$$u_2 = .52(\psi_2^{(\alpha)} + .36\psi_1^{(\alpha)} + (.36)^2)$$
$$= .52\{.66 + .36(1.07) + (.36)^2\} = .61,$$
$$\vdots$$

Having obtained the u_j and ψ_j weights, we can now calculate the one-step ahead forecast error variance using (14.4.8) and get

$$V(1) = E[Y_{41} - \hat{Y}_{40}(1)]^2 = \hat{\sigma}_a^2(u_0^2) + \hat{\sigma}_a^2(\psi_0)^2$$
$$= (52{,}601)(.52)^2 + (45{,}175)(1) = 59{,}398.31.$$

The forecasts $\hat{Y}_{40}(l)$ and their forecast error variances $V(l)$ for other l can be calculated similarly. Recall from Section 14.3.5 that in fitting the data set, the last 14 observations (1947–1960) were held back to examine the forecasting performance. For the 14 one-step ahead sales forecasts using the actual one-step ahead advertising value, Helmer and Johansson (1977) reported a mean squared forecasting error of 8912 for Equation (14.3.33) or (14.4.23) and 9155 for (14.3.34), both of which are considerably smaller than the MSE of 16,080 obtained from a univariate time series model of sales. Because of a feedback relationship from sales to advertising at lag 1, it does not seem appropriate to assume that the next period's advertising budget is known with certainty when forecasting the next period's sales. Rather than the actual advertising levels, if one-step ahead advertising levels from Equation (14.4.24) are used, the corresponding MSEs of the 14 one-step ahead sales forecasts are 15,177 for (14.3.33) or (14.4.23) and 14,044 for (14.3.34), respectively.

For more examples and discussion on forecasting using transfer function models, see Liu (1987).

14.5 Bivariate Frequency-Domain Analysis

14.5.1 Cross-Covariance Generating Functions and the Cross-Spectrum

Suppose that x_t and y_t are jointly stationary with the cross-covariance function $\gamma_{xy}(k)$ for $k = 0, \pm 1, \pm 2, \ldots$. Define the cross-covariance generating function as

$$\gamma_{xy}(B) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)B^k, \tag{14.5.1}$$

where the coefficient of B^k is the kth cross-covariance between x_t and y_t. This equation is the generalization of the autocovariance generating function introduced earlier. Specifically, $\gamma_{xx}(B) = \gamma_x(B)$ and $\gamma_{yy}(B) = \gamma_y(B)$ are autocovariance generating functions for x_t and y_t, respectively. If the cross-covariance sequence $\gamma_{xy}(k)$ is absolutely summable, i.e., $\sum_{k=-\infty}^{\infty}|\gamma_{xy}(k)| < \infty$, then its Fourier transform exists and is called the cross-spectrum (or cross-spectral density) between x_t and y_t and is given by

$$f_{xy}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_{xy}(k)e^{-i\omega k} \tag{14.5.2a}$$

$$= \frac{1}{2\pi}\gamma_{xy}(e^{-i\omega}), \tag{14.5.2b}$$

which is the generalization of the spectrum (or spectral density) for the univariate series as introduced in Section 12.2. Specifically, $f_{xx}(\omega) = f_x(\omega)$ and $f_{yy}(\omega) = f_y(\omega)$ are spectrums for x_t and y_t, respectively. Note that even when $\gamma_{xy}(k)$ is real for the real processes x_t and y_t, $f_{xy}(\omega)$ will in general be complex because $\gamma_{xy}(k) \neq \gamma_{xy}(-k)$. Hence, we can write

$$f_{xy}(\omega) = c_{xy}(\omega) - iq_{xy}(\omega), \tag{14.5.3}$$

where $c_{xy}(\omega)$ and $-q_{xy}(\omega)$ are the real and imaginary parts of $f_{xy}(\omega)$. More precisely,

$$c_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) \cos \omega k \tag{14.5.4}$$

and

$$q_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) \sin \omega k. \tag{14.5.5}$$

The function $c_{xy}(\omega)$ is called the cospectrum, and $q_{xy}(\omega)$ is called the quadrature spectrum, of x_t and y_t. These functions, however, are difficult to interpret. Alternatively, we can express $f_{xy}(\omega)$ in the polar form

$$f_{xy}(\omega) = A_{xy}(\omega) e^{i\phi_{xy}(\omega)}, \tag{14.5.6}$$

where

$$A_{xy}(\omega) = |f_{xy}(\omega)| = [c_{xy}^2(\omega) + q_{xy}^2(\omega)]^{1/2} \tag{14.5.7}$$

and

$$\phi_{xy}(\omega) = \tan^{-1}\left[\frac{-q_{xy}(\omega)}{c_{xy}(\omega)}\right]. \tag{14.5.8}$$

The functions $A_{xy}(\omega)$ and $\phi_{xy}(\omega)$ are known as the cross-amplitude spectrum and the phase spectrum, respectively. Two other useful functions are the gain function and the coherence. The gain function is defined as

$$G_{xy}(\omega) = \frac{|f_{xy}(\omega)|}{f_x(\omega)} = \frac{A_{xy}(\omega)}{f_x(\omega)}, \tag{14.5.9}$$

which is the ratio of the cross-amplitude spectrum to the input spectrum. The coherence (or the squared coherency) is defined by

$$K_{xy}^2(\omega) = \frac{|f_{xy}(\omega)|^2}{f_x(\omega) f_y(\omega)}, \tag{14.5.10}$$

which is essentially the standardized cross-amplitude spectrum.

In practice, several cross-spectral functions are needed to describe the relationship between two series in the frequency domain. In general, the cross-spectrum between two series is more easily analyzed through the gain function, the phase spectrum, and the coherence. Hence, they are possibly the three most commonly used cross-spectral functions in frequency domain analysis.

14.5.2 Interpretation of the Cross-Spectral Functions

To see the implications of the above functions associated with the cross-spectrum, we recall from Section 12.1.4 that any stationary zero mean process can be written in terms of the spectral or Cramer's representation. That is,

$$x_t = \int_{-\pi}^{\pi} e^{i\omega t}\, dU_x(\omega), \tag{14.5.11}$$

$$y_t = \int_{-\pi}^{\pi} e^{i\omega t}\, dU_y(\omega), \tag{14.5.12}$$

where $dU_x(\omega)$ and $dU_y(\omega)$ are each orthogonal processes such that for $\omega \neq \lambda$,

$$E[dU_x(\omega)\, dU_x^*(\lambda)] = E[dU_y(\omega)\, dU_y^*(\lambda)] = 0, \tag{14.5.13}$$

where $dU_x^*(\omega)$ and $dU_y^*(\omega)$ are the complex conjugates of $dU_x(\omega)$ and $dU_y(\omega)$, respectively. If x_t and y_t are also jointly stationary, we have

$$\begin{aligned}\gamma_{xy}(k) &= E(x_t y_{t+k}) \\ &= E(x_t y_{t+k}^*) \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i\omega t} e^{-i\lambda(t+k)} E[dU_x(\omega)\, dU_y^*(\lambda)] \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(\omega-\lambda)t} e^{-i\lambda k} E[dU_x(\omega)\, dU_y^*(\lambda)].\end{aligned} \tag{14.5.14}$$

Now, for the right-hand side to be a function of k only, the integrand in (14.5.14) must be zero when $\omega \neq \lambda$. Thus, in addition to (14.5.13), we also require cross-orthogonality, namely,

$$E[dU_x(\omega)\, dU_y^*(\lambda)] = 0, \qquad \text{for } \omega \neq \lambda. \tag{14.5.15}$$

Therefore, letting $\lambda = \omega$ in (14.5.14), we obtain

$$\gamma_{xy}(k) = \int_{-\pi}^{\pi} e^{i\omega k} E[dU_x(\omega)\, dU_y^*(\omega)]. \tag{14.5.16}$$

Assume that x_t and y_t are both purely nondeterministic and jointly stationary zero-mean processes with absolutely summable autocovariance and cross-covariance functions, i.e.,

$$\sum_{k=-\infty}^{\infty} |\gamma_x(k)| < \infty, \qquad \sum_{k=-\infty}^{\infty} |\gamma_y(k)| < \infty, \qquad \text{and} \qquad \sum_{k=-\infty}^{\infty} |\gamma_{xy}(k)| < \infty.$$

By the same argument that was used to derive (12.1.37), we have

$$f_x(\omega)\, d\omega = E\left[|dU_x(\omega)|^2\right], \tag{14.5.17}$$

$$f_y(\omega)\, d\omega = E\left[|dU_y(\omega)|^2\right], \tag{14.5.18}$$

and

$$f_{xy}(\omega)\, d\omega = E[dU_x(\omega)\, dU_y^*(\omega)], \tag{14.5.19}$$

where

$$E[dU_x(\omega)\, dU_x^*(\lambda)] = E[dU_y(\omega)\, dU_y^*(\lambda)] = 0$$

and

$$E[dU_x(\omega)\, dU_y^*(\lambda)] = 0, \qquad \text{for } \omega \neq \lambda.$$

Now, $E[dU_x(\omega)\, dU_y^*(\omega)] = \text{Cov}(dU_x(\omega), dU_y(\omega))$. Thus, from (14.5.19) we see that the cross-amplitude spectrum, $A_{xy}(\omega) = |f_{xy}(\omega)|$, measures the covariance between the ω-frequency component of x_t and the ω-frequency component of y_t. Also,

$$\begin{aligned} G_{xy}(\omega) &= \frac{|f_{xy}(\omega)|}{f_x(\omega)} = \frac{|E[dU_x(\omega)\, dU_y^*(\omega)]|}{E\left[|dU_x(\omega)|^2\right]} \\ &= \frac{|\text{Cov}(dU_x(\omega), dU_y(\omega))|}{\text{Var}[dU_x(\omega)]}, \end{aligned} \tag{14.5.20}$$

which implies that the gain function $G_{xy}(\omega)$ is simply the absolute value of the standard least squares regression coefficient of the ω-frequency of x_t. Similarly, we can write

$$\begin{aligned} K_{xy}^2(\omega) &= \frac{[f_{xy}(\omega)]^2}{f_x(\omega) f_y(\omega)} \\ &= \frac{\{\text{Cov}(dU_x(\omega), dU_y(\omega))\}^2}{\text{Var}[dU_x(\omega)]\, \text{Var}[dU_y(\omega)]}. \end{aligned} \tag{14.5.21}$$

Hence, $K_{xy}^2(\omega)$ is square of the correlation coefficient between the ω-frequency component of x_t and the ω-frequency component of y_t. Clearly, $0 \leq K_{xy}^2(\omega) \leq 1$. A value of $K_{xy}^2(\omega)$ close to one implies that the ω-frequency components of the two series are highly linearly related, and a value of $K_{xy}^2(\omega)$ near zero implies that they are only slightly linearly related.

Like the correlation coefficient between two random variables, the coherence is invariant under linear transformations. To see, suppose that

$$w_t = \sum_{k=-\infty}^{\infty} \alpha_k x_{t-k} = \alpha(B)x_t \tag{14.5.22}$$

and

$$v_t = \sum_{k=-\infty}^{\infty} \beta_k y_{t-k} = \beta(B)y_t. \tag{14.5.23}$$

Then, by (12.3.3), we have

$$f_w(\omega) = |\alpha(e^{-i\omega})|^2 f_x(\omega)$$
$$f_v(\omega) = |\beta(e^{-i\omega})|^2 f_y(\omega)$$

or, equivalently, by (14.5.17) and (14.5.18),

$$dU_w(\omega) = \alpha(e^{-i\omega})\, dU_x(\omega) \tag{14.5.24}$$

and

$$dU_v(\omega) = \beta(e^{-i\omega})\, dU_y(\omega). \tag{14.5.25}$$

Hence,

$$\begin{aligned} f_{wv}(\omega)\, d\omega &= E[dU_w(\omega)\, dU_v^*(\omega)] \\ &= \alpha(e^{-i\omega})\beta(e^{i\omega})E[dU_x(\omega)\, dU_y^*(\omega)] \\ &= \alpha(e^{-i\omega})\beta(e^{i\omega})f_{xy}(\omega)\, d\omega \end{aligned}$$

or

$$f_{wv}(\omega) = \alpha(e^{-i\omega})\beta(e^{i\omega})f_{xy}(\omega). \tag{14.5.26}$$

It follows that when $\alpha(e^{i\omega})$ and $\beta(e^{i\omega})$ are nonzero,

$$K_{wv}^2(\omega) = \frac{|f_{wv}(\omega)|^2}{f_w(\omega)f_v(\omega)} = \frac{|f_{xy}(\omega)|^2}{f_x(\omega)f_y(\omega)};$$

hence, the coherence is invariant under linear transformation.

To learn the interpretation of the phase spectrum, we recall from Section 12.1.4 that for any ω, $dU_x(\omega)$ and $dU_y(\omega)$ are complex valued random variables. Hence, we can write

$$dU_x(\omega) = A_x(\omega)e^{i\phi_x(\omega)}, \tag{14.5.27}$$

where $A_x(\omega)$ and $\phi_x(\omega)$ are the amplitude spectrum and the phase spectrum of the x_t series. Similarly,

$$dU_y(\omega) = A_y(\omega)e^{i\phi_y(\omega)}, \tag{14.5.28}$$

and $A_y(\omega)$ and $\phi_y(\omega)$ are the amplitude spectrum and the phase spectrum of the y_t series. Thus,

$$\begin{aligned} A_{xy}(\omega)e^{i\phi_{xy}(\omega)} &= f_{xy}(\omega)\, d\omega = E[dU_x(\omega)\, dU_y^*(\omega)] \\ &= E[A_x(\omega)A_y(\omega)]E\left[e^{i[\phi_x(\omega)-\phi_y(\omega)]}\right], \end{aligned} \tag{14.5.29}$$

where for simplicity we assume that the amplitude spectrum and the phase spectrum are independent. Consequently, $A_{xy}(\omega)$ can also be thought as the average value of the product of amplitudes of the ω-frequency components of x_t and y_t, and the phase spectrum $\phi_{xy}(\omega)$ represents the average phase shift, $[\phi_x(\omega) - \phi_y(\omega)]$, between the ω-frequency components of x_t and y_t. In terms of a causal model $y_t = \alpha x_{t-\tau} + e_t$ or $x_t = \beta y_{t-r} + a_t$, where there is no feedback relationship between x_t and y_t, the phase spectrum is a measure of the extent to which each frequency component of one series leads the other. The ω-frequency component of x_t leads the ω-frequency component of y_t if the phase $\phi_{xy}(\omega)$ is negative. The ω-frequency component of x_t lags the ω-frequency component of y_t if the phase $\phi_{xy}(\omega)$ is positive. For a given $\phi_{xy}(\omega)$, the shift in time units is $\phi_{xy}(\omega)/\omega$. Hence, the actual time delay of the component of y_t at frequency ω is equal to

$$\tau = -\frac{\phi_{xy}(\omega)}{\omega}, \tag{14.5.30}$$

which is not necessarily an integer.

From (14.5.6) and (14.5.8), it is clear that if $\phi'_{xy}(\omega)$ is a solution to (14.5.8), then $\phi_{xy}(\omega) = \phi'_{xy}(\omega) \pm 2\pi k$ for any integer k is also a solution of (14.5.8). Thus, the phase spectrum is defined only mod(2π). In practice, however, if we take the given function to be nonnegative, the range of $\phi_{xy}(\omega)$ is conventionally restricted to the interval from $-\pi$ to π. The value of $\phi_{xy}(\omega)$, which is the argument of $f_{xy}(\omega)$, i.e.,

$$\phi_{xy}(\omega) = \arg[f_{xy}(\omega)], \tag{14.5.31}$$

is then the angle between the positive half of the $c_{xy}(\omega)$ axis and the line joining the origin to the point $(c_{xy}(\omega), -q_{xy}(\omega))$. It follows that $\phi_{xy}(\omega)$ and $-q_{xy}(\omega)$ have the same sign. Furthermore, because $\phi_{xy}(\omega)$ is an odd symmetric function that is zero at $\omega = 0$, the phase spectrum is normally graphed only for ω over the range $0 \leq \omega \leq \pi$.

From (14.5.15) the components of x_t and y_t are independent at different frequencies. Thus, we need only examine the relationship between the corresponding components of x_t and y_t at each frequency as emphasized above in the phrase such as "the ω-frequency components of x_t and y_t."

14.5.3 Examples

EXAMPLE 14.3 Assume that x_t and y_t are jointly stationary and uncorrelated processes. Then $\gamma_{xy}(k) = 0$ for all k. Hence, by (14.5.2a), we have

$$f_{xy}(\omega) = 0, \qquad \text{for all } \omega,$$

which implies that $c_{xy}(\omega) = 0$, $q_{xy}(\omega) = 0$, $A_{xy}(\omega) = 0$, $\phi_{xy}(\omega) = 0$, $G_{xy}(\omega) = 0$, and $K^2_{xy}(\omega) = 0$ for all ω.

EXAMPLE 14.4 Consider the model

$$y_t = \alpha x_{t-m} + e_t, \tag{14.5.32}$$

where we assume that $m > 0$ and x_t and e_t are jointly independent zero mean stationary processes. Then

$$\begin{aligned}\gamma_{xy}(k) &= E(x_t y_{t+k}) \\ &= E[x_t(\alpha x_{t+k-m} + e_{t+k})] \\ &= \alpha\gamma_x(k - m).\end{aligned}$$

Hence, by (14.5.2a),

$$\begin{aligned}f_{xy}(\omega) &= \frac{\alpha}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_x(k - m)e^{-i\omega k} \\ &= \frac{\alpha}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_x(j)e^{-i\omega(j+m)} \\ &= \alpha e^{-i\omega m}f_x(\omega).\end{aligned} \tag{14.5.33}$$

Equation (14.5.33) implies that

$$\begin{aligned}c_{xy}(\omega) &= [\alpha\cos(\omega m)]f_x(\omega), \\ q_{xy}(\omega) &= [\alpha\sin(\omega m)]f_x(\omega), \\ A_{xy}(\omega) &= |\alpha|f_x(\omega), \\ \phi_{xy}(\omega) &= \tan^{-1}\left[\frac{-q_{xy}(\omega)}{c_{xy}(\omega)}\right] = -\omega m.\end{aligned}$$

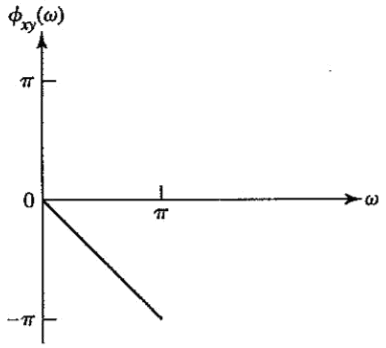


FIGURE 14.5 The phase spectrum for Example 14.4, with $m = 1$.

This phase spectrum shown in Figure 14.5 is a straight line with slope $(-m)$ and is worthy of further discussion. Because the phase function is negative for all ω, the x_t series leads the y_t series at all frequencies. By Equation (14.5.30), the actual time delay of the y_t series equals m at all frequencies. In general, the phase spectrum between two series may not be a straight line. Thus, the time delay between two series may change from frequency to frequency.

In Figure 14.5, if m instead equaled 4, then following the convention of restricting the phase to the interval $(-\pi, \pi)$ we would have a graph like Figure 14.6. In this case, the slope of each line is $(-m)$.

The gain function for this example is given by

$$G_{xy}(\omega) = \frac{A_{xy}(\omega)}{f_x(\omega)} = |\alpha|,$$

which equals, as expected, the absolute value of the least squares regression coefficient. For the coherence, we also need $f_y(\omega)$, which, by (12.2.22) and (12.3.3), equals

$$f_y(\omega) = \alpha^2 f_x(\omega) + f_e(\omega).$$

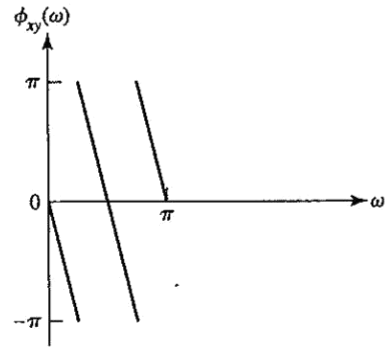


FIGURE 14.6 The phase spectrum for Example 14.4, with $m = 4$.

Hence,

$$K_{xy}^2(\omega) = \frac{|f_{xy}(\omega)|^2}{f_x(\omega)f_y(\omega)}$$
$$= \frac{[\alpha e^{i\omega m} f_x(\omega)][\alpha e^{-i\omega m} f_x(\omega)]}{f_x(\omega)[\alpha^2 f_x(\omega) + f_e(\omega)]}$$
$$= \left[1 + \frac{f_e(\omega)}{\alpha^2 f_x(\omega)}\right]^{-1}. \tag{14.5.34}$$

If $e_t \equiv 0$, then $K_{xy}^2(\omega) = 1$ for all frequencies ω. This result is also expected because, in this case, x_t and y_t satisfy the exact linear relationship $y_t = \alpha x_{t-m}$.

14.5.4 Estimation of the Cross-Spectrum

For a given set of time series data x_t and y_t, for $t = 1, 2, \ldots,$ and n, we have learned from Chapter 13 that the spectrum $f_x(\omega)$ for the x_t series can be estimated by the smoothed spectrum

$$\hat{f}_x(\omega) = \frac{1}{2\pi} \sum_{k=-M_x}^{M_x} W_x(k)\hat{\gamma}_x(k)e^{-i\omega k}, \tag{14.5.35}$$

where $\hat{\gamma}_x(k)$ is the sample autocovariance function, $W_x(k)$ is the lag window, and M_x is the truncation point. With similar notation, the smoothed spectrum for the y_t series is given by

$$\hat{f}_y(\omega) = \frac{1}{2\pi} \sum_{k=-M_y}^{M_y} W_y(k)\hat{\gamma}_y(k)e^{-i\omega k}. \tag{14.5.36}$$

Similarly, we can estimate the cross-spectrum by

$$\hat{f}_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-M_{xy}}^{M_{xy}} W_{xy}(k)\hat{\gamma}_{xy}(k)e^{-i\omega k}, \tag{14.5.37}$$

where $\hat{\gamma}_{xy}(k)$ is the sample cross-covariance function and $W_{xy}(k)$ and M_{xy} are the corresponding lag window and the truncation point for the smoothing. Clearly, we can write $\hat{f}_{xy}(\omega)$ as

$$\hat{f}_{xy}(\omega) = \hat{c}_{xy}(\omega) - i\hat{q}_{xy}(\omega), \tag{14.5.38}$$

where

$$\hat{c}_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-M_{xy}}^{M_{xy}} W_{xy}(k)\hat{\gamma}_{xy}(k) \cos \omega k$$
$$= \frac{1}{2\pi}\left\{W_{xy}(0)\hat{\gamma}_{xy}(0) + \sum_{k=1}^{M_{xy}} W_{xy}(k)[\hat{\gamma}_{xy}(k) + \hat{\gamma}_{xy}(-k)] \cos \omega k\right\} \tag{14.5.39}$$

and

$$\hat{q}_{xy}(\omega) = \frac{1}{2\pi}\sum_{k=-M_{xy}}^{M_{xy}} W_{xy}\hat{\gamma}_{xy}(k)\sin\omega k$$

$$= \frac{1}{2\pi}\left\{\sum_{k=1}^{M_{xy}} W_{xy}(k)\left[\hat{\gamma}_{xy}(k) - \hat{\gamma}_{xy}(-k)\right]\sin\omega k\right\} \tag{14.5.40}$$

are the estimated cospectrum and quadrature spectrum, respectively.

The truncation point M_{xy} and the lag window $W_{xy}(k)$ are chosen in a similar way to those used in the univariate spectral analysis. Rigorously, the truncation points M_x, M_y, and M_{xy} and the lag windows $W_x(k)$, $W_y(k)$, and $W_{xy}(k)$ do not necessarily have to be the same and should be allowed to vary depending on the characteristics of the univariate as well as the joint series. In practice, however, because of the simplification of the sampling properties of the spectrum estimates, they are often chosen to be the same. That is especially true for the lag window.

Having estimated the cospectrum and quadrature spectrum, we may estimate the cross-amplitude spectrum, the phase spectrum, the gain function, and the coherence as follows:

$$\hat{A}_{xy}(\omega) = \left[\hat{c}_{xy}^2(\omega) + \hat{q}_{xy}^2(\omega)\right]^{1/2}, \tag{14.5.41}$$

$$\hat{\phi}_{xy}(\omega) = \tan^{-1}\left[\frac{-\hat{q}_{xy}(\omega)}{\hat{c}_{xy}(\omega)}\right], \tag{14.5.42}$$

$$\hat{G}_{xy}(\omega) = \frac{\hat{A}_{xy}(\omega)}{\hat{f}_x(\omega)}, \tag{14.5.43}$$

and

$$K_{xy}^2(\omega) = \frac{|\hat{f}_{xy}(\omega)|^2}{\hat{f}_x(\omega)\hat{f}_y(\omega)}. \tag{14.5.44}$$

In the estimated cross-spectrum, the lag window places heavier weight on the sample cross-covariances around the zero lag. Thus, if one series leads another and the cross-covariance does not peak at zero lag, then the estimated cross-spectrum will be biased. This bias is especially severe for the estimated coherence. Because the coherence is invariant under the linear transformation, the bias can be reduced by the proper alignment of the series. That is, two series can be aligned by shifting one of the series by a time lag τ so that the peak in the cross-correlation function of the aligned series occurs at the zero lag. The time lag τ can be estimated by the lag corresponding to the maximum cross-correlation or by Equation (14.5.30) and the slope of the phase spectrum. It is also worth noting that the estimates of cross-amplitude, phase, and gain are unreliable when the coherence is small.

We do not discuss the sampling properties of cross-spectral estimates. Interested readers are referred to the books by Jenkins and Watts (1968), Hannan (1970), Brillinger (1975), Priestley (1981), and others. Instead, we discuss the relationship between the cross-spectrum and the transfer function model in the next section.

14.6 The Cross-Spectrum and Transfer Function Models

14.6.1 Construction of Transfer Function Models through Cross-Spectrum Analysis

Consider the two-sided general transfer function model

$$y_t = \sum_{j=-\infty}^{\infty} \nu_j x_{t-j} + n_t \tag{14.6.1a}$$

$$= \nu(B)x_t + n_t, \tag{14.6.1b}$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$ such that $\sum_{j=-\infty}^{\infty} |\nu_j| < \infty$. The series x_t and n_t are assumed to be independent and jointly stationary zero mean processes. Then,

$$\begin{aligned}\gamma_{xy}(k) &= E(x_t y_{t+k}) \\ &= E\left[x_t\left(\sum_{j=-\infty}^{\infty} \nu_j x_{t+k-j} + n_{t+k}\right)\right] \\ &= \sum_{j=-\infty}^{\infty} \nu_j \gamma_x(k-j) \\ &= \nu(B)\gamma_x(k).\end{aligned} \tag{14.6.2}$$

Multiplying both sides of (14.6.2) by B^k and summing, we get

$$\gamma_{xy}(B) = \nu(B)\gamma_x(B). \tag{14.6.3}$$

By (14.5.2b), we have

$$f_{xy}(\omega) = \nu(\omega) f_x(\omega). \tag{14.6.4}$$

Hence,

$$\nu(\omega) = \frac{f_{xy}(\omega)}{f_x(\omega)}, \qquad -\pi \le \omega \le \pi, \tag{14.6.5}$$

for all ω such that $f_x(\omega) \neq 0$. The function $\nu(\omega)$ is called the frequency response function of the linear (transfer function) system. Equation (14.6.5) shows that the frequency response function is the ratio of the cross-spectrum to the input spectrum. The impulse response function ν_k can be obtained from

$$\nu_k = \int_{-\pi}^{\pi} \nu(\omega) e^{i\omega k}\, d\omega. \tag{14.6.6}$$

To derive the noise spectrum (or error spectrum) $f_n(\omega)$, we again use the results from Equations (12.2.22) and (12.3.3) and obtain

$$f_y(\omega) = |\nu(\omega)|^2 f_x(\omega) + f_n(\omega). \tag{14.6.7}$$

Hence,

$$\begin{aligned} f_n(\omega) &= f_y(\omega) - |\nu(\omega)|^2 f_x(\omega) \\ &= f_y(\omega) - \frac{|f_{xy}(\omega)|^2}{f_x^2(\omega)} f_x(\omega) \\ &= f_y(\omega) - \frac{f_{xy}^*(\omega) f_{xy}(\omega)}{f_x(\omega)} \qquad (14.6.8a) \\ &= f_y(\omega) - f_{yx}(\omega) f_x^{-1}(\omega) f_{xy}(\omega), \qquad (14.6.8b) \end{aligned}$$

where we note, from (14.5.19), that $f_{xy}^*(\omega) = f_{yx}(\omega)$. The noise autocovariance function $\gamma_n(k)$ can then be obtained from

$$\gamma_n(k) = \int_{-\pi}^{\pi} f_n(\omega) e^{i\omega k} \, d\omega. \tag{14.6.9}$$

By substituting the estimates of $f_x(\omega)$, $f_y(\omega)$, and $f_{xy}(\omega)$ as described in Chapter 13 and Section 14.5.4, we obtain estimates of the impulse response function ν_k and the noise autocorrelation function $\rho_n(k)$. These estimates can then be used to identify the transfer function-noise model as discussed in Section 14.3. In other words, the transfer function model can also be constructed through the use of cross-spectrum analysis.

14.6.2 Cross-Spectral Functions of Transfer Function Models

The gain function between the input x_t and the output y_t can be easily seen from (14.5.9) and (14.6.5) as

$$G_{xy}(\omega) = |\nu(\omega)|. \tag{14.6.10}$$

Similarly, because $f_x(\omega)$ is nonnegative and real-valued, we have, from (14.5.31) and (14.6.5), that

$$\phi_{xy}(\omega) = \arg[f_{xy}(\omega)] = \arg[\nu(\omega)]. \tag{14.6.11}$$

Hence, the phase spectrum between the input series x_t and the output series y_t depends only on the phase shift of the frequency response function $\nu(\omega)$.

For the coherence, we note, from (14.6.4) and (14.6.7), that

$$\begin{aligned} K_{xy}^2(\omega) &= \frac{|f_{xy}(\omega)|^2}{f_x(\omega)f_y(\omega)} \\ &= \frac{|\nu(\omega)|^2 f_x(\omega)}{f_y(\omega)} \\ &= \frac{f_y(\omega) - f_n(\omega)}{f_y(\omega)} \\ &= 1 - \frac{f_n(\omega)}{f_y(\omega)}. \end{aligned} \tag{14.6.12}$$

Thus, the coherence is a measure of the ratio of the noise spectrum to the output spectrum at frequency ω. The coherence $K_{xy}^2(\omega)$ approaches unity as the ratio $f_n(\omega)/f_y(\omega)$ approaches zero. Conversely, the coherence approaches zero as the ratio $f_n(\omega)/f_y(\omega)$ approaches unity. From (14.6.7), because the output equals the buried signal plus the noise, we can form the noise-to-signal ratio

$$\frac{f_n(\omega)}{f_s(\omega)} = \frac{1 - K_{xy}^2(\omega)}{K_{xy}^2(\omega)}, \tag{14.6.13}$$

where $f_s(\omega) = |\nu(\omega)|^2 f_x(\omega)$ represents the signal of the system. The noise-to-signal ratio is large when the coherence is small, and vice versa. Equivalently, it can be shown that

$$K_{xy}^2(\omega) = \frac{f_s(\omega)/f_n(\omega)}{1 + f_s(\omega)/f_n(\omega)}. \tag{14.6.14}$$

Hence, the coherence is also a direct measure of the signal-to-noise ratio $f_s(\omega)/f_n(\omega)$ at frequency ω. The larger the coherence, the higher the signal-to-noise ratio.

The noise-to-signal ratio or, equivalently, the signal-to-noise ratio can be used as a criterion for model selection. Other things being equal, a model with a uniformly smaller noise-to-signal ratio and hence a larger signal-to-noise ratio is obviously more preferable.

14.7 Multiple-Input Transfer Function Models

More generally, the output series may be influenced by multiple input series, and we have the multiple-input causal model

$$y_t = \sum_{j=1}^{k} \nu_j(B)x_{jt} + n_t \tag{14.7.1}$$

or

$$y_t = \sum_{j=1}^{k} \frac{\omega_j(B)}{\delta_j(B)} B^{b_j} x_{jt} + \frac{\theta(B)}{\phi(B)} a_t, \tag{14.7.2}$$

where $\nu_j(B) = \omega_j(B)B^{b_j}/\delta_j(B)$ is the transfer function for the jth input series x_{jt} and a_t are assumed to be independent of each of the input series $x_{jt}, j = 1, 2, \ldots, k$. If the input series x_{it} and x_{jt} are uncorrelated for $i \neq j$, then all aspects of the time domain and frequency domain analysis of the single-input transfer function model discussed above can be extended and applied with no difficulty. For example, we can construct the transfer function $\nu_j(B)$ relating y_t and x_{jt} separately, one at a time. The noise model is then identified through the study of the generated noise series

$$\begin{aligned} n_t &= y_t - \hat{y}_t \\ &= y_t - \sum_{j=1}^{k} \hat{\nu}_j(B) x_{jt}. \end{aligned} \tag{14.7.3}$$

Again, if the original series are nonstationary, some transformations must be taken to produce stationarity.

The model building is much more complicated when the input series x_{it} and x_{jt} for $i \neq j$ are correlated. Liu and Hanssen (1982) suggested that a common filter be chosen from the AR polynomials of the input series with roots close to one to reduce their cross-correlations. We believe, however, that much more research is still needed in the model building of multiple-input transfer function models when input series are cross-correlated.

The intervention and outlier models introduced in Chapter 10 are special cases of a transfer function model. In other words, some input series x_{jt} could be intervention or outlier variables. As an example, we refer readers to Pack (1979), who fitted the Lydia Pinkham monthly data to a transfer function model and incorporated two price increases as isolated intervening events.

We close this chapter with the following remarks.

1. The reader may notice the close relationship between the transfer function models and the multiple regression models. In fact, one may think that the multiple regression models are special cases of the transfer function models because the noise processes in the regression models are often assumed to be uncorrelated white noise. The relationship is even more apparent if a transfer function model is written in terms of the difference equation form. There is, however, a vital difference between the transfer function and regression models in terms of modeling. The construction of both the transfer function and the noise models is important in the process. Thus, although the transfer function is linear, the resulting model is, in general, highly nonlinear in terms of its parameters, unlike the regression modeling, where linear models are usually the desired end products. In other words, linearity in parameter is a normal case with regression models but a special case with transfer function models.
2. One important element in a time series model is the time unit used in the model. To obtain proper and valid statistical inferences, the choice of an appropriate time unit is

very important. If the model for a phenomenon under investigation is regraded appropriate in terms of a certain time unit, say a month, then proper inferences about the underlying model should be drawn in terms of this same monthly time unit. Improper use of data in some larger time unit such as quarterly or annual quantities could lead to a substantial information loss in parameter estimation and forecasting. More seriously, it could cause a misspecification of lag structure in causality study. In fact, Tiao and Wei (1976) and Wei (1981) show that a causal relationship is often destroyed by temporal aggregation. This problem is addressed in more detail later in Chapter 20.

EXERCISES

14.1 Calculate and plot the impulse response weights for the following transfer functions:
(a) $v(B) = B/(1 - .8B)$,
(b) $v(B) = (3 + 2B)/(1 + .6B)$,
(c) $v(B) = (B + B^2)/(1 - .8B + .6B^2)$.

14.2 Consider the ARMA(1, 1) process

$$(1 - .4B)Z_t = (1 + .8B)a_t,$$

where a_t is a white noise series with zero mean and constant variance 1. Calculate the cross-correlation function between a_t and Z_t.

14.3 Suppose that $y_t = (2 + B)x_t + n_t$, where x_t and n_t are uncorrelated white noise processes, each with mean 0 and variance 3. Find the CCF between x_t and y_t.

14.4 Consider the transfer function model

$$y_t = \frac{\omega_0 - \omega_1 B}{(1 - \delta B)} x_{t-1} + (1 - \theta B)a_t$$

and $(1 - \phi B)x_t = \alpha_t$, where $\omega_0 = 1$, $\omega_1 = -2$, $\delta = .5$, $\phi = .4$, $\theta = .2$, $\sigma_a^2 = .53$, $\sigma_a^2 = .35$, $a_{99} = a_{100} = 0$, and the last few observations on x_t and y_t are as follows:

t	x_t	y_t
98	1.2	10.4
99	1.6	10.0
100	2.0	9.5

Calculate the forecasts $\hat{y}_{100}(l)$ and their forecast error variances for $l = 1, 2$, and 3.

14.5 Consider the following monthly housing sales X_t and housing starts Y_t between January 1965 and December 1975:

Housing starts Y_t (read across) in thousands of units.

52.149	47.205	82.150	100.931	98.408	97.351	96.489	88.830
80.876	85.750	72.351	61.198	46.561	50.361	83.236	94.343
84.748	79.828	69.068	69.362	59.404	53.530	50.212	37.972
40.157	40.274	66.592	79.839	87.341	87.594	82.344	83.712
78.194	81.704	69.088	47.026	45.234	55.431	79.325	97.983
86.806	81.424	86.398	82.522	80.078	85.560	64.819	53.847
51.300	47.909	71.941	84.982	91.301	82.741	73.523	69.465
71.504	68.039	55.069	42.827	33.363	41.367	61.879	73.835
74.848	83.007	75.461	77.291	75.961	79.393	67.443	69.041
54.856	58.287	91.584	116.013	115.627	116.946	107.747	111.663
102.149	102.882	92.904	80.362	76.185	76.306	111.358	119.840
135.167	131.870	119.078	131.324	120.491	116.990	97.428	73.195
77.105	73.560	105.136	120.453	131.643	114.822	114.746	106.806
84.504	86.004	70.488	46.767	43.292	57.593	56.946	102.237
96.340	99.318	90.715	79.782	73.443	69.460	57.898	41.041
39.791	39.959	62.498	77.777	92.782	90.284	92.782	90.655
84.517	93.826	71.646	55.650				

Housing sales X_t (read across) in thousands of units.

38	44	53	49	54	57	51	58	48	44	42	37
42	43	53	49	49	40	40	36	29	31	26	23
29	32	41	44	49	47	46	47	43	45	34	31
35	43	46	46	43	41	44	47	41	40	32	32
34	40	43	42	43	44	39	40	33	32	31	28
34	29	36	42	43	44	44	48	45	44	40	37
45	49	62	62	58	59	64	62	50	52	50	44
51	56	60	65	64	63	63	72	61	65	51	47
54	58	66	63	64	60	53	52	44	40	36	28
36	42	53	53	55	48	47	43	39	33	30	23
29	33	44	54	56	51	51	53	45	45	44	38

(a) Plot and examine the two series in the same graph.

(b) Prewhiten the X_t series and filter the Y_t series.

(c) Compute the sample CCF (for lags from -24 to 24) between the prewhitened and filtered series. Plot them together with ± 2 standard errors.

(d) Identify a tentatively transfer function–noise model for the data set.

(e) Estimate the parameters of your tentative model and perform diagnostic checking for model adequacy.

(f) Forecast housing starts and attach their 95% forecasts limits for the next 12 months.

14.6 Suppose that $y_t = x_t + .8x_{t-1} + e_t$ and $x_t = a_t$, where a_t and e_t are independent white noise processes, each with mean zero and variance 5. For the above model, find

(a) The cross-covariance function.

(b) The cross-spectrum.

(c) The amplitude.

(d) The phase.
(e) The gain function.
(f) The coherence.

14.7 Given

$$y_t = (x_t + x_{t-1} + x_{t-2} + x_{t-3}) + n_t,$$

where x_t and n_t are assumed to be independent and jointly stationary zero mean processes, find the frequency response function, gain function, and phase function.

14.8 For the data set given in Exercise 14.5, calculate, plot, and discuss
(a) The cospectrum.
(b) The phase spectrum.
(c) The gain function.
(d) The coherence.

15

Time Series Regression and Garch Models

Regression models are without a doubt the most commonly used statistical models. The standard assumptions associated with these models, however, are often violated when time series data are used. In this chapter, we illustrate the use of some time series techniques in regression analysis. In the process, we will introduce the autoregressive conditional heteroscedasticity (ARCH) and generalized autoregressive conditional heteroscedasticity (GARCH) models which have been found to be useful in many economic and financial studies.

15.1 Regression with Autocorrelated Errors

In standard regression analysis, we consider the model

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \varepsilon_t$$

or, equivalently,

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{15.1.1}$$

where Y_t is a dependent variable, $\mathbf{X}_t'$ is a $(1 \times k)$ vector of independent variables, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of parameters, and ε_t is an error term often assumed to be i.i.d. $N(0, \sigma_\varepsilon^2)$. Under these standard assumptions, it is well known that the OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is a minimum variance unbiased estimator. When $\mathbf{X}_t'$ is stochastic in (15.1.1), conditional on $\mathbf{X}_t'$, the results about the OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ also hold as long as ε_s and $\mathbf{X}_t'$ are independent for all s and t.

When $\mathbf{X}_t'$ is a vector of k lagged values of Y_t, i.e., $\mathbf{X}_t' = (Y_{t-1}, \ldots, Y_{t-k})$, $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_k)'$, and ε_t is white noise, the model in (15.1.1) becomes an AR(k) model

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_k Y_{t-k} + \varepsilon_t. \tag{15.1.2}$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is asymptotically unbiased and consistent. As shown in Section 7.4, however, this result no longer holds when the ε_t are autocorrelated. When that is the case, the estimator is no longer consistent. This caveat is important. When time series variables are used in a model, it is the norm rather than the exception that the error terms are autocorrelated. Even in univariate time series variables analysis when the underlying process is known to be an AR model as in (15.1.2), the error terms ε_t could still be autocorrelated unless the correct order of k is chosen. Thus, a residual analysis is an important step in regression analysis when time series variables are involved in the study.

Many methods can be used to test for autocorrelation. For example, we can either use the test based on the Durbin–Walson statistic or we can perform the residual autocorrelation analysis discussed in Chapter 6. The latter method is particularly useful. Based on the residuals from a preliminary analysis of (15.1.1), an underlying model can be identified for the error term using the identification techniques discussed in Chapter 6. A final analysis can then be performed on a modified model with a specified model substituted for the error term. In time series regression analysis, the following model is often proposed to account for the autocorrelated errors:

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t \tag{15.1.3}$$

for $t = 1, 2, \ldots, n$, where

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \cdots + \varphi_p\varepsilon_{t-p} + n_t \tag{15.1.4}$$

and the n_t are i.i.d. $N(0, \sigma^2)$.

Let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\xi} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The matrix form of the model in (15.1.3) is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{15.1.5}$$

where $\boldsymbol{\xi}$ follows a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$. When $\varphi_1, \ldots, \varphi_p$ and σ^2 are known in (15.1.4) and the model is stationary, $\boldsymbol{\Sigma}$ can be easily calculated. The diagonal element of $\boldsymbol{\Sigma}$ is the variance of ε_t, the jth off-diagonal element corresponds to the jth autocovariance of ε_t, and they can be easily computed as shown in Chapter 3. Given $\boldsymbol{\Sigma}$, the generalized least squares (GLS) estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{Y} \tag{15.1.6}$$

is known to be a minimum variance unbiased estimator.

Normally, we will not know the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\xi}$ because even if ε_t follows an AR(p) model given in (15.1.4), the σ^2 and AR parameters φ_j are usually unknown. As a remedy, the following iterative GLS is often used:

1. Calculate OLS residuals $\hat{\varepsilon}_t$ from OLS fitting of (15.1.3).
2. Estimate φ_j and σ^2 for the AR(p) model in (15.1.4) based on the OLS residuals, $\hat{\varepsilon}_t$, using any method discussed in Chapter 7. For example, a simple conditional OLS estimation can be used.
3. Compute Σ from (15.1.4) using the values of φ_j and σ^2 obtained in step 2.
4. Compute the GLS estimator, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}$, using Σ obtained in step 3. Compute the residuals $\hat{\varepsilon}_t$ from the GLS model fitting in step 4, and repeat steps 2 through 4 until some convergence criterion (such as the maximum change in the estimates between iterations becoming less than some specified quantity) is reached.

More generally, the error structure can be modified to include an ARMA model. The above GLS iterative estimation can still be used except that a nonlinear least squares estimation instead of OLS is needed to estimate the parameters in the error model. Alternatively, by substituting the error model in the regression equation, (15.1.3), we can also use the nonlinear estimation or maximum likelihood estimation discussed in Chapter 7 to jointly estimate the regression and error model parameters $\boldsymbol{\beta}$ and φ_j's as we did in Chapter 14.

Although the error term, ε_t, can be autocorrelated in the regression model, it should be stationary. A nonstationary error structure could produce a spurious regression where a significant regression can be achieved for totally unrelated series as shown in Granger and Newbold (1986), and Phillips (1986). In such a case, one should properly difference the series before estimating the regression.

15.2 ARCH and GARCH Models

One of the main assumptions of the standard regression analysis and regression models with autocorrelated errors introduced in Section 15.1 is that the variance, σ^2, of the errors n_t is constant. In many practical applications, this assumption may not be realistic. For example, in financial investment, it is generally agreed that stock markets' volatility is rarely constant over time. Indeed, the study of the market volatility as it relates to time is the main interest for many researchers and investors.

Such a model incorporating the possibility of a nonconstant error variance is called a heteroscedasticity model. Many approaches can be used to deal with heterocedasticity. For example, the weighted regression is often used if the error variance at different times is known. In practice, however, the error variance is normally unknown; therefore, models to account for the heteroscedasticity are needed.

Let us first consider the standard regression model with uncorrelated error

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{15.2.1}$$

where $\varepsilon_t = n_t$ and the n_t are uncorrelated but have variances that change over time. Following Engle (1982), let us assume that the error term can be modeled as

$$n_t = \sigma_t e_t, \tag{15.2.2}$$

where the e_t are i.i.d. random variables with mean 0 and variance 1, independent of past realizations of n_{t-i}, and

$$\sigma_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \cdots + \theta_s n_{t-s}^2. \tag{15.2.3}$$

Given all the information up to time $(t-1)$, the conditional variance of the n_t becomes

$$\begin{aligned} \text{Var}_{t-1}(n_t) &= E_{t-1}(n_t^2) = E(n_t^2 | n_{t-1}, n_{t-2}, \ldots) = \sigma_t^2 \\ &= \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \cdots + \theta_s n_{t-s}^2, \end{aligned} \tag{15.2.4}$$

which is related to the squares of past errors, and it changes over time. A large error through n_{t-i}^2 gives rise to the variance, which tends to be followed by another large error. This phenomenon of volatility clustering is common in many financial time series.

Using the forecasting results from Chapter 5, we see that Equation (15.2.4) is simply the optimal forecast of n_t^2 if n_t^2 follows the following AR(s) model

$$n_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \cdots + \theta_s n_{t-s}^2 + a_t, \tag{15.2.5}$$

where a_t is an $N(0, \sigma_a^2)$ white noise process. Thus, Engle (1982) called the model of the error term n_t with the variance specification given in (15.2.2) and (15.2.3) or, equivalently, in (15.2.5) the autoregressive conditional heteroscedasticity (ARCH) model. More specifically, it is an ARCH model of order s and is often denoted ARCH(s). The term has also been used for the model for Y_t in (15.2.1) with an error variance structure given in (15.2.2) and (15.2.3). In financial applications, because the mean return is often expected to be related to the variance, Engle, Lilien, and Robins (1987) generalize (15.2.1) to

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \alpha\sigma_t + \varepsilon_t$$

and call it the ARCH-in-mean or ARCH-M model.

To test for ARCH, one can first fit the ordinary least squares regression in (15.2.1) for $t = 1, 2, \ldots, n$. The first step is to compute the OLS residuals $\hat{\varepsilon}_t = \hat{n}_t$. Then, we can form the series $\hat{n}_t^2$ and examine its ACF and PACF to see whether it follows an AR model. Alternatively, we can, following Engle (1982), specify the following AR(s) model to $\hat{n}_t^2$:

$$\hat{n}_t^2 = \theta_0 + \theta_1 \hat{n}_{t-1}^2 + \theta_2 \hat{n}_{t-2}^2 + \cdots + \theta_s \hat{n}_{t-s}^2 + a_t \tag{15.2.6}$$

for $t = s+1, s+2, \ldots, n$ and test $H_0 : \theta_1 = \cdots = \theta_s = 0$. Under the null hypothesis that the n_t are i.i.d. $N(0, \sigma^2)$, we have $H_0 : \theta_1 = \cdots = \theta_s = 0$. Thus, if H_0 is true we expect that the estimates of θ_j's are close to zero and that the value of the coefficient of determination R^2 is small. Because

$$R^2 = \frac{\text{regression sum of squares}}{\text{total sum of squares}},$$

and under the null hypothesis

$$(n - s)R^2 \xrightarrow{d} \chi^2(s), \tag{15.2.7}$$

an ARCH phenomenon exists when the value of $(n - s)R^2$ is statistically significant.

A natural extension to the ARCH model is to consider that the conditional variance of the error process is related not only to the squares of past errors but also to the past conditional variances. Thus, we have the more general error process

$$n_t = \sigma_t e_t, \tag{15.2.8}$$

where the e_t are i.i.d. random variables with mean 0 and variance 1, independent of past realizations of n_{t-i},

$$\sigma_t^2 = \theta_0 + \phi_1\sigma_{t-1}^2 + \cdots + \phi_r\sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \cdots + \theta_s n_{t-s}^2, \tag{15.2.9}$$

and the roots of $(1 - \phi_1 B - \cdots - \phi_r B^r) = 0$ are outside the unit circle. To guarantee $\sigma_t^2 > 0$, we assume that $\theta_0 > 0$ and that ϕ_i and θ_j are nonnegative. The model for the error term given in (15.2.8) with the property (15.2.9) is known as the generalized autoregressive conditional heteroscedasticity (GARCH) model of order (r, s) and is denoted by GARCH(r, s). This model was first proposed by Bollerslev (1986). Clearly, when $r = 0$, the model reduces to the ARCH(s) process.

One should not be misled by (15.2.9) and conclude that the conditional variance follows an ARMA(r, s) model with r being the AR order and s being the MA order. In fact, the model in (15.2.9) is not a proper ARMA model. From Chapter 5, we see that in an ARMA process, the associated error is a white noise process corresponding to the one-step ahead forecast error. In (15.2.9), neither n_t^2 nor σ_t^2 plays this proper role. Let $a_t = (n_t^2 - \sigma_t^2)$ so that $\sigma_t^2 = (n_t^2 - a_t)$. We can rewrite (15.2.9) in terms of n_t^2 and a_t as follows:

$$n_t^2 = \theta_0 + \sum_{i=1}^{m}(\phi_i + \theta_i)n_{t-i}^2 + a_t - \sum_{j=1}^{r}\phi_i a_{t-j} \tag{15.2.10}$$

or

$$(1 - \alpha_1 B - \cdots - \alpha_m B^m)n_t^2 = \theta_0 + (1 - \phi_1 B - \cdots - \phi_r B^r)a_t, \tag{15.2.11a}$$

where $m = \max(r, s)$, $\phi_i = 0$ for $i > r$, $\theta_i = 0$ for $i > s$,

$$\alpha_i = (\phi_i + \theta_i), \tag{15.2.11b}$$

and

$$a_t = (n_t^2 - \sigma_t^2). \tag{15.2.11c}$$

We claim that a_t is the associated white noise process for the process of n_t^2; and therefore, (15.2.10) or (15.2.11a) is a proper ARMA model, which follows because $E_{t-1}(n_t^2) = \sigma_t^2, \sigma_t^2$

is the one-step ahead forecast of n_t^2, and a_t is the corresponding one-step ahead forecast error. In other words,

$$E_{t-1}(a_t) = E_{t-1}(n_t^2 - \sigma_t^2) = 0 = E(a_t)$$

and

$$\begin{aligned} E(a_i a_j) &= E(n_i^2 - \sigma_i^2)(n_j^2 - \sigma_j^2) = E(\sigma_i^2 e_i^2 - \sigma_i^2)(\sigma_j^2 e_j^2 - \sigma_j^2) \\ &= E[\sigma_i^2 \sigma_j^2 (e_i^2 - 1)(e_j^2 - 1)] \\ &= 0, \qquad \text{for } i \neq j, \end{aligned} \tag{15.2.12}$$

where we note that the e_t^2 are i.i.d. χ^2 (1). Thus, the GARCH(r, s) model in (15.2.8) and (15.2.9) implies that n_t^2 follows an ARMA(m, r) model in (15.2.11a) with the AR order being $m = \max(r, s)$.

From (15.2.10) or (15.2.11a), we see that this process will have a unit root if $(1 - \alpha_1 - \cdots - \alpha_m) = 0$, i.e., if $\sum_{j=1}^{m} \alpha_j = \sum_{j=1}^{m} (\phi_i + \theta_i) = \sum_{j=1}^{r} \phi_j + \sum_{j=1}^{s} \theta_j = 1$. In this case, the model will be called an integrated GARCH process, or simply IGARCH.

From (15.2.3) and (15.2.9), it is clear that in the specification of ARCH, GARCH, or IGARCH models, the effect of errors on the conditional variance is symmetric, i.e., a positive error has the same effect as a negative error of the same magnitude. Moreover, because σ_t^2 and n_t^2 are always nonnegative in (15.2.9), it is obvious that some restrictions on the coefficients θ_0, ϕ_i, and θ_j are needed. To accommodate the asymmetric relation between many financial variables and their volatility changes and to relax the restriction on the coefficients in the model, Nelson (1991) proposed that

$$\log(\sigma_t^2) = \gamma_t + \sum_{j=0}^{\infty} \psi_j g(a_{t-1-j}), \tag{15.2.13}$$

where $\psi_0 = 1$ and $a_t = n_t/\sigma_t$. The function g is chosen to allow for asymmetric changes, depending on the sign of a_t, in the conditional variance. For example, we may choose

$$g(a_t) = \delta a_t + \alpha(|a_t| - E|a_t|).$$

In this case, when $0 < a_t < \infty$, $g(a_t)$ is linear in a_t with slope $(\delta + \alpha)$; when $-\infty < a_t < 0$, $g(a_t)$ is linear with slope $(\delta - \alpha)$. Thus, $g(a_t)$ allows the conditional variance to respond asymmetrically to the rises and falls of the process. The specification of (15.2.13) implies that σ_t^2 will be positive regardless of the sign of the coefficients. The coefficients ψ_j are often assumed to relate to an ARMA specification to the model, i.e.,

$$\phi(B)(\log \sigma_t^2 - \gamma_t) = \theta(B) g(a_{t-1}),$$

where $\sum_{j=0}^{\infty} \psi_j B^j = \psi(B) = \theta(B)/\phi(B)$. The polynomial $\phi(B)$ is such that the roots of $\phi(B) = 0$ are all outside the unit circle and share no common factor with the polynomial $\theta(B)$. The model in which the evolution of the conditional variance satisfies (15.2.13) is called exponential GARCH, or simply EGARCH.

More generally, the regression model with autocorrelated error can be combined with the conditional heteroscedasticity model, i.e.,

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{15.2.14}$$

where

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \cdots + \varphi_p\varepsilon_{t-p} + n_t, \tag{15.2.15}$$

$$n_t = \sigma_t e_t, \tag{15.2.16}$$

$$\sigma_t^2 = \theta_0 + \phi_1\sigma_{t-1}^2 + \cdots + \phi_r\sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \cdots + \theta_s n_{t-s}^2, \tag{15.2.17}$$

and the e_t are i.i.d. $N(0, 1)$ and independent of past realizations of n_{t-i}. To test for heteroscedasticity in the error variance, we can perform the following steps:

1. Calculate OLS residuals $\hat{\varepsilon}_t$ from the OLS fitting of (15.2.14).
2. Fit an AR(p) model (15.2.15) to the $\hat{\varepsilon}_t$.
3. Obtain the residuals $\hat{n}_t$ from the AR fitting in (15.2.15).
4. Form the series $\hat{n}_t^2$ and compute its sample ACF, i.e.,

$$\hat{\rho}_i(\hat{n}_t^2) = \frac{\sum_{t=1}^{n-i}(\hat{n}_t^2 - \hat{\sigma}^2)(\hat{n}_{t+i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^{n}(\hat{n}_t^2 - \hat{\sigma}^2)^2}, \tag{15.2.18}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{n}_t^2. \tag{15.2.19}$$

Similarly, we can also compute its PACF. Then, we can check whether these ACF and PACF follow any pattern. A pattern of these ACF and PACF not only indicates ARCH or GARCH errors, but it also forms a good basis for their order specification.

To see how the ACF and PACF pattern can help identify a GARCH model, we first note that as shown in (15.2.11a), a general GARCH(r, s) model for σ_t^2 corresponds to an ARMA(m, r) model for n_t^2 with $m = \max(r, s)$. Thus, both ACF and PACF of $\hat{n}_t^2$ will show patterns of an exponential decay. The orders r and s can be identified using the similar techniques of identifying the orders of an ARMA model. On the other hand, for a pure ARCH(s) model, its ARMA representation, from (15.2.11a), reduces to the AR(s) model

$$(1 - \alpha_1 B - \cdots - \alpha_s B^s)n_t^2 = \theta_0 + a_t. \tag{15.2.20}$$

Hence, the corresponding PACF will cut off after lag s.

Alternatively, we can also use the following portmanteau Q statistic to test for $\rho_i(n_t^2) = 0, i = 1, 2, \ldots, k$ as suggested by McLeod and Li (1983):

$$Q(k) = n(n+2)\sum_{i=1}^{k}\frac{\hat{\rho}_i^2(\hat{n}_t^2)}{(n-i)}. \tag{15.2.21}$$

As in Chapter 7, under the null hypothesis that the n_t^2 are white noise, the $Q(k)$ statistic approximately follows a $\chi^2(k)$ distribution. In this case, the significance of the $Q(k)$ statistic occurring only for a small value of k indicates an ARCH model, and a persistent significance for a large value of k implies a GARCH model.

Recall that to test for an ARCH model, we can also fit model (15.2.6) and check the significance of the statistic given in (15.2.7).

15.3 Estimation of GARCH Models

Once a model is identified, the next step is to estimate its parameters. Let us consider the general regression model with autocorrelated error and conditional heterocedasticity GARCH structure for the error variance,

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{15.3.1}$$

where

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \cdots + \varphi_p\varepsilon_{t-p} + n_t, \tag{15.3.2}$$

$$n_t = \sigma_t e_t, \tag{15.3.3}$$

$$\sigma_t^2 = \theta_0 + \phi_1\sigma_{t-1}^2 + \cdots + \phi_r\sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \cdots + \theta_s n_{t-s}^2, \tag{15.3.4}$$

and the e_t are i.i.d. $N(0, 1)$ and independent of past realizations of n_{t-i}. The following methods can be used to estimate the parameters.

15.3.1 Maximum Likelihood Estimation

Note that the above model can be rewritten as

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + \frac{1}{(1 - \varphi_1 B - \cdots - \varphi_p B^p)}n_t \tag{15.3.5}$$

or

$$n_t = (1 - \varphi_1 B - \cdots - \varphi_p B^p)(Y_t - \mathbf{X}_t'\boldsymbol{\beta}). \tag{15.3.6}$$

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, and $\mathbf{Y}_0$ and $\mathbf{X}_0$ be some proper starting values required for computing n_t for $t = 1, \ldots n$ from (15.3.6). Then, similar to the estimation of transfer function models discussed in Section 14.3.3, we can obtain the maximum likelihood estimates of the parameters through maximizing the conditional likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}, \mathbf{Y}_0, \mathbf{X}_0) = \prod_{t=1}^{n} \left(\frac{1}{2\pi\sigma_t^2}\right)^{1/2} \exp\left(-\frac{n_t^2}{2\sigma_t^2}\right), \tag{15.3.7}$$

or the log-likelihood function

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}, \mathbf{Y}_0, \mathbf{X}_0) = \sum_{t=1}^{n} \frac{1}{2}\left(-\ln(2\pi) - \ln(\sigma_t^2) - \frac{n_t^2}{\sigma_t^2}\right), \tag{15.3.8}$$

where σ_t^2 is given in (15.3.4), n_t is given in (15.3.6), $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_r)$, and $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_s)$.

15.3.2 Iterative Estimation

Alternatively, because the GARCH(r, s) model for σ_t^2 in (15.3.4) is equivalent to an ARMA(m, r) model for n_t^2 with $m = \max(r, s)$, i.e.,

$$n_t^2 = \theta_0 + \sum_{i=1}^{m} \alpha_i n_{t-i}^2 + a_t - \sum_{j=1}^{r} \phi_j a_{t-j}, \tag{15.3.9}$$

where the a_t are zero mean white noise, we can estimate the parameters using the following steps:

1. Compute GLS estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ using the iterative GLS method discussed in Section 15.1. Let $\hat{\varepsilon}_t$ be the error series obtained from the last iteration of the GLS estimation.
2. Calculate the sample ACF and PACF of the $\hat{\varepsilon}_t$ series, and identify and fit its AR(p) representation.
3. Compute the residuals, $\hat{n}_t$, from the AR(p) fitting of the $\hat{\varepsilon}_t$ series in step 2.
4. Fit the identified ARMA (m, r) model on the squared residuals $\hat{n}_t^2$, i.e.,

$$\hat{n}_t^2 = \theta_0 + \sum_{i=1}^{m} \alpha_i \hat{n}_{t-i}^2 + a_t - \sum_{j=1}^{r} \phi_j a_{t-j}.$$

Let $\hat{\theta}_0$, $\hat{\alpha}_i$, and $\hat{\phi}_i$ be the estimates of the parameters in this ARMA(m, r) model. Then from (15.2.11a, b), we have

$$\begin{aligned} \hat{\theta}_0 &= \hat{\theta}_0, \\ \hat{\phi}_i &= \hat{\phi}_i, \qquad i = 1, \ldots, r, \\ \hat{\theta}_i &= \hat{\alpha}_i - \hat{\phi}_i, \qquad i = 1, \ldots, s. \end{aligned} \tag{15.3.10}$$

15.4 Computation of Forecast Error Variance

An important use of an ARCH or GARCH model is to find a forecast error variance that changes over time. We first note that the optimal l-step ahead forecast of Y_{t+l} given all information up to time t is the conditional expectation $E_t(Y_{t+l})$. The associated l-step ahead forecast error is $\varepsilon_{t+l} = Y_{t+l} - E_t(Y_{t+l})$. To calculate the conditional variance, we rewrite (15.3.2) as

$$(1 - \varphi_1 B - \cdots - \varphi_p B^p)\varepsilon_t = n_t. \tag{15.4.1}$$

Thus,

$$\varepsilon_t = \sum_{j=0}^{\infty} \psi_j B^j n_t, \tag{15.4.2}$$

where $\psi_0 = 1$ and

$$\sum_{j=0}^{\infty} \psi_j B^j = \frac{1}{(1 - \varphi_1 B - \cdots - \varphi_p B^p)}. \tag{15.4.3}$$

Using the result in Section 5.2.1, we see that given all the information up to time t, the l-step ahead forecast error variance is the conditional variance

$$\begin{aligned} \mathrm{Var}_t(\varepsilon_{t+l}) &= E_t[\varepsilon_{t+l} - E_t(\varepsilon_{t+l})]^2 = E_t\left[\sum_{j=0}^{l-1} \psi_j n_{t+l-j}\right]^2 \\ &= E_t\left[\sum_{j=0}^{l-1} \psi_j \sigma_{t+l-j}\, e_{t+l-j}\right]^2 = \sum_{j=0}^{l-1} \psi_j^2 \sigma_{t+l-j}^2, \end{aligned} \tag{15.4.4}$$

where $\sigma_{t+l-j}^2 = E_t(n_{t+l-j}^2)$, which can be calculated from (15.3.4) or, equivalently, (15.3.9) using the estimated parameters obtained earlier, i.e.,

$$\sigma_{t+l-j}^2 = E_t[n_{t+l-j}^2] = \theta_0 + \sum_{i=1}^{m} \alpha_i E_t[n_{t+l-j-i}^2] + E_t[a_{t+l-j}] - \sum_{i=1}^{r} \phi_i E_t[a_{t+l-j-i}]. \tag{15.4.5}$$

Clearly, in using (15.3.4) or (15.3.9), we may need σ_t^2 and n_t^2 for $t = -m + 1, \ldots,$ and 0, where $m = \max(r, s)$. Bollerslev (1986) suggests that these terms be chosen as

$$\sigma_t^2 = n_t^2 = \hat{\sigma}^2, \qquad \text{for } t = -m + 1, \ldots, 0, \tag{15.4.6}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}})^2. \tag{15.4.7}$$

As a special case, we note that in (15.3.2) that when $\varepsilon_t = n_t$, the model reduces to the GARCH(r, s) model of uncorrelated errors, and we have $\text{Var}_t(\varepsilon_{t+l}) = \sigma_{t+l}^2$. Thus, given the information up to time $(t-1)$, the one-step ahead conditional forecast error variance is simply σ_t^2, which changes over time.

15.5 Illustrative Examples

EXAMPLE 15.1 In this example, we consider the bivariate series listed as Series W15 in the appendix and plotted in Figure 15.1, which were originally presented in Bryant and Smith (1995). This series contains 46 monthly observations on Louisiana and Oklahoma spot prices for natural gas from January 1988 to October 1991, measured in dollars per million British thermal units ($/MBtu). The spot market for natural gas is the market in which natural gas is sold under a contract term of no more than one month. It was known

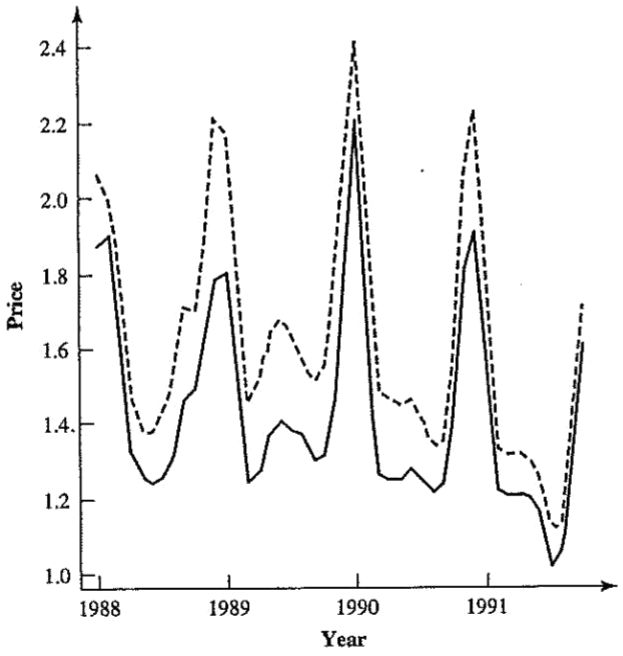


FIGURE 15.1 Spot prices of natural gas in Oklahoma (solid line) and Louisiana (dotted line) between January 1988 and October 1991.

that the spot price for Louisiana natural gas was known on or before the first day of trading on the Oklahoma market. To examine whether one can predict the Oklahoma spot price from the current spot price of Louisiana, we will try a time series regression model.

Preliminary Analysis We begin with a simple regression model with uncorrelated errors. That is, let O_t and L_t be the spot prices for Oklahoma and Louisiana, respectively. We consider the OLS estimation for the model

$$O_t = \beta_0 + \beta_1 L_t + \varepsilon_t, \tag{15.5.1}$$

which gives $R^2 = .9444$ and the following estimation result:

Parameter	Estimate	St.E.	t-value	P-value
β_0	.12	.049	2.43	.0193
β_1	.8	.029	27.34	.0000

Thus, the OLS regression equation is $\hat{O}_t = .12 + .8L_t$, and we can compute the OLS residuals, $\hat{\varepsilon}_t = O_t - \hat{O}_t$. The PACF of the $\hat{\varepsilon}_t$ series shows a significant spike at lag 1. Thus, we propose an AR(1) model for the error term, i.e.,

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + n_t. \tag{15.5.2}$$

The AR(1) fitting of the series $\hat{\varepsilon}_t$ gives $\hat{\varphi} = .502$. To check for heteroscedasticity in error variance, we compute the residuals, $\hat{n}_t = \hat{\varepsilon}_t - .502\hat{\varepsilon}_{t-1}$, and calculate the portmanteau $Q(k)$ statistic on the squared $\hat{n}_t^2$ series, with the results summarized in Table 15.1.

TABLE 15.1 Test for heteroscedasticity.

k	Q(k) Statistic	P-value
1	3.2921	.0696
2	3.3053	.1915
3	3.4124	.3323
4	3.7174	.4456
5	4.4355	.4486
6	4.5817	.5985
7	4.7803	.6867
8	5.4117	.7128
9	5.7651	.7632
10	7.1240	.7137
11	7.2210	.7809

Construction of a GARCH (r, s) Model At $\alpha = .05$, Table 15.1 shows that there is no evidence of heteroscedasticity. For $\alpha = .1$, however, these $Q(k)$ statistics clearly indicate an ARCH(1) or GARCH(0, 1) model. Thus, we will refit the spot prices with the model

$$O_t = \beta_0 + \beta_1 L_t + \varepsilon_t, \tag{15.5.3}$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + n_t, \tag{15.5.4}$$

$$n_t = \sigma_t e_t, \tag{15.5.5}$$

where the e_t are i.i.d. $N(0, 1)$, independent of n_{t-1}, and

$$\sigma_t^2 = \theta_0 + \theta_1 n_{t-1}^2. \tag{15.5.6}$$

The estimation results of this model are summarized in Table 15.2. From this table, we see that the estimate $\hat{\theta}_0$ is not significant but that $\hat{\theta}_1$ is highly significant. Hence,

$$\sigma_t^2 = 1.956 n_{t-1}^2 \tag{15.5.7}$$

or, equivalently, from (15.2.10),

$$n_t^2 = 1.956 n_{t-1}^2 + a_t, \tag{15.5.8}$$

where a_t is a zero mean white noise process.

Computation of the Conditional Forecast Error Variance To calculate the conditional variance, we note that from (15.4.2) and (15.4.3) that

$$\varepsilon_t = \frac{1}{(1 - .502B)} n_t = \sum_{j=0}^{\infty} \psi_j B^j n_t,$$

with $\psi_0 = 1$ $\psi_j = (.502)^j$ for $j = 1, 2, \ldots$. Thus, given all the information up to time t, by (15.4.4) the l-step ahead forecast error variance is the conditional variance

$$\mathrm{Var}_t(\varepsilon_{t+l}) = E_t[\varepsilon_{t+l} - E_t(\varepsilon_{t+l})]^2 = \sum_{j=0}^{l-1} (.502)^{2j} \sigma_{t+l-j}^2,$$

TABLE 15.2 Estimation of the GARCH(0, 1) model.

Coefficient	Value	St.E.	t-value	P-value
$\hat{\beta}_0$	.068	.0204	3.315	.0009
$\hat{\beta}_1$	.858	.0104	82.517	.0001
$\hat{\varphi}$	.502	.1289	3.894	.001
$\hat{\theta}_0$	.0002	.0002	1.363	.1728
$\hat{\theta}_1$	1.956	.6068	3.214	.0013

where $\sigma^2_{t+l-j} = E_t(n^2_{t+l-j})$. Specifically, following (15.4.4) and (15.4.5), given the information up to October 1991 that corresponds to $t = 46$, we can compute the conditional forecast error variances for $t = 47$ and $t = 48$ as follows:

$$\begin{aligned}
\text{Var}_{46}(\varepsilon_{47}) &= \sigma^2_{47} = E_{46}[n^2_{47}] = E_{46}[1.956n^2_{46} + a_{47}] = 1.956n^2_{46} \\
\text{Var}_{46}(\varepsilon_{48}) &= \sigma^2_{48} + (.502)^2\sigma^2_{47} = E_{46}[n^2_{48}] + (.502)^2E_{46}[n^2_{47}] \\
&= E_{46}[1.956n^2_{47} + a_{48}] + (.502)^2(1.956)n^2_{46} \\
&= 1.956E_{46}[n^2_{47}] + (.502)^2(1.956)n^2_{46} \\
&= (1.956)^2n^2_{46} + (.502)^2(1.956)n^2_{46} = 4.3189n^2_{46},
\end{aligned}$$

where n_{46} is the residual at $t = 46$ obtained from the autoregressive error fitting in (15.5.4). The other l-step ahead conditional forecast error variances can be calculated similarly.

Although we introduce the conditional variance through a regression model, the concept clearly applies to the univariate time series models. In other words, in the general ARMA model

$$(1 - \phi_1B - \cdots - \phi_pB^p)Z_t = \theta_0 + (1 - \theta_1B - \cdots - \theta_qB^q)a_t,$$

the white noise process a_t may follow a GARCH(r, s) model, i.e.,

$$a_t = \sigma_te_t,$$

where the e_t are i.i.d. random variables with mean 0 and variance 1, independent of past realizations of a_{t-l}, and

$$\sigma^2_t = \theta_0 + \phi_1\sigma^2_{t-1} + \cdots + \phi_r\sigma^2_{t-r} + \theta_1a^2_{t-1} + \cdots + \theta_sa^2_{t-s},$$

or an ARCH model if $r = 0$. Let us consider the following example.

EXAMPLE 15.2 We recall Series W6, the yearly US tobacco production from 1871 to 1984, first considered in Chapter 4 with its plot shown in Figure 4.2. The series is more volatile at some times than at others. Instead of taking a transformation, we now fit an ARIMA(0, 1, 1) model to the original series and try to examine its residuals. The estimated model is

$$\underset{(.0786)}{(1 - B)Z_t = (1 - .5619B)n_t},$$

where for consistency of notations used in this chapter we use n_t to denote the white noise error series. Traditionally, the variance of n_t is assumed to be constant. Examination of ACF and PACF of n^2_t, however, suggests the following ARCH(2) model:

$$\underset{(.089)}{n^2_t = .1729n^2_t} + \underset{(.094)}{.3698n^2_{t-2}} + a_t.$$

Given the information up to 1984 that corresponds to $t = 114$, we can compute the conditional error variances for $t = 115$ and $t = 116$ as follows:

$$\begin{aligned}
\text{Var}_{114}(n_{115}) &= \sigma_{115}^2 = E_{114}[n_{115}^2] = E_{114}[.1729n_{114}^2 + .3698\,n_{113}^2 + a_{115}] \\
&= .1729n_{114}^2 + .3698n_{113}^2, \\
\text{Var}_{114}(n_{116}) &= \sigma_{116}^2 = E_{114}[n_{116}^2] = E_{114}[.1729n_{115}^2 + .3698n_{114}^2 + a_{116}] \\
&= .1729E_{114}[n_{115}^2] + .3698n_{114}^2 \\
&= .1729[.1729n_{114}^2 + .3698n_{113}^2] + .3698n_{114}^2 \\
&= .3997n_{114}^2 + .0639n_{113}^2,
\end{aligned}$$

which shows that the variances are related to the squares of past errors and they change over time.

The investigation of conditional variance models has been one of the main areas of study in time series analysis of financial markets. Toward these ends, the GARCH model and its variations have been applied to many risk and volatility studies. We use the above simple examples to illustrate the procedure. Other examples include Engle, Ng, and Rothschild (1990), Day and Lewis (1992), and Lamoureux and Lastrapes (1993). These studies suggest that the selected model does not have to be complicated. Indeed, a very simple GARCH(r, s) model with $r \leq 2$ and $s \leq 2$ is often sufficient to provide a significant improvement over the traditional homoscedasticity models.

EXERCISES

15.1 Consider the ARCH(1) model $n_t = \sigma_t e_t$, where the e_t are i.i.d. random variable with mean 0 and variance 1, and $\sigma_t^2 = \theta_0 + \theta_1 n_{t-1}^2$. Show that the unconditional variance of n_t is $\text{Var}(n_t) = \theta_0/(1 - \theta_1)$.

15.2 Let n_t be the Gaussian white noise process with mean 0 and variance σ^2. If we fit the model by OLS

$$n_t^2 = \theta_0 + \theta_1 n_{t-1}^2 + \theta_2 n_{t-2}^2 + \cdots + \theta_s n_{t-s}^2 + a_t,$$

for $t = s + 1, s + 2, \ldots, N$, show that $(N - s)R^2 \xrightarrow{d} \chi^2(s)$, where R^2 is the multiple coefficient of determination.

15.3 Consider the general GARCH(r, s) model, $n_t = \sigma_t e_t$, where the e_t are i.i.d. random variables with mean 0 and variance 1, and

$$\sigma_t^2 = \theta_0 + \phi_1\sigma_{t-1}^2 + \cdots + \phi_r\sigma_{t-r}^2 + \theta_1 n_{t-1}^2 + \cdots + \theta_s n_{t-s}^2.$$

Show that for the unconditional variance of n_t to be positive and finite the coefficients need to satisfy the constraints $\theta_0 > 0$, $\phi_i \geq 0$, $\theta_j \geq 0$, and $\sum_{i=1}^{r}\phi_i + \sum_{j=1}^{s}\theta_j < 1$.

15.4 Consider the U.S. liquor sales between 1970 and 1980 given in Exercise 8.6 of Chapter 8.

(a) Build an appropriate ARIMA-GARCH model for the series.

(b) Compute the next 12-month forecasts and their conditional variances from the model.

(c) Compare and comment on the results obtained from the ARIMA-GARCH model and the standard ARIMA model.

15.5 Find a financial time series. Fit the series with a univariate model, perform an analysis on its squared residuals, and construct a GARCH model for the error variance if needed.

15.6 Find some related time series of your interest.

(a) Fit a regression model using OLS, perform a residual analysis, identify the error process, and construct a GARCH model if needed.

(b) Compute the one-step ahead and two-step ahead conditional forecast error variances.

16

Vector Time Series Models

The time series data in many empirical studies consist of observations from several variables. For example, in a sales performance study, the variables might include sales volume, prices, sales force, and advertising expenditures. In the transfer function model, we study a specific relationship between variables, i.e., an input-output relationship between one output variable and one or several input variables. In regression time series models, we also concentrate on relationship between a dependent variable and a set of independent variables. In many applications, however, transfer function and regression models may not be appropriate. In this chapter, we introduce a more general class of vector time series models to describe relationships among several time series variables. After an introduction of some basic concepts for vector processes, stationary and nonstationary vector ARMA models are discussed. Some useful notions such as partial autoregression and partial lag correlation matrix function are introduced. Detailed empirical studies are presented. The chapter ends with some descriptions of spectral properties of vector processes.

16.1 Covariance and Correlation Matrix Functions

Let $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]'$, $t = 0, \pm 1, \pm 2, \ldots$, be an m-dimensional jointly stationary real-valued vector process so that the mean $E(Z_{i,t}) = \mu_i$ is constant for each $i = 1, 2, \ldots, m$, and the cross-covariance between $Z_{i,t}$ and $Z_{j,s}$, for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$, are functions only of the time difference $(s - t)$. Hence, we have the mean vector

$$E(\mathbf{Z}_t) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \tag{16.1.1}$$

and the lag-k covariance matrix

$$\Gamma(k) = \text{Cov}\{\mathbf{Z}_t, \mathbf{Z}_{t+k}\} = E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+k} - \boldsymbol{\mu})']$$

$$= E\begin{bmatrix} Z_{1,t} - \mu_1 \\ Z_{2,t} - \mu_2 \\ \vdots \\ Z_{m,t} - \mu_m \end{bmatrix}[Z_{1,t+k} - \mu_1, Z_{2,t+k} - \mu_2, \ldots, Z_{m,t+k} - \mu_m]$$

$$= \begin{bmatrix} \gamma_{11}(k) & \gamma_{12}(k) & \cdots & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & \cdots & \gamma_{2m}(k) \\ \vdots & \vdots & \vdots & \\ \gamma_{m1}(k) & \gamma_{m2}(k) & \cdots & \gamma_{mm}(k) \end{bmatrix} = \text{Cov}\{\mathbf{Z}_{t-k}, \mathbf{Z}_t\}, \tag{16.1.2}$$

where

$$\gamma_{ij}(k) = E(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j) = E(Z_{i,t-k} - \mu_i)(Z_{j,t} - \mu_j)$$

for $k = 0, \pm 1, \pm 2, \ldots$, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, m$. As a function of k, $\Gamma(k)$ is called the covariance matrix function for the vector process $\mathbf{Z}_t$. For $i = j$, $\gamma_{ii}(k)$ is the autocovariance function for the ith component process $Z_{i,t}$; for $i \neq j$, $\gamma_{ij}(k)$ is the cross-covariance function between $Z_{i,t}$ and $Z_{j,t}$. The matrix $\Gamma(0)$ is easily seen to be the contemporaneous variance-covariance matrix of the process. A jointly stationary process implies that every univariate component process is stationary. A vector of univariate stationary processes, however, is not necessarily a jointly stationary vector process.

The correlation matrix function for the vector process is defined by

$$\boldsymbol{\rho}(k) = \mathbf{D}^{-1/2}\Gamma(k)\mathbf{D}^{-1/2} = [\rho_{ij}(k)] \tag{16.1.3}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$, where $\mathbf{D}$ is the diagonal matrix in which the ith diagonal element is the variance of the ith process; i.e.,

$$\mathbf{D} = \text{diag}[\gamma_{11}(0), \gamma_{22}(0), \ldots, \gamma_{mm}(0)].$$

Clearly, the ith diagonal element of $\boldsymbol{\rho}(k)$, $\boldsymbol{\rho}_{ii}(k)$, is the autocorrelation function for the ith component series $Z_{i,t}$ whereas the (i, j)th off-diagonal element of $\boldsymbol{\rho}(k)$,

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}, \tag{16.1.4}$$

represents the cross-correlation function between $Z_{i,t}$ and $Z_{j,t}$.

Like the univariate autocovariance and autocorrelation functions, the covariance and correlation matrix functions are also positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{\alpha}_i'\Gamma(t_i - t_j)\boldsymbol{\alpha}_j \geq 0 \tag{16.1.5}$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i'\boldsymbol{\rho}(t_i - t_j)\boldsymbol{\alpha}_j \geq 0 \tag{16.1.6}$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any set of real vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots,$ and $\boldsymbol{\alpha}_n$. The results follow immediately by evaluating the variance of $\sum_{i=1}^{n}\boldsymbol{\alpha}_i'\mathbf{Z}_{t_i}$ and its standardization. Note, however, that $\gamma_{ij}(k) \neq \gamma_{ij}(-k)$ for $i \neq j$ and also that $\boldsymbol{\Gamma}(k) \neq \boldsymbol{\Gamma}(-k)$. Instead, because $\gamma_{ij}(k) = E[(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j)] = E[(Z_{j,t+k} - \mu_j)(Z_{i,t} - \mu_i)] = \gamma_{ji}(-k)$, we have

$$\begin{cases}\boldsymbol{\Gamma}(k) = \boldsymbol{\Gamma}'(-k) \\ \boldsymbol{\rho}(k) = \boldsymbol{\rho}'(-k).\end{cases} \tag{16.1.7}$$

Sometimes, the covariance and correlation matrix functions are also called the autocovariance and autocorrelation matrix functions.

16.2 Moving Average and Autoregressive Representations of Vector Processes

An m-dimensional stationary vector process $\mathbf{Z}_t$ is said to be a linear process or a purely nondeterministic vector process if it can be written as a linear combination of a sequence of m-dimensional white noise random vectors, i.e.,

$$\begin{aligned}\mathbf{Z}_t &= \boldsymbol{\mu} + \mathbf{a}_t + \boldsymbol{\Psi}_1\mathbf{a}_{t-1} + \boldsymbol{\Psi}_2\mathbf{a}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{s=0}^{\infty}\boldsymbol{\Psi}_s\mathbf{a}_{t-s},\end{aligned} \tag{16.2.1}$$

where $\boldsymbol{\Psi}_0 = \boldsymbol{I}$ is the $m \times m$ identity matrix, the $\boldsymbol{\Psi}_j$'s are $m \times m$ coefficient matrices, and the $\mathbf{a}_t$'s are m-dimensional white noise random vectors with zero mean and covariance matrix structure

$$E[\mathbf{a}_t\mathbf{a}_{t+k}'] = \begin{cases}\boldsymbol{\Sigma}, & \text{if } k = 0, \\ \mathbf{0}, & \text{if } k \neq 0,\end{cases} \tag{16.2.2}$$

where $\boldsymbol{\Sigma}$ is a $m \times m$ symmetric positive definite matrix. Thus, although the elements of $\mathbf{a}_t$ at different times are uncorrelated, they may be contemporaneously correlated. Equivalently, using the backshift operator $B^s\mathbf{a}_t = \mathbf{a}_{t-s}$, we can write

$$\dot{\mathbf{Z}}_t = \boldsymbol{\Psi}(B)\mathbf{a}_t, \tag{16.2.3}$$

where $\dot{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu}$ and $\boldsymbol{\Psi}(B) = \sum_{s=0}^{\infty}\boldsymbol{\Psi}_sB^s$. The above representation is known as the moving average or Wold's representation.

Let $\Psi_s = [\psi_{ij,s}]$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$, where $\psi_{ij,0} = 1$ if $i = j$ and $\psi_{ij,0} = 0$ if $i \neq j$. We can write $\Psi(B) = [\psi_{ij}(B)]$, where $\psi_{ij}(B) = \sum_{s=0}^{\infty} \psi_{ij,s} B^s$. For the process to be stationary, we require that the coefficient matrices Ψ_s be square summable in the sense that each of the $m \times m$ sequences $\psi_{ij,s}$ is square summable, i.e., $\sum_{s=0}^{\infty} \psi_{ij,s}^2 < \infty$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$. Here and in the following an infinite sum of random variables is defined as the limit in quadratic mean of the finite partial sums. Thus, $\mathbf{Z}_t$ in (16.2.1) or (16.2.3) is defined such that

$$E\left[\left(\dot{\mathbf{Z}}_t - \sum_{j=0}^{n} \Psi_j \mathbf{a}_{t-j}\right)'\left(\dot{\mathbf{Z}}_t - \sum_{j=0}^{n} \Psi_j \mathbf{a}_{t-j}\right)\right] \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty. \quad (16.2.4)$$

Another useful form to express a vector process is the autoregressive representation, in which we regress the value of $\mathbf{Z}$ at time t on its own past values plus a vector of random shocks, i.e.,

$$\begin{aligned} \dot{\mathbf{Z}}_t &= \Pi_1 \dot{\mathbf{Z}}_{t-1} + \Pi_2 \dot{\mathbf{Z}}_{t-2} + \cdots + \mathbf{a}_t \\ &= \sum_{s=1}^{\infty} \Pi_s \mathbf{Z}_{t-s} + \mathbf{a}_t \end{aligned} \quad (16.2.5)$$

or, in terms of the backshift operator,

$$\Pi(B)\dot{\mathbf{Z}}_t = \mathbf{a}_t, \quad (16.2.6)$$

where

$$\Pi(B) = I - \sum_{s=1}^{\infty} \Pi_s B^s \quad (16.2.7)$$

and the Π_s are $m \times m$ autoregressive coefficient matrices. Let $\Pi_s = [\Pi_{ij,s}]$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$. Also, $\Pi_{ij,0} = 1$ if $i = j$ and $\Pi_{ij,0} = 0$ if $i \neq j$, i.e., $\Pi_0 = I$. We have $\Pi(B) = [\Pi_{ij}(B)]$, where $\Pi_{ij}(B) = \Pi_{ij,0} - \sum_{s=1}^{\infty} \Pi_{ij,s} B^s$. The vector process is said to be invertible if the autoregressive coefficient matrices Π_s are absolutely summable, i.e., $\sum_{s=0}^{\infty} |\Pi_{ij,s}| < \infty$ for all i and j.

An invertible process is not necessarily stationary. For a vector process with an invertible autoregressive representation to be stationary, we require that no zeros of the determinant of the autoregressive matrix polynomial, denoted by $|\Pi(B)|$, lie on or inside the unit circle, i.e., $|\Pi(B)| \neq 0$ for $|B| \leq 1$. Similarly, a stationary process is not necessarily invertible. For a vector process with a stationary moving average representation to be invertible, no zeros of the determinant of the moving average matrix polynomial should lie on or inside the unit circle, i.e., $|\Psi(B)| \neq 0$ for $|B| \leq 1$.

16.3 The Vector Autoregressive Moving Average Process

A useful class of parsimonious models is the vector autoregressive moving average ARMA(p, q) process

$$\Phi_p(B)\dot{\mathbf{Z}}_t = \Theta_q(B)\mathbf{a}_t, \tag{16.3.1}$$

where

$$\Phi_p(B) = \Phi_0 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$$

and

$$\Theta_q(B) = \Theta_0 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q$$

are the autoregressive and moving average matrix polynomials of orders p and q, respectively, and Φ_0 and Θ_0 are nonsingular $m \times m$ matrices. For any nondegenerate case where the covariance matrix Σ of $\mathbf{a}_t$ is positive definite, we assume in the following discussions, with no loss of generality, that $\Phi_0 = \Theta_0 = I$, the $m \times m$ identity matrix. When $p = 0$, the process becomes a vector MA(q) model

$$\dot{\mathbf{Z}}_t = \mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1} - \cdots - \Theta_q \mathbf{a}_{t-q}. \tag{16.3.2}$$

When $q = 0$, the process becomes a vector AR(p) model

$$\dot{\mathbf{Z}}_t = \Phi_1 \dot{\mathbf{Z}}_{t-1} + \cdots + \Phi_p \dot{\mathbf{Z}}_{t-p} + \mathbf{a}_t. \tag{16.3.3}$$

The process is stationary if the zeros of the determinantal polynomial $|\Phi_p(B)|$ are outside the unit circle. In this case, we can write

$$\dot{\mathbf{Z}}_t = \Psi(B)\mathbf{a}_t, \tag{16.3.4}$$

where

$$\begin{aligned}\Psi(B) &= [\Phi_p(B)]^{-1}\Theta_q(B) \\ &= \sum_{s=0}^{\infty} \Psi_s B^s\end{aligned} \tag{16.3.5}$$

such that the sequence Ψ_s is square summable. The process is invertible if the zeros of the determinantal polynomial $|\Theta_q(B)|$ are outside the unit circle, enabling us to write

$$\Pi(B)\dot{\mathbf{Z}}_t = \mathbf{a}_t \tag{16.3.6}$$

with

$$\Pi(B) = [\Theta_q(B)]^{-1}\Phi_p(B) = I - \sum_{s=1}^{\infty}\Pi_s B^s \tag{16.3.7}$$

such that the sequence Π_s is absolutely summable.

Recall from Section 3.2 that, for univariate time series, a given autocovariance or autocorrelation function may correspond to more than one ARMA(p, q) model. To ensure the unique representation, we have imposed invertibility conditions in the model selection. Similarly, in the vector ARMA model in (16.3.1), left-multiplying both sides of (16.3.1) by an arbitrary nonsingular matrix or by a matrix polynomial in B yields a class of equivalent models that all have identical covariance matrix structures. Thus, without further restrictions, the model in (16.3.1) is not identifiable in the sense that we cannot uniquely determine the values of p and q and the coefficient matrices Φ_i and Θ_j from the covariance matrices of $\mathbf{Z}_t$.

The problem of identifiability for vector ARMA models has been studied by Hannan (1969, 1970, 1976, 1979). Hannan considers the class of stationary processes satisfying the following conditions:

A. The only common left divisors of $\Phi_p(B)$ and $\Theta_q(B)$ are unimodular ones, i.e., if $\Phi_p(B) = C(B)H(B)$ and $\Theta(B) = C(B)K(B)$, then the determinant $|C(B)|$ is a constant; and
B. The zeros of $|\Phi_p(B)|$ (possibly after some proper differencing and transformations) are outside the unit circle and the zeros of $|\Theta_q(B)|$ are on or outside the unit circle.

He then argues that with the above conditions the model in (16.3.1) may be identified by any one of the following procedures:

1. From each class of equivalent models choose those with minimal moving average order q. Among these choose one with minimal autoregressive order p. The resulting representation will be unique if the rank $[\Phi_p\Theta_q]$ is equal to m.
2. Represent $\Phi_p(B)$ in lower triangular form. If the order of $\phi_{ij}(B)$, the (i, j) element of $\Phi_p(B)$, is less than or equal to the order of $\phi_{ii}(B)$ for all $i = 1, \ldots, m$, and $j = 1, \ldots, m$, then the model is identifiable.
3. Represent $\Phi_p(B)$ in a form such that $\Phi_p(B) = \phi_p(B)I$, where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is a univariate AR polynomial of order p. The model is identifiable if $\phi_p \neq 0$.

Procedure 1 is simpler and more commonly used than procedures 2 and 3. Other important phenomena on a vector ARMA representation are presented in Chapter 17.

Vector AR(1) Models The vector AR(1) model is given by

$$(I - \Phi_1 B)\dot{\mathbf{Z}}_t = \mathbf{a}_t \tag{16.3.8a}$$

or

$$\dot{\mathbf{Z}}_t = \Phi_1\dot{\mathbf{Z}}_{t-1} + \mathbf{a}_t. \tag{16.3.8b}$$

For $m = 2$,

$$\begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{bmatrix} - \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \dot{Z}_{1,t-1} \\ \dot{Z}_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \tag{16.3.9}$$

or

$$\begin{aligned} \dot{Z}_{1,t} &= \phi_{11}\dot{Z}_{1,t-1} + \phi_{12}\dot{Z}_{2,t-1} + a_{1t} \\ \dot{Z}_{2,t} &= \phi_{21}\dot{Z}_{1,t-1} + \phi_{22}\dot{Z}_{2,t-1} + a_{2t}. \end{aligned} \tag{16.3.10}$$

Thus, apart from current shocks, each $\dot{Z}_{i,t}$ depends not only lagged values of $\dot{Z}_{i,t}$ but also lagged values of other variables $\dot{Z}_{j,t}$. For example, if $\dot{Z}_{1,t}$ and $\dot{Z}_{2,t}$ represent the sales and the advertising expenditures (measured from some level μ) of a company at time t, then Equation (16.3.10) implies that current sales depend not only on the previous sales but also on the advertising expenditure at the last time period. Moreover, there is a feedback relationship between two series. Thus, the current advertising expenditure will also be influenced by the sales performance at the previous time period.

Vector ARMA and Transfer Function Models If $\phi_{12} = 0$ in (16.3.9), then we have

$$\begin{bmatrix} 1 - \phi_{11}B & 0 \\ -\phi_{21}B & 1 - \phi_{22}B \end{bmatrix} \begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \tag{16.3.11}$$

or

$$\begin{cases} \dot{Z}_{1,t} = \dfrac{1}{1 - \phi_{11}B} a_{1,t} \\ \dot{Z}_{2,t} = \dfrac{\phi_{21}B}{(1 - \phi_{22}B)} \dot{Z}_{1,t} + \dfrac{1}{(1 - \phi_{22}B)} a_{2,t}. \end{cases} \tag{16.3.12}$$

One should not, however, mistakenly regard (16.3.12) as a causal transfer function model relating the input $Z_{1,t}$ to the output $Z_{2,t}$ with one period delay with effect ϕ_{21}. This is so because the input series $Z_{1,t}$ and the noise series $a_{2,t}$ are correlated, which violates a basic assumption in the transfer function model. To reduce (16.3.12) to a causal transfer function model, we let

$$\begin{cases} a_{1,t} = b_{1,t} \\ a_{2,t} = \alpha a_{1,t} + b_{2,t}, \end{cases} \tag{16.3.13}$$

where α is the regression coefficient of $a_{2,t}$ on $a_{1,t}$. The error term $b_{2,t}$ is independent of $a_{1,t}$ and hence $b_{1,t}$. Thus, we have

$$\begin{aligned}
\dot{Z}_{2,t} &= \frac{\phi_{21}B}{(1-\phi_{22}B)}\dot{Z}_{1,t} + \frac{1}{(1-\phi_{22}B)}a_{2,t} \\
&= \frac{\phi_{21}B}{(1-\phi_{22}B)}\dot{Z}_{1,t} + \frac{1}{(1-\phi_{22}B)}(\alpha a_{1,t} + b_{2,t}) \\
&= \frac{\phi_{21}B}{(1-\phi_{22}B)}\dot{Z}_{1,t} + \frac{\alpha}{(1-\phi_{22}B)}a_{1,t} + \frac{1}{(1-\phi_{22}B)}b_{2,t} \\
&= \frac{\phi_{21}B}{(1-\phi_{22}B)}\dot{Z}_{1,t} + \frac{\alpha}{(1-\phi_{22}B)}(1-\phi_{11}B)\dot{Z}_{1,t} + \frac{1}{(1-\phi_{22}B)}b_{2,t} \\
&= \frac{\alpha + (\phi_{21} - \alpha\phi_{11})B}{(1-\phi_{22}B)}\dot{Z}_{1,t} + \frac{1}{(1-\phi_{22}B)}b_{2,t}. \qquad (16.3.14)
\end{aligned}$$

This result shows that the input series $\dot{Z}_{1,t}$ actually has an instantaneous effect α and a one-period delay effect $(\phi_{21} - \alpha\phi_{11})$. Let the variance-covariance matrix Σ of $[a_{1,t}, a_{2,t}]'$ be

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}. \qquad (16.3.15)$$

It can be easily shown that

$$\alpha = \frac{\sigma_{12}}{\sigma_{11}}, \qquad (16.3.16)$$

$$\text{Var}\begin{bmatrix} b_{1,t} \\ b_{2,t} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} - \dfrac{\sigma_{12}^2}{\sigma_{11}} \end{bmatrix}. \qquad (16.3.17)$$

Thus, treating the model in (16.3.12) as a transfer function model could be very misleading if $\sigma_{12} \neq 0$.

In summary, if the autoregressive matrix polynomial $\Phi_p(B)$ is triangular, a joint multivariate vector AR model can be reduced to a causal transfer function model. More generally, if $\Phi_p(B)$ is block triangular, then a joint multivariate vector model can be reduced to a block transfer function model where a feedback relationship could exist among series within each block.

Stationarity and Invertibility Conditions The vector AR(1) process is clearly invertible. For the process to be stationary, the zeros of the determinantal equation $|I - \Phi_1 B|$ must be outside the unit circle. Letting $\lambda = B^{-1}$, we have

$$|I - \Phi_1 B| = 0 \longleftrightarrow |\lambda I - \Phi_1| = 0. \qquad (16.3.18)$$

Thus, the zeros of $|I - \Phi_1 B|$ are related to the eigenvalues of Φ_1. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be the eigenvalues and $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m$ be the associated eigenvectors of Φ_1, such that $\Phi_1 \mathbf{h}_i = \lambda_i \mathbf{h}_i$ for $i = 1, 2, \ldots, m$. For simplicity, assume that the eigenvectors are linearly independent. Let $\Lambda = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_m]$ and $H = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m]$.

We have $\Phi_1 H = H\Lambda$ and

$$\Phi_1 = H\Lambda H^{-1}. \tag{16.3.19}$$

Now,

$$\begin{aligned} |I - \Phi_1 B| &= |I - H\Lambda H^{-1}B| = |I - H\Lambda B H^{-1}| = |I - \Lambda B| \\ &= \prod_{i=1}^{m}(1 - \lambda_i B). \end{aligned} \tag{16.3.20}$$

Hence, the zeros of $|I - \Phi_1 B|$ are outside the unit circle if and only if all the eigenvalues λ_i are inside the unit circle. This result implies that an equivalent condition for stationarity of a vector AR(1) process is that all the eigenvalues of Φ_1 lie inside the unit circle, i.e., $|\lambda_i| < 1$ for $i = 1, 2, \ldots, m$. In such case,

$$\lim_{n\to\infty} \Lambda^n = \lim_{n\to\infty} \text{diag}[\lambda_1^n, \lambda_2^n, \ldots, \lambda_m^n] = \mathbf{0} \tag{16.3.21}$$

and

$$\begin{aligned} \lim_{n\to\infty} \Phi_1^n &= \lim_{n\to\infty} H\Lambda H^{-1} \cdot H\Lambda H^{-1} \cdot \cdots \cdot H\Lambda H^{-1} \\ &= \lim_{n\to\infty} H\Lambda^n H^{-1} = \mathbf{0} \end{aligned} \tag{16.3.22}$$

It follows that

$$\lim_{n\to\infty}(I - \Phi_1 B)(I + \Phi_1 B + \Phi_1^2 B^2 + \cdots + \Phi_1^{n-1} B^{n-1}) = \lim_{n\to\infty}(I - \Phi_1^n B^n) = I$$

and

$$(I - \Phi_1 B)^{-1} = (I + \Phi_1 B + \Phi_1^2 B^2 + \cdots). \tag{16.3.23}$$

Thus, we can write

$$\dot{\mathbf{Z}}_t = [\Phi_1(B)]^{-1}\mathbf{a}_t \tag{16.3.24a}$$

$$= (I - \Phi_1 B)^{-1}\mathbf{a}_t \tag{16.3.24b}$$

$$= \sum_{s=0}^{\infty} \Phi_1^s \mathbf{a}_{t-s}, \tag{16.3.24c}$$

where $\Phi_0 = I$ and the Φ_1^s matrix decreases toward the $\mathbf{0}$ matrix as s increases. Note, however, that

$$[\Phi_1(B)]^{-1} = \frac{1}{|\Phi_1(B)|}\Phi_1^\dagger(B), \tag{16.3.25}$$

where $\Phi_1^\dagger(B)$ is the adjoint matrix of $\Phi_1(B)$. Hence, we can rewrite (16.3.24a) as

$$|\Phi_1(B)|\dot{\mathbf{Z}}_t = \Phi_1^\dagger(B)\mathbf{a}_t. \tag{16.3.26}$$

Because the order of the polynomial $|\Phi_1(B)|$ is m and the maximum order of the polynomial $\Phi_1^\dagger(B)$ is $(m-1)$, (16.3.26) implies that each $Z_{i,t}$ from a joint multivariate AR(1) model is marginally a univariate ARMA process of order up to $(m, m-1)$. Because cancellations may occur for some common factors between $|\Phi_1(B)|$ and $\Phi_{ij}^\dagger(B)$, the (i, j) element of the matrix $\Phi^\dagger(B)$, the autoregressive components of the univariate series may not be identical.

Recall that in the above derivation we assume that the eigenvectors of Φ_1 are linearly independent. In general, the eigenvectors of Φ_1 may not be all linearly independent. In such case, however, there still exists a nonsingular matrix H (see, e.g., Gantmacher [1960] and Noble [1969]) so that

$$\Lambda = \mathbf{H}^{-1}\Phi_1\mathbf{H} = \begin{bmatrix} \lambda_1 & \delta_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \delta_2 & \ddots & & 0 \\ & & & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \delta_{m-2} & 0 \\ & & & & \lambda_{m-1} & \delta_{m-1} \\ 0 & & \cdots & & 0 & \lambda_m \end{bmatrix}, \tag{16.3.27}$$

where the adjacent λ's may be equal and the δ's are either 0 or 1. The above results still follow if $|\lambda_i| < 1$ for $i = 1, 2, \ldots,$ and m.

16.3.1 Covariance Matrix Function for the Vector AR(1) Model

$$\begin{aligned} \Gamma(k) &= E[\dot{\mathbf{Z}}_{t-k}\dot{\mathbf{Z}}_t'] \\ &= E[\dot{\mathbf{Z}}_{t-k}(\Phi_1\dot{\mathbf{Z}}_{t-1} + \mathbf{a}_t)'] \\ &= E[\dot{\mathbf{Z}}_{t-k}\dot{\mathbf{Z}}_{t-1}'\Phi_1' + \dot{\mathbf{Z}}_{t-k}\mathbf{a}_t'] \\ &= \begin{cases} \Gamma(-1)\Phi_1' + \Sigma, & k = 0, \\ \Gamma(k-1)\Phi_1' = \Gamma(0)(\Phi_1')^k, & k \geq 1, \end{cases} \end{aligned} \tag{16.3.28}$$

where we note that $E(\dot{\mathbf{Z}}_{t-k}\mathbf{a}_t') = 0$ for $k \geq 1$. For $k = 1$, we have $\Gamma(1) = \Gamma(0)\Phi_1'$; hence,

$$\Phi_1 = \Gamma'(1)\Gamma^{-1}(0) \tag{16.3.29}$$

and

$$\begin{aligned}\Sigma &= \Gamma(0) - \Gamma(-1)\Gamma^{-1}(0)\Gamma(1) \\ &= \Gamma(0) - \Gamma'(1)\Gamma^{-1}(0)\Gamma(1).\end{aligned} \tag{16.3.30}$$

Given the covariance matrix function $\Gamma(k)$, Equations (16.3.29) and (16.3.30) can be used to calculate Σ and Φ_1. Conversely, given Φ_1 and Σ, we can compute $\Gamma(0)$ by noting that (16.3.30) implies that

$$\begin{aligned}\Sigma &= \Gamma(0) - \Gamma'(1)\Gamma^{-1}(0)\Gamma(0)\Gamma^{-1}(0)\Gamma(1) \\ &= \Gamma(0) - \Phi_1\Gamma(0)\Phi_1'.\end{aligned}$$

Thus,

$$\Gamma(0) = \Phi_1\Gamma(0)\Phi_1' + \Sigma. \tag{16.3.31}$$

Solving Equation (16.3.31) for $\Gamma(0)$, we let $\text{vec}(\boldsymbol{X})$ be the column vector formed by stacking the columns of a matrix $\boldsymbol{X}$. For example, if

$$\boldsymbol{X} = \begin{bmatrix} 2 & 6 \\ 3 & 5 \\ 1 & 4 \end{bmatrix},$$

then

$$\text{vec}(\boldsymbol{X}) = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 6 \\ 5 \\ 4 \end{bmatrix}. \tag{16.3.32}$$

Using the result (see Searle [1982, p. 333]),

$$\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\,\text{vec}(\boldsymbol{B}), \tag{16.3.33}$$

where $\otimes$ denotes the Kronecker product, we have from (16.3.31)

$$\text{vec}(\Gamma(0)) = \text{vec}(\Phi_1\Gamma(0)\Phi_1') + \text{vec}(\Sigma) = (\Phi_1 \otimes \Phi_1)\,\text{vec}(\Gamma(0)) + \text{vec}(\Sigma)$$

or

$$\text{vec}(\Gamma(0)) = [\mathbf{I} - \Phi_1 \otimes \Phi_1]^{-1}\,\text{vec}(\Sigma). \tag{16.3.34}$$

The use of (16.3.34) is given in Example 16.1. The function $\Gamma(k)$ for $k \geq 1$ can then be obtained from the second part of (16.3.28).

EXAMPLE 16.1 Consider the two-dimensional vector AR(1) process

$$(\mathbf{I} - \Phi_1 B)\dot{\mathbf{Z}}_t = \mathbf{a}_t,$$

where

$$\Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix},$$

$$\Sigma = E(\mathbf{a}_t \mathbf{a}_t') = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix},$$

and

$$\Gamma(0) = E(\dot{\mathbf{Z}}_t \dot{\mathbf{Z}}_t') = \begin{bmatrix} \gamma_{11}(0) & \gamma_{12}(0) \\ \gamma_{21}(0) & \gamma_{22}(0) \end{bmatrix}.$$

From (16.3.34), we have

$$\begin{bmatrix} \gamma_{11}(0) \\ \gamma_{21}(0) \\ \gamma_{12}(0) \\ \gamma_{22}(0) \end{bmatrix} = \left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi_{11}\phi_{11} & \phi_{11}\phi_{12} & \phi_{12}\phi_{11} & \phi_{12}\phi_{12} \\ \phi_{11}\phi_{21} & \phi_{11}\phi_{22} & \phi_{12}\phi_{21} & \phi_{12}\phi_{22} \\ \phi_{21}\phi_{11} & \phi_{21}\phi_{12} & \phi_{22}\phi_{11} & \phi_{22}\phi_{12} \\ \phi_{21}\phi_{21} & \phi_{21}\phi_{22} & \phi_{22}\phi_{21} & \phi_{22}\phi_{22} \end{bmatrix} \right]^{-1} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}.$$

Specifically, when

$$\Phi_1 = \begin{bmatrix} .5 & .4 \\ .1 & .8 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix},$$

the eigenvalues of Φ_1 are .4 and .9, which are inside the unit circle; hence, the process is stationary. Because the eigenvalues of $(A \otimes B)$ are products of eigenvalues of A with those of

B, it follows that all eigenvalues of $(\Phi_1 \otimes \Phi_1)$ are inside the unit circle; hence, the inverse in (16.3.34) exists. Thus,

$$\begin{bmatrix}\gamma_{11}(0)\\ \gamma_{21}(0)\\ \gamma_{12}(0)\\ \gamma_{22}(0)\end{bmatrix} = \left[\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} - \begin{bmatrix}.25&.2&.2&.16\\.05&.4&.04&.32\\.05&.04&.4&.32\\.01&.08&.08&.64\end{bmatrix}\right]^{-1}\begin{bmatrix}1\\.6\\.6\\1\end{bmatrix}$$

$$= \begin{bmatrix}1.4786&.85826&.85826&2.3114\\.21456&2.1271&.56456&2.6343\\.21456&.56456&2.1271&2.6343\\.14446&.65857&.65857&4.2488\end{bmatrix}\begin{bmatrix}1\\.6\\.6\\1\end{bmatrix} = \begin{bmatrix}4.8199\\4.4639\\4.4639\\5.1835\end{bmatrix},$$

$$\Gamma(0) = \begin{bmatrix}4.8199&4.4639\\4.4639&5.1835\end{bmatrix},$$

and, by (16.3.28),

$$\Gamma(k) = \begin{bmatrix}4.8199&4.4639\\4.4639&5.1835\end{bmatrix}\begin{bmatrix}.5&.1\\.4&.8\end{bmatrix}^k.$$

16.3.2 Vector AR(p) Models

The general vector AR(p) process

$$(I - \Phi_1 B - \cdots - \Phi_p B^p)\dot{\mathbf{Z}}_t = \mathbf{a}_t \tag{16.3.35a}$$

or

$$\dot{\mathbf{Z}}_t = \Phi_1 \dot{\mathbf{Z}}_{t-1} + \cdots + \Phi_p \dot{\mathbf{Z}}_{t-p} + \mathbf{a}_t \tag{16.3.35b}$$

is obviously invertible. For the process to be stationary we require that the zeros of $|I - \Phi_1 B - \cdots - \Phi_P B^P|$ lie outside the unit circle or, equivalently, the roots of

$$|\lambda^p I - \lambda^{p-1}\Phi_1 - \cdots - \Phi_p| = 0 \tag{16.3.36}$$

be inside the unit circle. In such case, we can write

$$\dot{\mathbf{Z}} = [\Phi_p(B)]^{-1}\mathbf{a}_t \tag{16.3.37a}$$

$$= \sum_{s=0}^{\infty} \Psi_s \mathbf{a}_{t-s}, \tag{16.3.37b}$$

where the Ψ_s weights are square summable and can be obtained by equating the coefficients of B^j in the matrix equation

$$(I - \Phi_1 B - \cdots - \Phi_p B^p)(I + \Psi_1 B + \Psi_2 B^2 + \cdots) = I, \tag{16.3.38}$$

i.e.,

$$\begin{aligned} \Psi_1 &= \Phi_1 \\ &\vdots \\ \Psi_j &= \Phi_1 \Psi_{j-1} + \cdots + \Phi_p \Psi_{j-p}, \qquad j \geq p. \end{aligned}$$

Because

$$[\Phi_p(B)]^{-1} = \frac{1}{|\Phi_p(B)|} \Phi_p^+(B), \tag{16.3.39}$$

where $\Phi_p^+(B) = [\Phi_{ij}^+(B)]$ is the adjoint matrix of $\Phi_p(B)$, we can write (16.3.37a) as

$$|\Phi_p(B)| \dot{\mathbf{Z}}_t = \Phi_p^+(B) \mathbf{a}_t. \tag{16.3.40}$$

Now, $|\Phi_P(B)|$ is a polynomial in B at most of order mp, and each $\phi_{ij}^+(B)$ is a polynomial in B at most of order $(m - 1)p$. Thus, each individual component series follows a univariate ARMA model up to a maximum order $(mp, (m - 1)p)$. The order can be much less if there are common factors between the AR and MA polynomials.

The covariance matrix function is obtained by left-multiplying $\dot{\mathbf{Z}}_{t-k}$ on both sides of the transposed equation of (16.3.35a) and then taking expectations, i.e.,

$$E[\dot{\mathbf{Z}}_{t-k}(\dot{\mathbf{Z}}_t' - \dot{\mathbf{Z}}_{t-1}'\Phi_1' - \cdots - \dot{\mathbf{Z}}_{t-p}'\Phi_p')] = E[\dot{\mathbf{Z}}_{t-k}\mathbf{a}_t']. \tag{16.3.41}$$

Thus,

$$\begin{aligned} &k = 0, \Gamma(0) - \Gamma'(1)\Phi_1' - \Gamma'(2)\Phi_2' - \Gamma'(3)\Phi_3' - \cdots - \Gamma'(p)\Phi_p' = \Sigma \\ &k = 1, \Gamma(1) - \Gamma(0)\Phi_1' - \Gamma'(1)\Phi_2' - \Gamma'(2)\Phi_3' - \cdots - \Gamma'(p-1)\Phi_p' = \mathbf{0} \\ &k = 2, \Gamma(2) - \Gamma(1)\Phi_1' - \Gamma(0)\Phi_2' - \Gamma'(1)\Phi_3' - \cdots - \Gamma'(p-2)\Phi_p' = \mathbf{0} \\ &\quad\vdots \qquad\qquad\qquad\qquad \vdots \\ &k = p, \Gamma(p) - \Gamma(p-1)\Phi_1' - \Gamma(p-2)\Phi_2' - \cdots - \Gamma(0)\Phi_p' = \mathbf{0} \\ &k \geq p, \Gamma(k) - \Gamma(k-1)\Phi_1' - \Gamma(k-2)\Phi_2' - \cdots - \Gamma(k-p)\Phi_p' = \mathbf{0}, \end{aligned}$$

where $\mathbf{0}$ is the $m \times m$ zero matrix. More compactly, for $k = 1, \ldots, p$, we obtain the following generalized Yule–Walker matrix equations:

$$\begin{bmatrix} \Gamma(0) & \Gamma'(1) & \Gamma'(2) & \cdots & \Gamma'(p-1) \\ \Gamma(1) & \Gamma(0) & \Gamma'(1) & \cdots & \Gamma'(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma(p-1) & \Gamma(p-2) & \Gamma(p-3) & \cdots & \Gamma(0) \end{bmatrix} \begin{bmatrix} \Phi_1' \\ \Phi_2' \\ \vdots \\ \Phi_p' \end{bmatrix} = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(p) \end{bmatrix}.$$

16.3.3 Vector MA(1) Models

The vector MA(1) models are given by

$$\dot{\mathbf{Z}}_t = (I - \Theta_1 B)\mathbf{a}_t, \tag{16.3.42}$$

where the $\mathbf{a}_t$ are $m \times 1$ vector white noise with mean zero and variance-covariance matrix Σ. For $m = 2$, we have

$$\begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-2} \end{bmatrix}. \tag{16.3.43}$$

The covariance matrix function of $\dot{\mathbf{Z}}_t$ is

$$\begin{aligned} \Gamma(0) &= E(\dot{\mathbf{Z}}_t \dot{\mathbf{Z}}_t') = E[(I - \Theta_1 B)\mathbf{a}_t][(I - \Theta_1 B)\mathbf{a}_t]' \\ &= E[(\mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1})(\mathbf{a}_t' - \mathbf{a}_{t-1}'\Theta_1')] \\ &= \Sigma + \Theta_1 \Sigma \Theta_1' \end{aligned} \tag{16.3.44}$$

$$\begin{aligned} \Gamma(k) &= E\left(\dot{\mathbf{Z}}_t \dot{\mathbf{Z}}_{t+k}'\right) \\ &= E[\mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1}][\mathbf{a}_{t+k}' - \mathbf{a}_{t+k-1}\Theta_1'] \\ &= \begin{cases} -\Sigma\Theta_1', & k = 1, \\ -\Theta_1\Sigma, & k = -1, \\ \mathbf{0}, & |k| > 1. \end{cases} \end{aligned} \tag{16.3.45}$$

Note that $\Gamma(-1) = \Gamma'(1)$ and that $\Gamma(k)$ cuts off after lag 1, a behavior that parallels the univariate MA(1) process.

To solve for Θ_1 and Σ in terms of $\Gamma(k)$'s, we note that, from (16.3.44), and (16.3.45)

$$\Sigma = \Gamma(0) - \Theta_1 \Sigma \Theta_1' = \Gamma(0) + \Gamma'(1)\Theta_1'$$
$$\Sigma\Theta_1' = \Gamma(0)\Theta_1' + \Gamma'(1)[\Theta_1']^2.$$

Hence,

$$\Theta_1^2\Gamma(1) + \Theta_1\Gamma(0) + \Gamma'(1) = \mathbf{0}, \tag{16.3.46}$$

which is a quadratic matrix equation in Θ_1. A method to solve such an equation can be found in Gantmacher (1960).

The vector MA(1) process is clearly stationary. For the process to be invertible, the zeros of $|I - \Theta_1 B|$ must be outside the unit circle or, equivalently, the eigenvalues of Θ_1 should be all less than 1 in absolute value. The proof is exactly identical to that of the stationarity condition for the vector AR(1) process. In such a case,

$$(I - \Theta_1 B)^{-1} = (I + \Theta_1 B + \Theta_1^2 B + \cdots) \tag{16.3.47}$$

and we can write

$$\mathbf{Z}_t + \sum_{s=1}^{\infty} \Theta_1^s \mathbf{Z}_{t-s} = \mathbf{a}_t, \tag{16.3.48}$$

where the Θ_1^s matrix decreases toward the $\mathbf{0}$ matrix as s increases.

16.3.4 Vector MA(q) Models

The general vector MA(q) process is given by

$$\dot{\mathbf{Z}}_t = (I - \Theta_1 B - \cdots - \Theta_q B^q)\mathbf{a}_t. \tag{16.3.49}$$

The covariance matrix function is

$$\begin{aligned}\Gamma(k) &= E[\mathbf{a}_t - \Theta_1\mathbf{a}_{t-1} - \cdots - \Theta_q\mathbf{a}_{t-q}][\mathbf{a}'_{t+k} - \mathbf{a}'_{t+k-1}\Theta'_1 - \cdots - \mathbf{a}'_{t+k-q}\Theta'_q] \\ &= \begin{cases} \sum_{j=0}^{q-k}\Theta_j\Sigma\Theta'_{j+k}, & \text{for } k = 0, 1, \ldots, q, \\ \mathbf{0}, & k > q, \end{cases}\end{aligned} \tag{16.3.50}$$

where $\Theta_0 = I$ and $\Gamma(-k) = \Gamma'(k)$. Note that $\Gamma(k)$ cuts off after lag q. The process is always stationary. If it is also invertible, then we can write

$$\Pi(B)\dot{\mathbf{Z}}_t = \mathbf{a}_t \tag{16.3.51}$$

with

$$\Pi(B) = [\Theta_q(B)]^{-1} = \frac{1}{|\Theta_q(B)|}\Theta_q^+(B) \tag{16.3.52}$$

so that the Π_s are absolutely summable. Because the elements of the adjoint matrix $\Theta_q^+(B)$ are polynomials in B of maximum order $(m-1)q$, the only requirement needed for the model to be invertible is that all the zeros of $|\Theta_q(B)|$ be outside the unit circle. The Π_s can be obtained by equating coefficients of B^j in

$$\Pi(B)\Theta_q(B) = I,$$
$$[I - \Pi_1 B - \Pi_2 B^2 - \cdots][I - \Theta_1 B - \cdots - \Theta_q B^q] = I. \tag{16.3.53}$$

Hence, Π_s can be calculated recursively from the Θ_j as

$$\Pi_s = \Pi_{s-1}\Theta_1 + \cdots + \Pi_{s-q}\Theta_q, \qquad s = 1, 2, \ldots, \tag{16.3.54}$$

where $\Pi_0 = -I$ and $\Pi_j = 0$ for $j < 0$.

16.3.5 Vector ARMA(1, 1) Models

The vector ARMA(1, 1) model is given by

$$(I - \Phi_1 B)\dot{\mathbf{Z}}_t = (I - \Theta_1 B)\mathbf{a}_t. \tag{16.3.55}$$

The model is stationary if the zeros of the determinantal polynomial $|I - \Phi_1 B|$ are outside the unit circle or if all the eigenvalues of Φ_1 are inside the unit circle. In such a case, we can write

$$\dot{\mathbf{Z}}_t = \sum_{s=0}^{\infty} \Psi_s \mathbf{a}_{t-s}, \tag{16.3.56}$$

where the Ψ_s weights are obtained by equating the coefficients of B^j in the matrix equation

$$(I - \Phi_1 B)(I + \Psi_1 B + \Psi_2 B^2 + \cdots) = (I - \Theta_1 B),$$

i.e.,

$$\Psi_j = \Phi_1 \Psi_{j-1} = \Phi_1^{j-1}(\Phi_1 - \Theta_1), \qquad j \geq 1. \tag{16.3.57}$$

The process is invertible if the zeros of $|I - \Theta_1 B|$ are outside the unit circle or if all the eigenvalues of Θ_1 are inside the unit circle. In such a case, we can write

$$\dot{\mathbf{Z}}_t = \Pi_1 \dot{\mathbf{Z}}_{t-1} + \Pi_2 \dot{\mathbf{Z}}_{t-2} + \cdots + \mathbf{a}_t \tag{16.3.58}$$

such that

$$(I - \Phi_1 B) = (I - \Theta_1 B)(I - \Pi_1 B - \Pi_2 B^2 - \cdots)$$

and

$$\Pi_j = \Theta_1^{j-1}(\Phi_1 - \Theta_1), \qquad j \geq 1. \tag{16.3.59}$$

For the covariance matrix function, we consider

$$E[\dot{\mathbf{Z}}_{t-k}(\dot{\mathbf{Z}}_t' - \dot{\mathbf{Z}}_{t-1}'\Phi_1')] = E[\dot{\mathbf{Z}}_{t-k}(\mathbf{a}_t' - \mathbf{a}_{t-1}'\Theta_1')]$$

and note that

$$\begin{aligned} E[\dot{\mathbf{Z}}_t(\mathbf{a}_{t-1}'\Theta_1')] &= E[(\Phi_1\dot{\mathbf{Z}}_{t-1} + \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1})(\mathbf{a}_{t-1}'\Theta_1')] \\ &= \Phi_1\Sigma\Theta_1' - \Theta_1\Sigma\Theta_1'. \end{aligned}$$

Hence, we obtain

$$\begin{cases} \Gamma(0) - \Gamma'(1)\Phi_1' = \Sigma - (\Phi_1 - \Theta_1)\Sigma\Theta_1', & k = 0, \\ \Gamma(1) - \Gamma(0)\Phi_1' = -\Sigma\Theta_1', & k = 1, \\ \Gamma(k) - \Gamma(k-1)\Phi_1' = \mathbf{0}, & k \geq 2. \end{cases} \tag{16.3.60}$$

Using the first two equations in (16.3.60), we have

$$\Gamma(0) - \Phi_1\Gamma(0)\Phi_1' = \Sigma + (\Phi_1 - \Theta_1)\Sigma(\Phi_1 - \Theta_1)' - \Phi_1\Sigma\Phi_1'$$

and

$$\Gamma(0) = \Phi_1\Gamma(0)\Phi_1' + \Sigma + (\Phi_1 - \Theta_1)\Sigma(\Phi_1 - \Theta_1)' - \Phi_1\Sigma\Phi_1'. \tag{16.3.61}$$

When the parameter matrices Φ_1, Θ_1, and Σ are known, Equation (16.3.61) is of the same form as Equation (16.3.31). Thus, $\Gamma(0)$ can be solved in the same way as illustrated in Example 16.1 and is given by

$$\operatorname{vec}(\Gamma(0)) = [I - \Phi_1 \otimes \Phi_1]^{-1} \operatorname{vec}(\Sigma + (\Phi_1 - \Theta_1)\Sigma(\Phi_1 - \Theta_1)' - \Phi_1\Sigma\Phi_1'). \tag{16.3.62}$$

The other covariance matrices $\Gamma(k)$ can then be calculated through (16.3.60).

Conversely, given the covariance matrix function $\Gamma(k)$, we can obtain the parameter matrices as follows. First, from the third equation of (16.3.60), for $k = 2$, we obtain

$$\Phi_1 = \Gamma'(2)[\Gamma'(1)]^{-1}. \tag{16.3.63}$$

Next, using the first two equations in (16.3.60), we get

$$\Sigma = \Gamma(0) - \Gamma'(1)\Phi_1' - (\Phi_1 - \Theta_1)[\Gamma(1) - \Gamma(0)\Phi_1']. \tag{16.3.64}$$

Hence,

$$\Theta_1\Sigma = \Theta_1[\Gamma(0) - \Gamma'(1)\Phi_1' - \Phi_1\Gamma(1) + \Phi_1\Gamma(0)\Phi_1'] + \Theta_1^2[\Gamma(1) - \Gamma(0)\Phi_1']. \tag{16.3.65}$$

Substituting the second equation of (16.3.60) into (16.3.65), we have

$$\Theta_1^2 A_1 + \Theta_1 A_2 + A_1' = \mathbf{0}, \tag{16.3.66}$$

where

$$A_1 = \Gamma(1) - \Gamma(0)\Phi_1',$$
$$A_2 = \Gamma(0) - \Gamma'(1)\Phi_1' - \Phi_1\Gamma(1) + \Phi_1\Gamma(0)\Phi_1'.$$

Equation (16.3.66) is of the same form as Equation (16.3.46) and can be solved similarly for Θ_1. Having Φ_1 and Θ_1, Σ can be calculated from (16.3.64)

16.4 Nonstationary Vector Autoregressive Moving Average Models

In the analysis of time series, it is very common to observe series that exhibit nonstationary behavior. One useful way to reduce nonstationary series to stationary series is by differencing. For example, in univariate time series, a nonstationary series Z_t is reduced to a stationary series $(1 - B)^d Z_t$ for some $d > 0$, so we can write

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t, \tag{16.4.1}$$

where $\phi_p(B)$ is a stationary AR operator. A seemingly natural extension of (16.4.1) to the vector process is

$$\Phi_p(B)(I - IB)^d \mathbf{Z}_t = \Theta_q(B)\mathbf{a}_t \tag{16.4.2}$$

or, equivalently,

$$\Phi_p(B)(1 - B)^d \mathbf{Z}_t = \Theta_q(B)\mathbf{a}_t. \tag{16.4.3}$$

This extension, however, implies that all component series are differenced the same number of times. This restriction is obviously unnecessary and undesirable. To be more flexible, we assume that even though $\mathbf{Z}_t$ may be nonstationary, it can be reduced to a stationary vector series by applying the differencing operator $\mathbf{D}(B)$, i.e.,

$$\mathbf{D}(B)\mathbf{Z}_t, \tag{16.4.4}$$

where

$$\mathbf{D}(B) = \begin{bmatrix} (1-B)^{d_1} & 0 & \cdots & 0 & 0 \\ 0 & (1-B)^{d_2} & \ddots & & \\ \vdots & \ddots & \ddots & & \vdots \\ & & & & 0 \\ 0 & \cdots & & 0 & (1-B)^{d_m} \end{bmatrix} \tag{16.4.5}$$

and $(d_1, d_2, \ldots, d_m)$ is a set of nonnegative integers. Thus, we have the nonstationary vector ARMA model for $\mathbf{Z}_t$

$$\Phi_p(B)\mathbf{D}(B)\mathbf{Z}_t = \Theta_q(B)\mathbf{a}_t, \tag{16.4.6}$$

where the zeros of $|\Phi_p(B)|$ and $|\Theta_q(B)|$ are outside the unit circle.

Differencing on vector time series is much more complicated and should be handled carefully. Overdifferencing may lead to complications in model fitting. One should be particularly careful when the orders of differencing for each component series are the same. Box and Tiao (1977) show that identical differencing imposed on a vector process may lead to a noninvertible representation. Furthermore, in some cases, as discussed in Chapter 17, a linear combination of nonstationary series could become stationary. In this circumstance, a model purely based on differences may not even exist. In such a case, one can use the following generalization of the vector ARMA model proposed by Tiao and Box (1981):

$$\Phi_p(B)\mathbf{Z}_t = \Theta_q(B)\mathbf{a}_t, \tag{16.4.7}$$

where the zeros of the determinantal polynomial $|\Phi_p(B)|$ are on or outside the unit circle.

16.5 Identification of Vector Time Series Models

In principle, identification of vector time series models is similar to the identification of univariate time series models discussed in Chapters 6 and 8. For a given observed vector time series $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$, we identify its underlying model from the pattern of its sample correlation and partial correlation matrices after proper transformations are applied to reduce a nonstationary series to a stationary series.

16.5.1 Sample Correlation Matrix Function

Given a vector time series of n observations $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$, and $\mathbf{Z}_n$, we can calculate the sample correlation matrix function

$$\hat{\boldsymbol{\rho}}(k) = [\hat{\rho}_{ij}(k)], \tag{16.5.1}$$

where the $\hat{\rho}_{ij}(k)$ are the sample cross-correlations for the ith and jth component series

$$\hat{\rho}_{ij}(k) = \frac{\sum_{t=1}^{n-k}\left(Z_{i,t} - \bar{Z}_i\right)\left(Z_{j,t+k} - \bar{Z}_j\right)}{\left[\sum_{t=1}^{n}\left(Z_{i,t} - \bar{Z}_i\right)^2 \sum_{t=1}^{n}\left(Z_{j,t} - \bar{Z}_j\right)^2\right]^{1/2}} \tag{16.5.2}$$

and $\bar{Z}_i$ and $\bar{Z}_j$ are the sample means of the corresponding component series. For a stationary vector process, Hannan (1970, p. 228) shows that $\hat{\boldsymbol{\rho}}(k)$ is a consistent estimator that is asymptotically normally distributed. The variance and covariance of these sample cross-correlation $\hat{\rho}_{ij}(k)$ obtained by Bartlett (1966) and given in (14.3.4) and (14.3.5) are rather complicated and depend on the underlying models. If $\rho_{ij}(k) = 0$ for $|k| > q$ for some q, then, from (14.3.5), we have

$$\operatorname{Var}[\hat{\rho}_{ij}(k)] \simeq \frac{1}{n-k}\left[1 + 2\sum_{s=1}^{q}\rho_{ii}(s)\,\rho_{jj}(s)\right], \qquad |k| > q. \tag{16.5.3}$$

When the $\mathbf{Z}_t$ series are white noise, Bartlett's approximation for the asymptotic covariance simplifies to

$$\operatorname{Cov}[\hat{\rho}_{ij}(k), \hat{\rho}_{ij}(k+s)] \simeq \frac{1}{n-k} \tag{16.5.4}$$

and

$$\operatorname{Var}[\hat{\rho}_{ij}(k)] \simeq \frac{1}{n-k}. \tag{16.5.5}$$

For large samples, $(n-k)$ is often replaced by n in the above expressions.

The sample correlation matrix function is very useful in identifying a finite-order moving average model as the correlation matrices are $\mathbf{0}$ beyond lag q for the vector MA(q) process. Unfortunately, the matrices are complex when the dimension of the vector is increased. The crowded numbers often make recognition of patterns difficult. To alleviate the problem, Tiao and Box (1981) introduced a convenient method of summarizing the sample correlations. Their method is to note the symbols $+$, $-$, and $\cdot$ in the (i, j) position of the sample correlation matrix, where $+$ denotes a value greater than 2 times the estimated standard errors, $-$ denotes a value less than -2 times the estimated standard errors, and $\cdot$ denotes a value within 2 estimated standard errors.

16.5.2 Partial Autoregression Matrices

A useful tool for identifying the order of a univariate AR model is the partial autocorrelation function. As defined in Section 2.3, it is the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, \ldots,$ and Z_{t+k-1} has been removed, i.e.,

$$\phi_{kk} = \frac{\operatorname{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{\operatorname{Var}(Z_t - \hat{Z}_t)}\sqrt{\operatorname{Var}(Z_{t+k} - \hat{Z}_{t+k})}}, \tag{16.5.6}$$

where $\hat{Z}_t$ and $\hat{Z}_{t+k}$ are the minimum mean squared error linear regression estimators of Z_t and Z_{t+k} based on $Z_{t+1}, Z_{t+2}, \ldots,$ and Z_{t+k-1}. As shown in Section 2.3, the partial autocorrelation between Z_t and Z_{t+k} is also equal to the last regression coefficient associated with Z_t when regressing Z_{t+k} on its k lagged variables $Z_{t+k-1}, Z_{t+k-2}, \ldots,$ and Z_t. The partial autocorrelation is useful for identifying the order of a univariate AR(p) model because ϕ_{kk} is zero for $|k| > p$. In this section, we consider a generalization of the concept to a vector time series process proposed by Tiao and Box (1981).

Tiao and Box (1981) define the partial autoregression matrix at lag s, denoted by $\mathcal{P}(s)$, to be the last matrix coefficient when the data are fitted to a vector autoregressive process of order s. This definition is a direct extension of the Box and Jenkins (1976, p. 64) definition of the partial autocorrelation function for univariate time series. That is, $\mathcal{P}(s)$ is equal to $\Phi_{s,s}$ in the multivariate linear regression

$$\mathbf{Z}_{t+s} = \Phi_{s,1}\mathbf{Z}_{t+s-1} + \Phi_{s,2}\mathbf{Z}_{t+s-2} + \cdots + \Phi_{s,s}\mathbf{Z}_t + \mathbf{e}_{s,t+s},$$

where $\mathbf{e}_{s,t+s}$ is the error term, and the $m \times m$ matrix coefficients $\Phi_{s,k}$, $k = 1, 2, \ldots, s$, are those that minimize

$$E\left[\left|\mathbf{Z}_{t+s} - \Phi_{s,1}\mathbf{Z}_{t+s-1} - \cdots - \Phi_{s,s}\mathbf{Z}_t\right|^2\right], \tag{16.5.7}$$

where $|\mathbf{V}_t|^2 = \mathbf{V}_t\mathbf{V}_t'$ for the m-dimensional vector $\mathbf{V}_t$.

Minimization of (16.5.7) leads to a multivariate generalization of the Yule–Walker equations in unnormalized form,

$$\begin{bmatrix} \Gamma(0) & \Gamma'(1) & \cdots & \Gamma'(s-1) \\ \Gamma(1) & \Gamma(0) & \cdots & \Gamma'(s-2) \\ \vdots & \vdots & & \vdots \\ \Gamma(s-1) & \Gamma(s-2) & \cdots & \Gamma(0) \end{bmatrix} \begin{bmatrix} \Phi_{s,1}' \\ \Phi_{s,2}' \\ \vdots \\ \Phi_{s,s}' \end{bmatrix} = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(s) \end{bmatrix}. \tag{16.5.8}$$

The Tiao and Box definition of $\mathcal{P}(s)$ leads to solving (16.5.8) for $\Phi_{s,s}'$, and determining the partial autoregression matrix requires solving (16.5.8) for successively higher orders of s. If, for $s \geq 2$, we let

$$\mathbf{A}(s) = \begin{bmatrix} \Gamma(0) & \Gamma'(1) & \cdots & \Gamma'(s-2) \\ \Gamma(1) & \Gamma(0) & \cdots & \Gamma'(s-3) \\ \vdots & \vdots & & \vdots \\ \Gamma(s-2) & \Gamma(s-3) & \cdots & \Gamma(0) \end{bmatrix},$$

$$\mathbf{b}(s) = \begin{bmatrix} \Gamma'(s-1) \\ \Gamma'(s-2) \\ \vdots \\ \Gamma'(1) \end{bmatrix}, \quad \mathbf{c}(s) = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(s-1) \end{bmatrix}, \tag{16.5.9}$$

and

$$\Phi_4'(s-1) = \begin{bmatrix} \Phi_{s,1}' \\ \Phi_{s,2}' \\ \vdots \\ \Phi_{s,s-1}' \end{bmatrix},$$

then the system (16.5.8) can be written as

$$\begin{bmatrix} \mathbf{A}(s) & \mathbf{b}(s) \\ \mathbf{b}'(s) & \Gamma(0) \end{bmatrix} \begin{bmatrix} \Phi_4'(s-1) \\ \Phi_{s,s}' \end{bmatrix} = \begin{bmatrix} \mathbf{c}(s) \\ \Gamma(s) \end{bmatrix}. \tag{16.5.10}$$

Hence,

$$\mathbf{A}(s)\Phi_4'(s-1) + \mathbf{b}(s)\Phi_{s,s}' = \mathbf{c}(s) \tag{16.5.11}$$

$$\mathbf{b}'(s)\Phi_4'(s-1) + \Gamma(0)\Phi_{s,s}' = \Gamma(s). \tag{16.5.12}$$

Equation (16.5.11) implies that

$$\Phi_4'(s-1) = [\mathbf{A}(s)]^{-1}\mathbf{c}(s) - [\mathbf{A}(s)]^{-1}\mathbf{b}(s)\Phi_{s,s}'. \tag{16.5.13}$$

Substituting (16.5.13) into (16.5.12) and solving for $\Phi_{s,s}'$ we have

$$\Phi_{s,s}' = \{\Gamma(0) - \mathbf{b}'(s)[\mathbf{A}(s)]^{-1}\mathbf{b}(s)\}^{-1}\{\Gamma(s) - \mathbf{b}'(s)[\mathbf{A}(s)]^{-1}\mathbf{c}(s)\}. \tag{16.5.14}$$

The partial autoregression matrix function is therefore defined to be

$$\mathcal{P}(s) = \begin{cases} \Gamma'(1)[\Gamma(0)]^{-1}, & s = 1, \\ \{\Gamma'(s) - \mathbf{c}'(s)[\mathbf{A}(s)]^{-1}\mathbf{b}(s)\}\{\Gamma(0) - \mathbf{b}'(s)[\mathbf{A}(s)]^{-1}\mathbf{b}(s)\}^{-1}, & s > 1. \end{cases} \tag{16.5.15}$$

The $\mathcal{P}(s)$ so defined has the property that if the model is a vector AR(p), then

$$\mathcal{P}(s) = \begin{cases} \Phi_p, & s = p, \\ \mathbf{0}, & s > p. \end{cases}$$

Like the partial autocorrelation function for the univariate case, the partial autoregression matrix function, $\mathcal{P}(s)$, has the cutoff property for vector AR processes.

Although the autoregression matrix function, $\mathcal{P}(s)$, is motivated from consideration of vector AR models, the function is defined in terms of the $\Gamma(k)$'s through (16.5.15) and hence can be applied to any stationary process. Unlike the univariate partial autocorrelation function, however, the elements of $\mathcal{P}(s)$ are not proper correlation coefficients.

Sample Partial Autoregression Matrices Sample estimates of the $\mathcal{P}(s)$ can be obtained by replacing the unknown $\Gamma(s)$ in (16.5.15) by the sample covariance matrices $\hat{\Gamma}(s)$:

$$\hat{\Gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_{t+s} - \bar{\mathbf{Z}})', \qquad s = 1, 2, \ldots, \tag{16.5.16}$$

where $\bar{\mathbf{Z}} = (\bar{Z}_1, \bar{Z}_2, \ldots, \bar{Z}_m)'$ is the sample mean vector. Alternatively, estimates of the $\mathcal{P}(s)$ can be obtained by fitting vector autoregressive models of successively higher orders by standard multivariate linear regression models. First, we note that if a process is a stationary vector AR(p), we can write it as

$$\mathbf{Z}_t' = \boldsymbol{\tau}' + \sum_{j=1}^{p}\mathbf{Z}_{t-j}'\boldsymbol{\Phi}_j' + \mathbf{a}_t', \tag{16.5.17}$$

where τ' is a constant vector. Thus, given n observations, and for $t = p + 1, \ldots, n$, we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{16.5.18}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Z}_{p+1}' \\ \vdots \\ \mathbf{Z}_n' \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & \mathbf{Z}_p' & \cdots & \mathbf{Z}_1' \\ \vdots & \vdots & & \vdots \\ 1 & \mathbf{Z}_{n-1}' & \cdots & \mathbf{Z}_{n-p}' \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\tau}' \\ \boldsymbol{\Phi}_1' \\ \vdots \\ \boldsymbol{\Phi}_p' \end{bmatrix}, \qquad \text{and} \qquad \mathbf{e} = \begin{bmatrix} \mathbf{a}_{p+1}' \\ \mathbf{a}_{p+2}' \\ \vdots \\ \mathbf{a}_n' \end{bmatrix}.$$

When the $\mathbf{a}_t$ are $N(\mathbf{0}, \boldsymbol{\Sigma})$ Gaussian vector white noise, we have, from the discussion of multivariate linear models given in Supplement 16.A, the following results.

1. $$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y} \tag{16.5.19}$$

   is joint multivariate normal with the mean and variance-covariance matrix

   $$\begin{cases} E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \\ \text{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} \otimes (X'X)^{-1}, \end{cases} \tag{16.5.20}$$

   where $\otimes$ denotes the Kronecker product.
2. The maximum likelihood estimate of $\boldsymbol{\Sigma}$ is given by

   $$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\mathbf{S}(p), \tag{16.5.21}$$

where $N = n - p$ and $\mathbf{S}(p)$ is the residual sum of squares and cross-products, i.e.,

$$\mathbf{S}(p) = \sum_{t=p+1}^{n} (\mathbf{Z}_t - \hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\Phi}}_1 \mathbf{Z}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_p \mathbf{Z}_{t-p}) \times (\mathbf{Z}_t - \hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\Phi}}_1 \mathbf{Z}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_p \mathbf{Z}_{t-p})'. \tag{16.5.22}$$

The estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Sigma}} \otimes (X'X)^{-1}. \tag{16.5.23}$$

The estimated variances and correlations of the elements of $\hat{\boldsymbol{\beta}}$ such as $\hat{\boldsymbol{\Phi}}_1, \ldots,$ and $\hat{\boldsymbol{\Phi}}_p$ are provided by many statistical software that contain multivariate linear models.

3. $\mathbf{S}(p)$ follows a Wishart distribution with $(N - pm - 1)$ d.f. and is denoted as $W_{N-pm-1}(\boldsymbol{\Sigma})$. Furthermore, $\hat{\boldsymbol{\beta}}$ is independent of $\mathbf{S}(P)$.
4. The maximum likelihood function is given by

$$L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}) = \frac{e^{-Nm/2}}{(2\pi)^{Nm/2} |\hat{\boldsymbol{\Sigma}}|^{N/2}}. \tag{16.5.24}$$

Summary Statistics for Partial Autoregression To test for the significance of the sample partial autoregression, we can use an overall χ^2 test. To do so, we note that Equation (16.5.18) can also be written as

$$\mathbf{Y} = X_1 \boldsymbol{\beta}_1 + X_2 \boldsymbol{\Phi}_p' + \mathbf{e}, \tag{16.5.25}$$

where

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \boldsymbol{\tau}' \\ \boldsymbol{\Phi}_1' \\ \vdots \\ \boldsymbol{\Phi}_{p-1}' \end{bmatrix}, \quad X_2 = \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_2' \\ \vdots \\ \mathbf{Z}_{n-p}' \end{bmatrix},$$

and

$$X_1 = \begin{bmatrix} 1 & \mathbf{Z}_p' & \cdots & \mathbf{Z}_2' \\ \vdots & \vdots & & \vdots \\ 1 & \mathbf{Z}_{n-1}' & \cdots & \mathbf{Z}_{n-p+1}' \end{bmatrix}.$$

To test the hypothesis

$$\begin{cases} H_0: & \boldsymbol{\Phi}_p = \mathbf{0} \\ H_1: & \boldsymbol{\Phi}_p \neq \mathbf{0}, \end{cases} \tag{16.5.26}$$

we note that under H_0, the estimate of $\boldsymbol{\Sigma}$ is given by

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{N}\mathbf{S}(p-1), \tag{16.5.27}$$

where $\mathbf{S}(p-1)$ is the residual sum of squares and cross-products under H_0, i.e.,

$$\begin{aligned}\mathbf{S}(p-1) = \sum_{t=p+1}^{n} &(\mathbf{Z}_t - \hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\Phi}}_1\mathbf{Z}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_{p-1}\mathbf{Z}_{t-p+1}) \\ &\cdot(\mathbf{Z}_t - \hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\Phi}}_1\mathbf{Z}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_{p-1}\mathbf{Z}_{t-p+1})'.\end{aligned} \tag{16.5.28}$$

The corresponding maximum likelihood function under H_0 is given by

$$L(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\Sigma}}_1) = \frac{e^{-Nm/2}}{(2\pi)^{Nm/2}|\hat{\boldsymbol{\Sigma}}_1|^{N/2}}. \tag{16.5.29}$$

Thus, the likelihood ratio statistic for testing $\boldsymbol{\Phi}_p = \mathbf{0}$ against $\boldsymbol{\Phi}_p \neq \mathbf{0}$ is

$$U = \frac{|\mathbf{S}(p)|}{|\mathbf{S}(p-1)|}. \tag{16.5.30}$$

Following the result given in Supplement 16.A, we see that under H_0, the statistic

$$M(p) = -\left(N - \frac{1}{2} - pm\right)\ln U \tag{16.5.31}$$

asymptotically has a chi-square distribution with m^2 degrees of freedom. We reject H_0 if U is small and hence when $M(p)$ is large.

One can also use Akaike's information criterion (AIC) to select a model. Some software programs provide both $M(p)$ and AIC. For example, the software SCA computes AIC as

$$\text{AIC} = -\frac{2}{n}\ln(\text{maximum likelihood}) + \frac{2}{n}(\text{the number of parameters}).$$

For an m-dimensional vector AR(p) model, it becomes

$$AIC(p) = \ln(|\mathbf{S}(p)|) + \frac{2pm^2}{n},$$

and one selects a model that gives the smallest AIC value. Ideally, different criteria will reach the same model. Otherwise, use of good judgment from a model builder may be required.

16.5.3 Partial Lag Correlation Matrix Function

By definition, the partial autocorrelation function between Z_t and Z_{t+s} for the univariate time series is the correlation coefficient between Z_t and Z_{t+s} after the linear dependence of each on the intervening variables $Z_{t+1}, \ldots, Z_{t+s-1}$ is removed (see, for example, Hannan [1970, p. 22]). As shown in Section 2.3, this partial autocorrelation between Z_t and Z_{t+s} can also be obtained as the last regression coefficient associated with Z_t when regressing Z_{t+s} on its s lagged variables Z_{t+s-1}, $Z_{t+s-2}, \ldots, Z_t$. Tiao and Box (1981) extend the latter result to the vector case and introduce the partial autoregression matrix function $\mathcal{P}(s)$. As we have noted, however, unlike the univariate partial autocorrelation function, the elements of $\mathcal{P}(s)$ are not proper correlation coefficients.

Heyse and Wei (1985a, b) extend the definition of univariate partial autocorrelation to vector time series and derive the correlation matrix between $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$ after removing the linear dependence of each on the intervening vectors $\mathbf{Z}_{t+1}, \mathbf{Z}_{t+2}, \ldots$, and $\mathbf{Z}_{t+s-1}$. This correlation matrix is defined as the correlation between the residual vectors

$$\begin{aligned}\mathbf{u}_{s-1,t+s} &= \mathbf{Z}_{t+s} - \boldsymbol{\alpha}_{s-1,1}\mathbf{Z}_{t+s-1} - \cdots - \boldsymbol{\alpha}_{s-1,s-1}\mathbf{Z}_{t+1} \\ &= \begin{cases} \mathbf{Z}_{t+s} - \sum_{k=1}^{s-1} \boldsymbol{\alpha}_{s-1,k}\mathbf{Z}_{t+s-k}, & s \geq 2, \\ \mathbf{Z}_{t+1}, & s = 1, \end{cases} \end{aligned} \tag{16.5.32}$$

and

$$\begin{aligned}\mathbf{v}_{s-1,t} &= \mathbf{Z}_t - \boldsymbol{\beta}_{s-1,1}\mathbf{Z}_{t+1} - \cdots - \boldsymbol{\beta}_{s-1,s-1}\mathbf{Z}_{t+s-1} \\ &= \begin{cases} \mathbf{Z}_t - \sum_{k=1}^{s-1} \boldsymbol{\beta}_{s-1,k}\mathbf{Z}_{t+k}, & s \geq 2, \\ \mathbf{Z}_t, & s = 1. \end{cases} \end{aligned} \tag{16.5.33}$$

The multivariate linear regression coefficient matrices $\boldsymbol{\alpha}_{s-1,k}$ and $\boldsymbol{\beta}_{s-1,k}$ are those that minimize $E[|\mathbf{u}_{s-1,t+s}|^2]$ and $E[|\mathbf{v}_{s-1,t}|^2]$, respectively. Note that (16.5.32) is the residual from a regression of $\mathbf{Z}_{t+s}$ on the carriers $\mathbf{Z}_{t+s-1}, \ldots, \mathbf{Z}_{t+1}$, and that (16.5.33) is the residual from a regression of $\mathbf{Z}_t$ on the same set of carriers.

For $s \geq 2$, we let

$$\boldsymbol{\alpha}'(s) = \begin{bmatrix} \boldsymbol{\alpha}'_{s-1,1} \\ \boldsymbol{\alpha}'_{s-1,2} \\ \vdots \\ \boldsymbol{\alpha}'_{s-1,s-1} \end{bmatrix}, \qquad \mathbf{Z}_t(s) = \begin{bmatrix} \mathbf{Z}_{t+s-1} \\ \mathbf{Z}_{t+s-2} \\ \vdots \\ \mathbf{Z}_{t+1} \end{bmatrix}, \tag{16.5.34}$$

and use $\mathbf{A}(s)$ and $\mathbf{c}(s)$ from (16.5.9) and consider the minimization of

$$\begin{aligned} E[|\mathbf{u}_{s-1,t+s}|^2] &= E[|\mathbf{Z}_{t+s} - \boldsymbol{\alpha}(s)\mathbf{Z}_t(s)|^2] \\ &= E[(\mathbf{Z}_{t+s} - \boldsymbol{\alpha}(s)\mathbf{Z}_t(s))(\mathbf{Z}_{t+s} - \boldsymbol{\alpha}(s)\mathbf{Z}_t(s))'] \\ &= \Gamma(0) - \boldsymbol{\alpha}(s)\mathbf{c}(s) - \mathbf{c}'(s)\boldsymbol{\alpha}'(s) + \boldsymbol{\alpha}(s)\mathbf{A}(s)\boldsymbol{\alpha}'(s). \end{aligned} \tag{16.5.35}$$

Taking the derivative of (16.5.35) with respect to the elements of $\boldsymbol{\alpha}'(s)$ and setting the resulting equations equal to $\mathbf{0}$ gives (Graham [1981, p. 54])

$$\mathbf{A}(s)\boldsymbol{\alpha}'(s) = \mathbf{c}(s), \tag{16.5.36}$$

which are recognized as the multivariate normal equations for the regression of $\mathbf{Z}_{t+s}$ on $\mathbf{Z}_{t+s-1}, \ldots, \mathbf{Z}_{t+1}$, i.e.,

$$\begin{bmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}'(1) & \cdots & \boldsymbol{\Gamma}'(s-2) \\ \boldsymbol{\Gamma}(1) & \boldsymbol{\Gamma}(0) & \cdots & \boldsymbol{\Gamma}'(s-3) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}(s-2) & \boldsymbol{\Gamma}(s-3) & \cdots & \boldsymbol{\Gamma}(0) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}'_{s-1,1} \\ \boldsymbol{\alpha}'_{s-1,2} \\ \vdots \\ \boldsymbol{\alpha}'_{s-1,s-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}(1) \\ \boldsymbol{\Gamma}(2) \\ \vdots \\ \boldsymbol{\Gamma}(s-1) \end{bmatrix}. \tag{16.5.37}$$

Solving the system (16.5.36) or (16.5.37) gives the familiar multivariate linear regression coefficient matrices

$$\boldsymbol{\alpha}'(s) = [\mathbf{A}(s)]^{-1}\mathbf{c}(s). \tag{16.5.38}$$

The linear combination $\boldsymbol{\alpha}(s)\mathbf{Z}_t(s)$ defines the linear projection of $\mathbf{Z}_{t+s}$ onto the space spanned by $\mathbf{Z}_{t+s-1}, \ldots, \mathbf{Z}_{t+1}$. Because

$$\begin{aligned} E[\mathbf{Z}_t(s)(\mathbf{Z}_{t+s} - \boldsymbol{\alpha}(s)\mathbf{Z}_t(s))'] &= E[\mathbf{Z}_t(s)\mathbf{Z}'_{t+s} - \mathbf{Z}_t(s)\mathbf{Z}'_t(s)\boldsymbol{\alpha}'(s)] \\ &= \boldsymbol{c}(s) - \mathbf{A}(s)[\mathbf{A}(s)]^{-1}\mathbf{c}(s) \\ &= \mathbf{0}, \end{aligned}$$

we have that $\mathbf{u}_{s-1,t+s}$ and $\mathbf{Z}_t(s)$ are uncorrelated and that

$$\begin{aligned} \operatorname{var}(\mathbf{Z}_{t+s}) &= \operatorname{var}(\mathbf{u}_{s-1,t+s}) + \operatorname{var}[\boldsymbol{\alpha}(s)\mathbf{Z}_t(s)] \\ &= \operatorname{var}(\mathbf{u}_{s-1,t+s}) + \mathbf{c}'(s)[\mathbf{A}(s)]^{-1}\mathbf{c}(s), \end{aligned}$$

so

$$\begin{aligned} \operatorname{var}(\mathbf{u}_{s-1,t+s}) &= \boldsymbol{\Gamma}(0) - \mathbf{c}'(s)[\mathbf{A}(s)]^{-1}\mathbf{c}(s) \\ &= \boldsymbol{\Gamma}(0) - \boldsymbol{\alpha}(s)\boldsymbol{c}(s). \end{aligned} \tag{16.5.39}$$

Similarly, letting

$$\boldsymbol{\beta}'(s) = \begin{bmatrix} \boldsymbol{\beta}'_{s-1,s-1} \\ \boldsymbol{\beta}'_{s-1,s-2} \\ \vdots \\ \boldsymbol{\beta}'_{s-1,1} \end{bmatrix}$$

and using $\mathbf{A}(s)$ and $\mathbf{b}(s)$ from (16.5.9) and $\mathbf{Z}_t(s)$ from (16.5.34), we can write the normal equations for the $\boldsymbol{\beta}_{s-1,k}$ of (16.5.33) as

$$\mathbf{A}(s)\boldsymbol{\beta}'(s) = \mathbf{b}(s), \tag{16.5.40}$$

which has the solution

$$\boldsymbol{\beta}'(s) = [\mathbf{A}(s)]^{-1}\mathbf{b}(s).$$

Also, in a similar manner to (16.5.39),

$$\begin{aligned} \text{Var}(\mathbf{v}_{s-1,t}) &= \Gamma(0) - \mathbf{b}'(s)[\mathbf{A}(s)]^{-1}\mathbf{b}(s) \\ &= \Gamma(0) - \boldsymbol{\beta}(s)\mathbf{b}(s). \end{aligned} \tag{16.5.41}$$

The covariance between $\mathbf{v}_{s-1,t}$ and $\mathbf{u}_{s-1,t+s}$ is

$$\begin{aligned} \text{Cov}(\mathbf{v}_{s-1,t}, \mathbf{u}_{s-1,t+s}) &= E[(\mathbf{Z}_t - \boldsymbol{\beta}(s)\mathbf{Z}_t(s))(\mathbf{Z}_{t+s} - \boldsymbol{\alpha}(s)\mathbf{Z}_t(s))'] \\ &= \Gamma(s) - \mathbf{b}'(s)[\mathbf{A}(s)]^{-1}\mathbf{c}(s) \\ &= \Gamma(s) - \mathbf{b}'(s)\boldsymbol{\alpha}'(s). \end{aligned} \tag{16.5.42}$$

We denote $\text{var}(\mathbf{u}_{s-1,t+s})$ by $\mathbf{V}_\mathbf{u}(s)$, $\text{var}(\mathbf{v}_{s-1,t})$ by $\mathbf{V}_\mathbf{v}(s)$, and $\text{cov}(\mathbf{v}_{s-1,t}, \mathbf{u}_{s-1,t+s})$ by $\mathbf{V}_{\mathbf{vu}}(s)$. Note that $\text{cov}(\mathbf{u}_{s-1,t+s}, \mathbf{v}_{s-1,t})$ is equal to $\mathbf{V}'_{\mathbf{vu}}(s)$.

For the case $s = 1$, it is easily seen from (16.5.32) and (16.5.33) that $\mathbf{V}_\mathbf{u}(1) = \mathbf{V}_\mathbf{v}(1) = \Gamma(0)$ and that $\mathbf{V}_{\mathbf{vu}}(1) = \Gamma(1)$ because there are no intervening vectors between $\mathbf{Z}_t$ and $\mathbf{Z}_{t+1}$.

Heyse and Wei (1985a, b) define the partial lag autocorrelation matrix at lag s to be

$$\mathbf{P}(s) = [\mathbf{D}_\mathbf{v}(s)]^{-1}\mathbf{V}_{\mathbf{vu}}(s)[\mathbf{D}_u(s)]^{-1}, \tag{16.5.43}$$

where $\mathbf{D}_\mathbf{v}(s)$ is a diagonal matrix in which the ith diagonal element is the square root of the ith diagonal element of $\mathbf{V}_\mathbf{v}(s)$ and $\mathbf{D}_\mathbf{u}(s)$ is similarly defined for $\mathbf{V}_\mathbf{u}(s)$. The term *partial lag autocorrelation matrix* is used because it is the autocorrelation matrix between the elements of $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$, after the removal of their linear dependence on the vectors at the intervening lags, $\mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+s-1}$. They call $\mathbf{P}(s)$, as a function of s, the partial lag autocorrelation matrix function. $\mathbf{P}(s)$, as a matrix of autocorrelation coefficients, is a vector extension of the partial autocorrelation function in the same manner as the autocorrelation matrix function is a vector extension of the autocorrelation function. Recognizing that the univariate case corresponds to the special case when $m = 1$, they recommend referring to $\mathbf{P}(s)$ as the partial lag autocorrelation function even in the univariate case. In this book, we simply refer to $\mathbf{P}(k)$ as the partial lag correlation matrix function.

Note from (16.5.15) that Tiao and Box's partial autoregression matrix at lag s equals

$$\mathcal{P}(s) = \mathbf{V}'_{\mathbf{vu}}(s)[\mathbf{V}_\mathbf{v}(s)]^{-1}. \tag{16.5.44}$$

The following is a summary of the relationships and differences between our partial lag correlation matrix $\mathbf{P}(s)$ and Tiao and Box's partial autoregression matrix $\mathcal{P}(s)$

1. From (16.5.43) and (16.5.44), we see that $\mathbf{P}(s) = \mathcal{P}(s) = \mathbf{0}$ if and only if $\mathbf{V}_{\mathbf{vu}}(s) = \mathbf{0}$. Thus, $\mathbf{P}(s)$ and $\mathcal{P}(s)$ share the same cut-off property for autoregressive processes.
2. When $m = 1$, it is easily shown by reversing the order of the equations in either (16.5.36) or (16.5.40) that

   (a) $\alpha_{s-1,k} = \beta_{s-1,k}$, for $k = 1, 2, \ldots, s-1$,

   and

   (b) $\mathbf{V}_{\mathbf{u}}(s) = \mathbf{V}_{\mathbf{v}}(s)$.

   Thus,

   $$\mathbf{P}(s) = \frac{\mathbf{V}_{\mathbf{vu}}(s)}{\mathbf{V}_{\mathbf{v}}(s)} = \mathcal{P}(s).$$

   Hence, the partial lag correlation matrix and Tiao and Box's partial autoregression matrix are equal in the univariate case.
3. The partial lag correlation matrix $\mathbf{P}(s)$ is a correlation coefficient matrix because each of its elements is a properly normalized correlation coefficient. Here, however, $\mathcal{P}(s)$ is not equal to the correlation coefficient matrix between the elements $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$, allowing for $\mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+s-1}$, except in the univariate case mentioned in step 2. This property is one of the reasons that Tiao and Box called $\mathcal{P}(s)$ the partial autoregression matrix.
4. Even for a general m-dimensional vector process, when $s = 1$, we see from (16.5.43) that $\mathbf{P}(1) = \boldsymbol{\rho}(1)$. That is, the partial lag correlation matrix at lag 1 reduces to the regular correlation matrix at lag 1, which is to be expected because in this case there are no intervening vectors between $\mathbf{Z}_t$ and $\mathbf{Z}_{t+1}$. This natural property that also holds for the univariate partial autocorrelation function ϕ_{kk}, however, is not shared by $\mathcal{P}(s)$ except in the univariate case when $m = 1$.

Recursive Algorithm for Computing P(s) In this section, we discuss a recursive procedure for computing the partial lag correlation matrices. This method is also mentioned in Ansley and Newbold (1979a) and provides a vector generalization of Durbin's (1960) recursive computational procedure for univariate partial autocorrelations. Box, Jenkins, and Reinsel (1994, p. 87) give a nice summary of Durbin's procedure.

Given the regressions from (16.5.32) and (16.5.33), for $s \geq 2$,

$$\mathbf{Z}_{t+s} = \sum_{k=1}^{s-1} \alpha_{s-1,k} \mathbf{Z}_{t+s-k} + \mathbf{u}_{s-1,t+s}, \tag{16.5.45}$$

$$\mathbf{Z}_t = \sum_{k=1}^{s-1} \beta_{s-1,k} \mathbf{Z}_{t+k} + \mathbf{v}_{s-1,t}, \tag{16.5.46}$$

we want to consider the regressions

$$\mathbf{Z}_{t+s+1} = \sum_{k=1}^{s} \boldsymbol{\alpha}_{s,k}\mathbf{Z}_{t+s+1-k} + \mathbf{u}_{s,t+s+1}, \tag{16.5.47}$$

$$\mathbf{Z}_t = \sum_{k=1}^{s} \boldsymbol{\beta}_{s,k}\mathbf{Z}_{t+k} + \mathbf{v}_{s,t}. \tag{16.5.48}$$

Corresponding to the definition of the partial lag correlation matrix, our interest is in the correlation between $\mathbf{v}_{s,t}$ and $\mathbf{u}_{s,t+s+1}$; thus, we require the computation of the multivariate linear regression coefficient matrices $\boldsymbol{\alpha}_{s,k}$ and $\boldsymbol{\beta}_{s,k}$ in (16.5.47) and (16.5.48). Let

$$\mathbf{u}_{s-1,t+s} = \boldsymbol{\alpha}_s^{\dagger}\mathbf{v}_{s-1,t} + \mathbf{u}_{t+s}^{\dagger} \tag{16.5.49}$$

be the regression of the residual $\mathbf{u}_{s-1,t+s}$ from (16.5.45) on the residual $\mathbf{v}_{s-1,t}$ from (16.5.46), where the multivariate linear regression coefficient is given by

$$\begin{aligned}\boldsymbol{\alpha}_s^{\dagger} &= \mathrm{Cov}(\mathbf{u}_{s-1,t+s}, \mathbf{v}_{s-1,t})[\mathrm{Var}(\mathbf{v}_{s-1,t})]^{-1} \\ &= \mathbf{V}'_{\mathbf{vu}}(s)[\mathbf{V}_{\mathbf{v}}(s)]^{-1}.\end{aligned}$$

Note from (16.5.44) that $\boldsymbol{\alpha}_s^{\dagger}$ is, in fact, equal to the partial autoregression matrix at lag s. From (16.5.47), however, the partial autoregression matrix at lag s is equal to $\boldsymbol{\alpha}_{s,s}$. Therefore, in the regression relation (16.5.49), $\boldsymbol{\alpha}_s^{\dagger} = \boldsymbol{\alpha}_{s,s}$ and $\mathbf{u}_{t+s}^{\dagger} = \mathbf{u}_{s,t+s+1}$. Substituting for $\mathbf{u}_{s-1,t+s}$ and $\mathbf{v}_{s-1,t}$ from (16.5.45) and (16.5.46) in the regression (16.5.49) gives

$$\mathbf{Z}_{t+s} - \sum_{k=1}^{s-1} \boldsymbol{\alpha}_{s-1,k}\mathbf{Z}_{t+s-k} = \boldsymbol{\alpha}_{s,s}\left(\mathbf{Z}_t - \sum_{k=1}^{s-1} \boldsymbol{\beta}_{s-1,k}\mathbf{Z}_{t+k}\right) + \mathbf{u}_{s,t+s+1}.$$

Hence,

$$\mathbf{Z}_{t+s} = \sum_{k=1}^{s-1} (\boldsymbol{\alpha}_{s-1,k} - \boldsymbol{\alpha}_{s,s}\boldsymbol{\beta}_{s-1,s-k})\mathbf{Z}_{t+s-k} + \boldsymbol{\alpha}_{s,s}\mathbf{Z}_t + \mathbf{u}_{s,t+s+1}.$$

Because the multivariate linear regression coefficient matrix that defines the regression of $\mathbf{Z}_{t+s}$ on the carriers $\mathbf{Z}_{t+s-1}, \ldots, \mathbf{Z}_t$ is the same as those that define the regression of $\mathbf{Z}_{t+s+1}$ on the carriers $\mathbf{Z}_{t+s}, \ldots, \mathbf{Z}_{t+1}$, the coefficients $\boldsymbol{\alpha}_{s,k}$ in (16.5.47) can be computed from

$$\boldsymbol{\alpha}_{s,s} = \mathbf{V}'_{\mathbf{vu}}(s)[\mathbf{V}_{\mathbf{v}}(s)]^{-1}$$

and

$$\boldsymbol{\alpha}_{s,k} = \boldsymbol{\alpha}_{s-1,k} - \boldsymbol{\alpha}_{s,s}\boldsymbol{\beta}_{s-1,s-k}. \tag{16.5.50}$$

Similarly, the multivariate linear regression coefficients $\boldsymbol{\beta}_{s,k}$ in (16.5.48) can be found by considering the regression of $\mathbf{v}_{s-1,t}$ on $\mathbf{u}_{s-1,t+s}$, i.e.,

$$\boldsymbol{\beta}_{s,s} = \mathbf{V}_{\mathbf{vu}}(s)[\mathbf{V}_{\mathbf{u}}(s)]^{-1}$$

and

$$\boldsymbol{\beta}_{s,k} = \boldsymbol{\beta}_{s-1,k} - \boldsymbol{\beta}_{s,s}\boldsymbol{\alpha}_{s-1,s-k}. \tag{16.5.51}$$

This development leads naturally to the following recursive procedure for computing $\mathbf{P}(s)$ in which $\mathbf{V}_u(s)$, $\mathbf{V}_v(s)$, and $\mathbf{V}_{vu}(s)$ are obtained using (16.5.39), (16.5.41), and (16.5.42), respectively. For $s = 1$,

$$\begin{aligned}
\mathbf{V}_u(1) &= \mathbf{V}_v(1) = \Gamma(0),\\
\mathbf{V}_{vu}(1) &= \Gamma(1),\\
\boldsymbol{\alpha}_{1,1} &= \Gamma'(1)[\Gamma(0)]^{-1},\\
\boldsymbol{\beta}_{1,1} &= \Gamma(1)[\Gamma(0)]^{-1}.
\end{aligned}$$

For $s \geq 2$ and $k = 1, \ldots, s-1$,

$$\mathbf{V}_u(s) = \Gamma(0) - \sum_{k=1}^{s-1} \boldsymbol{\alpha}_{s-1,k}\Gamma(k) \qquad \text{[from (16.5.39)]},$$

$$\mathbf{V}_v(s) = \Gamma(0) - \sum_{k=1}^{s-1} \boldsymbol{\beta}_{s-1,k}\Gamma'(k) \qquad \text{[from (16.5.41)]},$$

$$\mathbf{V}_{vu}(s) = \Gamma(s) - \sum_{k=1}^{s-1} \Gamma(s-k)\boldsymbol{\alpha}'_{s-1,k} \qquad \text{[from (16.5.42)]},$$

$$\begin{aligned}
\boldsymbol{\alpha}_{s,s} &= \mathbf{V}'_{vu}(s)[\mathbf{V}_v(s)]^{-1},\\
\boldsymbol{\alpha}_{s,k} &= \boldsymbol{\alpha}_{s-1,k} - \boldsymbol{\alpha}_{s,s}\boldsymbol{\beta}_{s-1,s-k},\\
\boldsymbol{\beta}_{s,s} &= \mathbf{V}_{vu}(s)[\mathbf{V}_u(s)]^{-1},\\
\boldsymbol{\beta}_{s,k} &= \boldsymbol{\beta}_{s-1,k} - \boldsymbol{\beta}_{s,s}\boldsymbol{\alpha}_{s-1,s-k}.
\end{aligned}$$

The partial lag correlation matrix at lag s is, from (16.5.43),

$$\mathbf{P}(s) = [\mathbf{D}_v(s)]^{-1}\mathbf{V}_{vu}(s)[\mathbf{D}_u(s)]^{-1},$$

where $\mathbf{D}_v(s)$ is a diagonal matrix whose ith diagonal element is the square root of the ith diagonal element of $\mathbf{V}_v(s)$ and $\mathbf{D}_u(s)$ is similarly defined for $\mathbf{V}_u(s)$.

In addition, we note that this algorithm also provides $\mathcal{P}(k)$, the partial autoregression matrix. The matrix coefficients for fitting $\mathbf{Z}_t$ to a vector autoregressive process of order s can be computed from $\boldsymbol{\Theta}_{s,k} = \boldsymbol{\alpha}_{s,k}$ as in (16.5.50).

Following Quenouille (1957, p. 40), Ansley and Newbold (1979a) define the multivariate partial autocorrelation matrix at lag s to be

$$\mathbf{Q}(s) = [\mathbf{W}_u(s)]^{-1}\mathbf{V}'_{vu}(s)[\mathbf{W}_v(s)]^{-1}, \tag{16.5.52}$$

where $\mathbf{W}_u(s)$ and $\mathbf{W}_v(s)$ are the symmetric square roots of $\mathbf{V}_u(s)$ and $\mathbf{V}_v(s)$, defined such that $[\mathbf{W}_u(s)]^2 = \mathbf{V}_u(s)$ and $[\mathbf{W}_v(s)]^2 = \mathbf{V}_v(s)$. Like $\mathcal{P}(s)$, however, $\mathbf{Q}(s)$ is not equal to the correlation coefficient matrix between the elements of $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$, allowing for $\mathbf{Z}_{t+1}, \ldots,$ and $\mathbf{Z}_{t+s-1}$. In fact, its elements are not correlation coefficients except in the univariate case in which $\mathcal{P}(s) = \mathbf{Q}(s) = \mathbf{P}(s)$.

Sample Partial Lag Correlation Matrices Estimates of the partial lag correlation matrices, denoted $\hat{\mathbf{P}}(s)$, are obtained by using $\hat{\Gamma}(j)$ in place of $\hat{\Gamma}(j)$ for $j = 0, 1, \ldots, s-1$ in $\mathbf{P}(s)$. Because $\hat{\mathbf{P}}(s)$ is a proper correlation matrix, the results of sample correlation matrices can be used for its inference. Under the null hypothesis that $\mathbf{Z}_t$ is an autoregressive process of order $s-1$, the two residuals $\mathbf{u}_{s-1,\,t+s}$ and $\mathbf{v}_{s-1,\,t}$ are uncorrelated white noise series. Using Quenouille (1957, p. 41) and Hannan (1970, p. 400), the elements of $\hat{\mathbf{P}}(s)$, denoted $\hat{P}_{ij}(s)$, are independent and asymptotically normally distributed with mean 0 and variance $1/n$. Thus, we can use Tiao and Box's $+$, $-$, $\cdot$ notation introduced at the end of Section 16.5.1 to indicate values of $\hat{P}_{ij}(s)$ that are larger than $2/\sqrt{n}$, less than $-2/\sqrt{n}$, or between $-2/\sqrt{n}$ and $2/\sqrt{n}$. In addition, $n[\hat{P}_{ij}(s)]^2$ is asymptotically distributed as a χ^2 with one degree of freedom, which implies that

$$X(s) = n\sum_{i=1}^{m}\sum_{j=1}^{m}[\hat{P}_{ij}(s)]^2 \tag{16.5.53}$$

is asymptotically distributed as a χ^2 with m^2 degrees of freedom. The term $X(s)$ provides a diagnostic aid for determining the order of an autoregressive model.

16.6 Model Fitting and Forecasting

Once a vector ARMA model in (16.3.1) is identified, the methods discussed in Chapter 7 for the univariate case can be generalized to estimate the associated parameter matrices $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$, $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q)$, and $\boldsymbol{\Sigma}$. For example, given $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n)$ from a zero mean Gaussian vector ARMA(p, q) process, the log-likelihood function is given by

$$\begin{aligned}\ln L(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}|\mathbf{Z}) &= -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{n}\mathbf{a}_t'\,\boldsymbol{\Sigma}^{-1}\mathbf{a}_t \\ &= -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}S(\boldsymbol{\Phi}, \boldsymbol{\Theta}),\end{aligned} \tag{16.6.1}$$

where

$$\mathbf{a}_t = \mathbf{Z}_t - \boldsymbol{\Phi}_1\mathbf{Z}_{t-1} - \cdots - \boldsymbol{\Phi}_p\mathbf{Z}_{t-p} + \boldsymbol{\Theta}_1\mathbf{a}_{t-1} + \cdots + \boldsymbol{\Theta}_q\mathbf{a}_{t-q} \tag{16.6.2}$$

and

$$\mathbf{S}(\Phi, \Theta) = \sum_{t=1}^{n} \mathbf{a}_t \mathbf{a}_t'. \tag{16.6.3}$$

Alternatively, because $\mathbf{Z}_t$ for $t \leq 0$ are not available, (16.6.1) can be replaced by the conditional log-likelihood function

$$\ln L_*(\Phi, \Theta, \Sigma|\mathbf{Z}|) = -\frac{n-p}{2} \ln |(2\pi)^m \Sigma| - \frac{1}{2} \operatorname{tr} \Sigma^{-1} \mathbf{S}_*(\Phi, \Theta), \tag{16.6.4}$$

where

$$\mathbf{S}_*(\Phi, \Theta) = \sum_{t=p+1}^{n} \mathbf{a}_t \mathbf{a}_t' \tag{16.6.5}$$

and the $\mathbf{a}_{p+1-q}, \ldots, \mathbf{a}_p$ are assumed to be zero. Clearly, maximizing the log-likelihood function in general leads to a nonlinear estimation problem.

Properties of the maximum likelihood estimators obtained from (16.6.4) have been discussed by Nicholls (1976, 1977) and Anderson (1980). Let $\boldsymbol{\eta}$ be the vector of all parameters and $\hat{\boldsymbol{\eta}}$ be the corresponding maximum likelihood estimator. They have shown that $\hat{\boldsymbol{\eta}}$ is asymptotically unbiased and consistent and that $\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ has an asymptotic multivariate normal distribution, i.e.,

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\hat{\eta}}) \tag{16.6.6}$$

where $\xrightarrow{d}$ denotes convergence in distribution, $\Sigma_{\hat{\eta}}$ is the inverse of the information matrix

$$\Sigma_{\hat{\eta}} = \left[-E\left(\frac{\partial^2 \ln L_*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \, \partial \boldsymbol{\eta}'} \right) \right]^{-1}, \tag{16.6.7}$$

and the derivatives are evaluated at the true parameter vector $\boldsymbol{\eta}$.

If the process $\mathbf{Z}_t$ has a nonzero mean vector $\boldsymbol{\mu}$, it can be estimated by the sample mean, i.e.,

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{Z}}, \tag{16.6.8}$$

where $\overline{\mathbf{Z}}$ is the vector of sample means. The sample mean is then subtracted from the series prior to estimating the parameters $\boldsymbol{\eta}$. It can be shown that

$$\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\hat{\mu}}), \tag{16.6.9}$$

where

$$\Sigma_{\hat{\mu}} = [\Phi(1)]^{-1} \Theta(1) \Sigma \Theta'(1) [\Phi'(1)]^{-1} \tag{16.6.10}$$

and $\hat{\boldsymbol{\mu}} = \overline{\mathbf{Z}}$ is asymptotically independent of $\hat{\boldsymbol{\eta}}$. Hence, asymptotically, this result is equivalent to jointly estimating $\boldsymbol{\mu}$ with the other parameters by the maximum likelihood method.

The exact likelihood function for vector time series has been studied by Osborn (1977), Phadke and Kedem (1978), Hillmer and Tiao (1979), Nicholls and Hall (1979), and Hannan, Dunsmuir, and Deistler (1980). The estimation algorithms are complicated and in general very slow. Tiao and Box recommend using the conditional method in the preliminary stages of model building and later switching to the exact method.

Estimation of a vector time series model often takes multiple stages. Initially, we would estimate every element of a parameter matrix. Based on the initial unrestricted fit, we might consider the restricted fit by retaining only the statistically significant coefficients with the nonsignificant coefficients set to zero.

Once the parameters have been estimated, one should check the adequacy of the fitted model through a careful diagnostic analysis of the residuals

$$\hat{\mathbf{a}}_t = \dot{\mathbf{Z}}_t - \hat{\Phi}_1 \dot{\mathbf{Z}}_{t-1} - \cdots - \hat{\Phi}_p \dot{\mathbf{Z}}_{t-p} + \hat{\Theta}_1 \hat{\mathbf{a}}_{t-1} + \cdots + \hat{\Theta}_q \hat{\mathbf{a}}_{t-q}, \quad (16.6.11)$$

where $\dot{\mathbf{Z}}_t$ is now used to denote $\mathbf{Z}_t$ if $\boldsymbol{\mu} = \mathbf{0}$ and to denote $(\mathbf{Z}_t - \hat{\boldsymbol{\mu}})$ otherwise and where the $\hat{\Phi}_i$ and $\hat{\Theta}_j$ are estimates of the parameters Φ_i and Θ_j. For an adequate model, the residuals should be white noise. Hence, the correlation matrices of $\hat{\mathbf{a}}_t$ should be insignificant and have no pattern. Overall χ^2 tests based on the sample correlations of the residuals are suggested by Hosking (1980) and Li and McLeod (1981). A close inspection of the correlation structure, however, may reveal subtle relationships that suggest directions of improvement in the proposed model.

Forecasting for a vector ARMA model may be carried out in exactly the same way as in the univariate case discussed in Chapter 5. The recursion (5.3.3) for computing forecasts, the computation of the Ψ_j weights (5.2.3) or (5.2.26), the forecast error variance (5.2.9), the forecast limits of (5.2.10), the autoregressive representation of forecasts (5.4.3) and (5.4.4), forecast updating (5.5.5), and the eventual forecast function (5.6.2) are generalized directly to the vector case simply by replacing scalars by vectors and matrices. For example, the ℓ-step ahead minimum mean square error forecast can be recursively calculated by

$$\hat{\mathbf{Z}}_n(\ell) = \Phi_1 \hat{\mathbf{Z}}_n(\ell - 1) + \cdots + \Phi_p \hat{\mathbf{Z}}_n(\ell - p) - \Theta_1 E(\mathbf{a}_{n+\ell-1}) - \cdots - \Theta_q E(\mathbf{a}_{n+\ell-q}), \quad (16.6.12)$$

where $\hat{\mathbf{Z}}_n(j) = \mathbf{Z}_{n+j}$ for $j \leq 0$, $E(\mathbf{a}_{n+j}) = \mathbf{a}_{n+j}$ for $j \leq 0$, and $E(\mathbf{a}_{n+j}) = \mathbf{0}$ for $j > 0$. The forecast error variance-covariance matrix is given by

$$\text{Var}(\mathbf{e}_n(\ell)) = \sum_{j=0}^{\ell-1} \Psi_j \Sigma \Psi_j', \quad (16.6.13)$$

where the Ψ_j matrices are computed using (16.3.5) and $\Psi_0 = I$.

16.7 An Empirical Example

Recall that the analysis of the Lydia Pinkham data performed in Section 14.3.5 indicates that there is a feedforward and feedback relationship between advertising and sales. We now apply the joint vector time series approach introduced in the preceding sections to analyze this data set.

Only a few time series software programs are available for modeling vector ARMA models. These include SCA, MTS, and the newer version of SAS. SCA 1997 implements the partial autoregression matrices proposed by Tiao and Box (1981), and MTS, a multivariate time series program developed by Automatic Forecasting Systems, Inc. (1987), implements the partial lag correlation matrices suggested by Heyse and Wei (1985a, b). We use MTS in the following analysis.

16.7.1 Model Identification

Let $Z_{1,t}$ be the series of annual advertising expenditures and $Z_{2,t}$ be the series of annual sales for the Lydia Pinkham data between 1907 and 1960 as plotted in Figure 14.3. As shown in Section 14.3.5, there is a need for differencing the data. Table 16.1 shows and Figure 16.1 plots the sample correlation matrices along with their indicator symbols for the differenced series of advertising and sales from lag 1 through 10. We note that the (i, j) element of the lag k matrix, $\hat{\boldsymbol{\rho}}(k)$, is the estimate of $\rho_{ij}(k)$ if $i \leq j$ and of $\rho_{ij}(-k)$ if $i > j$. The pattern of $\hat{\rho}_{11}(k)$ indicates a possible low-order AR model for the series $(1 - B)Z_{1,t}$ and the pattern of $\hat{\rho}_{22}(k)$ implies a possible MA(1) model for the series $(1 - B)Z_{2,t}$.

For more clues, we examine the partial lag correlation matrix function together with the indicator symbol and summary statistics shown in Table 16.2. The result indicates a possible AR(2) model for the data, i.e.,

$$(I - \Phi_1 B - \Phi_2 B^2)(1 - B)\mathbf{Z}_t = \mathbf{a}_t, \tag{16.7.1}$$

where $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}]'$.

16.7.2 Parameter Estimation

Once a tentative model is identified, efficient estimates of the parameter matrices Φ_i and $\boldsymbol{\Sigma}$ are determined by maximizing the likelihood function. The results of the estimation for Equation (16.7.1) are shown in Table 16.3. The parameters were estimated in three stages. In the first

TABLE 16.1 Sample correlation matrices for the differenced Lydia Pinkham data.

k	1	2	3	4	5
$\hat{\boldsymbol{\rho}}(k)$	$\begin{bmatrix} .05 & .31 \\ .22 & .43 \end{bmatrix}$	$\begin{bmatrix} -.40 & -.15 \\ -.14 & .10 \end{bmatrix}$	$\begin{bmatrix} -.09 & .07 \\ -.06 & .07 \end{bmatrix}$	$\begin{bmatrix} .43 & .28 \\ .16 & -.11 \end{bmatrix}$	$\begin{bmatrix} .04 & .19 \\ -.14 & .03 \end{bmatrix}$
	$\begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} + & + \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$
k	6	7	8	9	10
$\hat{\boldsymbol{\rho}}(k)$	$\begin{bmatrix} -.34 & -.19 \\ -.30 & -.16 \end{bmatrix}$	$\begin{bmatrix} -.07 & .01 \\ -.19 & -.13 \end{bmatrix}$	$\begin{bmatrix} .16 & .09 \\ .03 & -.14 \end{bmatrix}$	$\begin{bmatrix} .09 & .02 \\ .03 & -.21 \end{bmatrix}$	$\begin{bmatrix} -.20 & .05 \\ .20 & -.08 \end{bmatrix}$
	$\begin{bmatrix} - & \cdot \\ - & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$

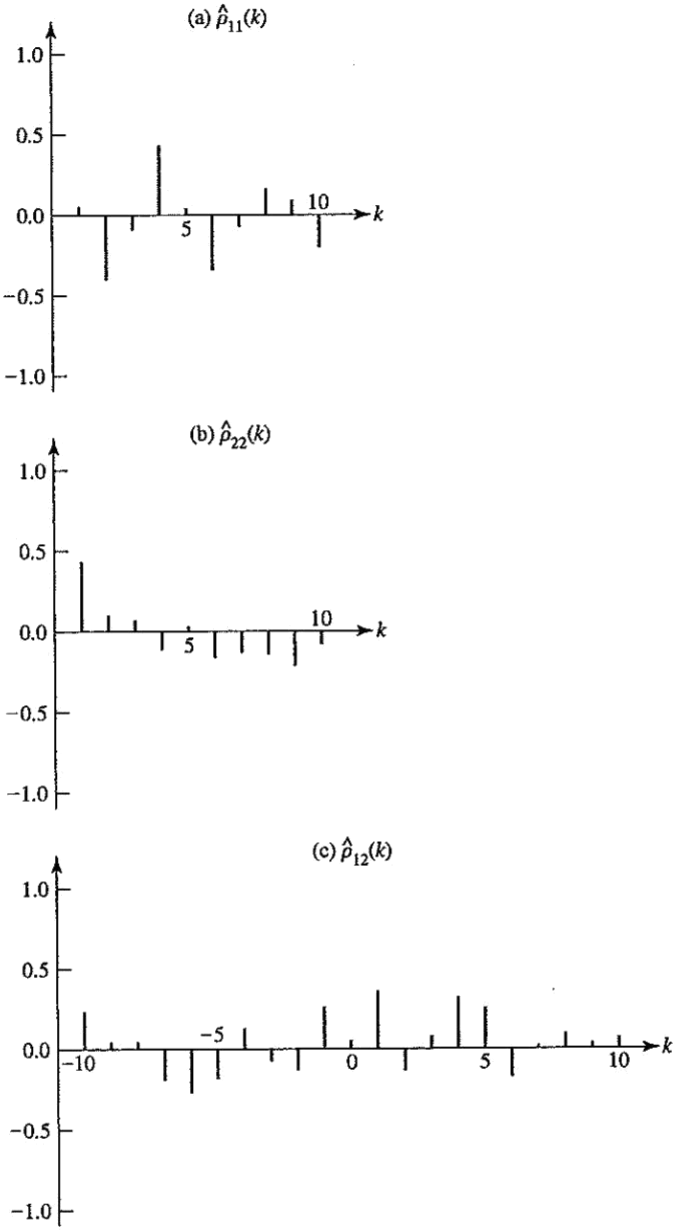


FIGURE 16.1 Sample auto- and cross-correlations for Lydia Pinkham data.

stage, all elements of Φ_1 and Φ_2 were estimated with their estimated standard errors. Those elements of Φ_2 that were less than 1.5 times their estimated standard errors were set to zero for the second stage. They included $\Phi_2(1, 2)$ and $\Phi_2(2, 2)$. After the second estimation stage, attention focused on $\hat{\Phi}_1$. Because $\Phi_1(2, 1)$ was less than 1.5 times its estimated standard error, it too was dropped from the model. The third stage estimates represent our fitted model

TABLE 16.2 Sample partial lag correlation matrix function for differenced Lydia Pinkham data.

k	$\hat{P}(k)$	Indicator symbols	$X(k)$	P-value
1	$\begin{bmatrix} .05 & .31 \\ .22 & .43 \end{bmatrix}$	$\begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix}$	17.59	.001
2	$\begin{bmatrix} -.53 & -.28 \\ -.32 & -.10 \end{bmatrix}$	$\begin{bmatrix} - & - \\ - & \cdot \end{bmatrix}$	25.00	.000
3	$\begin{bmatrix} -.12 & -.04 \\ .21 & .09 \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	3.61	.479
4	$\begin{bmatrix} .24 & .05 \\ .21 & -.01 \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	5.53	.242
5	$\begin{bmatrix} -.11 & -.32 \\ .16 & -.02 \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ \cdot & \cdot \end{bmatrix}$	7.45	.115
6	$\begin{bmatrix} -.07 & -.20 \\ -.06 & -.05 \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	2.70	.618

TABLE 16.3 Estimation results for Equation (16.7.1).

	$\hat{\Phi}_1$	$\hat{\Phi}_2$	$\hat{\Sigma}$
(1) Full model	$\begin{bmatrix} -.26 & .55 \\ (.14) & (.14) \\ -.09 & .51 \\ (.17) & (.16) \end{bmatrix}$	$\begin{bmatrix} -.49 & -.03 \\ (.14) & (.14) \\ -.32 & .08 \\ (.16) & (.17) \end{bmatrix}$	$\begin{bmatrix} 31.28 & \\ 17.16 & 42.92 \end{bmatrix}$
(2) Intermediate	$\begin{bmatrix} -.26 & .54 \\ (.14) & (.13) \\ -.08 & .53 \\ (.16) & (.16) \end{bmatrix}$	$\begin{bmatrix} -.51 & 0 \\ (.12) & \\ -.27 & 0 \\ (.13) & \end{bmatrix}$	$\begin{bmatrix} 31.32 & \\ 17.07 & 43.11 \end{bmatrix}$
(3) Final model	$\begin{bmatrix} -.26 & .54 \\ (.14) & (.13) \\ 0 & .48 \\ & (.13) \end{bmatrix}$	$\begin{bmatrix} -.51 & 0 \\ (.12) & \\ -.27 & 0 \\ (.13) & \end{bmatrix}$	$\begin{bmatrix} 31.32 & \\ 17.05 & 43.30 \end{bmatrix}$

$$\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -.26 & .54 \\ 0 & .48 \end{bmatrix}B - \begin{bmatrix} -.51 & 0 \\ -.27 & 0 \end{bmatrix}B^2\right\}\begin{bmatrix} (1-B)Z_{1,t} \\ (1-B)Z_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \quad (16.7.2)$$

and

$$\hat{\Sigma} = \begin{bmatrix} 31.32 & 17.05 \\ 17.05 & 43.30 \end{bmatrix}.$$

TABLE 16.4 Residual correlation matrices for Equation (16.7.2).

k	1	2	3	4
$\hat{\rho}(k)$	$\begin{bmatrix} -.04 & -.01 \\ -.08 & -.01 \end{bmatrix}$	$\begin{bmatrix} .03 & -.04 \\ -.01 & .01 \end{bmatrix}$	$\begin{bmatrix} -.19 & .07 \\ -.15 & .04 \end{bmatrix}$	$\begin{bmatrix} .16 & .09 \\ -.04 & .01 \end{bmatrix}$
	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$

16.7.3 Diagnostic Checking

To guard against model misspecification, a detailed diagnostic analysis of the residual series $\hat{\mathbf{a}}_t = \mathbf{Z}_t - \hat{\Phi}_1\mathbf{Z}_{t-1} - \hat{\Phi}_2\mathbf{Z}_{t-2}$ is necessary. Such an analysis includes plots of standardized residuals against time and analysis of residual correlation matrices. Table 16.4 gives the residual correlation matrices along with their indicator symbols for lags 1 through 4 that indicate that the restricted AR(2) model (16.7.2) provides an adequate representation of the data.

16.7.4 Forecasting

For illustration, we use Equation (16.7.2) to compute the one-step ahead forecast of $\mathbf{Z}_{55}$ from the forecast origin $t = 54$. First, we note that after multiplying the matrices and rearranging terms, we can rewrite Equation (16.7.2) as

$$Z_{1,t} = .74Z_{1,t-1} + .54Z_{2,t-1} - .25Z_{1,t-2} - .54Z_{2,t-2} + .51Z_{1,t-3} + a_{1,t} \quad (16.7.3)$$

and

$$Z_{2,t} = 1.48Z_{2,t-1} - .27Z_{1,t-2} - .48Z_{2,t-2} + .27Z_{1,t-3} + a_{2,t}. \quad (16.7.4)$$

Hence, given $Z_{1,1} = 608, \ldots, Z_{1,52} = 639, Z_{1,53} = 644, Z_{1,54} = 564, Z_{2,1} = 1016, \ldots, Z_{2,52} = 1390, Z_{2,53} = 1387$, and $Z_{2,54} = 1289$, we have

$$\begin{aligned}
\hat{Z}_{1,54}(1) &= E(Z_{1,55}|Z_{1,t}, Z_{2,t}, t \le 54) \\
&= .74Z_{1,54} + .54Z_{2,54} - .25Z_{1,53} - .54Z_{2,53} + .51Z_{1,52} \\
&= .74(564) + .54(1289) - .25(644) - .54(1387) + .51(639) \\
&= 529.33, \\
\hat{Z}_{2,54}(1) &= E(Z_{2,55}|Z_{1,t}, Z_{2,t}, t \le 54) \\
&= 1.48Z_{2,54} - .27Z_{1,53} - .48Z_{2,53} + .27Z_{1,52} \\
&= 1.48(1289) - .27(644) - .48(1387) + .27(639) \\
&= 1240.61.
\end{aligned}$$

Let $e_{1,1} = Z_{1,55} - \hat{Z}_{1,54}(1)$ and $e_{2,1} = Z_{2,55} - \hat{Z}_{2,54}(1)$ be the one-step ahead forecast errors. Note that $e_{1,1} = a_{1,55}$ and $e_{2,1} = a_{2,55}$, and the forecast error variance for $[e_{1,1}, e_{2,1}]'$ is given by

$$\hat{\Sigma} = \begin{bmatrix} 31.32 & 17.05 \\ 17.05 & 43.30 \end{bmatrix}.$$

Thus, $\text{Var}(e_{1,1}) = 31.32$, $\text{Var}(e_{2,1}) = 43.30$, and $\text{Cov}(e_{1,1}, e_{2,1}) = 17.05$.

16.7.5 Further Remarks

From Equation (16.7.2), we conclude that $\Phi_1(1, 2) > 0$. This conclusion confirms our suspicions from the cross-correlation functions estimated from the residual series of the transfer function model, shown in Table 14.9, that a feedback relationship with sales leading advertising existed in the data. This feedback relationship is even more explicit from the forecast equation of the advertising given in (16.7.3).

One should not conclude from (16.7.3) and (16.7.4) that there is no contemporaneous relationship between sales and advertising. In our form of ARMA models, we choose $\Phi_0 = \Theta_0 = I$, but a general form of the variance-covariance matrix Σ for the innovation $\mathbf{a}_t$. Hence, the contemporaneous relationship between components of vector series is modeled through the off-diagonal elements of Σ. From $\hat{\Sigma}$ in (16.7.2), we can estimate the correlation between the residuals $\hat{a}_{1,t}$ and $\hat{a}_{2,t}$, which equals $\hat{\rho}_{\hat{a}_{1,t}\hat{a}_{2,t}} = 17.05/\sqrt{(31.32)(43.30)} = .46$. This estimate indicates that advertising and sales were also contemporaneously related.

The advantage of a joint vector model over a transfer function model is evident when there is a feedback relationship between variables. It should be emphasized, however, that the above results by no means imply that the proposed vector model is the best choice for describing the advertising and sales phenomena of Lydia Pinkham Medicine Company. Model building is a delicate process that also requires proper understanding of related subject matters under study. In this case, these matters include the sales, marketing, and operation of the company. The above example is used to illustrate a vector time series model and its implications for the data set when only the past time series observations are used. For an interesting study of building a model for the Lydia Pinkham data using other external information, the reader is referred to Palda (1964), among others.

16.8 Spectral Properties of Vector Processes

The spectral results in Sections 12.1 and 14.5 for the univariate and bivariate processes are readily generalized to the m-dimensional vector process. For a jointly stationary m-dimensional vector process $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]'$, the spectral representation of $\mathbf{Z}_t$ is given by

$$\mathbf{Z}_t = \int_{-\pi}^{\pi} e^{i\omega t}\, d\mathbf{U}(\omega), \tag{16.8.1}$$

where $dU(\omega) = [dU_1(\omega), dU_2(\omega), \ldots, dU_m(\omega)]'$ and the $dU_i(\omega)$, $i = 1, 2, \ldots, m$, are both orthogonal as well as cross-orthogonal. The spectral representation of the covariance matrix function is given by

$$\Gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k}\, dF(\omega), \tag{16.8.2}$$

where

$$\begin{aligned} dF(\omega) &= E\{dU(\omega)\, dU^*(\omega)\} \\ &= [E\{dU_i(\omega)\, dU_j^*(\omega)\}] \\ &= [dF_{ij}(\omega)], \end{aligned} \tag{16.8.3}$$

$dU^*(\omega)$ is the transpose of the complex conjugate of $dU(\omega)$, and $F(\omega)$ is the spectral distribution matrix function of $\mathbf{Z}_t$. The diagonal elements $F_{ii}(\omega)$ are the spectral distribution functions of the $Z_{i,t}$ and the off-diagonal elements $F_{ij}(\omega)$ are the cross-spectral distribution functions between the $Z_{i,t}$ and the $Z_{j,t}$.

If the covariance matrix function is absolutely summable in the sense that each of the $m \times m$ sequences $\gamma_{ij}(k)$ is absolutely summable, then the spectrum matrix or the spectral density matrix function exists and is given by

$$\begin{aligned} f(\omega)\, d\omega &= dF(\omega) \\ &= [dF_{ij}(\omega)] \\ &= [f_{ij}(\omega)\, d\omega]. \end{aligned} \tag{16.8.4}$$

Thus, we can write

$$\Gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k} f(\omega)\, d\omega \tag{16.8.5}$$

and

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Gamma(k) e^{-i\omega k}. \tag{16.8.6}$$

The element $f_{ii}(\omega)$ in (16.8.4) is the spectrum or the spectral density of $Z_{i,t}$, and the element $f_{ij}(\omega)$ in (16.8.4) is the cross-spectrum or the cross-spectral density of $Z_{i,t}$ and $Z_{j,t}$. It is easily seen that the spectral density matrix function $f(\omega)$ is positive semi-definite, i.e., $c'f(\omega)c \geq 0$ for any nonzero m-dimensional complex vector c. Also, the matrix $f(\omega)$ is Hermitian, i.e.,

$$f^*(\omega) = f(\omega). \tag{16.8.7}$$

Hence, $f_{ij}(\omega) = f_{ji}^*(\omega)$ for all i and j.

The results of bivariate frequency-domain analysis can be used for the multivariate case. Specifically, the cospectrum, phase spectrum, gain function, and coherence may now be defined for any pair of component processes in $\mathbf{Z}_t$.

Supplement 16.A Multivariate Linear Regression Models

Let $Y_1, \ldots, Y_m$ be the m response (dependent) variables and $X_1, \ldots, X_r$ be a common set of r predictor (independent) variables. Assume that each response follows its own multiple regression such that

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11} X_1 + \cdots + \beta_{r1} X_r + a_1 \\ Y_2 &= \beta_{02} + \beta_{12} X_1 + \cdots + \beta_{r2} X_r + a_2 \\ &\vdots \\ Y_m &= \beta_{0m} + \beta_{1m} X_1 + \cdots + \beta_{rm} X_r + a_m, \end{aligned} \tag{16.A.1}$$

where the error term $\mathbf{a} = [a_1, \ldots, a_m]'$ has mean $E(\mathbf{a}) = \mathbf{0}$ and variance-covariance $\text{Var}(\mathbf{a}) = \boldsymbol{\Sigma}$. Now suppose that there are a total of N trials in the system of (16.A.1) and that we let $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{im}]'$ denote the responses for the ith trial, $[1, X_{i1}, \ldots, X_{ir}]$ denote the constant term and predictor variables for the ith trial, and $\mathbf{a}_i = [a_{i1}, \ldots, a_{im}]'$ denote the corresponding errors. We can represent the system in the matrix form

$$\begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ & \vdots & & \\ Y_{N1} & Y_{N2} & \cdots & Y_{Nm} \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1r} \\ & \vdots & & \\ 1 & X_{N1} & \cdots & X_{Nr} \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ & \vdots & & \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ & \vdots & & \\ a_{N1} & a_{N2} & \cdots & a_{Nm} \end{bmatrix} \tag{16.A.2}$$

or, more compactly,

$$\underset{N\times m}{\mathbf{Y}} = \underset{N\times(r+1)}{\mathbf{X}} \underset{(r+1)\times m}{\boldsymbol{\beta}} + \underset{N\times m}{\boldsymbol{\varepsilon}}, \tag{16.A.3}$$

where $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ denote the matrices of responses, predictors, parameters, and errors, respectively. For clarity, we also list their dimensions underneath. If $\mathbf{a}_i$ and $\mathbf{a}_j$ are uncorrelated for $i \neq j$, then the error specification becomes

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \tag{16.A.4a}$$

and

$$\text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_N. \tag{16.A.4b}$$

The model given in (16.A.2) or (16.A.3) is the so-called multivariate linear regression model.

Assume further that rank $(\mathbf{X}) = (r + 1) < N$ and $\mathbf{a}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, N$. We have the following results:

1. The MLE for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
2. $\hat{\boldsymbol{\beta}} \sim \mathbf{N}[\boldsymbol{\beta}, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}]$.

3. The MLE for Σ is $\hat{\Sigma} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{N} = \dfrac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{N}$,

   and $N\hat{\Sigma} \sim W_{N-r-1}(\Sigma)$, a Wishart distribution with $(N - r - 1)$ degrees of freedom.
4. $\hat{\beta}$ and $\hat{\Sigma}$ are independent.
5. $L(\hat{\beta}, \hat{\Sigma}) = \dfrac{e^{-Nm/2}}{(2\pi)^{Nm/2}|\hat{\Sigma}|^{N/2}}$, where $L(\cdot)$ is the likelihood function.

In certain applications, it may be of interest to test the null hypothesis that some of the predictor variables are unnecessary. Let us label these predictors $[X_{k+1}, X_{k+2}, \ldots, X_r]$ and label the corresponding parameters for the ith response $[\beta_{(k+1)i}, \beta_{(k+2)i}, \ldots, \beta_{ri}]$, $i = 1, \ldots, m$. By partitioning the matrices $\mathbf{X}$ and $\boldsymbol{\beta}$ for the N trials presented in (16.A.3) accordingly, i.e.,

$$\mathbf{X} = \underset{N\times(k+1)}{\mathbf{X}_1} \quad \underset{N\times(r-k)}{\mathbf{X}_2},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \underset{(k+1)\times m}{\boldsymbol{\beta}_{(1)}} \\ \underset{(r-k)\times m}{\boldsymbol{\beta}_{(2)}} \end{bmatrix},$$

we can rewrite (16.A.3) as

$$\underset{N\times m}{\mathbf{Y}} = \underset{N\times(r+1)}{\mathbf{X}} \quad \underset{(r+1)\times m}{\boldsymbol{\beta}} + \underset{N\times m}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{(1)} \\ \boldsymbol{\beta}_{(2)} \end{bmatrix} + \boldsymbol{\varepsilon}. \tag{16.A.5}$$

The hypothesis that the predictors in $\mathbf{X}_2$ are unnecessary is equivalent to testing that $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$ versus $H_a : \boldsymbol{\beta}_{(2)} \neq \mathbf{0}$. Under the hypothesis we have

$$\underset{N\times m}{\mathbf{Y}} = \underset{N\times(k+1)}{\mathbf{X}_1} \quad \underset{(k+1)\times m}{\boldsymbol{\beta}_{(1)}} + \underset{N\times m}{\boldsymbol{\varepsilon}}. \tag{16.A.6}$$

Let

$$\hat{\boldsymbol{\beta}}_{(1)} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \tag{16.A.7}$$

be the estimate of $\boldsymbol{\beta}$ for the reduced model of (16.A.6). The estimate of Σ in this case now becomes

$$\hat{\Sigma}_1 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N} = \frac{(\mathbf{Y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_{(1)})'(\mathbf{Y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_{(1)})}{N}, \tag{16.A.8}$$

and

$$L(\hat{\boldsymbol{\beta}}_{(1)}, \hat{\Sigma}_1) = \frac{e^{-Nm/2}}{(2\pi)^{Nm/2}|\hat{\Sigma}_1|^{N/2}}. \tag{16.A.9}$$

Let $\boldsymbol{\Omega}$ be the original parameter space for (16.A.3) and $\boldsymbol{\Omega}_0$ be the reduced parameter space under H_0. To test the null hypothesis H_0, we can apply the likelihood ratio test

$$\Lambda = \left[\frac{\max\limits_{\boldsymbol{\Omega}_0} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\max\limits_{\boldsymbol{\Omega}} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})}\right] = \left[\frac{L(\hat{\boldsymbol{\beta}}_{(1)}, \hat{\boldsymbol{\Sigma}}_1)}{L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})}\right] = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_1|}\right)^{N/2}.$$

It is well known that

$$-2\ln \Lambda = -2\ln\left[\frac{\max\limits_{\boldsymbol{\Omega}_0} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\max\limits_{\boldsymbol{\Omega}} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})}\right] = -2\ln\left[\frac{L(\hat{\boldsymbol{\beta}}_{(1)}, \hat{\boldsymbol{\Sigma}}_1)}{L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})}\right] = -N\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_1|}\right), \quad (16.A.10)$$

which follows approximately a chi-square distribution with the degree of freedom equal to $[\dim(\boldsymbol{\Omega}) - \dim(\boldsymbol{\Omega}_0)]$, i.e., $\chi^2((r-k)m)$. For a better approximation, one can use the modified statistic due to Bartlett (1947),

$$-\left[N - \frac{1}{2}(m + r + k + 1)\right]\ln\left[\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_1|}\right]. \quad (16.A.11)$$

It is clear that $0 < \Lambda \leq 1$. We reject the null hypothesis when Λ is small or $-2\ln\Lambda$ and Bartlett's modified statistic is large.

In the vector AR(p) model, we have

$$\mathbf{Z}_t = \boldsymbol{\tau} + \boldsymbol{\Phi}_1\mathbf{Z}_{t-1} + \cdots + \boldsymbol{\Phi}_p\mathbf{Z}_{t-p} + \mathbf{a}_t, \quad (16.A.12)$$

where $\mathbf{Z}_t = [Z_{1,t}, \ldots, Z_{m,t}]'$, $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_m]'$, the $\boldsymbol{\Phi}_i$ are $m \times m$ coefficient matrices, and the $\mathbf{a}_t$ are Gaussian vector white noise $\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Thus,

$$\begin{aligned}
Z_{1,t} &= \tau_1 + \phi_{1,11}Z_{1,t-1} + \cdots + \phi_{1,1m}Z_{m,t-1} + \cdots + \phi_{p,11}Z_{1,t-p} + \cdots \\
&\quad + \phi_{p,1m}Z_{m,t-p} + a_{1,t}, \\
&\vdots \\
Z_{m,t} &= \tau_m + \phi_{1,m1}Z_{1,t-1} + \cdots + \phi_{1,mm}Z_{m,t-1} + \cdots + \phi_{p,m1}Z_{1,t-p} + \cdots \\
&\quad + \phi_{p,mm}Z_{m,t-p} + a_{m,t}.
\end{aligned} \quad (16.A.13)$$

When testing the hypothesis

$$\begin{cases} H_0: \boldsymbol{\Phi}_p = \mathbf{0} \\ H_a: \boldsymbol{\Phi}_p \neq \mathbf{0}, \end{cases} \quad (16.A.14)$$

we have $N = (n - p)$, $r = pm$, and $k = (p-1)m$.

EXERCISES

16.1 Determine the stationarity and invertibility of the following two-dimensional vector models:

(a) $(I - \Phi_1 B)\mathbf{Z}_t = \mathbf{a}_t$, where $\Phi_1 = \begin{bmatrix} .8 & .3 \\ .1 & .6 \end{bmatrix}$, and $\Sigma = I$.

(b) $(I - \Phi_1 B)\mathbf{Z}_t = \mathbf{a}_t$, where $\Phi_1 = \begin{bmatrix} .4 & .2 \\ -.2 & .8 \end{bmatrix}$, and $\mathbf{a}_t \sim N\left(\mathbf{0}, \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}\right)$.

(c) $\mathbf{Z}_t = (I - \Theta_1 B)\mathbf{a}_t$, where $\Theta_1 = \begin{bmatrix} .6 & 1.2 \\ .4 & .8 \end{bmatrix}$, and $\Sigma = I$.

(d) $(I - \Phi_1 B)\mathbf{Z}_t = (I - \Theta_1 B)\mathbf{a}_t$, where $\Phi_1 = \begin{bmatrix} .8 & .3 \\ .1 & .6 \end{bmatrix}$, $\Theta_1 = \begin{bmatrix} .4 & .2 \\ .3 & .6 \end{bmatrix}$, and $\Sigma = I$.

(e) $(I - \Phi_1 B)\mathbf{Z}_t = (I - \Theta_1 B)\mathbf{a}_t$ and $\mathbf{a}_t \sim N(\mathbf{0}, \Sigma)$, where

$$\Phi_1 = \begin{bmatrix} 1 & .5 \\ 1 & -.5 \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} .5 & .6 \\ .7 & .8 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}.$$

16.2 Consider the following simple demand and supply model:

$$Q_t = \alpha_0 + \alpha_1 P_t + a_{1,t} \qquad \text{(demand)}$$
$$Q_t = \beta_0 + \beta_1 P_{t-1} + a_{2,t} \qquad \text{(supply)},$$

where Q_t represents quantity demanded at time t and P_t represents the price at time t, and $\mathbf{a}_t = [a_{1,t}, a_{2,t}]'$ is a two-dimensional stationary vector process.

(a) Express the model for $[Q_t, P_t]'$ as a joint vector process if $\mathbf{a}_t$ is a vector white noise process.

(b) Repeat the problem in part (*a*) if $a_{1,t} = (1 - \theta_{11}B)b_{1,t}$ and $a_{2,t} = (1 - \theta_{22}B)b_{2,t}$, where $\mathbf{b}_t = [b_{1,t}, b_{2,t}]'$ is a vector white noise process.

16.3 Find the marginal process $Z_{1,t}$ for the models given in Exercise 16.1.

16.4 Consider the models given in Exercise 16.1.

(a) Find the covariance matrix function, $\Gamma(k)$, for $k = 0, \pm 1, \pm 2$, and ± 3.

(b) Find the correlation matrix function, $\rho(k)$, for $k = \pm 1, \pm 2$, and ± 3.

16.5 Find the AR and MA representations for the models given in parts (d) and (e) of Exercise 16.1.

16.6 Consider the process

$$Z_{1,t} = Z_{1,t-1} + a_{1,t} + \theta a_{1,t-1}$$
$$Z_{2,t} = \phi Z_{1,t} + a_{2,t},$$

where $|\phi| < 1$, $|\theta| < 1$, and $\mathbf{a} = [a_{1,t}, a_{2,t}]' \sim N(\mathbf{0}, \Sigma)$.

(a) Write the process in a vector form.

(b) Is the process $[Z_{1,t}, Z_{2,t}]'$ stationary? Invertible?

(c) Write down the model for the vector of first differences $(\mathbf{I} - \mathbf{I}B)\mathbf{Z}_t$, where $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}]'$. Is the resulting model stationary? Invertible?

16.7 Calculate the partial autoregression matrices $\mathcal{P}(k)$ with $k = 1, 2, \ldots, 10$, for the models given in parts (a) and (c) of Exercise 16.1.

16.8 Calculate the partial lag correlation matrix function $\mathbf{P}(k)$ with $k = 1, 2, \ldots, 10$, for the models given in parts (a) and (c) of Exercise 16.1.

16.9 Calculate the one-step ahead forecast $\hat{\mathbf{Z}}_t(1)$ and the forecast error variance for the model given in part (d) of Exercise 16.1.

16.10 Consider the U.S. housing sales and housing starts given in Exercise 14.5.

(a) Calculate the sample correlation matrix function $\hat{\boldsymbol{\rho}}(k)$, with $k = 1, 2, \ldots, 15$, for the original series between 1965 and 1974.

(b) Repeat part (a) for the properly differenced series.

16.11 Repeat Exercise 16.10 for the partial lag correlation matrix functions $\mathbf{P}(k)$.

16.12 Complete the construction of a joint vector model for the data of U.S. housing sales and starts between 1965 and 1974 from the information obtained in Exercises 16.10 and 16.11.

16.13 Forecast the housing starts for 1975 using the joint vector model and compare the result with the forecast obtained in Exercise 14.5.

16.14 Build a joint vector model for the five series of U.S. hog data given in the appendix as Series W13.

16.15 Let n be the number of observations for the vector AR(p) model given in (16.5.17). Show that to properly estimate the parameters in the model, it is required that $n > \left(p + \frac{1}{2}\right)m + \frac{3}{2}$.

17

More on Vector Time Series

Although the basic procedures of model building between univariate time series and vector time series are the same, some important phenomena are unique in vector time series. In this chapter, we introduce concepts of cointegration, partial processes, and equivalent representations of a vector time series model. Understanding these concepts is important in the analysis and model building of vector time series.

17.1 Unit Roots and Cointegration in Vector Processes

A univariate nonstationary process or series Z_t is said to be integrated of order d, denoted as $I(d)$, if its $(d-1)$th difference is nonstationary but the dth difference, $\Delta^d Z_t = (1-B)^d Z_t$, is stationary. Thus, Z_t is $I(1)$ if Z_t is nonstationary but its first difference, $\Delta Z_t = (1-B)Z_t$, is stationary. Under this notion, an $I(0)$ process will be stationary. Because the difference of a stationary process is stationary, however, to restrict ourselves to a real integrated process we define an $I(0)$ process to be a stationary process where in terms of its MA representation, $Z_t = \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, we have $\psi(1) \neq 0$. Under this restriction, the differenced series, $\Delta Z_t = (1-B)Z_t$, of a stationary process, $Z_t = \theta(B)a_t$, where $\theta(\mathrm{B}) = \sum_{\mathrm{j}=1}^{\infty} \theta_{\mathrm{j}} B^j$, and $\sum_{\mathrm{j}=1}^{\infty} |\theta_{\mathrm{j}}| < \infty$, is not I(0) because the resulting MA representation of the differenced series is $\Delta Z_t = \theta(B)(1-B)a_t = \psi(B)a_t$ and $\psi(1) = 0$. Thus, for Z_t to be $I(1)$ we mean that $\Delta Z_t = (1-B)Z_t$ is not only stationary but is actually $I(0)$.

As a nonstationary process, an integrated series can wander extensively. Economic and market forces, however, tend to keep many of these integrated series together and form equilibrium relationships. Examples include short-term and long-term interest rates, income and consumption, and nominal GNP. This leads to the concept of cointegration.

An $(m \times 1)$ vector time series $\mathbf{Z}_t$ is said to be cointegrated of order (d, b), denoted as $CI(d, b)$, where $0 < b \leq d$, if each of its component series $Z_{i,t}$ is $I(d)$ but some linear combination of the series $\boldsymbol{\beta}'\mathbf{Z}_t$ is $I(d-b)$ for some nonzero $(1 \times \mathrm{m})$ constant vector $\boldsymbol{\beta}'$. In such a case, $\boldsymbol{\beta}'$ is called a cointegrating vector which is sometimes also called the long-run parameter. The cointegrating vector is clearly not unique because if $\boldsymbol{\beta}'\mathbf{Z}_t$ is stationary, then so too is $c\boldsymbol{\beta}'\mathbf{Z}_t$ for any nonzero constant c. Hence, $c\boldsymbol{\beta}'$ is also a cointegrating vector. The most common case is $d = b = 1$, and it will be used in our discussions.

The dangers of differencing all the components of a vector time series was reported by Box and Tiao (1977), who showed that nonstationarity of a vector process can be caused by a small number of nonstationary components. Granger (1986) and Engle and Granger (1987) further developed these ideas and proposed the concept of cointegration. This concept has been widely used in many interesting applications, especially in the area of economics and finance. See, for example, Kremers (1989).

As an illustration, consider the two-dimensional vector AR(1) process

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}. \tag{17.1.1}$$

Clearly, the component $Z_{1,t}$ is a random walk,

$$(1 - B)Z_{1,t} = a_{1,t}. \tag{17.1.2}$$

For the component $Z_{2,t}$, we have $Z_{2,t} = -\phi Z_{1,t-1} + a_{2,t}$. After differencing, this equation becomes

$$\begin{aligned} (1 - B)Z_{2,t} &= -\phi(1 - B)Z_{1,t-1} + (1 - B)a_{2,t} \\ &= -\phi a_{1,t-1} + a_{2,t} - a_{2,t-1}, \end{aligned} \tag{17.1.3}$$

which follows an MA(1) process. Thus, each component series is an $I(1)$ process. However, let $\boldsymbol{\beta}' = [\phi \quad 1]$, then the linear combination

$$\boldsymbol{\beta}'\mathbf{Z}_t = \phi Z_{1,t} + Z_{2,t} = \phi Z_{1,t} - \phi Z_{1,t-1} + a_{2,t} = \phi a_{1,t} + a_{2,t}$$

is clearly $I(0)$. Hence, $\mathbf{Z}_t$ is cointegrated with a cointegrating vector $\boldsymbol{\beta}' = [\phi \quad 1]$. This interesting phenomenon is shown in Figure 17.1, where $Z_{1,t}$ and $Z_{2,t}$ are each nonstationary, but their linear combination is stationary.

More generally, if the $(m \times 1)$ vector series $\mathbf{Z}_t$ contains more than two components, each being $I(1)$, then there may exist k $(< m)$ linearly independent $(1 \times m)$ vectors $\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_k'$ such that $\boldsymbol{\beta}'\mathbf{Z}_t$ is a stationary $(k \times 1)$ vector process, where

$$\boldsymbol{\beta}' = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \\ \vdots \\ \boldsymbol{\beta}_k' \end{bmatrix}$$

is a $(k \times m)$ cointegrating matrix. Clearly, vectors $\boldsymbol{\beta}_i'$ are not unique, neither is $\boldsymbol{\beta}'$. If for any other $(1 \times m)$ vector $\mathbf{b}'$ that is linearly independent of the rows of $\boldsymbol{\beta}'$, however, we have that $\mathbf{b}'\mathbf{Z}_t$ is nonstationary, then the $\mathbf{Z}_t$ is said to be cointegrated of rank k. The vectors $\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_k'$ form a basis for the space of the cointegrating vectors known as the cointegration space.

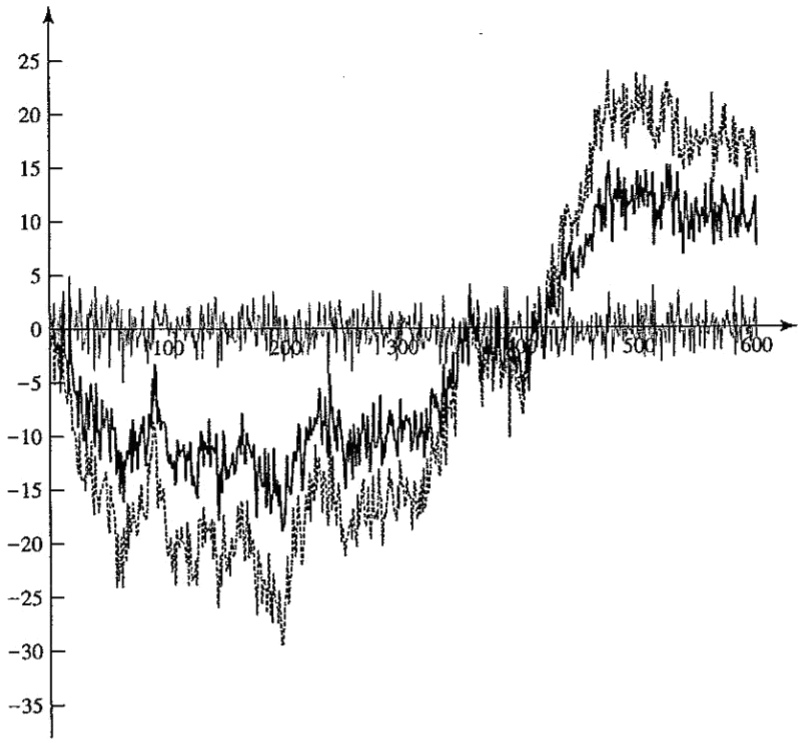


FIGURE 17.1 A cointegrated process where $Z_{1,t}$ (a heavy solid line) and $Z_{2,t}$ (a dotted line) are each nonstationary, but their linear combination Y_t (a thin solid line) is stationary.

17.1.1 Representations of Nonstationary Cointegrated Processes

MA Representation Because each component of $\mathbf{Z}_t$ is $I(1)$, the implication is that $\Delta\mathbf{Z}_t = (\mathbf{I} - \mathbf{I}B)\mathbf{Z}_t$ is stationary. Let $\boldsymbol{\theta}_0 = E(\Delta\mathbf{Z}_t)$. We can write $\Delta\mathbf{Z}_t$ in the MA form

$$\Delta\mathbf{Z}_t - \boldsymbol{\theta}_0 = \boldsymbol{\Psi}(B)\mathbf{a}_t, \tag{17.1.4}$$

where $\mathbf{a}_t$ is the vector white noise process with mean $E(\mathbf{a}_t) = \mathbf{0}$, variance-covariance matrix $E(\mathbf{a}_t\mathbf{a}_t') = \boldsymbol{\Sigma}$, $\boldsymbol{\Psi}(B) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j B^j$, $\boldsymbol{\Psi}_0 = \boldsymbol{I}$, and the coefficient matrices $\boldsymbol{\Psi}_j$ is absolutely summable.

Let $\mathbf{X}_t = \boldsymbol{\Psi}(B)\mathbf{a}_t$. Then, $\Delta\mathbf{Z}_t - \boldsymbol{\theta}_0 = \mathbf{X}_t$ and

$$\mathbf{Z}_t = \mathbf{Z}_{t-1} + \boldsymbol{\theta}_0 + \mathbf{X}_t. \tag{17.1.5}$$

Applying a direct substitution for $\mathbf{Z}_{t-1}$ in (17.1.5) gives

$$\mathbf{Z}_t = \mathbf{Z}_0 + \boldsymbol{\theta}_0 t + \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_t. \tag{17.1.6}$$

A straightforward extension of Equations (9.4.6) and (9.4.7) to vector processes gives

$$\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_t = \Psi(1)[\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t] + \mathbf{Y}_t - \mathbf{Y}_0, \tag{17.1.7}$$

where $\mathbf{Y}_t = \sum_{j=0}^{\infty} \Psi_j^* \mathbf{a}_{t-j}$ and $\Psi_j^* = -\sum_{i=1}^{\infty} \Psi_{j+i}$ which is an absolutely summable matrix sequence. Hence, the vector process $\mathbf{Y}_t$ is stationary. Combining (17.1.6) and (17.1.7), we get

$$\mathbf{Z}_t = \mathbf{Z}_0 + \boldsymbol{\theta}_0 t + \Psi(1)[\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t] + \mathbf{Y}_t - \mathbf{Y}_0 \tag{17.1.8}$$

and

$$\boldsymbol{\beta}'\mathbf{Z}_t = \boldsymbol{\beta}'(\mathbf{Z}_0 - \mathbf{Y}_0) + \boldsymbol{\beta}'\boldsymbol{\theta}_0 t + \boldsymbol{\beta}'\Psi(1)[\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t] + \boldsymbol{\beta}'\mathbf{Y}_t. \tag{17.1.9}$$

Clearly, the process $\mathbf{b}'[\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t]$ is nonstationary for any nonzero $(1 \times m)$ vector $\mathbf{b}'$; hence, (17.1.9) implies that $\boldsymbol{\beta}'\mathbf{Z}_t$ will be stationary if and only if

$$\boldsymbol{\beta}'\boldsymbol{\theta}_0 = \mathbf{0} \tag{17.1.10}$$

and

$$\boldsymbol{\beta}'\Psi(1) = \mathbf{0}. \tag{17.1.11}$$

Equation (17.1.10) places a restriction on the nonzero drift across the components, whereas Equation (17.1.11) says that the determinant $|\Psi(B)| = 0$ at $B = 1$; hence, $\Psi(B)$ is noninvertible. By (17.1.4), this implies that we cannot invert the MA form and hence can never represent a cointegrated process with a vector AR form in $\Delta\mathbf{Z}_t$. The vector AR representation of a cointegrated process must be in terms of $\mathbf{Z}_t$ directly.

EXAMPLE 17.1 It can be easily shown from (17.1.2) and (17.1.3) that the MA representation of the two-dimensional cointegrated process given in (17.1.1) is given by

$$\begin{bmatrix} \Delta Z_{1,t} \\ \Delta Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi B & (1 - B) \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}.$$

Note that

$$\Psi(B) = \begin{bmatrix} 1 & 0 \\ -\phi B & (1 - B) \end{bmatrix} \quad \text{and} \quad \Psi(1) = \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} \neq \mathbf{0},$$

but $|\Psi(1)| = 0$. Thus, $\Psi(B)$ is not invertible, and we cannot represent the process in an AR form in terms of $\Delta\mathbf{Z}_t$.

AR Representation Now, suppose that $\mathbf{Z}_t$ is nonstationary and can be represented as an AR(p) model

$$\Phi_p(B)\mathbf{Z}_t = \boldsymbol{\theta}_0 + \mathbf{a}_t \tag{17.1.12}$$

such that $|\Phi_p(B)| = 0$ contains some unit roots, where $\Phi_p(B) = \mathbf{I} - \Phi_1 B - \cdots - \Phi_p B^p$. If each component of $\mathbf{Z}_t$ is $I(1)$, then, from (17.1.4),

$$(1 - B)\mathbf{Z}_t = \boldsymbol{\theta}_0 + \Psi(B)\mathbf{a}_t. \tag{17.1.13}$$

Hence,

$$\begin{aligned}(1 - B)\Phi_p(B)\mathbf{Z}_t &= \Phi_p(B)\boldsymbol{\theta}_0 + \Phi_p(B)\Psi(B)\mathbf{a}_t \\ &= \Phi_p(1)\boldsymbol{\theta}_0 + \Phi_p(B)\Psi(B)\mathbf{a}_t. \end{aligned} \tag{17.1.14}$$

Multiplying $(1 - B)$ on both sides of (17.1.12) and comparing with (17.1.14), we have

$$(1 - B)\mathbf{a}_t = \Phi_p(1)\boldsymbol{\theta}_0 + \Phi_p(B)\Psi(B)\mathbf{a}_t, \tag{17.1.15}$$

where we note that $(1 - B)\boldsymbol{\theta}_0 = \mathbf{0}$. Because Equation (17.1.15) holds for any $\mathbf{a}_t$, however, it implies that

$$\Phi_p(1)\boldsymbol{\theta}_0 = \mathbf{0} \tag{17.1.16}$$

and

$$(1 - B)\mathbf{I} = \Phi_p(B)\Psi(B)$$

for any B. Hence,

$$\Phi_p(1)\Psi(1) = \mathbf{0}. \tag{17.1.17}$$

Comparing (17.1.10) and (17.1.11) with (17.1.16) and (17.1.17), we see that in terms of an AR representation, the matrix $\Phi_p(1)$ must belong to the space spanned by the rows of $\boldsymbol{\beta}'$. That is,

$$\Phi_p(1) = \mathbf{M}\boldsymbol{\beta}' \tag{17.1.18}$$

for some $(m \times k)$ matrix $\mathbf{M}$.

EXAMPLE 17.2 Although an integrated process cannot be represented in an AR form in terms of $\Delta\mathbf{Z}_t$, its AR representation does exist in terms of $\mathbf{Z}_t$ itself. Recall the two-dimensional cointegrated process and its AR presentation in terms of $\mathbf{Z}_t$ given in (17.1.1):

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}.$$

In this case,

$$\Phi_1(B) = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} B \right)$$

and the cointegrating matrix

$$\boldsymbol{\beta}' = [\phi \quad 1].$$

Thus, $\mathbf{Z}_t$ is a cointegrated process of rank $k = 1$. Note that

$$\Phi_1(1) = \begin{bmatrix} 0 & 0 \\ \phi & 1 \end{bmatrix} = \mathbf{M}\boldsymbol{\beta}' = \mathbf{M}[\phi \quad 1],$$

where

$$\mathbf{M} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Error-Correction Representation Next, note that in the AR representation, the matrix polynomial $\Phi_p(B)$ can be written as

$$\Phi_p(B) = \mathbf{I} - \Phi_1 \mathrm{B} - \cdots - \Phi_p B^p = (\mathbf{I} - \boldsymbol{\lambda} B) - (\boldsymbol{\alpha}_1 B + \cdots + \boldsymbol{\alpha}_{p-1} B^{p-1})(1 - B), \tag{17.1.19}$$

where $\boldsymbol{\lambda} = \Phi_1 + \cdots + \Phi_p$ and $\boldsymbol{\alpha}_j = -(\Phi_{j+1} + \cdots + \Phi_p)$ for $j = 1, 2, \ldots, p-1$. Thus, (17.1.12) can be rewritten as

$$(\mathbf{I} - \boldsymbol{\lambda} B)\mathbf{Z}_t - (\boldsymbol{\alpha}_1 B + \cdots + \boldsymbol{\alpha}_{p-1} B^{p-1})\Delta \mathbf{Z}_t = \boldsymbol{\theta}_0 + \mathbf{a}_t$$

or

$$\mathbf{Z}_t = \boldsymbol{\theta}_0 + \boldsymbol{\lambda}\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_1 \Delta \mathbf{Z}_{t-1} + \cdots + \boldsymbol{\alpha}_{p-1}\Delta \mathbf{Z}_{t-p+1} + \mathbf{a}_t. \tag{17.1.20}$$

Subtracting $\mathbf{Z}_{t-1}$ from both sides of (17.1.20) gives

$$\Delta \mathbf{Z}_t = \boldsymbol{\theta}_0 + \boldsymbol{\gamma}\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_1 \Delta \mathbf{Z}_{t-1} + \cdots + \boldsymbol{\alpha}_{p-1}\Delta \mathbf{Z}_{t-p+1} + \mathbf{a}_t, \tag{17.1.21}$$

where $\boldsymbol{\gamma} = \boldsymbol{\lambda} - \mathbf{I} = -\Phi_p(1) = -\mathbf{M}\boldsymbol{\beta}'$ by (17.1.18) and (17.1.19). Hence,

$$\Delta \mathbf{Z}_t = \boldsymbol{\theta}_0 - \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\alpha}_1 \Delta \mathbf{Z}_{t-1} + \cdots + \boldsymbol{\alpha}_{p-1}\Delta \mathbf{Z}_{t-p+1} + \mathbf{a}_t \tag{17.1.22}$$

for some $(m \times k)$ matrix $\mathbf{M}$, where $\mathbf{Y}_{t-1} = \boldsymbol{\beta}'\mathbf{Z}_{t-1}$ is a $(k \times 1)$-dimensional stationary process. Equations (17.1.21) and (17.1.22) imply that the differenced series $\Delta \mathbf{Z}_t$ of a cointegrated process $\mathbf{Z}_t$ cannot be described using only the values of past lagged differences, $\Delta \mathbf{Z}_j$, for $j < t$. The process must include an "error-correction" term, $\mathbf{M}\mathbf{Y}_{t-1} = \mathbf{M}\boldsymbol{\beta}'\mathbf{Z}_{t-1}$. If we regard the relation $\Delta \mathbf{Z}_t$ in terms of its own past lagged values $\Delta \mathbf{Z}_j$ for $j < t$ as a long-run equilibrium, then the term $\mathbf{Y}_{t-1} = \boldsymbol{\beta}'\mathbf{Z}_{t-1}$ can be viewed as an error from the equilibrium and the coefficient $\mathbf{M}$ is the required adjustment or correction for this error. Accordingly,

Equation (17.1.21) or (17.1.22) is known as the error-correction representation of the cointegrating system. The representation was first proposed by Davidson et al. (1978).

EXAMPLE 17.3 For the cointegrated process given in (17.1.1),

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix},$$

we have $p = 1$, $\boldsymbol{\theta}_0 = \mathbf{0}$, $\boldsymbol{\Phi}_j = \mathbf{0}$ for $j \geq 2$, and hence $\boldsymbol{\alpha}_j = \mathbf{0}$ for $j \geq 1$. From Example 17.2,

$$\mathbf{M} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad Y_{t-1} = \boldsymbol{\beta}'\mathbf{Z}_{t-1} = [\phi \quad 1] \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} = \phi Z_{1,t-1} + Z_{2,t-1}.$$

Thus, the error-correction representation of (17.1.1) is given by

$$\begin{aligned} \Delta \mathbf{Z}_t = \begin{bmatrix} 0 & 0 \\ -\phi & -1 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} &= -\begin{bmatrix} 0 \\ 1 \end{bmatrix} [\phi \quad 1] \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \\ &= -\begin{bmatrix} 0 \\ 1 \end{bmatrix} Y_{t-1} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \end{aligned}$$

where $Y_{t-1} = \phi Z_{1,t-1} + Z_{2,t-1}$ is the stationary error-correction term.

17.1.2 Decomposition of $\mathbf{Z}_t$

Assume that nonstationarity of $\mathbf{Z}_t$ is due to unit roots in the AR matrix polynomial. That is, in the AR representation

$$\boldsymbol{\Phi}_p(B)\mathbf{Z}_t = \boldsymbol{\theta}_0 + \mathbf{a}_t, \tag{17.1.23}$$

some of the roots of $|\boldsymbol{\Phi}_p(B)| = 0$ are equal to 1. More specifically, assume that there are h ($\leq m$) unit roots and that all other roots of $|\boldsymbol{\Phi}_p(B)| = 0$ are outside the unit circle. Thus, $|\boldsymbol{\Phi}_p(1)| = |\mathbf{I} - \sum_{j=1}^{p} \boldsymbol{\Phi}_j| = 0$; the matrix $\sum_{j=1}^{p} \boldsymbol{\Phi}_j$ has h eigenvalues equal to 1 and $(m - h)$ other eigenvalues less than 1 in absolute value. Let $\mathbf{P}$ be an $(m \times m)$ matrix such that $\mathbf{P}^{-1}[\sum_{j=1}^{p} \boldsymbol{\Phi}_j]\mathbf{P} = \text{diag}[\mathbf{I}_h \quad \boldsymbol{\Lambda}_k] = \mathbf{J}$, where $k = (m - h)$ and $\mathbf{J}$ is the Jordan canonical form of $\sum_{j=1}^{p} \boldsymbol{\Phi}_j$. Then,

$$\mathbf{P}^{-1}[-\boldsymbol{\Phi}_p(1)]\mathbf{P} = \mathbf{P}^{-1}\left[\sum_{j=1}^{p} \boldsymbol{\Phi}_j - \mathbf{I}\right]\mathbf{P} = \mathbf{J} - \mathbf{I} = \text{diag}[\mathbf{0} \quad \boldsymbol{\Lambda}_k - \mathbf{I}_k]. \tag{17.1.24}$$

Let $\mathbf{P} = [\mathbf{S} \quad \mathbf{R}]$ and $(\mathbf{P}^{-1})' = [\mathbf{H} \quad \boldsymbol{\beta}]$, where $\mathbf{S}$ and $\mathbf{H}$ are $(m \times h)$ matrices, and $\mathbf{R}$ and $\boldsymbol{\beta}$ are $(m \times k)$ matrices so that

$$-\boldsymbol{\Phi}_p(1) = \mathbf{P}[\mathbf{J} - \mathbf{I}]\mathbf{P}^{-1} = [\mathbf{S} \quad \mathbf{R}] \begin{bmatrix} \mathbf{0}_{h\times h} & \mathbf{0}_{h\times k} \\ \mathbf{0}_{k\times h} & \boldsymbol{\Lambda}_k - \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \mathbf{H}' \\ \boldsymbol{\beta}' \end{bmatrix} = \mathbf{R}(\boldsymbol{\Lambda}_k - \mathbf{I}_k)\boldsymbol{\beta}'. \tag{17.1.25}$$

Recall that $\gamma = -\Phi_p(1)$ in (17.1.21), which can therefore also be written as

$$\Delta \mathbf{Z}_t = \boldsymbol{\theta}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_1\Delta\mathbf{Z}_{t-1} + \cdots + \boldsymbol{\alpha}_{p-1}\Delta\mathbf{Z}_{t-p+1} + \mathbf{a}_t, \quad (17.1.26)$$

where $\boldsymbol{\alpha} = \mathbf{R}(\Lambda_k - \mathbf{I}_k) = -\mathbf{M}$. It follows that although $\mathbf{Z}_t$ is nonstationary, the k $(= (m - h))$ linear combinations $\boldsymbol{\beta}'\mathbf{Z}_t$ are stationary, where $k = \text{rank}(\Phi_p(1)) = \text{rank}(\boldsymbol{\beta}')$. That is, $\mathbf{Z}_t$ is cointegrated of rank k.

Let $\mathbf{Y}_{1,t} = \mathbf{H}'\mathbf{Z}_t$. Because $\mathbf{P}^{-1}$ is nonsingular, the rows of $\mathbf{H}'$ are linearly independent of the rows of $\boldsymbol{\beta}'$. Thus, $\mathbf{Y}_{1,t}$ must be nonstationary. In fact, $\text{rank}(\mathbf{H}') = h$, which also equals the number of unit roots. This h-dimensional $\mathbf{Y}_{1,t}$ process is actually the underlying common stochastic trend related to the h unit roots, which is also known as the common factors in the process of $\mathbf{Z}_t$. It is the common driving force for the nonstationary phenomenon in each component $Z_{i,t}$ of $\mathbf{Z}_t$.

Let $\mathbf{Y}_{2,t} = \boldsymbol{\beta}'\mathbf{Z}_t$. Then,

$$\begin{bmatrix}\mathbf{Y}_{1,t}\\ \mathbf{Y}_{2,t}\end{bmatrix} = \begin{bmatrix}\mathbf{H}'\\ \boldsymbol{\beta}'\end{bmatrix}\mathbf{Z}_t = \mathbf{P}^{-1}\mathbf{Z}_t \quad (17.1.27)$$

and

$$\mathbf{Z}_t = \mathbf{P}\begin{bmatrix}\mathbf{Y}_{1,t}\\ \mathbf{Y}_{2,t}\end{bmatrix} = \mathbf{S}\mathbf{Y}_{1,t} + \mathbf{R}\mathbf{Y}_{2,t}. \quad (17.1.28)$$

Hence, the cointegrated process $\mathbf{Z}_t$ is a linear combination of the h-dimensional purely nonstationary component, $\mathbf{Y}_{1,t} = \mathbf{H}'\mathbf{Z}_t$, and the k-dimensional stationary component, $\mathbf{Y}_{2,t} = \boldsymbol{\beta}'\mathbf{Z}_t$, where $k = (\text{m} - h)$. The concepts discussed in this section are closely related to the common factor analysis studied by Pena and Box (1987), Stock and Watson (1988), and many others.

17.1.3 Testing and Estimating Cointegration

Following the above discussion, we see that the first step in testing cointegration is to test the null hypothesis of a unit root in each component series $Z_{i,t}$ individually using the unit root tests discussed in Chapter 9. If the hypothesis is not rejected, then the next step is to test cointegration among the components, i.e., to test whether $\mathbf{Y}_t = \boldsymbol{\beta}'\mathbf{Z}_t$ is stationary for some matrix or vector $\boldsymbol{\beta}'$. Sometimes the choice of the matrix/vector $\boldsymbol{\beta}'$ is based on some theoretical consideration. For example, if $\mathbf{Z}_t = [Z_{1,t} \quad Z_{2,t}]'$, where $Z_{1,t}$ represents revenue and $Z_{2,t}$ represents expenditure, we may want to test whether the revenues and expenditures are in some long-run equilibrium relation and therefore whether $Y_t = Z_{1,t} - Z_{2,t}$ is stationary. In such a case, we choose $\boldsymbol{\beta}' = [1 \quad -1]$.

In testing for cointegration in $\mathbf{Z}_t$ with a known cointegrating vector $\boldsymbol{\beta}'$, we can formulate the null hypothesis to test whether the process $Y_t = \boldsymbol{\beta}'\mathbf{Z}_t$ contains a unit root so that we can again use the tests discussed in Chapter 9. We will conclude that $\mathbf{Z}_t$ is cointegrated if the null hypothesis is rejected.

When the cointegrating vector is unknown, we can use the following methods to test and estimate the cointegration.

The Regression Method First, note that if the m-dimensional vector process $\mathbf{Z}_t' = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]$ is cointegrated, then there is a nonzero $m \times 1$ vector $\boldsymbol{\beta}' = [c_1, c_2, \ldots, c_m]$ such that $\boldsymbol{\beta}'\mathbf{Z}_t$ is stationary. With no loss of generality, say $c_1 \neq 0$. Then, $(1/c_1)\boldsymbol{\beta}'$ is also a cointegrating vector. Thus, a very natural approach that was also suggested by Engle and Granger (1987) to test and estimate cointegration is to consider the regression model for $Z_{1,t}$

$$Z_{1,t} = \phi_1 Z_{2,t} + \cdots + \phi_{m-1} Z_{m,t} + \varepsilon_t \tag{17.1.29}$$

and check whether the error series ε_t is $I(1)$ or $I(0)$. If ε_t is $I(1)$ and contains a unit root, then $\mathbf{Z}_t$ cannot be cointegrated. Otherwise, if ε_t is $I(0)$, then $\mathbf{Z}_t$ is cointegrated with a normalizing cointegrating vector given by $\boldsymbol{\beta}' = [1, \phi_1, \ldots, \phi_m]$. The results remain the same for the above model with a constant term included.

There are many subtle points for the regression model in (17.1.29). In drawing inferences using the OLS estimated model, the following should be noted:

1. In testing the error series for nonstationarity, we calculate the OLS estimate, $\hat{\boldsymbol{\beta}}' = [1, \hat{\phi}_1, \ldots, \hat{\phi}_m]$ and use the residual series $\hat{\varepsilon}_t$ for the test.
2. The estimates $\hat{\phi}_i$ do not have asymptotic t-distributions, and the standard t-test for $\phi_i = 0$ is not applicable unless ε_t is $I(0)$. As indicated in Chapter 15, the regression in (17.1.29) will produce spurious regression if ε_t is nonstationary.
3. If $\hat{\varepsilon}_t$ is $I(0)$ and $\mathbf{Z}_t$ is cointegrated, then we can let $\mathbf{Y}_{t-1} = \hat{\boldsymbol{\beta}}'\mathbf{Z}_{t-1}$ in (17.1.22) and perform the estimation of the error-correction model.
4. We estimate the cointegrating vector by normalizing the first element to be unity. Clearly, we can normalize any nonzero element c_i and regress $Z_{i,t}$ on other variables in estimating the regression. Although most of the time the results are consistent, inconsistent conclusions could sometimes occur, which is a weakness of the approach. Because of its simplicity, however, the approach is still very commonly used.

To test whether the error series ε_t is $I(1)$, Engle and Granger (1987) proposed several test statistics, including the standard Dickey–Fuller test and the augmented Dickey–Fuller test. For these later tests, recall from Equation (9.4.22), in terms of the error term process in (17.1.29), that to test for a unit root in the general case we either test $H_0 : \varphi = 1$ against $H_1 : \varphi < 1$ for the model

$$\varepsilon_t = \varphi\varepsilon_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta\varepsilon_{t-j} + a_t \tag{17.1.30}$$

or, equivalently, test $H_0 : \lambda = 0$ against $H_1 : \lambda < 0$ for the model

$$\Delta\varepsilon_t = \lambda\varepsilon_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta\varepsilon_{t-j} + a_t. \tag{17.1.31}$$

In testing, note that we use the residuals $\hat{\varepsilon}_t$ obtained from the OLS fitting of (17.1.29), but they are only estimates and not the actual error terms. To adjust for the difference, instead of using

TABLE 17.1 Critical values of T, for $H_0 : \lambda = 0$ against $H_1 : \lambda < 0$.

Level of significance:	1%	5%
$p = 1$	−4.07	−3.37
$p > 1$	−3.73	−3.17

the standard Dickey-Fuller tables, Engle and Granger (1987) conducted a Monte Carlo experiment on the model

$$\Delta\hat{\varepsilon}_t = \lambda\hat{\varepsilon}_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta\hat{\varepsilon}_{t-j} + a_t \tag{17.1.32}$$

and produced the critical values given in Table 17.1 for the t-statistic, $T = \hat{\lambda}/S_{\hat{\lambda}}$. One should reject the null hypothesis if the T-value is less than the critical value and conclude that cointegration exists among the variables in the vector.

EXAMPLE 17.4 From the discussion of this section, it is clear that in modeling vector time series, one should not automatically difference the component series even if they are each nonstationary. In fact, in terms of a vector AR model, the results of this section show that if the component series are cointegrated, then a stationary vector AR model for the differenced series does not exist. One should either construct a nonstationary AR vector model for the original series or construct an error-correction model for the differenced series. Let us now further examine the Lydia Pinkham data discussed in Chapter 16 and test whether the advertising and sales series are cointegrated.

Let $Z_{1,t}$ be the advertising series and $Z_{2,t}$ be the sales series. To test whether they are each $I(d)$, first, we obtain the following separate OLS estimated regression for each series:

$$\begin{aligned} Z_{1,t} &= \underset{(80.53)}{163.86} + \underset{(.07965)}{.8251}\ Z_{1,t-1} \\ Z_{2,t} &= \underset{(98.74)}{148.3} + \underset{(.05079)}{.9222}\ Z_{2,t-1} \end{aligned}$$

where the values in the parentheses under estimates are their standard errors. Thus, $T = (.8251 - 1)/.07965 = -2.196$ and $T = (.9222 - 1)/.05079 = -1.532$, respectively. At 5%, the critical value from Table G of the appendix is −2.93. Thus, the hypothesis of a unit root is not rejected for either series, and both $Z_{1,t}$ and $Z_{2,t}$ are $I(d)$, with $d \geq 1$. To find the order of d for the two series, we now consider the following OLS estimated regressions on their differences:

$$\begin{aligned} \Delta Z_{1,t} &= \underset{(31.56)}{2.16} + \underset{(.1409)}{.0253}\ \Delta Z_{1,t-1} \\ \Delta Z_{2,t} &= \underset{(30.03)}{4.02} + \underset{(.1278)}{.429}\ \Delta Z_{2,t-1} \end{aligned}$$

where the values in the parentheses under estimates are their standard errors. Now, $T = (.0253 - 1)/.1409 = -6.918$, and $T = (.429 - 1)/.1278 = -4.468$, respectively. At 5%, the critical value from Table G is -2.93. Thus, the hypothesis of a unit root is rejected for either differenced series, and we conclude that both $Z_{1,t}$ and $Z_{2,t}$ are $I(1)$.

Next, we consider the regression model

$$Z_{1,t} = \alpha + \phi Z_{2,t} + \varepsilon_t.$$

The OLS estimated equation is

$$\underset{}{Z_{1,t} =} \underset{(90.07)}{45.2} + \underset{(.0457)}{.4904}\ Z_{2,t},$$

with the estimated regression of (17.1.32) for $p = 1$,

$$\Delta\hat{\varepsilon}_t = \underset{(.1177)}{-.421}\ \hat{\varepsilon}_{t-1}.$$

From Table 17.1, because the value of $T = -.421/.1177 = -3.58$ is less than the critical value -3.37 at the 5% level, we reject the hypothesis of a unit root. Thus, we conclude that the residual series is $I(0)$ and the advertising and sales are cointegrated.

It will be interesting to see whether the same conclusion will be reached when sales is regressed on advertising. In this case, we consider the regression model

$$Z_{2,t} = \alpha + \phi Z_{1,t} + \varepsilon_t.$$

The OLS estimated equation is

$$\underset{}{Z_{2,t} =} \underset{(127.4)}{488.8} + \underset{(.1269)}{1.4346}\ Z_{1,t},$$

with the following estimated regression of $\Delta\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}$,

$$\Delta\hat{\varepsilon}_t = \underset{(.0972)}{-.3005}\ \hat{\varepsilon}_{t-1}.$$

From Table 17.1, because the value of $T = -.3005/.0972 = -3.09$ is not less than the critical values -3.37 at the 5% level, we cannot reject the hypothesis of a unit root. Thus, the residual series is not $I(0)$, and the advertising and sales series are not cointegrated, which is different from the earlier conclusion. It should be noted, however, at 1% significance level, the critical value is -4.07 from Table 17.1. Both cases become insignificant, and we reach the same conclusion that the advertising and sales series are not cointegrated. This may be expected. For most companies, especially growth companies, we expect their revenues after expenditures to grow.

The Likelihood Ratio Test Let $\mathbf{Z}_t$ be a nonstationary $(m \times 1)$ vector process that can be represented as a vector AR(p) form

$$\Phi_p(B)\mathbf{Z}_t = \boldsymbol{\theta}_0 + \mathbf{a}_t, \tag{17.1.33}$$

where $\Phi_p(B) = \mathbf{I} - \Phi_1 B - \cdots - \Phi_p B^p$. When $\mathbf{Z}_t$ is cointegrated, by (17.1.21), it can be written as

$$\Delta\mathbf{Z}_t = \boldsymbol{\theta}_0 + \boldsymbol{\gamma}\mathbf{Z}_{t-1} + \boldsymbol{\alpha}_1\Delta\mathbf{Z}_{t-1} + \cdots + \boldsymbol{\alpha}_{p-1}\Delta\mathbf{Z}_{t-p+1} + \mathbf{a}_t. \tag{17.1.34}$$

If in (17.1.33) $|\Phi_p(B)| = 0$ contains h unit roots and $k = (m - h)$ roots outside the unit circle, then, from the discussion of Sections 17.1.1 and 17.1.2, it is equivalent to stating that the process $\mathbf{Z}_t$ contains k cointegrating relations. Under this null hypothesis H_0, we have

$$\boldsymbol{\gamma} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{17.1.35}$$

for some $(m \times k)$ matrix $\boldsymbol{\alpha}$, and $\mathbf{Y}_t = \boldsymbol{\beta}'\mathbf{Z}_t$ is a $(k \times 1)$-dimensional stationary process. Thus, only k linear combinations, $\boldsymbol{\beta}'\mathbf{Z}_{t-1}$, of $\mathbf{Z}_{t-1}$, which are stationary, will actually appear in (17.1.34).

If the vector white noise process $\mathbf{a}_t$ is Gaussian, i.e., if the $\mathbf{a}_t$ are i.i.d. $N(\mathbf{0}, \boldsymbol{\Sigma})$, then given a sample of $(n + p)$ observations on $\mathbf{Z}_t$, i.e., $\{\mathbf{Z}_{-p+1}, \mathbf{Z}_{-p+2}, \ldots, \mathbf{Z}_1, \ldots, \mathbf{Z}_n\}$, the likelihood function of $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n\}$ conditional on $\{\mathbf{Z}_{-p+1}, \mathbf{Z}_{-p+2}, \ldots, \mathbf{Z}_0\}$ is

$$\begin{aligned} L(\boldsymbol{\theta}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{p-1}, \boldsymbol{\Sigma}) &= (2\pi)^{-nm/2}|\boldsymbol{\Sigma}|^{-n/2} \\ &\times \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}[\Delta\mathbf{Z}_t - \boldsymbol{\theta}_0 - \boldsymbol{\gamma}\mathbf{Z}_{t-1} - \boldsymbol{\alpha}_1\Delta\mathbf{Z}_{t-1} - \cdots - \boldsymbol{\alpha}_{p-1}\Delta\mathbf{Z}_{t-p+1}]'\right. \\ &\left.\boldsymbol{\Sigma}^{-1}[\Delta\mathbf{Z}_t - \boldsymbol{\theta}_0 - \boldsymbol{\gamma}\mathbf{Z}_{t-1} - \boldsymbol{\alpha}_1\Delta\mathbf{Z}_{t-1} - \cdots - \boldsymbol{\alpha}_{p-1}\Delta\mathbf{Z}_{t-p+1}]\right\}. \end{aligned} \tag{17.1.36}$$

The MLE of $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{p-1}, \boldsymbol{\Sigma})$ will be chosen so that the likelihood function is maximized subject to the constraint that $\boldsymbol{\gamma}$ can be written in the form of (17.1.35) or, equivalently, subject to the reduced-rank restriction on $\boldsymbol{\gamma}$ so that $\text{rank}(\boldsymbol{\gamma}) = k$.

To test the null hypothesis, one can use the following likelihood ratio

$$\Lambda = \frac{\sup_{H_0} L(\boldsymbol{\theta}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{p-1}, \boldsymbol{\Sigma})}{\sup L(\boldsymbol{\theta}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{p-1}, \boldsymbol{\Sigma})}. \tag{17.1.37}$$

Under the null hypothesis, some linear combinations of $\mathbf{Z}_t$ components are nonstationary, so the asymptotic distribution of $-2\ln\Lambda$ is no longer a chi-square distribution. This particular distribution was studied by Johansen (1988, 1991) and Reinsel and Ahn (1992). They show that

$$-2\ln\Lambda = -n\sum_{i=k+1}^{m}\ln(1 - \hat{\rho}_i^2), \tag{17.1.38}$$

where $\hat{\rho}_{k+1}, \ldots,$ and $\hat{\rho}_m (\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq \cdots \geq \hat{\rho}_m^2)$ are the $(m-k) = h$ smallest sample partial canonical correlations between the components of $\Delta \mathbf{Z}_t$ and $\mathbf{Z}_{t-1}$ given $\Delta \mathbf{Z}_{t-1}, \ldots, \Delta \mathbf{Z}_{t-p+1}$. Furthermore, when (17.1.34) does not contain a constant term,

$$-2 \ln \Lambda \xrightarrow{D} \operatorname{tr}\left\{\left[\int_0^1 \mathbf{W}(x)[d\mathbf{W}(x)]'\right]'\left[\int_0^1 \mathbf{W}(x)[\mathbf{W}(x)]'\,dx\right]^{-1}\left[\int_0^1 \mathbf{W}(x)[d\mathbf{W}(x)]'\right]\right\}, \tag{17.1.39}$$

where $\mathbf{W}(x)$ is a h-dimensional standard Brownian motion process, which depends only on h and not on the order p of the AR model. The asymptotic distribution of $-2 \ln \Lambda$ when the model contains a constant term can be derived similarly, and we refer readers to Reinsel and Ahn (1992) for details. We reject the null hypothesis when Λ is small or, equivalently, when $-2 \ln \Lambda$ is large. The approximate critical values of the asymptotic distribution of $-2 \ln \Lambda$ for the likelihood ratio test of H_0: rank$(\gamma) \leq k$ against a general alternative, for both cases without and with a constant term, based on Monte Carlo simulations have been obtained by Johansen (1988) and Reinsel and Ahn (1992). Some commonly used values selected from Reinsel and Ahn are shown in Table 17.2. The estimate of the cointegrating matrix can be obtained from the estimation of the error-correction model in (17.1.34).

Note that from the discussion in Section 17.1.2, it is clear that testing for k cointegrating relations is equivalent to testing for $h = (m-k)$ unit roots in the process. When $h = 1$, the asymptotic distribution in (17.1.39) reduces to

$$\frac{\left\{\frac{1}{2}[[W(1)]^2 - 1]\right\}^2}{\int_0^1 [W(x)]^2 dx}, \tag{17.1.40}$$

TABLE 17.2 Critical values for the likelihood ratio test statistic under the null hypothesis of rank$(\gamma) = k = (m-h)$, against the alternative of rank $(\gamma) > k$, where h is the number of unit roots.

		Probability of a Smaller Value		
Case 1: Without a Constant	h	.90	.95	.99
	1	2.94	4.10	6.97
	2	10.45	12.30	16.41
	3	21.70	24.24	29.83
	4	37.00	40.18	46.30
	5	56.12	63.15	67.62
Case 2: With a Constant	h	.90	.95	.99
	1	6.59	8.16	11.65
	2	15.80	17.97	22.79
	3	28.87	31.73	37.38
	4	45.82	49.35	56.25
	5	66.98	71.14	79.23

TABLE 17.3 Critical values for the likelihood ratio test statistic under the null hypothesis of rank(γ) = $k = (m - h)$ against the alternative of rank(γ) = $k + 1$, where h is the number of unit roots.

	Probability of a Smaller Value		
Case 1: Without a Constant	h	.9	.95	.99
	1	2.86	3.84	6.51
	2	9.52	11.44	15.69
	3	15.59	17.89	22.99
	4	21.58	23.80	28.82
	5	27.62	30.04	35.17
Case 2: With a Constant	h	.9	.95	.99
	1	6.69	8.08	11.58
	2	12.78	14.60	18.78
	3	18.96	21.28	26.15
	4	24.92	27.34	32.62
	5	30.82	33.26	38.86

which can be seen to be the square of the asymptotic distribution of T statistic given in (9.3.7) that we have used to test the unit root for a univariate time series model.

We may also consider the likelihood ratio test of the null hypothesis of k cointegrating relations against the alternative of $(k + 1)$ cointegrating relations. The test statistic has been derived by Johansen (1991) and is given by

$$-2 \ln \Lambda = -n \ln(1 - \hat{\rho}_{k+1}^2), \tag{17.1.41}$$

where $\hat{\rho}_{k+1}$ is the $(k + 1)^{th}$ largest sample partial canonical correlation between the components of $\Delta \mathbf{Z}_t$ and $\mathbf{Z}_{t-1}$ given $\Delta \mathbf{Z}_{t-1}, \ldots, \Delta \mathbf{Z}_{t-p+1}$. The critical values of the asymptotic distribution for the test statistic have been obtained through simulations by Johansen and Juselius (1990), and some commonly used values are given in Table 17.3.

EXAMPLE 17.5 For the Lydia Pinkham data, we have shown in Example 17.4 that the advertising and sales series are each $I(1)$. To illustrate the use of the likelihood ratio test, following (17.1.34) and the fitted model (16.7.1) discussed in Section 16.7, we now consider the error-correction model

$$\Delta \mathbf{Z}_t = \boldsymbol{\theta}_0 + \boldsymbol{\gamma} \mathbf{Z}_{t-1} + \boldsymbol{\alpha}_1 \Delta \mathbf{Z}_{t-1} + \boldsymbol{\alpha}_2 \Delta \mathbf{Z}_{t-2} + \mathbf{a}_t.$$

The squared partial canonical correlations between $\Delta \mathbf{Z}_t$ and $\mathbf{Z}_{t-1}$ given $\Delta \mathbf{Z}_{t-1}$ and $\Delta \mathbf{Z}_{t-2}$ are obtained as $\hat{\rho}_1^2 = .2466$ and $\hat{\rho}_2^2 = .1166$. Because $m = 2$, the rank (γ) to be tested will be either 0 or 1. We begin with testing the hypothesis that rank (γ) = $k = 0$ or equivalently, $h = 2$. Now,

$$-2 \ln \Lambda = -n \sum_{i=k+1}^{m} \ln(1 - \hat{\rho}_i^2) = -54 [\ln(1 - .2466) + \ln(1 - .1166)] = 21.99,$$

which is larger than the 5% critical value 17.97 from Table 17.2. Thus, we reject the null hypothesis and conclude that the series of advertising and sales are possibly cointegrated. Next, we test the null hypothesis that rank $(\gamma) = k = 1$ or equivalently, $h = 1$. In this case,

$$-2 \ln \Lambda = -n \sum_{i=k+1}^{m} \ln(1 - \hat{\rho}_i^2) = -54[\ln(1 - .1166)] = 6.69,$$

which is less than the 5% critical value 8.16 from Table 17.2. Thus, we don't reject the null hypothesis and conclude that the series of advertising and sales are cointegrated. In testing rank $(\gamma) = k = 1$, because $m = 2$, we can also use $h = 1$ in Table 17.3. The 5% critical value is 8.08. The minor difference between Tables 17.2 and Table 17.3 for $h = 1$ comes from different simulations.

If 1% level of significance is used in testing the hypothesis that rank$(\gamma) = k = 0$ or equivalently, $h = 2$, the value of the test statistic 21.99 is less than the critical value 22.79 from Table 17.2. The hypothesis is not rejected; the testing stops, and we conclude that advertising and sales series are not cointegrated. Thus, at 1% level of significance, the regression method and the likelihood ratio test reach the same conclusion.

17.2 Partial Process and Partial Process Correlation Matrices

In Section 16.5.3, we discussed the partial lag correlation matrix at lag s as the correlation (expressed as an $m \times m$ matrix) between the vectors $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$, after removing the linear dependence of each on the vectors at intervening lags $\mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+s-1}$. In this section, we introduce another form of partial correlation proposed by Heyse and Wei (1984), in which $\mathbf{Z}_t$ is partitioned into three subvectors, $\mathbf{Z}_t = (\mathbf{Z}_{1,t}', \mathbf{Z}_{2,t}', \mathbf{Z}_{3,t}')'$ of dimensions m_1, m_2, and m_3, respectively. That is, we consider the correlation (expressed as a matrix of dimension $m - m_3$) between the subvectors $(\mathbf{Z}_{1,t}', \mathbf{Z}_{2,t}')'$ and $(\mathbf{Z}_{1,t+s}', \mathbf{Z}_{2,t+s}')'$ after removing the linear dependence of each on the third component series $\mathbf{Z}_{3,t}$. Heyse and Wei (1984) call this the partial process correlation matrix at lag s. These correlations are useful in determining whether an apparent linear relationship between $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ is due to an intrinsic relationship of each with $\mathbf{Z}_{3,t}$.

17.2.1 Covariance Matrix Generating Function

Consider a zero mean covariance stationary vector time series $\mathbf{Z}_t$ of dimension m having moving average representation

$$\begin{aligned} \mathbf{Z}_t &= \sum_{s=0}^{\infty} \Psi_s \mathbf{a}_{t-s} \\ &= \Psi(B)\mathbf{a}_t, \end{aligned} \tag{17.2.1}$$

where $\Psi(B) = (I - \Psi_1 B - \Psi_2 B^2 - \cdots)$ is an $m \times m$ matrix polynomial in the backshift operator B, such that the coefficients, Ψ_s, are square matrices satisfying the square summability condition.

The covariance matrix generating function for the process $\mathbf{Z}_t$ is defined as

$$G_{\mathbf{Z}}(B) = \sum_{s=-\infty}^{\infty} \Gamma_{\mathbf{Z}}(s)B^s. \tag{17.2.2}$$

Extending a result in Box, Jenkins, and Reinsel (1994, p. 85), it can be shown that

$$G_{\mathbf{Z}}(B) = \Psi(F)\Sigma\Psi'(B), \tag{17.2.3}$$

where $F = B^{-1}$. Writing $G_{\mathbf{Z}}(B) = [\gamma^{ij}(B)]$, we note that

$$\gamma^{ii}(B) = \sum_{s=-\infty}^{\infty} \gamma_{ii}(s)B^s, \qquad i = 1, 2, \ldots, m, \tag{17.2.4}$$

is the autocovariance generating function for the ith component process of $\mathbf{Z}_t$. The variance of the process $\mathbf{Z}_{i,t}$, $\gamma_{ii}(0)$, is the coefficient of $B^0 = 1$, and the autocovariance at lag s, $\gamma_{ii}(s)$, is the coefficient of B^s in (17.2.4). Similarly, $\gamma^{ij}(B) = \sum_{s=-\infty}^{\infty} \gamma_{ij}(s)B^s$, $i = 1, \ldots, m, j = 1, \ldots, m$, and $i \neq j$, is the cross-covariance generating function between the ith and jth components of $\mathbf{Z}_t$.

17.2.2 Partial Covariance Matrix Generating Function

Suppose that $\mathbf{Z}_t$ is suitably partitioned into 3 subvectors, $\mathbf{Z}_t = (\mathbf{Z}_{1,t}', \mathbf{Z}_{2,t}', \mathbf{Z}_{3,t}')'$, of dimensions m_1, m_2, and m_3 with $m_1 + m_2 + m_3 = m$. Our interest is in the correlation matrix at lag s for the process $[\mathbf{Z}_{1,t}', \mathbf{Z}_{2,t}']'$, after removing the linear dependence on the process $\mathbf{Z}_{3t}$. Partitioning conformably the $\Psi(B)$ and $\mathbf{a}_t$ the moving average representation for $\mathbf{Z}_t$, (17.2.1), can be written as

$$\begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \mathbf{Z}_{3,t} \end{bmatrix} = \begin{bmatrix} \psi_{11}(B) & \psi_{12}(B) & \psi_{13}(B) \\ \psi_{21}(B) & \psi_{22}(B) & \psi_{23}(B) \\ \psi_{31}(B) & \psi_{32}(B) & \psi_{33}(B) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1,t} \\ \mathbf{a}_{2,t} \\ \mathbf{a}_{3,t} \end{bmatrix},$$

where the components $\psi_{ij}(B)$ are matrix polynomials in B with $\psi_{ii}(0) = \mathbf{I}_{m_i}$ and $\psi_{ij}(0) = \mathbf{0}_{m_i \times m_j}$ for $i \neq j$. The lag s covariance matrix can be similarly partitioned as

$$\Gamma_{\mathbf{Z}}(s) = \begin{bmatrix} \Gamma_{11}(s) & \Gamma_{12}(s) & \Gamma_{13}(s) \\ \Gamma_{21}(s) & \Gamma_{22}(s) & \Gamma_{23}(s) \\ \Gamma_{31}(s) & \Gamma_{32}(s) & \Gamma_{33}(s) \end{bmatrix},$$

where $\Gamma_{ij}(s) = E[\mathbf{Z}_{i,t}\mathbf{Z}_{j,t+s}']$. The covariance matrix generating function (17.2.2) for this particular partition of $\mathbf{Z}_t$ can be written as

$$G_{\mathbf{Z}}(B) = \begin{bmatrix} G_{11}(B) & G_{12}(B) & G_{13}(B) \\ G_{21}(B) & G_{22}(B) & G_{23}(B) \\ G_{31}(B) & G_{32}(B) & G_{33}(B) \end{bmatrix}, \tag{17.2.5}$$

where

$$G_{ij}(B) = \sum_{s=-\infty}^{\infty} \Gamma_{ij}(s)B^s.$$

Note that because $\Gamma_{ij}(s) = \Gamma'_{ji}(-s)$, the components of $G_{\mathbf{Z}}(B)$ are related as

$$G_{ij}(B) = G'_{ji}(B^{-1}) = G'_{ji}(F).$$

To find the correlation between $(\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t})'$ and $(\mathbf{Z}'_{1,t+s}, \mathbf{Z}'_{2,t+s})'$ after removing the linear dependence of each on the series $\mathbf{Z}_{3,t}$, we let $\hat{\mathbf{Z}}_{1,t}$ and $\hat{\mathbf{Z}}_{2,t}$ be the linear projections of $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ on $\mathbf{Z}_{3,t}$, respectively, and consider the correlation matrix function for the residual process

$$\begin{bmatrix} \mathbf{Z}_{1.3,t} \\ \mathbf{Z}_{2.3,t} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{1,t} - \hat{\mathbf{Z}}_{1,t} \\ \mathbf{Z}_{2,t} - \hat{\mathbf{Z}}_{2,t} \end{bmatrix}. \tag{17.2.6}$$

That is, let

$$\begin{aligned} \hat{\mathbf{Z}}_{1,t} &= \boldsymbol{\alpha}(B)\mathbf{Z}_{3,t} \\ &= \sum_{i=-\infty}^{\infty} \boldsymbol{\alpha}_i \mathbf{Z}_{3,t-i} \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbf{Z}}_{2,t} &= \boldsymbol{\beta}(B)\mathbf{Z}_{3,t} \\ &= \sum_{i=-\infty}^{\infty} \boldsymbol{\beta}_i \mathbf{Z}_{3,t-i}, \end{aligned}$$

where the $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are chosen to minimize, respectively,

$$E[|\mathbf{Z}_{1,t} - \boldsymbol{\alpha}(B)\mathbf{Z}_{3,t}|^2] \tag{17.2.7}$$

and

$$E[|\mathbf{Z}_{2,t} - \boldsymbol{\beta}(B)\mathbf{Z}_{3,t}|^2]. \tag{17.2.8}$$

Taking the partial derivative of (17.2.7) with respect to $\boldsymbol{\alpha}_j$ (Graham [1981, p. 54]) and setting the resulting derivative equal to $\mathbf{0}$ leads to the normal equations

$$\sum_{i=-\infty}^{\infty} \Gamma_{33}(j-i)\boldsymbol{\alpha}'_i = \Gamma_{31}(j) \qquad \text{for } j = 0, \pm 1, \pm 2, \ldots. \tag{17.2.9}$$

Multiplying the jth equation in (17.2.9) by B^j and summing over j gives

$$\sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty}\Gamma_{33}(j-i)\boldsymbol{\alpha}_i' B^j = \sum_{j=-\infty}^{\infty}\Gamma_{31}(j)B^j.$$

Hence,

$$\boldsymbol{G}_{33}(B)\boldsymbol{\alpha}'(B) = \boldsymbol{G}_{31}(B),$$

from which we have the solution

$$\boldsymbol{\alpha}(B) = \boldsymbol{G}_{13}(F)\boldsymbol{G}_{33}^{-1}(F).$$

The covariance matrix at lag s for the residual process $(\mathbf{Z}_{1t} - \hat{\mathbf{Z}}_{1t})$ is equal to

$$\begin{aligned}\Gamma_{11.3}(s) &= E[(\mathbf{Z}_{1,t} - \boldsymbol{\alpha}(B)\mathbf{Z}_{3,t})(\mathbf{Z}_{1,t+s} - \boldsymbol{\alpha}(B)\mathbf{Z}_{3,t+s})'] \\ &= \Gamma_{11}(s) - \Gamma_{13}(s)\boldsymbol{\alpha}'(B) - \boldsymbol{\alpha}(B)\Gamma_{31}(s) + \boldsymbol{\alpha}(B)\Gamma_{33}(s)\boldsymbol{\alpha}'(B). \quad (17.2.10)\end{aligned}$$

The covariance matrix generating function for the residual process can be determined by multiplying (17.2.10) by B^s and summing over s, i.e.,

$$\begin{aligned}G_{11.3}(B) &= \sum_{s=-\infty}^{\infty}\Gamma_{11.3}(s)B^s \\ &= \sum_{s=-\infty}^{\infty}[\Gamma_{11}(s)B^s - \Gamma_{13}(s)B^s\boldsymbol{\alpha}'(B) - \boldsymbol{\alpha}(B)\Gamma_{31}(s)B^s + \boldsymbol{\alpha}(B)\Gamma_{33}(s)B^s\boldsymbol{\alpha}'(B)] \\ &= \boldsymbol{G}_{11}(B) - \boldsymbol{G}_{13}(B)\boldsymbol{G}_{33}^{-1}(B)\boldsymbol{G}_{31}(B) - \boldsymbol{G}_{13}(F)\boldsymbol{G}_{33}^{-1}(F)\boldsymbol{G}_{31}(B) \\ &\quad + \boldsymbol{G}_{13}(F)\boldsymbol{G}_{33}^{-1}(F)\boldsymbol{G}_{33}(B)\boldsymbol{G}_{33}^{-1}(B)\boldsymbol{G}_{31}(B) \\ &= \boldsymbol{G}_{11}(B) - \boldsymbol{G}_{13}(B)\boldsymbol{G}_{33}^{-1}(B)\boldsymbol{G}_{31}(B). \quad (17.2.11)\end{aligned}$$

Similarly, $E[|\mathbf{Z}_{2,t} - \boldsymbol{\beta}(B)\mathbf{Z}_{3,t}|^2]$ is minimized by choosing

$$\boldsymbol{\beta}(B) = \boldsymbol{G}_{23}(F)\boldsymbol{G}_{33}^{-1}(F).$$

The covariance matrix at lag s for the residual process $(\mathbf{Z}_{2,t} - \hat{\mathbf{Z}}_{2,t})$ is equal to

$$\begin{aligned}\Gamma_{22.3}(s) &= E[(\mathbf{Z}_{2,t} - \boldsymbol{\beta}(B)\mathbf{Z}_{3,t})(\mathbf{Z}_{2,t+s} - \boldsymbol{\beta}(B)\mathbf{Z}_{3,t+s})'] \\ &= \Gamma_{22}(s) - \Gamma_{23}(s)\boldsymbol{\beta}'(B) - \boldsymbol{\beta}(B)\Gamma_{32}(s) + \boldsymbol{\beta}(B)\Gamma_{33}(s)\boldsymbol{\beta}'(B),\end{aligned}$$

from which the covariance matrix generating function can be determined as

$$G_{22.3}(B) = G_{22}(B) - G_{23}(B)G_{33}^{-1}(B)G_{32}(B). \tag{17.2.12}$$

Also,

$$\begin{aligned}\Gamma_{12.3}(s) &= E[(\mathbf{Z}_{1,t} - \boldsymbol{\alpha}(B)\mathbf{Z}_{3,t})(\mathbf{Z}_{2,t+s} - \boldsymbol{\beta}(B)\mathbf{Z}_{3,t+s})'] \\ &= \Gamma_{12}(s) - \Gamma_{13}(s)\boldsymbol{\beta}'(B) - \boldsymbol{\alpha}(B)\Gamma_{32}(s) + \boldsymbol{\alpha}(B)\Gamma_{33}(s)\boldsymbol{\beta}'(B)\end{aligned}$$

and

$$\Gamma_{21.3}(s) = \Gamma_{12.3}'(-s).$$

Thus,

$$\begin{aligned}G_{12.3}(B) &= G_{12}(B) - G_{13}(B)G_{33}^{-1}(B)G_{32}(B) - G_{13}(F)G_{33}^{-1}(F)G_{32}(B) \\ &\quad + G_{13}(F)G_{33}^{-1}(F)G_{33}(B)G_{33}^{-1}(B)G_{32}(B) \\ &= G_{12}(B) - G_{13}(B)G_{33}^{-1}(B)G_{32}(B)\end{aligned} \tag{17.2.13}$$

and

$$G_{21.3}(B) = G_{21}(B) - G_{23}(B)G_{33}^{-1}(B)G_{31}(B). \tag{17.2.14}$$

Note that from (17.2.11) to (17.2.14) we have the covariance matrix generating function for the residual process $(\mathbf{Z}_{1.3,t}', \mathbf{Z}_{2.3,t}')'$

$$\begin{aligned}G_{12.3}^{\dagger}(B) &= \begin{bmatrix} G_{11.3}(B) & G_{12.3}(B) \\ G_{21.3}(B) & G_{22.3}(B) \end{bmatrix} \\ &= \begin{bmatrix} G_{11}(B) & G_{12}(B) \\ G_{21}(B) & G_{22}(B) \end{bmatrix} - \begin{bmatrix} G_{13}(B) \\ G_{23}(B) \end{bmatrix} [G_{33}(B)]^{-1} [G_{31}(B) \quad G_{32}(B)].\end{aligned} \tag{17.2.15}$$

If we let

$$\Gamma_{12.3}^{\dagger}(s) = \begin{bmatrix} \Gamma_{11.3}(s) & \Gamma_{12.3}(s) \\ \Gamma_{21.3}(s) & \Gamma_{22.3}(s) \end{bmatrix}$$

be the covariance matrix for the residual process $(\mathbf{Z}_{1.3,t}', \mathbf{Z}_{2.3,t}')'$ at lag s, then $\Gamma_{12.3}^{\dagger}(s)$ is obtained directly as the coefficient of B^s in $G_{12.3}^{\dagger}(B)$. Specifically, $\Gamma_{12.3}^{\dagger}(0)$ is the coefficient of B^0 in (17.2.15) and represents the variance-covariance matrix for the residual process.

Let $\boldsymbol{D}$ be the diagonal matrix, with its ith diagonal element being the square root of the ith diagonal element of $\Gamma^{\dagger}_{12.3}(0)$; then we can define the correlation matrix at lag s for the residual process as

$$\begin{aligned}\boldsymbol{\rho}^{\dagger}_{12.3}(s) &= \boldsymbol{D}^{-1}\boldsymbol{\Gamma}^{\dagger}_{12.3}(s)\boldsymbol{D}^{-1} \\ &= \begin{bmatrix} \boldsymbol{\rho}_{11.3}(s) & \boldsymbol{\rho}_{12.3}(s) \\ \boldsymbol{\rho}_{21.3}(s) & \boldsymbol{\rho}_{22.3}(s) \end{bmatrix}. \end{aligned} \tag{17.2.16}$$

Heyse and Wei (1984) call the residual process $(\mathbf{Z}'_{1.3,t}, \mathbf{Z}'_{2.3,t})'$ in (17.2.6) the partial process of $(\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t})'$, allowing for $\mathbf{Z}_{3,t}$, because it represents the remaining stochastic process in $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ after the linear relationship of each with the third stochastic component series $\mathbf{Z}_{3t}$ has been removed. They call $\boldsymbol{G}^{\dagger}_{12.3}(s)$ in (17.2.15) the partial process covariance matrix generating function and $\boldsymbol{\rho}^{\dagger}_{12.3}(s)$ in (17.2.16) the partial process correlation matrix function for the partial process of $(\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t})'$, allowing for $\mathbf{Z}_{3,t}$.

17.2.3 Partial Process Sample Correlation Matrix Functions

The partial process correlation matrix function, $\boldsymbol{\rho}^{\dagger}_{12.3}(s)$, for the $(m_1 + m_2)$-dimensional $(\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t})'$ component of $\mathbf{Z}_t$, after removing the linear dependence on $\mathbf{Z}_{3,t}$, is the correlation matrix function for the partial process. If a sample of the partial process $\mathbf{Z}^{\dagger}_t = [\mathbf{Z}'_{1.3,t}, \mathbf{Z}'_{2.3,t}]'$ were observable, then clearly $\boldsymbol{\rho}^{\dagger}_{12.3}(s)$ could be estimated directly using the procedures discussed in Section 16.5.1. Thus, one intuitively reasonable estimation procedure would be first to compute the residuals from the regression equations

$$\hat{\mathbf{Z}}_{1.3,t} = \mathbf{Z}_{1,t} - \sum_{k=-P_1}^{Q_1} \hat{\boldsymbol{\alpha}}_k \mathbf{Z}_{3,t-k}, \qquad t = Q_1 + 1, Q_1 + 2, \ldots, n - P_1 \tag{17.2.17}$$

and

$$\hat{\mathbf{Z}}_{2.3,t} = \mathbf{Z}_{2,t} - \sum_{l=-P_2}^{Q_2} \hat{\boldsymbol{\beta}}_l \mathbf{Z}_{3,t-l}, \qquad t = Q_2 + 1, Q_2 + 2, \ldots, n - P_2 \tag{17.2.18}$$

using multivariate least squares estimates for $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_l$, where n is the number of observed values of $\mathbf{Z}_t$.

The partial process correlation matrix at lag s, $\boldsymbol{\rho}^{\dagger}_{12.3}(s)$, can be estimated by the sample correlation matrix at lag s for the series

$$\hat{\mathbf{Z}}^{\dagger}_t = (\hat{\mathbf{Z}}'_{1.3,t}, \hat{\mathbf{Z}}'_{2.3,t})',$$

where $t = Q + 1, Q + 2, \ldots, n - P$, with $Q = \max(Q_1, Q_2)$ and $P = \max(P_1, P_2)$. That is,

$$\hat{\boldsymbol{\rho}}^{\dagger}_{12.3}(s) = [\hat{\boldsymbol{D}}]^{-1}\hat{\boldsymbol{\Gamma}}^{\dagger}_{12.3}(s)[\hat{\boldsymbol{D}}]^{-1}, \tag{17.2.19}$$

where

$$\hat{\Gamma}^{\dagger}_{12.3}(s) = [n_0]^{-1} \sum_{t=Q+1}^{n-P-s} (\hat{\mathbf{Z}}^{\dagger}_t - \overline{\mathbf{Z}}^{\dagger})(\hat{\mathbf{Z}}^{\dagger}_{t+s} - \overline{\mathbf{Z}}^{\dagger})',$$

$n_0 = (n - P - Q)$ is the effective sample size, $\overline{\mathbf{Z}}^{\dagger}$ is the vector of means for $\hat{\mathbf{Z}}^{\dagger}_t$, and $\hat{\boldsymbol{D}}$ is the diagonal matrix in which the kth diagonal element is the square root of the corresponding element of $\hat{\Gamma}^{\dagger}_{12.3}(0)$.

Let $\widetilde{\Gamma}^{\dagger}_{12.3}(s)$ be the partial process sample covariance matrix at lag s for the unobservable partial process $\mathbf{Z}^{\dagger}_t$; for $i = 1, \ldots, (m_1 + m_2)$, and $j = 1, \ldots, (m_1 + m_2)$, let $\widetilde{\gamma}^{\dagger}_{ij}(s)$ be the (i, j)th element of $\widetilde{\Gamma}^{\dagger}_{12.3}(s)$ Also denote the (i, j)th element of $\hat{\Gamma}^{\dagger}_{12.3}(s)$ by $\hat{\gamma}^{\dagger}_{ij}(s)$, which we recall was computed from the series of estimated residuals $\hat{\mathbf{Z}}^{\dagger}_t(s)$ in (17.2.17) and (17.2.18). Hannan (1970, p. 453) shows that the $\hat{\gamma}^{\dagger}_{ij}(s)$ differ from the $\widetilde{\gamma}^{\dagger}_{ij}(s)$ by quantities that are of the order $(n_0 - s)^{-1}$ and that $\sqrt{n_0}[\hat{\gamma}^{\dagger}_{ij}(s) - \widetilde{\gamma}^{\dagger}_{ij}(s)]$ converges in probability to zero, which implies that $\sqrt{n_0}\hat{\gamma}^{\dagger}_{ij}(s)$ has the same asymptotic distribution as $\sqrt{n_0}\,\widetilde{\gamma}^{\dagger}_{ij}(s)$. In other words, for large n and moderate choices for P_1, Q_1, P_2, and Q_2, elements of the partial process sample correlation matrix obtained by computing the sample correlation matrix for the estimated residuals (17.2.17) and (17.2.18) can be interpreted in the usual way.

One problem with this procedure for computing $\hat{\boldsymbol{\rho}}^{\dagger}_{12.3}(s)$ is with the choices of P_1, Q_1, P_2, and Q_2 for computing the residuals in (17.2.17) and (17.2.18). This problem is common in building statistical models, and we don't seek to solve it here. It is apparent that the choice of the P's and Q's involves a trade-off. If they are too small, then some effects of the third series will remain in $\hat{\boldsymbol{\rho}}^{\dagger}_{12.3}(s)$. If they are too large, then the effective sample size n_0 will be too small. In dealing with these choices, we must remind ourselves that our interest is *not* in estimating the linear regression coefficients $\boldsymbol{\alpha}_k$, $k = -P_1, -(P_1 - 1), \ldots, Q_1$, and $\boldsymbol{\beta}_l$, $l = -P_2, -(P_2 - 1), \ldots, Q_2$. Instead, our interest is in removing the linear dependence of $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ on $\mathbf{Z}_{3,t}$ for the purpose of estimating the remaining correlation structure in $(\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t})$. In practice, diagnostic aids that examine the statistical significance of the parameters are helpful. Nonetheless, it is generally advisable to choose several values for the limits to be sure that the resulting estimates of $\boldsymbol{\rho}^{\dagger}_{12.3}(s)$ do not change dramatically. Note also that because the effective sample size n_0 uses $\max(P_1, P_2)$ and $\max(Q_1, Q_2)$, one can choose both sets of limits at the maximum in (17.2.17) and (17.2.18) without reducing n_0.

17.2.4 An Empirical Example: The U.S. Hog Data

The classical U.S. hog data consist of five series of 82 annual observations for the years 1867 to 1948. For convenience, the data are listed in the appendix as Series W13. The five series are

H_t: Hog numbers recorded on January 1 in the Census of Agriculture.
P_t: Hog prices in dollars per head on January 1.
Q_t: Corn prices in dollars per bushel on December 1.
C_t: Corn supply in bushels produced during the year.
W_t: Farm wage rate.

TABLE 17.4 Estimated cross-correlation for hog numbers H_t and hog prices P_{t+s} at lag s.

Lag s	$\hat{\rho}_{12}(s)$	Indicator	Lag s	$\hat{\rho}_{12}(s)$	Indicator
0	.62	+			
−1	.65	+	1	.60	+
−2	.62	+	2	.66	+
−3	.56	+	3	.70	+
−4	.50	+	4	.69	+
−5	.43	+	5	.58	+
−6	.35	+	6	.48	+
−7	.29	+	7	.42	+
−8	.24	+	8	.38	+
−9	.17	·	9	.34	+
−10	.10	·	10	.33	+
−11	.08	·	11	.35	+
−12	.10	·	12	.38	+

Note: 2 × estimated standard error ≈ .22.
+ indicates cross-correlations greater than .22.
· indicates cross-correlations between −.22 and .22.

A description of the data appears in Quenouille (1957), Box and Tiao (1977), and Tiao and Tsay (1983), who fit the data to a five-dimensional vector autoregressive moving average model. The data were logged and linearly coded by Quenouille, and this form of the data was used for the present analysis.

Our attention focuses on the relationship between hog numbers H_t and hog prices P_t. In terms of our development thus far, we are considering the case where $\mathbf{Z}_{1,t} = H_t$, $\mathbf{Z}_{2,t} = P_t$, and $\mathbf{Z}_{3,t} = (Q_t, C_t, W_t)'$.

The estimated cross-correlations, $\hat{\rho}_{12}(s)$, for hog numbers H_t and hog prices P_{t+s} are shown in Table 17.4 for lags $s = 0, \pm 1, \ldots, \pm 12$. These values were taken from the (1, 2) element of the sample correlation matrix $\hat{\rho}(s)$ for the five series. These cross-correlations are persistently high and provide very little information regarding the relationship between hog numbers and prices, except possibly that the series are nearly nonstationary or are commonly related to a third series.

Following our definition of the partial process and the development in Section 17.2.3, we first want to compute the residuals

$$\hat{H}_t^{\dagger} = H_t - \sum_{k=-P_1}^{Q_1} (\hat{\alpha}_{1k}, \hat{\alpha}_{2k}, \hat{\alpha}_{3k})(Q_{t+k}, C_{t+k}, W_{t+k})' \tag{17.2.20}$$

and

$$\hat{P}_t^{\dagger} = P_t - \sum_{l=-P_2}^{Q_2} (\hat{\beta}_{1l}, \hat{\beta}_{2l}, \hat{\beta}_{3l})(Q_{t+l}, C_{t+l}, W_{t+l})' \tag{17.2.21}$$

TABLE 17.5 Estimated partial process cross-correlations for hog numbers H_t and hog prices P_{t+s} at lag s.

Lag s	$\hat{\rho}^{\dagger}_{12.3}$	Indicator	Lag s	$\hat{\rho}^{\dagger}_{12.3}(s)$	Indicator
0	−.31	−			
−1	.23	+	1	−.51	−
−2	.03	·	2	−.16	·
−3	−.18	·	3	.07	·
−4	−.01	·	4	−.03	·
−5	.21	·	5	−.05	·
−6	.10	·	6	.00	·
−7	−.04	·	7	.12	·
−8	−.14	·	8	.04	·
−9	−.16	·	9	−.14	·
−10	−.20	·	10	−.10	·
−11	−.30	−	11	−.02	·
−12	−.20	·	12	−.16	·

Note: 2 × estimated standard error ≈ .23.
+ indicates cross-correlations greater than .23.
− indicates cross-correlations less than −.23.
· indicates cross-correlations between −.23 and .23.

using ordinary least squares estimates for the $\hat{\alpha}_{ik}$ and $\hat{\beta}_{il}$. To observe the effect of increasing the limits in the regressions (17.2.20) and (17.2.21), we used the symmetric limits $L = P_1 = Q_1 = P_2 = Q_2 = 0, 1, \ldots, 8$. The partial process sample correlation matrix function $\hat{\rho}^{\dagger}_{12.3}(s)$ was then computed from the estimated residual series $(\hat{H}^{\dagger}_t, \hat{P}^{\dagger}_t)'$ for each of these choices and was used to estimate the cross-correlation between hog numbers and prices after removing the linear dependence on corn price, corn supply, and wage rates. These values are shown in Table 17.5 for the choice $L = 3$ along with the +, −, · symbols indicating whether correlation coefficients exceed two times their estimated standard errors $(1/75)^{1/2}$.

Interpretation of $\hat{\rho}^{\dagger}_{12.3}(s)$ in Table 17.5 indicates that hog numbers and prices were related in the period under study beyond their common dependence on corn prices, corn supply, and wage rates. Only four values of $\hat{\rho}^{\dagger}_{12.3}(s)$ are larger than twice their estimated standard errors, however, and of these we consider only $\hat{\rho}^{\dagger}_{12.3}(0)$, $\hat{\rho}^{\dagger}_{12.3}(1)$, and $\hat{\rho}^{\dagger}_{12.3}(-1)$ because the meaning of $\hat{\rho}^{\dagger}_{12.3}(-11) = -.30$ is not clear and may be due to random variation. Here, $\hat{\rho}^{\dagger}_{12.3}(0) = -.31$ indicates that hog numbers H_t and hog prices P_t at time t were negatively correlated contemporaneously, $\hat{\rho}^{\dagger}_{12.3}(1) = -.51$ shows that hog numbers H_t at time t and hog prices P_{t+1} at time $t + 1$ were negatively correlated, and $\hat{\rho}^{\dagger}_{12.3}(-1) = .23$ indicates that hog prices P_t at time t were positively correlated with hog numbers H_{t+1} at time $t + 1$.

Next, we examine the effect of increasing the limits of the regressions (17.2.20) and (17.2.21) on the three partial process sample cross-correlations $\hat{\rho}^{\dagger}_{12.3}(-1)$, $\hat{\rho}^{\dagger}_{12.3}(0)$, and $\hat{\rho}^{\dagger}_{12.3}(1)$. In Table 17.6, the three estimates are shown for each of the choices $L = 0, 1, \ldots, 8$. The value $L =$ null represents the null case, i.e., the sample cross-correlations $\hat{\rho}_{12}(-1)$, $\hat{\rho}_{12}(0)$, and $\hat{\rho}_{12}(1)$.

TABLE 17.6 Estimated partial process cross-correlations for increasing values of the limits $P = Q$ in (17.2.20) and (17.2.21).

$P = Q$	$\hat{\rho}^{\dagger}_{12.3}(-1)$	$\hat{\rho}^{\dagger}_{12.3}(0)$	$\hat{\rho}^{\dagger}_{12.3}(1)$
Null	.65(+)	.62(+)	.60(+)
0	.06(·)	−.42(−)	−.60(−)
1	.28(+)	−.42(−)	−.53(−)
2	.30(+)	−.24(−)	−.53(−)
3	.23(+)	−.31(−)	−.51(−)
4	.32(+)	−.24(−)	−.44(−)
5	.28(+)	−.07(·)	−.37(−)
6	.51(+)	−.06(·)	−.22(·)
7	.60(+)	.13(·)	−.21(·)
8	.53(+)	.01(·)	−.39(·)

Note: Null indicates sample cross-correlations.
+ and − indicate partial process cross-correlations which exceed two times their estimated standard errors.

The estimates $\hat{\rho}^{\dagger}_{12.3}(-1)$, $\hat{\rho}^{\dagger}_{12.3}(0)$, and $\hat{\rho}^{\dagger}_{12.3}(1)$ are plotted against the limit L in Figure 17.2(a), (b), and (c), respectively.

The results of this analysis indicate that the effect of increasing the limit L on the estimates $\hat{\rho}^{\dagger}_{12.3}(-1)$ and $\hat{\rho}^{\dagger}_{12.3}(1)$ is minimal, although $\hat{\rho}^{\dagger}_{12.3}(0)$ does diminish to values closer to zero as the limit increases. Hence, our interpretation of the contemporaneous cross-correlation between hog numbers H_t and hog prices P_t from Table 17.5 may be questionable. The additional evidence enhances our conclusion that hog numbers H_t at time t and hog prices P_{t+1} at time $t + 1$ were negatively correlated and that hog prices P_t at time t and hog numbers H_{t+1} at time $t + 1$ were positively correlated for the period under study beyond their common dependence on corn prices, corn supply, and wage rates. This result certainly corresponds to simple economic theories relating supply and prices. This phenomenon becomes apparent for this data, however, only through the study of the partial process and the partial process correlation function.

17.3 Equivalent Representations of a Vector ARMA Model

An ARMA model can often be represented by different forms. For example, with no loss of generality, consider the zero mean first-order univariate autoregressive (AR(1)) model

$$\phi(B)Z_t = a_t, \tag{17.3.1}$$

where $\phi(B) = (1 - \phi B)$, B is the backshift operator such that $BZ_t = Z_{t-1}$, and a_t is a white noise series with mean zero and constant variance σ^2. If the zero of the polynomial $\phi(B)$ lies outside the unit circle, then the model can also be written as the infinite moving average (MA) form

$$Z_t = \psi(B)a_t, \tag{17.3.2}$$

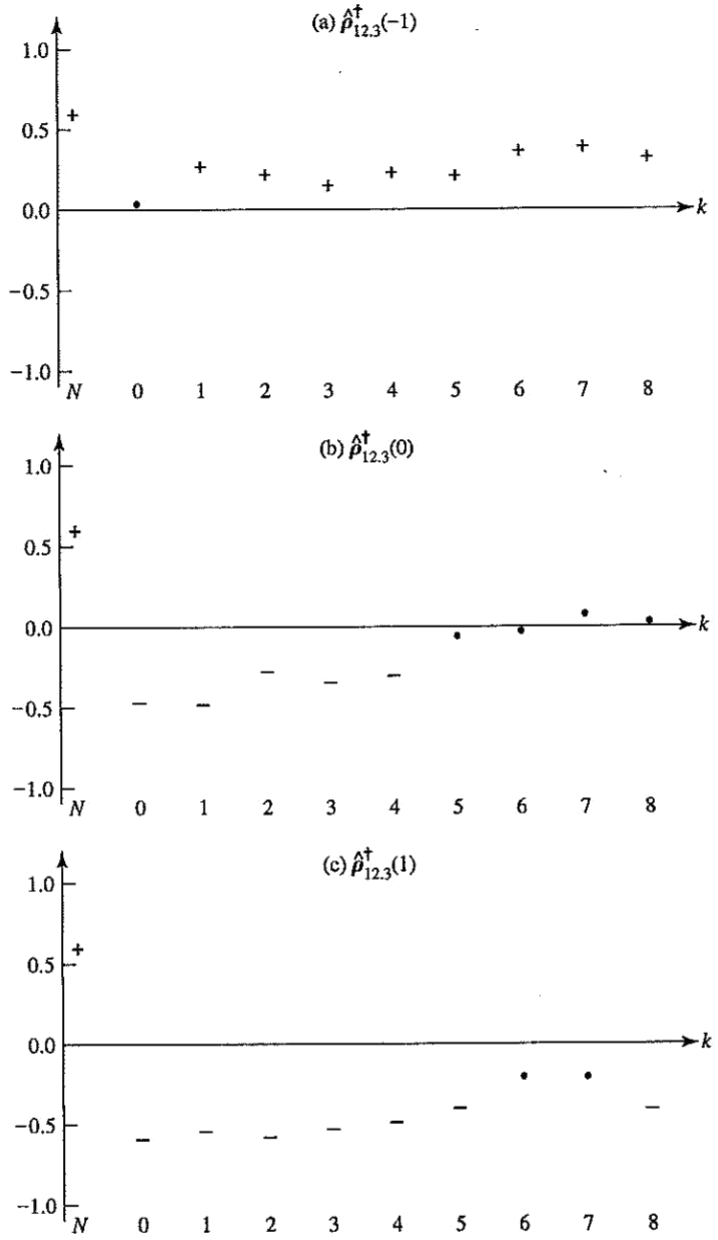


FIGURE 17.2 $\hat{\boldsymbol{\rho}}^{\dagger}_{12.3}(j)$ for increasing values of the limits in the regressions (17.2.20) and (17.2.21).

where $\psi(B) = [\phi(B)]^{-1} = (1 + \phi B + \phi^2 B^2 + \cdots)$. Similarly, the first-order univariate moving average (MA(1)) model

$$Z_t = \theta(B)a_t, \tag{17.3.3}$$

with $\theta(B) = (1 - \theta B)$ and its zero outside the unit circle, can alternatively be expressed as the infinite autoregressive (AR) form

$$\pi(B)Z_t = a_t, \tag{17.3.4}$$

with $\pi(B) = [\theta(B)]^{-1} = (1 + \theta B + \theta^2 B^2 + \cdots)$. The infinite forms in (17.3.2) and (17.3.4) indeed contain an infinite number of terms as long as the parameters ϕ and θ are not zero. More generally, there are interesting dual representations. That is, for a univariate time series, a finite-order stationary AR process can be represented by an infinite-order MA form, and a finite-order invertible MA process can be represented by an infinite-order AR form. Thus, a finite-order AR process corresponds to an infinite-order MA process, and a finite-order MA process corresponds to an infinite-order AR process.

These alternative representations are useful. For example, the AR form is useful in expressing the forecast as a weighted average of previous observations, and the MA form makes the calculation of the forecast variance especially simple.

Now, suppose that $\mathbf{Z}_t = [Z_{1,t}, \ldots, Z_{m,t}]'$ is an $m \times 1$ vector. The generalization of (17.3.1) is the vector AR(1) process

$$\Phi(B)\mathbf{Z}_t = \mathbf{a}_t, \tag{17.3.5}$$

where $\Phi(B) = (\mathbf{I} - \Phi B)$, $\mathbf{I}$ is an $m \times m$ identity matrix, $\Phi = [\phi_{ij}]$ is a nonzero $m \times m$ coefficient matrix with its (i, j) element $\phi_{i,j}$, and $\mathbf{a}_t = [a_{1,t}, \ldots, a_{m,t}]'$ is an $m \times 1$ vector white noise process with mean vector $\mathbf{0}$ and positive definite variance-covariance matrix Σ. Because a vector time series is a generalization of a univariate time series, it is intuitively reasonable to expect that the above dual representations also hold for a vector process. In other words, we expect that a finite-order vector AR process will correspond to an infinite-order vector MA process and that a finite-order vector MA process will correspond to an infinite-order vector AR process. This conjecture is not necessarily true, however. The possibility of multiple representations for a vector ARMA model has been known in the literature (see, for example, Hannan and Deistler [1988]). In Section 17.3.1, we present some simple results showing that under certain conditions a vector process can be simultaneously represented by a finite-order vector AR model, a finite-order vector MA model, or a finite-order vector autoregressive moving average (ARMA) model. We also establish the maximum order of these representations. In Section 17.3.2, we discuss some of the implications of these special types of vector time series models.

17.3.1 Finite-Order Representations of a Vector Time Series Process

For the vector AR model given in (17.3.5), it is known from Section 16.3.1 that when roots of the determinantal polynomial $|\Phi(B)| = 0$ lie outside the unit circle or, equivalently, when all the eigenvalues of Φ lie inside the unit circle, the model can be written as the MA form

$$\begin{aligned}\mathbf{Z}_t &= [\Phi(B)]^{-1}\mathbf{a}_t \\ &= (\mathbf{I} - \Phi B)^{-1}\mathbf{a}_t \\ &= \sum_{j=0}^{\infty} \Phi^j \mathbf{a}_{t-j},\end{aligned} \tag{17.3.6}$$

where $\Phi_0 = \mathbf{I}$. Is the MA form in (17.3.6) always an infinite order? To answer the question, let us consider the three-dimensional vector AR(1) model taken from Newbold (1982),

$$(\mathbf{I} - \Phi B)\mathbf{Z}_t = \mathbf{a}_t, \tag{17.3.7}$$

with

$$\Phi^j = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Because $\Phi^2 \neq \mathbf{0}$ and $\Phi^j = \mathbf{0}$ for $j > 2$, the form in (17.3.6) represents a vector MA(2) model. In fact, any vector AR(1) model with a nilpotent coefficient matrix can also be written as a finite-order vector MA model. To see why that happens, let $|\mathbf{A}|$ denote the determinant of a matrix $\mathbf{A}$ and adj($\mathbf{A}$) denote the adjoint matrix of $\mathbf{A}$. We note that in the vector AR(1) model given in (17.3.5),

$$[\Phi(B)]^{-1} = \frac{1}{|\Phi(B)|}\operatorname{adj}[\Phi(B)]. \tag{17.3.8}$$

Now, $\Phi(B) = (\mathbf{I} - \Phi B)$ is a matrix polynomial in B of order 1. The order of the resulting matrix polynomial in B of its adjoint matrix clearly will be finite. Thus, the inverse of a nondegenerate AR(1) matrix polynomial (i.e., $\Phi(B) \neq \mathbf{I}$) will be of a finite order if the determinant $|\Phi(B)|$ is independent of B. For the example given in (17.3.7), we note that the determinant $|\mathbf{I} - \Phi B| = 1$. To understand this special feature of a vector ARMA model, we investigate in this section conditions for these seemingly different but equivalent representations.

From the above observations, we have the following results (more detailed proofs can be found in Shen and Wei [1995]):

1. Given the m-dimensional vector AR(1) model

$$(\mathbf{I} - \Phi B)\mathbf{Z}_t = \mathbf{a}_t$$

defined in (17.3.5), the following conditions are equivalent:

(a) The vector AR(1) process can be expressed as a finite-order vector MA(q) model.
(b) The determinantal polynomial $|\mathbf{I} - \Phi B|$ is independent of B. Especially, with $B = 0$, we have $|\mathbf{I} - \Phi B| = 1$.
(c) $|\Phi| = 0$, $\sum_{i=1}^{m}|\Phi(i)| = 0$, $\sum_{1\le i\le j\le m}|\Phi(i,j)| = 0, \ldots,$ $\sum_{i=1}^{m}\phi_{ii} = 0$, where $\Phi(i_1, \ldots, i_k)$ is the matrix of Φ with its $i_1, \ldots, i_k$ columns and rows removed, and $\phi_{i,j}$ denotes the ith diagonal element of Φ.
(d) The eigenvalues of Φ are all 0.
(e) The matrix Φ is nilpotent, i.e., there exists an integer k such that $\Phi^k \neq \mathbf{0}$, and $\Phi^j = \mathbf{0}$ for $j > k$. In fact,

$$\mathbf{Z}_t = [\mathbf{I} - \Phi B]^{-1}\mathbf{a}_t = \frac{1}{|\mathbf{I} - \Phi B|}\operatorname{adj}[\mathbf{I} - \Phi B]\mathbf{a}_t. \tag{17.3.9}$$

Because $|\mathbf{I} - \Phi B|$ is a constant, $(\mathbf{I} - \Phi B)$ is an $m \times$ m matrix with its elements being polynomials in B of order 1, and the elements of the adjoint matrix adj$[\mathbf{I} - \Phi B]$ are polynomials in B of order less than or equal to $(m - 1)$, the m-dimensional vector AR(1) model can be expressed as a finite m-dimensional vector MA(q) model with $q \leq (m - 1)$.

2. Similarly, the m-dimensional vector MA(1) model

$$\mathbf{Z}_t = (\mathbf{I} - \Theta B)\mathbf{a}_t, \tag{17.3.10}$$

where $\Theta = [\theta_{ij}]$ is an $m \times m$ coefficient matrix with (i, j) element $\theta_{i,j}$, can be expressed as a finite-order vector AR(p) model, with $p \leq (m - 1)$, if any one of the following holds:

(a) The determinantal polynomial $|\mathbf{I} - \Theta B|$ is independent of B.

(b) $|\Theta| = 0$, $\sum_{i=1}^{m} |\Theta(i)| = 0$, $\sum_{1 \leq i < j \leq m} |\Theta(i, j)| = 0, \ldots,$ $\sum_{i=1}^{m} \theta_{ii} = 0,$ where $\Theta(i_1, \ldots, i_k)$ is the matrix of Θ with its $i_1, \ldots, i_k$ columns and rows removed and $\theta_{i,i}$ denotes the ith diagonal element of Θ.

(c) The eigenvalues of Θ are all 0.

(d) The matrix Θ is nilpotent.

3. Next, let us consider the m-dimensional vector ARMA(p, q) process

$$\Phi_p(B)\mathbf{Z}_t = \Theta_q(B)\mathbf{a}_t, \tag{17.3.11}$$

where $\Phi_p(B) = (\mathbf{I} - \Phi_1 B - \cdots - \Phi_p B^p)$ and $\Theta_q(B) = (\mathbf{I} - \Theta_1 B - \cdots - \Theta_q B^q)$ are matrix polynomials in B of order p and q, respectively. Because

$$\mathbf{Z}_t = \frac{1}{|\Phi_p(B)|} \text{adj}[\Phi_p(B)]\Theta_q(B)\mathbf{a}_t, \tag{17.3.12}$$

the process can be represented as a finite-order vector MA(k) model with $k \leq [(m - 1)p + q]$ if the determinantal polynomial $|\Phi_p(B)|$ is independent of B. On the other hand, because

$$\frac{1}{|\Theta_q(B)|} \text{adj}[\Theta_q(B)]\Phi_p(B)\mathbf{Z}_t = \mathbf{a}_t, \tag{17.3.13}$$

the process can be represented as a finite-order vector AR(n) model with $n \leq [(m - 1)q + p]$ if the determinantal polynomial $|\Theta_q(B)|$ is independent of B.

4. More generally, if, in (17.3.11), $\Phi_p(B) = \boldsymbol{H}_{p_1}(B)\Omega_{p-p_1}(B)$ such that $|\boldsymbol{H}_{p_1}(B)|$ is independent of B, then the process can be written as the finite-order vector ARMA($p - p_1$, k) model with $p_1 \leq p$ and $k \leq [(m - 1)p_1 + q]$

$$\Omega_{p-p_1}(B)\mathbf{Z}_t = \frac{1}{|\mathbf{H}_{p_1}(B)|} \text{adj}[\mathbf{H}_{p_1}(B)]\Theta_q(B)\mathbf{a}_t. \tag{17.3.14}$$

On the other hand, if, in (17.3.11), $\Theta_q(B) = V_{q_1}(B)\Gamma_{q-q_1}(B)$ such that $|V_{q_1}(B)|$ is independent of B, then the process can be written as the finite-order vector ARMA(n, $q - q_1$) model with $q_1 \leq q$ and $n \leq [(m-1)q_1 + p]$

$$\frac{1}{|\mathbf{V}_{q_1}(B)|}\text{adj}[\mathbf{V}_{q_1}(B)]\Phi_p(B)\mathbf{Z}_t = \Gamma_{q-q_1}(B)\mathbf{a}_t. \tag{17.3.15}$$

EXAMPLE 17.6 Consider the two-dimensional vector ARMA(2, 2) process

$$(\mathbf{I} - \Phi_1 B - \Phi_2 B^2)\mathbf{Z}_t = (\mathbf{I} - \Theta_1 B - \Theta_2 B^2)\mathbf{a}_t, \tag{17.3.16}$$

where

$$\Phi_1 = \begin{bmatrix} .8 & 1.3 \\ .1 & .6 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -.1 & -.6 \\ 0 & 0 \end{bmatrix},$$
$$\Theta_1 = \begin{bmatrix} .4 & .2 \\ 1.3 & .6 \end{bmatrix}, \quad \text{and} \quad \Theta_2 = \begin{bmatrix} 0 & 0 \\ -.4 & -.2 \end{bmatrix}.$$

It can be easily shown that $|\mathbf{I} - \Phi_1 B - \Phi_2 B^2| = (1 - 1.4B + .45B^2)$, which is not independent of B. Hence, Model in (17.3.16) cannot be represented as a finite-order vector MA model. It can be shown, however, that the

$$(\mathbf{I} - \Phi_1 B - \Phi_2 B^2) = (\mathbf{I} - \mathbf{H}B)(\mathbf{I} - \Omega B),$$

with

$$\mathbf{H} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \Omega = \begin{bmatrix} .8 & .3 \\ .1 & .6 \end{bmatrix}, \quad \text{and} \quad |\mathbf{H}_1(B)| = |\mathbf{I} - \mathbf{H}B| = 1.$$

Hence, by (17.3.14), (17.3.16) can also be written as the vector ARMA(1, 3) model

$$(\mathbf{I} - \Omega B)\mathbf{Z}_t = (\mathbf{I} - \Psi_1 B - \Psi_2 B^2 - \Psi_3 B^3)\mathbf{a}_t, \tag{17.3.17}$$

where

$$\Psi_1 = \begin{bmatrix} .4 & -.8 \\ 1.3 & .6 \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} 1.3 & .6 \\ -.4 & -.2 \end{bmatrix}, \quad \text{and} \quad \Psi_3 = \begin{bmatrix} -.4 & -.2 \\ 0 & 0 \end{bmatrix}.$$

Similarly, $|\mathbf{I} - \Theta_1 B - \Theta_2 B^2| = (1 - B + .18B^2)$, which is not independent of B; hence, the process cannot be represented by a finite-order purely AR vector model. Because, however,

$$(\mathbf{I} - \Theta_1 B - \Theta_2 B^2) = (\mathbf{I} - \mathbf{V}B)(\mathbf{I} - \Gamma B),$$

with

$$\mathbf{V} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} .4 & .2 \\ .3 & .6 \end{bmatrix}, \qquad \text{and} \qquad |\mathbf{I} - \mathbf{V}B| = 1,$$

by (17.3.15), (17.3.16) can also be written as the vector ARMA(3, 1) model

$$(\mathbf{I} - \Pi_1 B - \Pi_2 B^2 - \Pi_3 B^3)\mathbf{Z}_t = (\mathbf{I} - \Gamma B)\mathbf{a}_t, \tag{17.3.18}$$

where

$$\Pi_1 = \begin{bmatrix} .8 & 1.3 \\ -.9 & .6 \end{bmatrix}, \qquad \Pi_2 = \begin{bmatrix} -.1 & -.6 \\ .8 & 1.3 \end{bmatrix}, \qquad \text{and} \qquad \Pi_3 = \begin{bmatrix} 0 & 0 \\ -.1 & -.6 \end{bmatrix}.$$

The models in (17.3.16), (17.3.17), and (17.3.18) are equivalent. Tiao and Tsay (1989) also call these equivalent models exchangeable models.

17.3.2 Some Implications

The results in Section 17.3.1 indicate that for a vector time series, a certain process can be simultaneously represented as a finite-order vector AR model, a finite-order vector MA model, or a finite-order vector ARMA model, a phenomenon that cannot exist in the univariate time series. In this section, we examine some implications of these representations. Some implications are direct consequences of the representations.

Four implications are as follows:

1. A finite-order vector AR or ARMA model is a finite period memory model if it can also be represented as a finite-order vector MA model. That is, the optimal forecast of the process after a finite period of time is the mean vector of the process. This property is in drastic contrast to the univariate time series process, where the mean becomes the optimal forecast for an AR or an ARMA model only as a limit when the forecast period tends to infinity.
2. In terms of stationarity in the wide sense, an m-dimensional stationary vector process $\mathbf{Z}_t$ can be written in an MA form

$$\mathbf{Z}_t = \sum_{k=0}^{\infty} \Psi_k \mathbf{a}_{t-k} \tag{17.3.19}$$

such that the $m \times m$ coefficient matrices $\Psi_k = [\psi_{ij,k}]$ are square summable in the sense that each of the $m \times m$ sequences is square summable, i.e. $\sum_{k=0}^{\infty} \psi_{ij,k}^2 < \infty$ for

$i = 1, \ldots, m$ and $j = 1, \ldots, m$. Thus, no nonstationary vector process can be represented by both a finite-order AR vector model and a finite-order vector MA model. An m-dimensional vector process $\mathbf{Z}_t$ is said to be invertible if it can be written in an AR form

$$\mathbf{Z}_t = \sum_{k=0}^{\infty} \Pi_k \mathbf{Z}_{t-k} + \mathbf{a}_t \tag{17.3.20}$$

such that the $m \times m$ coefficient matrices $\Pi_k = [\pi_{ij,k}]$ are absolutely summable, i.e., $\sum_{k=0}^{\infty} |\pi_{ij,k}| < \infty$ for $i = 1, \ldots, m$ and $j = 1, \ldots, m$. Hence, no noninvertible vector process can be simultaneously represented by both a finite-order vector AR model and a finite-order vector MA model. In other words, if a vector time series process can be simultaneously represented by a finite-order AR model and a finite-order MA model, then the process must be both stationary and invertible.

3. A nonstationary process $\mathbf{Z}_t$ is said to be integrated of order d, denoted as $I(d)$, if its $(d-1)$th difference is nonstationary but its dth difference is stationary. As discussed in Section 17.1, an $m \times 1$ vector process $\mathbf{Z}_t = (Z_{1,t}, \ldots, Z_{m,t})'$ is said to be cointegrated if each of its component series $Z_{i,t}$ is $I(1)$, but some linear combination of $\mathbf{Z}_t$, i.e., $\mathbf{C}'\mathbf{Z}_t$ for some $k \times m$ matrix $\mathbf{C}'$ with $k < m$, is a stationary $k \times 1$ vector process. The discussion in item 2 implies that a cointegrated vector process $\mathbf{Z}_t$ cannot be simultaneously represented by a finite-order vector AR, a finite-order vector MA, or a finite-order vector ARMA process. If multiple representations occur, then they must be in terms of its linear combination $\mathbf{C}'\mathbf{Z}_t$ for some matrix $\mathbf{C}'$.
4. If a vector time series process can simultaneously be represented by a finite-order AR model and a finite-order MA model, then the process can be reduced to an input-output type of transfer function model discussed in Chapter 14. We illustrate the phenomenon through a first-order vector process. The results for other cases are similar.

   First, let us consider an m-dimensional vector AR(1) process

$$(I - \Phi B)\mathbf{Z}_t = \mathbf{a}_t. \tag{17.3.21}$$

If the process can also be written as a finite-order vector MA model, then by earlier results the eigenvalues of Φ are all zero. Hence, there exists a nonsingular matrix $\mathbf{H}$ such that $\mathbf{H}\Phi\mathbf{H}^{-1} = \mathbf{J}$, where $\mathbf{J}$ is a Jordan matrix whose elements are zero except for those on the first superdiagonal, which are equal to 0 or 1, i.e.,

$$\mathbf{H}\Phi\mathbf{H}^{-1} = \begin{bmatrix} 0 & \delta & 0 & \cdots & 0 \\ & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & & 0 \\ & & & & \delta \\ 0 & & \cdots & & 0 \end{bmatrix} = \mathbf{J}, \tag{17.3.22}$$

where $\delta = 0$ or 1. Multiplying $\mathbf{H}$ on both sides of (17.3.21) and letting $\mathbf{Y}_t = \mathbf{H}\mathbf{Z}_t$ and $\mathbf{e}_t = \mathbf{H}\mathbf{a}_t$, gives

$$\mathbf{H}(\mathbf{I} - \Phi B)\mathbf{H}^{-1}\mathbf{H}\mathbf{Z}_t = \mathbf{H}\mathbf{a}_t$$

or

$$(\mathbf{I} - \mathbf{J}B)\mathbf{Y}_t = \mathbf{e}_t, \tag{17.3.23}$$

where $\mathbf{e}_t$ is the m-dimensional white noise process with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{H}\Sigma\mathbf{H}'$. $(\mathbf{I} - \mathbf{J}B)$ is a square matrix whose elements are zero except for those on the principal diagonal, which are all equal to 1, and those on the first super-diagonal, which are all equal to $-\delta B$. Thus, (17.3.23) becomes

$$\begin{bmatrix} 1 & -\delta B & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & 0 \\ \vdots & & & & -\delta B \\ 0 & 0 & & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{m,t} \end{bmatrix} = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{m,t} \end{bmatrix}. \tag{17.3.24}$$

Next, consider an m-dimensional vector MA(1) process

$$\mathbf{Z}_t = (\mathbf{I} - \Theta B)\mathbf{a}_t. \tag{17.3.25}$$

If the process can also be written as a finite-order vector AR model, then there exists an integer p so that $\Theta^p \neq \mathbf{0}$ but $\Theta^n = \mathbf{0}$ for $n > p$, and

$$(\mathbf{I} + \Theta B + \Theta^2 B^2 + \cdots + \Theta^p B^p)\mathbf{Z}_t = \mathbf{a}_t. \tag{17.3.26}$$

Because Θ is nilpotent, there exists a nonsingular matrix $\mathbf{R}$ such that $\mathbf{R}\Theta\mathbf{R}^{-1} = \mathbf{J}$, where $\mathbf{J}$ is a Jordan matrix defined in (17.3.22). Multiplying $\mathbf{R}$ on both sides of (17.3.26) and letting $\mathbf{W}_t = \mathbf{R}\mathbf{Z}_t$ and $\mathbf{r}_t = \mathbf{R}\mathbf{a}_t$ gives

$$\mathbf{R}(\mathbf{I} + \Theta B + \Theta^2 B^2 + \cdots + \Theta^p B^p)\mathbf{R}^{-1}\mathbf{R}\mathbf{Z}_t = \mathbf{R}\mathbf{a}_t$$

or

$$(\mathbf{I} + \mathbf{R}\Theta\mathbf{R}^{-1}B + \cdots + \mathbf{R}\Theta^p\mathbf{R}^{-1}B^p)\mathbf{W}_t = \mathbf{r}_t.$$

Because $\mathbf{R}\Theta^k\mathbf{R}^{-1} = \mathbf{J}^k$,

$$(\mathbf{I} + \mathbf{J}B + \cdots + \mathbf{J}^pB^p)\mathbf{W}_t = \mathbf{r}_t, \tag{17.3.27}$$

where $\mathbf{r}_t$ is the m-dimensional white noise process with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{R}\Sigma\mathbf{R}'$. More explicitly, (17.3.27) becomes

$$\begin{bmatrix} 1 & \delta\mathrm{B} & \delta^2B^2 & \cdots & \delta^{p-1}B^{p-1} & \delta^pB^p \\ 0 & 1 & \delta\mathrm{B} & \ddots & \cdots & \delta^{p-1}B^{p-1} \\ & & & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \delta\mathrm{B} & \delta^2B^2 \\ & & & & 1 & \delta\mathrm{B} \\ 0 & & \cdots & & 0 & 1 \end{bmatrix} \begin{bmatrix} W_{1,t} \\ W_{2,t} \\ \vdots \\ W_{m,t} \end{bmatrix} = \begin{bmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{m,t} \end{bmatrix}. \tag{17.3.28}$$

These examples show that if a vector time series process can simultaneously be represented by a finite-order vector AR model and a finite-order vector MA model, then it can be written in a vector AR form with a triangular coefficient matrix. Thus, the process can be reduced to an input-output type of transfer function model as discussed in Section 16.3.1.

EXERCISES

17.1 Consider the process, $Z_t = \Psi(B)a_t = (1 - B)a_t = a_t - a_{t-1}$, where the a_t are i.i.d. random variables with mean 0 and variance 1. Note that $\Psi(1) = 0$; hence, it is not $I(0)$. Find the process of its sum, and show that the resulting process is asymptotically stationary and thus is not an $I(1)$ integrated process.

17.2 Consider the two-dimensional vector AR(1) process

$$\mathbf{Z}_t = \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \phi \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}.$$

Verify that each component series of $\mathbf{Z}_t$ is $I(1)$.

17.3 Consider the bivariate process given in Exercise 17.2.

(a) Why is $\mathbf{Z}_t$ cointegrated? Find its cointegrating vector.

(b) Find its MA representation.

(c) Find its AR representation.

(d) Find its error-correction representation.

17.4 Find a vector time series with its component series being nonstationary.

(a) Build a vector ARMA model for the series, and examine the interrelationship among the component series.

(b) Perform the unit root test to verify whether each component series is $I(1)$.

(c) Test whether the components are cointegrated.

(d) If the components are cointegrated, find an error-correction model for the series that describes a long-term relationship among the components.

17.5 Let $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})'$ and

$$\begin{aligned} Z_{1,t} &= Z_{3,t-1} + a_{1,t} \\ Z_{2,t} &= a_{2,t} \\ Z_{3,t} &= Z_{2,t-1} + a_{3,t}, \end{aligned}$$

where $\mathbf{a}_t = (a_{1,t}, a_{2,t}, a_{3,t})'$ is a vector of white noise process with mean zero and variance-covariance matrix $\Sigma = \boldsymbol{I}$.

(a) Find the covariance matrix generating function $\boldsymbol{G}_{\mathbf{Z}}(B)$ of $\mathbf{Z}_t$.

(b) Find the covariance matrix function $\Gamma_{\mathbf{Z}}(k)$ and correlation matrix function $\boldsymbol{\rho}_{\mathbf{Z}}(k)$ of $\mathbf{Z}_t$.

(c) Find the partial process covariance matrix generating function $G^{\dagger}_{12.3}(B)$ for $(Z_{1,t}, Z_{2,t})'$ allowing for $Z_{3,t}$.

(d) Find the partial process covariance matrix function $\Gamma^{\dagger}_{12.3}(k)$ and correlation matrix function $\boldsymbol{\rho}^{\dagger}_{12.3}(k)$.

(e) Comment on your findings.

17.6 Let $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})'$ and

$$\begin{aligned} Z_{1,t} &= a_{1,t} \\ Z_{2,t} &= a_{2,t} \\ Z_{3,t} &= Z_{1,t-1} + a_{3,t}, \end{aligned}$$

where $\mathbf{a}_t = (a_{1,t}, a_{2,t}, a_{3,t})'$ is a vector of white noise process with mean zero and variance-covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

(a) Find the covariance matrix generating function $\boldsymbol{G}_{\mathbf{Z}}(B)$ of $\mathbf{Z}_t$.

(b) Find correlation matrix function $\boldsymbol{\rho}_{12}(k)$ for $(Z_{1,t}, Z_{2,t})'$.

(c) Find the partial process covariance matrix generating function $\boldsymbol{G}^{\dagger}_{[2.3]}(B)$ for $(Z_{1,t}, Z_{2,t})'$ allowing for $Z_{3,t}$.

(d) Find the partial process correlation matrix function $\boldsymbol{\rho}^{\dagger}_{[2.3]}(k)$ for $(Z_{1,t}, Z_{2,t})'$ allowing for $Z_{3,t}$.

(e) Comment on your findings.

17.7 Find a vector time series and perform a partial process analysis as discussed in Section 17.2.4.

17.8 Let $\mathbf{Z}_t$ be the m-dimensional vector AR(1) model

$$(\mathbf{I} - \Phi B)\mathbf{Z}_t = \mathbf{a}_t,$$

where $\mathbf{a}_t$ is the m-dimensional white noise vector process with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Show that the following conditions are equivalent:

(a) The vector AR(1) process can be expressed as a finite-order vector MA(q) model.

(b) The determinantal polynomial $|\mathbf{I} - \Phi B|$ is independent of B. Especially, when $B = 0$, we have $|\mathbf{I} - \Phi B| = 1$.

(c) The eigenvalues of Φ are all 0.

(d) The matrix Φ is nilpotent, i.e., there exists an integer k such that $\Phi^k \neq \mathbf{0}$, and $\Phi^j = \mathbf{0}$ for $j > k$.

17.9 (a) Find an example of a vector process that has both finite AR and finite MA representations.

(b) Express the model in part (a) in a transfer function model.

18

State Space Models and the Kalman Filter

Having presented various univariate and multivariate time series models in previous chapters, we now give a brief introduction to the state space model and Kalman filtering. This very general approach in time series modeling can be used for both univariate and multivariate time series.

18.1 State Space Representation

The state space representation of a system is a fundamental concept in modern control theory. The state of a system is defined to be a minimum set of information from the present and past such that the future behavior of the system can be completely described by the knowledge of the present state and the future input. Thus, the state space representation is based on the Markov property, which implies that given the present state, the future of a system is independent of its past. Consequently, the state space representation of a system is also called the Markovian representation of the system. Let $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$ be the outputs of a system to the inputs $\mathbf{X}_{1t}$ and $\mathbf{X}_{2t}$, respectively. A system is said to be linear if and only if a linear combination of the inputs, $a\mathbf{X}_{1t} + b\mathbf{X}_{2t}$, produces the same linear combination of the outputs, $a\mathbf{Y}_{1t} + b\mathbf{Y}_{2t}$, for any constants a and b. A system is said to be time-invariant if the characteristics of a system do not change with time so that if the input $\mathbf{X}_t$ produces the output $\mathbf{Y}_t$, then the input $\mathbf{X}_{t-t_0}$ will produce the output $\mathbf{Y}_{t-t_0}$. A system is said to be linear time-invariant if it is both linear and time-invariant. Many physical processes can be modeled by linear time-invariant systems that contain stationary processes discussed in previous chapters as special cases.

For a linear time-invariant system, its state space representation is described by the state equation

$$\mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{X}_{t+1} \tag{18.1.1}$$

and the output equation

$$\mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t, \tag{18.1.2}$$

where $\mathbf{Y}_t$ is a state vector of dimension k, $\mathbf{A}$ is a $k \times k$ transition matrix, $\mathbf{G}$ is a $k \times n$ input matrix, $\mathbf{X}_t$ is an $n \times 1$ vector of the input to the system, $\mathbf{Z}_t$ is an $m \times 1$ vector of the output, and $\mathbf{H}$ is an $m \times k$ output or observation matrix. If both the input $\mathbf{X}_t$ and the output $\mathbf{Z}_t$ are stochastic processes, then the state space representation is given by

$$\mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1} \tag{18.1.3}$$

$$\mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t + \mathbf{b}_t, \tag{18.1.4}$$

where $\mathbf{a}_{t+1} = \mathbf{X}_{t+1} - E(\mathbf{X}_{t+1} \mid \mathbf{X}_i, i \leq t)$ is the $n \times 1$ vector of one-step ahead forecast error of the input process $\mathbf{X}_t$ and $\mathbf{b}_t$ is an $m \times 1$ vector of disturbances assumed to be independent of $\mathbf{a}_t$. The vector $\mathbf{a}_{t+1}$ is also known as the innovation of the input $\mathbf{X}_t$ at time $(t + 1)$. When $\mathbf{Z}_t = \mathbf{X}_t$, $\mathbf{b}_t$ vanishes from (18.1.4), the state space representation of a stationary stochastic process $\mathbf{Z}_t$ becomes

$$\begin{cases} \mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1} \\ \mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t. \end{cases} \tag{18.1.5}$$

Thus, the process $\mathbf{Z}_t$ is the output of a time-invariant linear stochastic system driven by a white noise input $\mathbf{a}_t$. The $\mathbf{Y}_t$ is known as the state of the process.

The state equation is also called the system equation or the transition equation, and the output equation is also referred to as the measurement equation or the observation equation. The state space representation of a system is related to the Kalman filter and was originally developed by control engineers (see Kalman [1960], Kalman and Bucy [1961], and Kalman, Falb, and Arbib [1969]). The system is also known as a state space model. Akaike (1974a) seems to have been the first to apply the concept directly to the analysis of ARMA models.

18.2 The Relationship between State Space and ARMA Models

To see the relationship between state space and ARMA models for either the univariate or multivariate case, consider the zero mean stationary m-dimensional vector ARMA(p, q) model

$$\Phi(B)\mathbf{Z}_t = \Theta(B)\mathbf{a}_t \tag{18.2.1a}$$

or

$$\mathbf{Z}_t = \Phi_1\mathbf{Z}_{t-1} + \cdots + \Phi_p\mathbf{Z}_{t-p} + \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1} - \cdots - \Theta_q\mathbf{a}_{t-q}, \tag{18.2.1b}$$

where $\Phi(B) = (\mathbf{I} - \Phi_1 B - \cdots - \Phi_p B^p)$, $\Theta(B) = (\mathbf{I} - \Theta_1 B - \cdots - \Theta_q B^q)$, and $\mathbf{a}_t$ is an m-dimensional multivariate zero mean white noise process.

Rewriting (18.2.1a) in a moving average form

$$\begin{aligned}\mathbf{Z}_t &= \Phi^{-1}(B)\Theta(B)\mathbf{a}_t \\ &= \sum_{j=0}^{\infty}\Psi_j\mathbf{a}_{t-j},\end{aligned} \tag{18.2.2}$$

where $\Psi_0 = \mathbf{I}$, we have

$$\mathbf{Z}_{t+i} = \sum_{j=0}^{\infty}\Psi_j\mathbf{a}_{t+i-j}. \tag{18.2.3}$$

Let

$$\mathbf{Z}_{t+i|t} = E[\mathbf{Z}_{t+i} \mid \mathbf{Z}_k, k \le t]. \tag{18.2.4}$$

Then

$$\mathbf{Z}_{t+i|t} = \sum_{j=i}^{\infty}\Psi_j\mathbf{a}_{t+i-j}.$$

Now,

$$\begin{aligned}\mathbf{Z}_{t+i|t+1} &= E(\mathbf{Z}_{t+i} \mid \mathbf{Z}_k, k \le t+1) \\ &= \sum_{j=(i-1)}^{\infty}\Psi_j\mathbf{a}_{t+i-j} \\ &= \sum_{j=i}^{\infty}\Psi_j\mathbf{a}_{t+i-j} + \Psi_{i-1}\mathbf{a}_{t+1} \\ &= \mathbf{Z}_{t+i|t} + \Psi_{i-1}\mathbf{a}_{t+1}.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{Z}_{t+1|t+1} &= \mathbf{Z}_{t+1|t} + \mathbf{a}_{t+1} \\ \mathbf{Z}_{t+2|t+1} &= \mathbf{Z}_{t+2|t} + \Psi_1\mathbf{a}_{t+1} \\ \mathbf{Z}_{t+3|t+1} &= \mathbf{Z}_{t+3|t} + \Psi_2\mathbf{a}_{t+1} \\ &\vdots \\ \mathbf{Z}_{t+p-1|t+1} &= \mathbf{Z}_{t+p-1|t} + \Psi_{p-2}\mathbf{a}_{t+1} \\ \mathbf{Z}_{t+p|t+1} &= \mathbf{Z}_{t+p|t} + \Psi_{p-1}\mathbf{a}_{t+1} \\ &= \Phi_p\mathbf{Z}_{t|t} + \Phi_{p-1}\mathbf{Z}_{t+1|t} + \cdots \\ &\quad + \Phi_1\mathbf{Z}_{t+p-1|t} + \Psi_{p-1}\mathbf{a}_{t+1}.\end{aligned}$$

We assume, without loss of generality, that $p > q$ by adding $\Phi_i = \mathbf{0}$ in (18.2.1b) if necessary. Then, from (18.2.1b), we have

$$\begin{aligned}\mathbf{Z}_{t+p|t} &= \Phi_1\mathbf{Z}_{t+p-1|t} + \cdots + \Phi_p\mathbf{Z}_t \\ \mathbf{Z}_{t+p+1|t} &= \Phi_1\mathbf{Z}_{t+p|t} + \cdots + \Phi_p\mathbf{Z}_{t+1|t} \\ &= f(\mathbf{Z}_t, \mathbf{Z}_{t+1|t}, \ldots, \mathbf{Z}_{t+p-1|t}).\end{aligned}$$

In fact, it is clear that $\mathbf{Z}_{t+p+i|t}$ for $i \geq 0$ is a function of $\mathbf{Z}_t$, $\mathbf{Z}_{t+1|t}$, ..., and $\mathbf{Z}_{t+p-1|t}$. Hence, the state vector is $\{\mathbf{Z}_t, \mathbf{Z}_{t+1|t}, \ldots, \mathbf{Z}_{t+p-1|t}\}$, and the state space representation of the vector ARMA(p, q) model is given by

$$\begin{bmatrix}\mathbf{Z}_{t+1|t+1}\\ \mathbf{Z}_{t+2|t+1}\\ \vdots\\ \mathbf{Z}_{t+p|t+1}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0}\\ \vdots & & & & & \vdots\\ \Phi_p & \Phi_{p-1} & & \cdots & & \Phi_1\end{bmatrix}\begin{bmatrix}\mathbf{Z}_t\\ \mathbf{Z}_{t+1|t}\\ \vdots\\ \mathbf{Z}_{t+p-1|t}\end{bmatrix} + \begin{bmatrix}\mathbf{I}\\ \Psi_1\\ \vdots\\ \Psi_{p-1}\end{bmatrix}\mathbf{a}_{t+1} \tag{18.2.5}$$

and

$$\mathbf{Z}_t = [\mathbf{I}_m \mathbf{0} \quad \cdots \quad \mathbf{0}]\begin{bmatrix}\mathbf{Z}_t\\ \mathbf{Z}_{t+1|t}\\ \vdots\\ \mathbf{Z}_{t+p-1|t}\end{bmatrix}. \tag{18.2.6}$$

Note that in the above representation, the first m components of the state vector are equal to $\mathbf{Z}_t$. Equation (18.2.6) is usually suppressed in the representation of an ARMA model.

Now suppose that a process $\mathbf{Z}_t$ has a state space representation

$$\begin{cases}\mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1}\\ \mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t,\end{cases} \tag{18.2.7}$$

where it is assumed that $\mathbf{Y}_t$ is a $mp \times 1$ vector of the state and a_t is the innovation of $\mathbf{Z}_t$. If the characteristic polynomial of $\mathbf{A}$ is given by $|\lambda\mathbf{I} - \mathbf{A}| = \sum_{i=0}^{p}\phi_i\lambda^{p-i}$, where $\phi_0 = 1$, then by the Cayley–Hamilton theorem (see, for example, Noble [1969] and Searle [1982])

$$\sum_{i=0}^{p}\phi_i\mathbf{A}^{p-i} = \mathbf{0}. \tag{18.2.8}$$

By successive substitution in the first equation of (18.2.7), we have, for $i > 0$,

$$\begin{aligned}\mathbf{Y}_{t+i} &= \mathbf{A}\mathbf{Y}_{t+i-1} + \mathbf{G}\mathbf{a}_{t+i} \\ &= \mathbf{A}(\mathbf{A}\mathbf{Y}_{t+i-2} + \mathbf{G}\mathbf{a}_{t+i-1}) + \mathbf{G}\mathbf{a}_{t+i} \\ &= \mathbf{A}^2\mathbf{Y}_{t+i-2} + \mathbf{A}\mathbf{G}\mathbf{a}_{t+i-1} + \mathbf{G}\mathbf{a}_{t+i} \\ &\vdots \\ &= \mathbf{A}^i\mathbf{Y}_t + \mathbf{A}^{i-1}\mathbf{G}\mathbf{a}_{t+1} + \cdots + \mathbf{G}\mathbf{a}_{t+i}. \end{aligned} \tag{18.2.9}$$

Now,

$$\begin{aligned}\mathbf{Z}_{t+p} &= \mathbf{H}\mathbf{Y}_{t+p} \\ &= \mathbf{H}(\mathbf{A}^p\mathbf{Y}_t + \mathbf{A}^{p-1}\mathbf{G}\mathbf{a}_{t+1} + \cdots + \mathbf{G}\mathbf{a}_{t+p}) \\ \phi_1\mathbf{Z}_{t+p-1} &= \mathbf{H}\phi_1\mathbf{Y}_{t+p-1} \\ &= \mathbf{H}\phi_1(\mathbf{A}^{p-1}\mathbf{Y}_t + \mathbf{A}^{p-2}\mathbf{G}\mathbf{a}_{t+1} + \cdots + \mathbf{G}\mathbf{a}_{t+p-1}) \\ &\vdots \\ \phi_{p-1}\mathbf{Z}_{t+1} &= \mathbf{H}\phi_{p-1}\mathbf{Y}_{t+1} \\ &= \mathbf{H}\phi_{p-1}(\mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1}) \\ \phi_p\mathbf{Z}_t &= \mathbf{H}\phi_p\mathbf{Y}_t. \end{aligned}$$

Hence, $\mathbf{Z}_t$ has an ARMA representation

$$\begin{aligned}&\mathbf{Z}_{t+p} + \phi_1\mathbf{Z}_{t+p-1} + \cdots + \phi_{p-1}\mathbf{Z}_{t+1} + \phi_p\mathbf{Z}_t \\ &\quad = \mathbf{H}(\mathbf{A}^p + \phi_1\mathbf{A}^{p-1} + \cdots + \phi_{p-1}\mathbf{A} + \phi_p\mathbf{I})\mathbf{Y}_t \\ &\qquad + \mathbf{H}(\mathbf{A}^{p-1} + \phi_1\mathbf{A}^{p-2} + \cdots + \phi_{p-1}\mathbf{I})\mathbf{G}\mathbf{a}_{t+1} \\ &\qquad + \cdots + \mathbf{H}\mathbf{G}\mathbf{a}_{t+p} \\ &\quad = \Theta_0\mathbf{a}_{t+p} + \Theta_1\mathbf{a}_{t+p-1} + \cdots + \Theta_{p-1}\mathbf{a}_{t+1}, \end{aligned} \tag{18.2.10}$$

where $\mathbf{H}(\mathbf{A}^p + \phi_1\mathbf{A}^{p-1} + \cdots + \phi_{p-1}\mathbf{A} + \phi_p\mathbf{I})\mathbf{Y}_t = \mathbf{0}$ by (18.2.8) and $\Theta_i = \mathbf{H}(\mathbf{A}^i + \phi_1\mathbf{A}^{i-1} + \cdots + \phi_i\mathbf{I})\mathbf{G}$.

In the representation (18.2.10), we use the Cayley–Hamilton theorem and obtain the AR polynomial as $\Phi_p(B) = \phi_p(B)\mathbf{I}$ where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is a univariate polynomial of order p. Recall from Section 16.3, this is one of possible representations of an identifiable ARMA model.

Note that in the state space representation (18.2.5) of an ARMA model, because the first m components of the state vector are equal to $\mathbf{Z}_t$, we can also recast the state representation into an ARMA model by directly solving the state space system of equations for the first m components. We illustrate this method in the following examples.

EXAMPLE 18.1 Consider the univariate ARMA(2, 1) model,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t - \theta_1 a_{t-1}. \tag{18.2.11}$$

By rewriting (18.2.11) in the moving average form

$$\begin{aligned} Z_t &= (1 - \phi_1 B - \phi_2 B^2)^{-1}(1 - \theta_1 B)a_t \\ &= \sum_{j=0}^{\infty} \Psi_j a_{t-j}, \end{aligned}$$

where $\Psi_0 = 1$, $\Psi_1 = \phi_1 - \theta_1, \ldots$, we have, from (18.2.5) and (18.2.6), the following state space representation of (18.2.11):

$$\begin{bmatrix} Z_{t+1|t+1} \\ Z_{t+2|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{bmatrix} \begin{bmatrix} Z_t \\ Z_{t+1|t} \end{bmatrix} + \begin{bmatrix} 1 \\ \Psi_1 \end{bmatrix} a_{t+1}. \tag{18.2.12}$$

To recast the state space representation (18.2.12) into an ARMA model for Z_t, we note that (18.2.12) implies that

$$Z_{t+1|t+1} = Z_{t+1|t} + a_{t+1}, \tag{18.2.13}$$

$$Z_{t+2|t+1} = \phi_1 Z_{t+1|t} + \phi_2 Z_t + \Psi_1 a_{t+1}. \tag{18.2.14}$$

By (18.2.13), we have

$$Z_{t+2|t+2} = Z_{t+2|t+1} + a_{t+2}. \tag{18.2.15}$$

Because $Z_{t+1|t+1} = Z_{t+1}$ and $Z_{t+2|t+2} = Z_{t+2}$, by substituting (18.2.13) into (18.2.14) and (18.2.14) into (18.2.15), we obtain

$$\begin{aligned} Z_{t+2} &= \phi_1(Z_{t+1} - a_{t+1}) + \phi_2 Z_t + \Psi_1 a_{t+1} + a_{t+2} \\ &= \phi_1 Z_{t+1} + \phi_2 Z_t + a_{t+2} + (\Psi_1 - \phi_1)a_{t+1} \end{aligned}$$

or

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t - \theta_1 a_{t-1},$$

which is the ARMA(2, 1) model given in (18.2.11).

For a multivariate vector process, it is possible for some components of $Z_{t+i|t}$ to be linearly combinations of the remaining components. In such cases, the state vector consists of a subset of the possible components of $\{Z_{t+i|t} \mid i = 0, 1, \ldots, (p-1)\}$. Thus, depending on whether the redundant components, which are linearly dependent on others, have been eliminated, seemingly different but equivalent structure may arise. We illustrate this point in the following example.

EXAMPLE 18.2 Consider the following two-dimensional vector ARMA(1, 1) model:

$$Z_t = \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} - \begin{bmatrix} \theta_{11} & \theta_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \quad (18.2.16)$$

By (18.2.5), the state space representation of (18.2.16) is given by

$$\begin{bmatrix} Z_{t+1|t+1} \\ Z_{t+2|t+1} \end{bmatrix} = \begin{bmatrix} 0 & I \\ \Phi_2 & \Phi_1 \end{bmatrix} \begin{bmatrix} Z_t \\ Z_{t+1|t} \end{bmatrix} + \begin{bmatrix} I \\ \Psi_1 \end{bmatrix} a_{t+1}, \quad (18.2.17)$$

where $\Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ and $\Phi_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a zero matrix added to change the ARMA (1, 1) model to an ARMA(2, 1) model so that $p > q$. By the above remarks, the state vector is equal to $[Z_t', Z_{t+1|t}']'$ if all components of this vector are linearly independent and equal to a subset of $[Z_t', Z_{t+1|t}']'$ if some components of the vector are linearly dependent. To examine the linear dependency of the components of the possible state vector $[Z_t', Z_{t+1|t}']'$, we note that $[Z_t', Z_{t+1|t}']' = [Z_{1,t}, Z_{2,t}, Z_{1,t+1|t}, Z_{2,t+1|t}]'$. Now,

$$Z_{1,t} = \phi_{11}Z_{1,t-1} + \phi_{12}Z_{2,t-1} + a_{1,t} - \theta_{11}a_{1,t-1} - \theta_{12}a_{2,t-1}, \quad (18.2.18)$$

$$Z_{2,t} = \phi_{21}Z_{1,t-1} + \phi_{22}Z_{2,t-1} + a_{2,t}, \quad (18.2.19)$$

$$Z_{1,t+1|t} = \phi_{11}Z_{1,t} + \phi_{12}Z_{2,t} - \theta_{11}a_{1,t} - \theta_{12}a_{2,t}, \quad (18.2.20)$$

$$Z_{2,t+1|t} = \phi_{21}Z_{1,t} + \phi_{22}Z_{2,t}. \quad (18.2.21)$$

Hence, the state vector is $[Z_{1,t}, Z_{2,t}, Z_{1,t+1|t}]'$. Because the component $Z_{2,t+1|t}$ is a linear combination of $Z_{1,t}$ and $Z_{2,t}$, as shown in (18.2.21), it is not included in the state vector.

The representation (18.2.17) can now be simplified by representing $Z_{1,t+1}$, $Z_{2,t+1}$, and $Z_{1,t+2|t+1}$ in terms of the state vector $[Z_{1,t}, Z_{2,t}, Z_{1,t+1|t}]'$ and the input noises $a_{1,t+1}$ and $a_{2,t+1}$. Clearly, from (18.2.18), (18.2.19), and (18.2.20), we have

$$\begin{aligned} Z_{1,t+1} &= Z_{1,t+1|t+1} \\ &= Z_{1,t+1|t} + a_{1,t+1} \quad (18.2.22) \\ Z_{2,t+1} &= Z_{2,t+1|t+1} \\ &= \phi_{21}Z_{1,t} + \phi_{22}Z_{2,t} + a_{2,t+1} \quad (18.2.23) \\ Z_{1,t+2|t+1} &= \phi_{11}Z_{1,t+1} + \phi_{12}Z_{2,t+1} - \theta_{11}a_{1,t+1} - \theta_{12}a_{2,t+1} \\ &= \phi_{11}(Z_{1,t+1|t} + a_{1,t+1}) + \phi_{12}(\phi_{21}Z_{1,t} + \phi_{22}Z_{2,t} \\ &\quad + a_{2,t+1}) - \theta_{11}a_{1,t+1} - \theta_{12}a_{2,t+1} \\ &= \phi_{12}\phi_{21}Z_{1,t} + \phi_{12}\phi_{22}Z_{2,t} + \phi_{11}Z_{1,t+1|t} \\ &\quad + (\phi_{11} - \theta_{11})a_{1,t+1} + (\phi_{12} - \theta_{12})a_{2,t+1}. \quad (18.2.24) \end{aligned}$$

Combining (18.2.22), (18.2.23), and (18.2.24), we obtain the reduced state space representation

$$\begin{bmatrix} Z_{1,t+1} \\ Z_{2,t+1} \\ Z_{1,t+2|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ \phi_{21} & \phi_{22} & 0 \\ \phi_{12}\phi_{21} & \phi_{12}\phi_{22} & \phi_{11} \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{1,t+1|t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ (\phi_{11} - \theta_{11}) & (\phi_{12} - \theta_{12}) \end{bmatrix} \begin{bmatrix} a_{1,t+1} \\ a_{2,t+1} \end{bmatrix}. \quad (18.2.25)$$

We may recast the state space form (18.2.25) into ARMA form by expressing the system of equations in (18.2.25) in terms of $Z_{1,t}$ and $Z_{2,t}$. Using (18.2.22), we can write the conditional expectations in (18.2.24) as

$$\begin{aligned}(Z_{1,t+2} - a_{1,t+2}) &= \phi_{12}\phi_{21}Z_{1,t} + \phi_{12}\phi_{22}Z_{2,t} + \phi_{11}(Z_{1,t+1} - a_{1,t+1}) \\ &\quad + (\phi_{11} - \theta_{11})a_{1,t+1} + (\phi_{12} - \theta_{12})a_{2,t+1}\end{aligned}$$

or

$$\begin{aligned}Z_{1,t+2} &= \phi_{11}Z_{1,t+1} + \phi_{12}(\phi_{21}Z_{1,t} + \phi_{22}Z_{2,t} + a_{2,t+1}) \\ &\quad + a_{1,t+2} - \theta_{11}a_{1,t+1} - \theta_{12}a_{2,t+1} \\ &= \phi_{11}Z_{1,t+1} + \phi_{12}Z_{2,t+1} + a_{1,t+2} - \theta_{11}a_{1,t+1} - \theta_{12}a_{2,t+1}. \quad (18.2.26)\end{aligned}$$

Hence,

$$Z_{1,t+1} = \phi_{11}Z_{1,t} + \phi_{12}Z_{2,t} + a_{1,t+1} - \theta_{11}a_{1,t} - \theta_{12}a_{2,t}. \quad (18.2.27)$$

Combining (18.2.27) and (18.2.23), we get the vector ARMA(1, 1) model given in (18.2.16).

Thus, theoretically there is no distinction between the state space and the ARMA representations of a stationary process. The results obtained from one representation can be used for the other representation. Unless all components of a state vector are linearly independent, however, when converting a state space form into an ARMA form, there will be nontrivial common factors in the autoregressive and moving average polynomials. Hence, the resulting ARMA model may have superficially inflated the orders of p and q.

18.3 State Space Model Fitting and Canonical Correlation Analysis

The state space representation given in Sections 18.1 and 18.2 is clearly not unique. For example, given (18.1.5), we can form a new state vector $\mathbf{V}_t = \mathbf{M}\mathbf{Y}_t$ for any nonsingular matrix $\mathbf{M}$ and obtain a new state space representation

$$\mathbf{V}_{t+1} = \mathbf{A}_1\mathbf{V}_t + \mathbf{G}_1\mathbf{a}_{t+1} \tag{18.3.1}$$

and

$$\mathbf{Z}_t' = \mathbf{H}_1\mathbf{V}_t, \tag{18.3.2}$$

where $\mathbf{A}_1 = \mathbf{MAM}^{-1}$, $\mathbf{G}_1 = \mathbf{MG}$, and $\mathbf{H}_1 = \mathbf{HM}^{-1}$. Through a canonical representation as shown in Akaike (1976), however, we can obtain a unique solution. In the canonical correlation representation, the state vector is uniquely determined through the canonical correlation analysis between the set of current and past observations $(\mathbf{Z}_n, \mathbf{Z}_{n-1}, \ldots)$ and the set of current and future values $(\mathbf{Z}_n, \mathbf{Z}_{n+1|n}, \ldots)$. In terms of an AR($p$) model, because the eventual forecast function is determined by the AR polynomial and $(\mathbf{Z}_n, \mathbf{Z}_{n-1}, \ldots, \mathbf{Z}_{n-p})$ contains essentially all the information relevant for the future values of the process, the canonical correlation analysis is simply performed between the data space

$$\mathcal{D}_n = (\mathbf{Z}_n', \mathbf{Z}_{n-1}', \ldots, \mathbf{Z}_{n-p}')' \tag{18.3.3}$$

and the predictor space

$$\mathcal{F}_n = (\mathbf{Z}_n', \mathbf{Z}_{n+1|n}', \ldots, \mathbf{Z}_{n+p|n}')' \tag{18.3.4}$$

Consider the following block Hankel matrix of the covariance between $\mathcal{D}_n = (\mathbf{Z}_n', \mathbf{Z}_{n-1}', \ldots, \mathbf{Z}_{n-p}')'$ and $\mathcal{F}_n = (\mathbf{Z}_n', \mathbf{Z}_{n+1|n}', \ldots, \mathbf{Z}_{n+p|n}')'$ defined by

$$\Gamma = \begin{bmatrix} \Gamma(0) & \Gamma(1) & \cdots & \Gamma(p) \\ \Gamma(1) & \Gamma(2) & \cdots & \Gamma(p+1) \\ \vdots & \vdots & \vdots & \\ \Gamma(p) & \Gamma(p+1) & \cdots & \Gamma(2p) \end{bmatrix}, \tag{18.3.5}$$

where we use a property of conditional expectations and hence that $\mathrm{Cov}(\mathbf{Z}_{n-i}, \mathbf{Z}_{n+j|n}) = \mathrm{Cov}(\mathbf{Z}_{n-i}, \mathbf{Z}_{n+j})$. Akaike (1974a, 1976) shows that under the nonsingularity assumption of $\Gamma(0)$, for general vector ARMA models, the rank of Γ is equal to the dimension of the state vector and thus is also equal to the number of nonzero canonical correlations between $\mathcal{D}_n$ and $\mathcal{F}_n$.

When the model is unknown, the choice of the order p is obtained from the optimal AR fitting of the data, which is often based on AIC discussed in Chapter 7. For the vector process, AIC is defined as

$$AIC = n \ln|\Sigma_p| + 2pm^2, \tag{18.3.6}$$

where

n = the number of observations,

$|\mathbf{\Sigma}_p|$ = determinant of the covariance matrix for the innovations or the white noise series in the AR(p) fitting, and

m = the dimensions of the vector process $\mathbf{Z}_t$.

The optimal AR order p is chosen so that AIC is minimum. Thus, the canonical correlation analysis will be based on the block Hankel matrix of sample covariances, i.e.,

$$\hat{\Gamma} = \begin{bmatrix} \hat{\Gamma}(0) & \hat{\Gamma}(1) & \cdots & \hat{\Gamma}(p) \\ \hat{\Gamma}(1) & \hat{\Gamma}(2) & \cdots & \hat{\Gamma}(p+1) \\ \vdots & \vdots & & \vdots \\ \hat{\Gamma}(p) & \hat{\Gamma}(p+1) & \cdots & \hat{\Gamma}(2p) \end{bmatrix}, \tag{18.3.7}$$

where $\hat{\Gamma}(j), j = 0, 1, \ldots, 2p$, are the sample covariance matrices as defined in (16.5.16).

As discussed in Example 18.2, because some components of a prediction vector $\mathbf{Z}_{n+i|n}$ may be linearly combinations of remaining components, the canonical correlation analysis is performed between all the components of the data space

$$\mathcal{D}_n = [Z_{1,n}, Z_{2,n}, \ldots, Z_{m,n}, Z_{1,n-1}, Z_{2,n-1}, \ldots, Z_{m,n-1}, \ldots, Z_{1,n-p}, Z_{2,n-p}, \ldots, Z_{m,n-p}]' \tag{18.3.8}$$

and the components of the predictor space

$$\begin{aligned} \mathcal{F}_n = [&Z_{1,n}, Z_{2,n}, \ldots, Z_{m,n}, Z_{1,n+1|n}, Z_{2,n+1|n}, \ldots, Z_{m,n+1|n}, \ldots, \\ &Z_{1,n+p|n}, Z_{2,n+p|n}, \ldots Z_{m,n+p|n}]. \end{aligned}' \tag{18.3.9}$$

Because the state vector is known to be a subset of the predictor space, a sequence of potential state vectors, $\mathbf{Y}_n^j$, is determined through a canonical correlation analysis between a sequence, $\mathcal{F}_n^j$, of subsets of $\mathcal{F}_n$, and the data space $\mathcal{D}_n$, based on the submatrix $\hat{\Gamma}^j$ forming from the columns of $\hat{\Gamma}$, which correspond to the components of $\mathcal{D}_n$ and $\mathcal{F}_n^j$.

More specifically, because the canonical correlation between $\mathbf{Z}_n = [Z_{1,n}, Z_{2,n}, \ldots, Z_{m,n}]'$ and $\mathcal{D}_n$ are $1, \ldots, 1$, which are clearly nonzero, the current state vector is set to $\mathbf{Z}_n$ and the first subset in the sequence, $\mathcal{F}_n^1$, is set to $[Z_{1,n}, Z_{2,n}, \ldots, Z_{m,n}, Z_{1,n+1|n}]'$. If the smallest canonical correlation of $\hat{\Gamma}^1$ is judged to be zero, then a linear combination of $\mathcal{F}_n^1$ is uncorrelated with the data space $\mathcal{D}_n$. Thus, the component $Z_{1,n+1|n}$ and any $Z_{1,n+i|n}$ are excluded from further consideration for components in the state vector. If the smallest canonical correlation is nonzero, then $Z_{1,n+1|n}$ is added to the current state vector. The sequence $\mathcal{F}_n^j$ is now generalized by adding to the current state vector the next component of $\mathcal{F}_n$ that does not correspond to the component that has previously failed to be included

in the state vector. The smallest canonical correlation of $\hat{\Gamma}^j$ is computed and tested for its significance. If it is significally different from zero, then the component is added to the state vector. Otherwise, the component is excluded from the current state vector and also from any further consideration. The state vector selection is complete when no elements of $\mathcal{F}_n$ are left to be added to the current state vector.

For each step in the sequence of canonical correlation analysis, the significance of the smallest canonical correlation, denoted by $\hat{\rho}_{\min}$, is based on the following information criterion from Akaike (1976):

$$C = -n \ln(1 - \hat{\rho}_{\min}^2) - 2[m(p + 1) - q + 1], \tag{18.3.10}$$

where q is the dimension of $\mathcal{F}_n^j$ at the current step. If $C \leq 0$, $\rho_{\min}$ is taken to be zero; otherwise, it is taken to be nonzero. To test the significance of the canonical correlation $\hat{\rho}_{\min}$, one can also follow Supplement 18A and use the approximate χ^2 test given by Bartlett (1941), who shows that under the null hypothesis of zero canonical correlation, the statistic

$$\chi^2 = -(n - 1/2[m(p + 1) + q + 1]) \ln(1 - \hat{\rho}_{\min}^2) \tag{18.3.11}$$

has an approximate chi-square distribution with $[m(p + 1) - q + 1]$ degrees of freedom, i.e., $\chi^2(m(p + 1) - q + 1)$.

Once the state vector is identified, we have the canonical representation of a state space model

$$\begin{cases} \mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1} \\ \mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t, \end{cases} \tag{18.3.12}$$

where $\mathbf{a}_t$ is a Gaussian vector white noise series with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, i.e., $N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\mathbf{H} = [\mathbf{I}_m, \mathbf{0}]$ where $\mathbf{I}_m$ is the $m \times m$ identity matrix. Clearly, from the results of Section 18.2, estimates for the matrices $\mathbf{A}$, $\mathbf{G}$, and $\boldsymbol{\Sigma}$ can be obtained from the parameter estimates of the optimal AR model fitting. More natural estimates of the elements for the transition matrix A, however, are obtained from the canonical correlation analysis. For example, let k be the number of components of the final state vector $\mathbf{Y}_t$, hence, $\mathbf{A}$ is a $k \times k$ transition matrix. From (18.3.9), it is known that $m \leq k \leq m(p + 1)$. We now illustrate how the estimates in the first row of $\mathbf{A}$ are derived, which are related to the first step of the sequence of canonical correlation analyses when $Z_{1, n+1|n}$ is added to the vector $\mathbf{Z}_n$ to form the first subset $\mathcal{F}_n^1$ in deciding whether $Z_{1, n+1|n}$ should be included in the state vector. When the smallest canonical correlation between $\mathcal{F}_n^1$ and $\mathcal{D}_n$ is judged to be nonzero, $Z_{1, n+1|n}$ becomes the $(m + 1)$th component of the state vector. Thus, the first row of $\mathbf{A}$ will have 1 in the $(m + 1)$th column and 0 elsewhere. When the smallest canonical correlation is judged to be zero, a linear combination of $\mathcal{F}_n^1$ is uncorrelated with the data space $\mathcal{D}_n$, and $Z_{1, n+1|n}$ is excluded from the state vector. Because the determinant of $\hat{\Gamma}(0)$ is nonzero, we can take the coefficient of $Z_{1, n+1|n}$ in this linear combination to be unity. Thus, we have the

relationship $Z_{1,n+1|n} = \boldsymbol{\alpha}'\mathbf{Z}_n$; the coefficients of the vector α are used as estimates of the first m columns of the first row of the transition matrix $\mathbf{A}$, and the remaining $(k - m)$ columns of the first row are zero. The estimates in the other rows of $\mathbf{A}$ are obtained similarly.

Alternatively, once a state space model (18.3.12) is identified, one can use the maximum likelihood procedure to obtain more efficient estimates of $\mathbf{A}$, $\mathbf{G}$, and $\boldsymbol{\Sigma}$. For a given sequence of n observations $\mathbf{Z}_1, \mathbf{Z}_2, \ldots,$ and $\mathbf{Z}_n$, because

$$\mathbf{Y}_t = (\mathbf{I} - \mathbf{AB})^{-1}\mathbf{G}\mathbf{a}_t, \tag{18.3.13}$$

we have

$$\mathbf{Z}_t = \mathbf{H}(\mathbf{I} - \mathbf{AB})^{-1}\mathbf{G}\mathbf{a}_t \tag{18.3.14}$$

and

$$\mathbf{a}_t = [\mathbf{H}(\mathbf{I} - \mathbf{AB})^{-1}\mathbf{G}]^{-1}\mathbf{Z}_t. \tag{18.3.15}$$

Thus, the log-likelihood function, subject to additive constants, becomes

$$\ln L(\mathbf{A}, \mathbf{G}, \boldsymbol{\Sigma} \,|\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \propto -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{A}, \mathbf{G}), \tag{18.3.16}$$

where

$$\mathbf{S}(\mathbf{A}, \mathbf{G}) = \sum_{t=1}^{n}\mathbf{a}_t\mathbf{a}_t'. \tag{18.3.17}$$

The standard maximum likelihood estimation, as discussed in Section 16.6, can now be used to derive estimates of $\mathbf{A}$, $\mathbf{G}$, and $\boldsymbol{\Sigma}$. As pointed out in Section 16.6, the maximum likelihood estimation involves a highly nonlinear iterative procedure. The estimates obtained from the canonical correlations analysis may serve as initial estimates in this more efficient estimation procedure. In the process of estimation, one can set certain elements of $\mathbf{A}$ and $\mathbf{G}$ to some constants such as zero or one.

18.4 Empirical Examples

We now give two examples of the state space modeling. The first one is a univariate time series, and the second is a multivariate vector process. SAS/ETS of the SAS Institute (1999) is one of a few computer software that can be used to perform state space modeling. Its STATESPACE procedure is used in the analysis of both examples. We note, however, in Equation (18.3.11), the STATESPACE procedure uses a multiplicative

TABLE 18.1 AIC of AR models for the blowfly data.

p	0	1	2	3	4	5
AIC	386.22	326.07	327.41	329.27	331.13	333.07
p	6	7	8	9	10	
AIC	333.93	335.56	337.37	338.93	340.45	

constant, $-(n - 1/2[m(p + 1) - q + 1])$, instead of $-(n - 1/2[m(p + 1) + q + 1])$, although in a large sample, the difference in using these two different multiplicative constants will not affect the conclusion in most cases.

EXAMPLE 18.3 The univariate time series used in this illustration is the square root transformed blowfly data of 82 observations (Series W3), which we analyzed in Chapters 6 and 7.

The first step in the state space modeling is to find an optimal AR fitting of the data. From Table 18.1, the AIC is minimum at $p = 1$. Hence, the optimal AR order p is chosen to be 1, which corresponds to the same model we identified for the series in Chapter 6 using the sample ACF and PACF.

With $p = 1$, we apply the canonical correlation analysis between the data space $\{Z_n, Z_{n-1}\}$ and the predictor space $\{Z_n, Z_{n+1|n}\}$. The minimum canonical correlation coefficient equals .0891, and the information criterion equals

$$\begin{aligned} C &= -82 \ln[1 - (.0891)^2] - 2[1(1 + 1) - 2 + 1] \\ &= -1.35, \end{aligned} \tag{18.4.1}$$

which is negative. Hence, $Z_{n+1|n}$ is not included in the state vector. This result is also supported by Bartlett's chi-square test

$$\begin{aligned} \chi^2 &= -\left\{82 - \frac{1}{2}[1(1 + 1) + 2 + 1]\right\} \ln[1 - (.0891)^2] \\ &= .63, \end{aligned} \tag{18.4.2}$$

which is not significant as $\chi^2 = .63 < \chi^2_{.05}(1) = 3.04$. The above canonical correlation data analysis is summarized in Table 18.2.

TABLE 18.2 Canonical correlation analysis for blowfly data.

State vector	Correlations		Infor.(C)	Chi-square	DF
$Z_t, Z_{t+1\|t}$	1	.0891	−1.35	.63	1

Thus, the state vector consists only of Z_t and the canonical representation of the state space model for the data is given by

$$Z_{t+1} = AZ_t + a_{t+1}, \tag{18.4.3}$$

which is the AR(1) model we fitted in Chapter 7. The estimate of the transition matrix A is .73, which is a scalar and is equal to $\hat{\rho}_1$. Because the AR(1) is a Markov process, this result is certainly expected.

EXAMPLE 18.4 As a second illustration, consider the bivariate series of a leading indicator X_t and sales Y_t taken from Series M in Box, Jenkins, and Reinsel (1994). They fitted the data with the transfer function model

$$\begin{cases} x_t = (1 - .32B)\alpha_t \\ y_t = .035 + \dfrac{.482}{(1 - .72B)}x_{t-3} + (1 - .54B)a_t, \end{cases} \tag{18.4.4}$$

where $x_t = (1 - B)X_t$, $y_t = (1 - B)Y_t$, $\hat{\sigma}_\alpha^2 = .0676$, and $\hat{\sigma}_a^2 = .0484$. We now apply the state space procedure to construct a state space model for the series $\mathbf{Z}_t = [x_t, y_t]'$ for $t = 1, 2, \ldots, 149$.

Again, in constructing a state space model for the data, we first search for the optimal order in the AR fitting of the data. From Table 18.3 because AIC is minimum at $p = 5$, the optimal AR order is chosen to be 5.

Based on $p = 5$, we next apply the canonical correlation analysis between the data space $\{\mathbf{Z}_n, \mathbf{Z}_{n-1}, \ldots, \mathbf{Z}_{n-5}\}$ and the predictor space $\{\mathbf{Z}_n, \mathbf{Z}_{n+1|n}, \ldots, \mathbf{Z}_{n+5|n}\}$. As mentioned earlier, however, that $\mathbf{Z}_{n+1|n}$ through $\mathbf{Z}_{n+5|n}$ are potential elements of the state vector does not mean that all their components $x_{n+i|n}$, $y_{n+i|n}$, for $i = 1, \ldots, 5$ are needed in the final state vector. Thus, the canonical correlation analysis is performed between

$$\{x_{n-i}, y_{n-i} \mid i = 0, 1, 2, 3, 4, 5\}$$

and

$$\{x_{n+i|n}, y_{n+i|n} \mid i = 0, 1, 2, 3, 4, 5\}$$

TABLE 18.3 AIC of AR models for the differences of series M in Box, Jenkins, and Reinsel

p	0	1	2	3	4
AIC	−230.24	−271.62	−308.88	−608.97	−665.00
p	5	6	7	8	9
AIC	−673.53	−669.53	−667.06	−663.53	−661.81

TABLE 18.4 Canonical correlation analysis for the differences of Series M in Box, Jenkins, and Reinsel.

State vector	Correlation	Infor.(C)	Chi-square	DF
$x_n, y_n, x_{n+1\|n}$	1, 1, .22	−12.63	7.12	10
$x_n, y_n, y_{n+1\|n}$	1, 1, .97	410.55	416.10	10
$x_n, y_n, y_{n+1\|n}, y_{n+2\|n}$	1, 1, .98, .95	335.40	342.72	9
$x_n, y_n, y_{n+1\|n}, y_{n+2\|n}, y_{n+3\|n}$	1, 1, .98, .97, .20	−9.80	6.03	8

as shown in Table 18.4. Because the C value equals -12.63, which is negative when the component $x_{n+1|n}$ is considered, $x_{n+i|n}$ for $i \geq 1$ are excluded from the state vector. The C values are positive when the components $y_{n+1|n}$ and $y_{n+2|n}$ are considered but negative when $y_{n+3|n}$ is considered. Hence, the final state vector, from Table 18.4, is

$$\begin{bmatrix} x_t \\ y_t \\ y_{t+1|t} \\ y_{t+2|t} \end{bmatrix}. \tag{18.4.5}$$

The resulting state space model is

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ y_{t+2|t+1} \\ y_{t+3|t+1} \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_t \\ y_t \\ y_{t+1|t} \\ y_{t+2|t} \end{bmatrix} + \mathbf{G} \begin{bmatrix} a_{1,t+1} \\ a_{2,t+1} \end{bmatrix}, \tag{18.4.6}$$

where $[a_{1,t+1}, a_{2,t+1}]'$ is an $N(\mathbf{0}, \boldsymbol{\Sigma})$ vector white noise process.

The estimates of $\mathbf{A}$, $\mathbf{G}$, and $\boldsymbol{\Sigma}$ are given by

$$\hat{\mathbf{A}} = \begin{bmatrix} -.448 & .021 & 0 & 0 \\ (.073) & (.016) & & \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 4.654 & .031 & -.042 & .739 \\ (.126) & (.020) & (.029) & (.022) \end{bmatrix}, \hat{\mathbf{G}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -.100 & -.103 \\ (.097) & (.080) \\ .001 & .104 \\ (.097) & (.060) \end{bmatrix},$$

and

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} .079 & .0005 \\ .0005 & .108 \end{bmatrix},$$

where the value in parentheses below an estimate is the standard error of the estimate. Because the t-value of the (4, 1) element of the input matrix is much less than 1 (t = .001/.097 = .04), we set it to zero and reestimate the restricted model. The final fitted state space model becomes

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ y_{t+2|t+1} \\ y_{t+3|t+1} \end{bmatrix} = \begin{bmatrix} -.448 & .021 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 4.657 & .031 & -.042 & .739 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ y_{t+1|t} \\ y_{t+2|t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -.100 & -.103 \\ 0 & .104 \end{bmatrix} \begin{bmatrix} a_{1,t+1} \\ a_{2,t+1} \end{bmatrix}, \tag{18.4.7}$$

and

$$\hat{\Sigma} = \begin{bmatrix} .079 & .0005 \\ .0005 & .108 \end{bmatrix}.$$

Diagnostic tests indicate that the residuals are not different from white noise series, and we conclude that the model is adequate. It would be interesting to compare the state space model (18.4.7) and the transfer function model (18.4.4) in terms of variable relationships and forecasting. We leave it as an exercise.

18.5 The Kalman Filter and Its Applications

Consider a general state space model of a stochastic system given by (18.1.3) and (18.1.4), i.e.,

$$\begin{cases} \mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{G}\mathbf{a}_{t+1} \\ \mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t + \mathbf{b}_t, \end{cases} \tag{18.5.1}$$

where $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{H}$ are fixed matrices of order $k \times k$, $k \times n$, and $m \times k$, respectively. The vector $\mathbf{a}_t$ is the system noise, which is an i.i.d. Gaussian white noise $N(\mathbf{0}, \boldsymbol{\Sigma})$ series, and $\mathbf{b}_t$ is the measurement noise, which is an i.i.d. Gaussian white noise $N(\mathbf{0}, \boldsymbol{\Omega})$ series. Furthermore, it is assumed that the two noise series $\mathbf{a}_t$ and $\mathbf{b}_t$ are independent of each other. More precisely, $\mathbf{a}_t$ and $\mathbf{b}_t$ are assumed to have a joint multivariate normal distribution with mean $\mathbf{0}$ and a block diagonal covariance matrix, denoted by

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \right). \tag{18.5.2}$$

Once a state space model is constructed, the l-step ahead forecasts from the forecast origin time t can be calculated as follows:

$$\begin{aligned}\hat{\mathbf{Y}}_t(l) &= E(\mathbf{Y}_{t+l} \mid \mathbf{Y}_j, j \le t) \\ &= \mathbf{A}\hat{\mathbf{Y}}_t(l-1) \\ &= \mathbf{A} \cdot \mathbf{A}\hat{\mathbf{Y}}_t(l-2) \\ &\vdots \\ &= \mathbf{A}^l \hat{\mathbf{Y}}_t. \end{aligned} \tag{18.5.3}$$

Hence,

$$\begin{aligned}\hat{\mathbf{Z}}_t(l) &= E(\mathbf{Z}_{t+l} \mid \mathbf{Z}_j, j \le t) \\ &= \mathbf{H}\hat{\mathbf{Y}}_t(l) \\ &= \mathbf{H}\mathbf{A}^l \hat{Y}_t, \end{aligned} \tag{18.5.4}$$

where

$$\hat{\mathbf{Y}}_t = E(\mathbf{Y}_t \mid \mathbf{Y}_j, j \le t) = \mathbf{Y}_t.$$

Clearly, from (18.5.4), the accuracy of the forecasts $\hat{\mathbf{Z}}_t(l)$ depends on the quality of the estimate $\hat{\mathbf{Y}}_t$ of the state vector $\mathbf{Y}_t$, which summarizes the information from the past that is needed to forecast the future. To improve forecasts, when a new observation becomes available, it should be used to update the state vector and hence to update the forecasts. In this section, we introduce the Kalman filter, which is a recursive procedure used to make inferences about the state vector $\mathbf{Y}_t$.

To specify the distribution of the state vectors $\mathbf{Y}_t$ for $t = 1, 2, \ldots$, in (18.5.1), we assume that the distribution $p(\mathbf{Y}_0)$ of the state vector $\mathbf{Y}_0$ at time zero has mean $\hat{\mathbf{Y}}_0$ and covariance V_0, i.e.,

$$\mathbf{Y}_0 \sim N(\hat{\mathbf{Y}}_0, \mathbf{V}_0). \tag{18.5.5}$$

This equation, together with the system equation in (18.5.1), determines the distribution of the state vectors $p(\mathbf{Y}_t)$, for $t = 1, 2, \ldots$. These distributions present our belief prior to having observed the data $\mathbf{Z}_t^*$ and are referred to as the prior distributions. After observing the data, we revise our prior distribution and obtain the posterior distribution of the state vector, which is the conditional distribution $p(\mathbf{Y}_t \mid \mathbf{Z}_t^*)$, where

$$\mathbf{Z}_t^* = (\mathbf{Z}_j | 1 \le j \le t). \tag{18.5.6}$$

By Bayes's theorem,

$$p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_{t+1}^*) \propto p(\mathbf{Z}_{t+1} \mid \mathbf{Y}_{t+1}, \mathbf{Z}_t^*) p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_t^*). \tag{18.5.7}$$

Hence, a posterior distribution $p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_t^*)$ at an earlier time becomes a prior distribution for calculating a new posterior distribution $p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_{t+1}^*)$ when a new observation $\mathbf{Z}_{t+1}$ becomes available.

Now, suppose that the posterior distribution, $p(\mathbf{Y}_t \mid \mathbf{Z}_t^*)$, of the state vector $\mathbf{Y}_t$ at time t is a normal distribution with a mean $\hat{\mathbf{Y}}_t$ and a covariance matrix $\mathbf{V}_t$, i.e.,

$$(\mathbf{Y}_t \mid \mathbf{Z}_t^*) \sim N(\hat{\mathbf{Y}}_t, \mathbf{V}_t). \tag{18.5.8}$$

At time $(t + 1)$, when a new observation $\mathbf{Z}_{t+1}$ becomes available, we would like to update the state vector and derive the new posterior distribution $p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_{t+1}^*)$. Because

$$\begin{aligned} \mathbf{e}_{t+1} &= \mathbf{Z}_{t+1} - \hat{\mathbf{Z}}_t(1) \\ &= \mathbf{Z}_{t+1} - \mathbf{H}\mathbf{A}\hat{\mathbf{Y}}_t \\ &= \mathbf{H}\mathbf{Y}_{t+1} + \mathbf{b}_{t+1} - \mathbf{H}\mathbf{A}\hat{\mathbf{Y}}_t \\ &= \mathbf{H}(\mathbf{Y}_{t+1} - \mathbf{A}\hat{\mathbf{Y}}_t) + \mathbf{b}_{t+1}, \end{aligned} \tag{18.5.9}$$

knowing $\mathbf{Z}_{t+1}$ is equivalent to knowing the one-step ahead forecast error $\mathbf{e}_{t+1}$. Thus, to derive the posterior distribution $p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_{t+1}^*) = p(\mathbf{Y}_{t+1} \mid \mathbf{Z}_{t+1}, \mathbf{Z}_t^*)$, we need only to find the posterior distribution $p(\mathbf{Y}_{t+1} \mid \mathbf{e}_{t+1}, \mathbf{Z}_t^*)$ of $(\mathbf{Y}_{t+1} \mid \mathbf{e}_{t+1}, \mathbf{Z}_t^*)$. The conditional distribution of $(\mathbf{Y}_{t+1} \mid \mathbf{e}_{t+1}, \mathbf{Z}_t^*)$, however, follows immediately if we can find the joint normal distribution of $[(\mathbf{Y}_{t+1}, \mathbf{e}_{t+1})' \mid \mathbf{Z}_t^*]$.

Note that at time $(t + 1)$, having observed $\mathbf{Z}_t^*$ but prior to observing $\mathbf{Z}_{t+1}$, we have, via the system equation in (18.5.1), the prior distribution

$$(\mathbf{Y}_{t+1} \mid \mathbf{Z}_t^*) \sim N(\mathbf{A}\hat{\mathbf{Y}}_t, \mathbf{R}_{t+1}), \tag{18.5.10}$$

where

$$\mathbf{R}_{t+1} = \mathbf{A}\mathbf{V}_t\mathbf{A}' + \mathbf{G}\Sigma\mathbf{G}'. \tag{18.5.11}$$

After observing $\mathbf{Z}_{t+1}$, from (18.5.9), we have the posterior or conditional distribution

$$(\mathbf{e}_{t+1} \mid \mathbf{Y}_{t+1}, \mathbf{Z}_t^*) \sim N(\mathbf{H}(\mathbf{Y}_{t+1} - \mathbf{A}\hat{\mathbf{Y}}_t), \Omega). \tag{18.5.12}$$

Recall that for any two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$,

$$P(\mathbf{X}_1, \mathbf{X}_2) = p(\mathbf{X}_2)P(\mathbf{X}_1 | \mathbf{X}_2).$$

In terms of a joint multivariate normal distribution, we have

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{18.5.13}$$

if and only if

$$\begin{cases} \mathbf{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) & (18.5.14) \\ \mathbf{X}_1 | \mathbf{X}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}). & (18.5.15) \end{cases}$$

Now, let $\mathbf{X}_2$ correspond to $\mathbf{Y}_{t+1}$ and $\mathbf{X}_1$ correspond to $\mathbf{e}_{t+1}$. Then, from (18.5.10) and (18.5.14),

$$\begin{cases} \boldsymbol{\mu}_2 = \mathbf{A}\hat{\mathbf{Y}}_t \\ \boldsymbol{\Sigma}_{22} = \mathbf{R}_{t+1}, \end{cases} \tag{18.5.16}$$

and by (18.5.12), (18.5.15), and (18.5.16), we have

$$\begin{aligned} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\mathbf{R}_{t+1}^{-1}(\mathbf{Y}_{t+1} - \mathbf{A}\hat{\mathbf{Y}}_t) \\ &= \mathbf{H}(\mathbf{Y}_{t+1} - \mathbf{A}\hat{\mathbf{Y}}_t). \end{aligned}$$

Hence,

$$\begin{cases} \boldsymbol{\mu}_1 = \mathbf{0} \\ \boldsymbol{\Sigma}_{12} = \mathbf{H}\mathbf{R}_{t+1}. \end{cases} \tag{18.5.17}$$

Also, by the same equations,

$$\begin{aligned} \boldsymbol{\Omega} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ &= \boldsymbol{\Sigma}_{11} - \mathbf{H}\mathbf{R}_{t+1}\mathbf{R}_{t+1}^{-1}\mathbf{R}_{t+1}\mathbf{H}', \end{aligned}$$

we get

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega} + \mathbf{H}\mathbf{R}_{t+1}\mathbf{H}'. \tag{18.5.18}$$

Thus, by (18.5.13),

$$\left(\begin{bmatrix}\mathbf{Y}_{t+1}\\ \mathbf{e}_{t+1}\end{bmatrix}\middle|\mathbf{Z}_t\right) \sim N\left(\begin{bmatrix}\mathbf{A}\hat{\mathbf{Y}}_t\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{R}_{t+1} & \mathbf{R}_{t+1}\mathbf{H}'\\ \mathbf{H}\mathbf{R}_{t+1} & \boldsymbol{\Omega}+\mathbf{H}\mathbf{R}_{t+1}\mathbf{H}'\end{bmatrix}\right). \tag{18.5.19}$$

It follows, from (18.5.15), that

$$(\mathbf{Y}_{t+1}|\mathbf{e}_{t+1}, \mathbf{Z}_t) \sim N(\hat{\mathbf{Y}}_{t+1}, \mathbf{V}_{t+1}), \tag{18.5.20}$$

where

$$\begin{aligned}\hat{\mathbf{Y}}_{t+1} &= \mathbf{A}\hat{\mathbf{Y}}_t + \mathbf{R}_{t+1}\mathbf{H}'(\boldsymbol{\Omega}+\mathbf{H}\mathbf{R}_{t+1}\mathbf{H}')^{-1}\mathbf{e}_{t+1}\\ &= \mathbf{A}\hat{\mathbf{Y}}_t + \mathbf{K}_{t+1}(\mathbf{Z}_{t+1} - \hat{\mathbf{Z}}_t(1)),\end{aligned} \tag{18.5.21}$$

$$\begin{aligned}\mathbf{V}_{t+1} &= \mathbf{R}_{t+1} - \mathbf{R}_{t+1}\mathbf{H}'(\boldsymbol{\Omega}+\mathbf{H}\mathbf{R}_{t+1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{R}_{t+1}\\ &= \mathbf{R}_{t+1} - \mathbf{K}_{t+1}\mathbf{H}\mathbf{R}_{t+1}\\ &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H})(\mathbf{A}\mathbf{V}_t\mathbf{A}'+\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}'),\end{aligned} \tag{18.5.22}$$

with

$$\begin{aligned}\mathbf{K}_{t+1} &= \mathbf{R}_{t+1}\mathbf{H}'(\boldsymbol{\Omega} + \mathbf{H}\mathbf{R}_{t+1}\mathbf{H}')^{-1},\\ \mathbf{R}_{t+1} &= \mathbf{A}\mathbf{V}_t\mathbf{A}' + \mathbf{G}\boldsymbol{\Sigma}\mathbf{G}'.\end{aligned} \tag{18.5.23}$$

Equations (18.5.21) and (18.5.22) are the basic recursive formula used to update the mean and the covariance matrix and hence the distribution of the state vector, $\mathbf{Y}_{t+1}$, after the new observation, $\mathbf{Z}_{t+1}$, has become available. The updated estimates $\hat{\mathbf{Y}}_{t+1}$ of the state is, from (18.5.21), the sum of the projected estimate using observations up to time t, and the one-step ahead forecast error $\mathbf{e}_{t+1} = \mathbf{Z}_{t+1} - \hat{\mathbf{Z}}_t(1)$. The matrix $\mathbf{K}_{t+1}$ is called the Kalman gain, which determines the weight given to the forecast error. In fact, it can be easily seen from (18.5.21) that the Kalman gain is simply the coefficient of the least squares regression of $\mathbf{Y}_{t+1}$ on the forecast error $\mathbf{e}_{t+1}$ conditional on $\mathbf{Z}_t$. Thus, Kalman filtering is a recursive updating procedure that consists of forming a preliminary estimate of the state and then revising the estimate by adding a correction to this preliminary estimate. The magnitude of the correction is determined by how well the preliminary estimate predicted the new observation. The Kalman filter will tend toward a steady state when the Kalman gain approaches a limit.

Clearly, for the given state space model (18.5.1), the recursive equations to compute $\hat{\mathbf{Y}}_{t+1}$ in (18.5.21) and $\mathbf{V}_{t+1}$ in (18.5.22) require initial values $\hat{\mathbf{Y}}_0$ and $\mathbf{V}_0$ specified in (18.5.5). These initial values can be the values from the previous studies. Otherwise, some guess values from experience can be used. In that case, following a principle of uninformative prior, the $\mathbf{V}_0$ is often chosen as a diagonal matrix with large elements. Thus, for a large n, the results will be dominated by the information from the data.

The methods of Kalman filtering were originally developed by Kalman (1960) within the context of linear systems. Because of the ease of implementation of the algorithm on digital computers, however, these methods have now become well known and widely used in many areas of applications. For example, consider the simple state space model

$$\begin{cases} \mathbf{Z}_t = \mathbf{H}\mathbf{Y}_t + \mathbf{b}_t \\ \mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{a}_t, \end{cases} \tag{18.5.24}$$

where

$$\begin{bmatrix} \mathbf{b}_t \\ \mathbf{a}_t \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Omega & \mathbf{0} \\ \mathbf{0} & \Sigma \end{bmatrix} \right). \tag{18.5.25}$$

The commonly used multiple regression model, $\mathbf{Z}_t = \mathbf{X}\boldsymbol{\beta} + \mathbf{b}_t$, can be easily seen as the state space model with $\mathbf{H} = \mathbf{X}$ (a $l \times m$ vector of independent variables), $\mathbf{A} = \mathbf{I}$, $\mathbf{Y}_t = \boldsymbol{\beta}_t$ (an $m \times 1$ vector of m coefficents), and $\Sigma = \mathbf{0}$. The time-varying regression model

$$\begin{cases} \mathbf{Z}_t = \mathbf{X}\boldsymbol{\beta}_t + \mathbf{b}_t \\ \boldsymbol{\beta}_t = \mathbf{A}\boldsymbol{\beta}_{t-1} + \mathbf{a}_t \end{cases} \tag{18.5.26}$$

is the above state space model with $\mathbf{H} = \mathbf{X}$, $\mathbf{Y}_t = \boldsymbol{\beta}_t$, and $\Sigma \neq \mathbf{0}$. The advantage of this formulation is that one can estimate the time-varying coefficients, $\boldsymbol{\beta}_t$, using Kalman filtering.

Other applications include time series forecasting (Mehra [1979]), Bayesian forecasting (Harrison and Stevens [1976]), the analysis of time series models with time-varying coefficients (Young [1974]; Bohlin [1976]), regression analysis (O'Hagan [1978]), missing observations (Kohn and Ansley [1983]), quality control (Phadke [1981]), temporal aggregation (Schmidt and Gamerman [1997]), and data disaggregation (Gudmundsson [1999]).

The recursive updating equations can be derived using different approaches. For example, Priestley and Subba Rao (1975) derive the Kalman filter algorithm through the statistical technique of factor analysis. An expository article is given by Meinhold and Singpurwalla (1983). For further references, see Kalman (1960), Kalman and Bucy (1961), Jazwinski (1970), and others.

Supplement 18.A Canonical Correlations

Let $\mathbf{X}$ be a $q \times 1$ vector and $\mathbf{Y}$ be an r $\times$ 1 vector with the variance-covariance matrix Σ given by

$$\Sigma = \mathrm{Var}\left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \right) = \begin{bmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{Y}} \end{bmatrix}. \tag{18.A.1}$$

With no loss of generality, assume that $q \leq r$. In many applications, we are interested in finding the correlation between the linear combination of the variables in $\mathbf{X}$ and the linear

combination of variables in $\mathbf{Y}$, i.e., the correlation between $\boldsymbol{\alpha}'\mathbf{X}$ and $\boldsymbol{\beta}'\mathbf{Y}$ for some nonzero $q \times 1$ vector $\boldsymbol{\alpha}$ and $r \times 1$ vector $\boldsymbol{\beta}$.

Let $\mathbf{U}_1 = \boldsymbol{\alpha}_1'\mathbf{X}$ and $\mathbf{V}_1 = \boldsymbol{\beta}_1'\mathbf{Y}$. The first canonical correlation, $\rho(1)$, is defined to be the maximum correlation between $\mathbf{U}_1$ and $\mathbf{V}_1$ over all possible $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$, i.e.,

$$\rho(1) = \text{corr}(\mathbf{U}_1, \mathbf{V}_1) = \max_{\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1} \{\text{corr}\,(\boldsymbol{\alpha}_1'\mathbf{X}, \boldsymbol{\beta}_1'\mathbf{Y})\} \tag{18.A.2}$$

such that $\text{Var}\,(\boldsymbol{\alpha}_1'\mathbf{X}) = 1$ and $\text{Var}\,(\boldsymbol{\beta}_1'\mathbf{Y}) = 1$. Thus, we need to maximize

$$F = \max_{\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1} \{\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{XY}}\,\boldsymbol{\beta}_1\} \tag{18.A.3}$$

such that $\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\beta}_1 = 1$. Using Lagrange multipliers, we maximize the function:

$$L = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\beta}_1 - \phi_1(\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\alpha}_1 - 1) - \phi_2(\boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\beta}_1 - 1). \tag{18.A.4}$$

Taking partial derivatives with respect to $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$, ϕ_1, and ϕ_2 and setting them to 0, we get

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\alpha}_1} &= \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\beta}_1 - 2\phi_1\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\alpha}_1 = \mathbf{0}, \\
\frac{\partial L}{\partial \boldsymbol{\beta}_1} &= \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\alpha}_1 - 2\phi_2\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\beta}_1 = \mathbf{0}, \\
\frac{\partial L}{\partial \phi_1} &= \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\alpha}_1 - 1 = 0, \\
\frac{\partial L}{\partial \phi_2} &= \boldsymbol{\beta}_1'\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\beta}_1 - 1 = 0.
\end{aligned} \tag{18.A.5}$$

Multiplying the first equation in (18.A.5) by $\boldsymbol{\alpha}_1'$, the second equation by $\boldsymbol{\beta}_1'$, and using the constraints in the last two equations, the system to solve for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ in (18.A.5) becomes

$$\begin{bmatrix} -\rho\boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{XY}} \\ \boldsymbol{\Sigma}_{\mathbf{YX}} & -\rho\boldsymbol{\Sigma}_{\mathbf{Y}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\beta}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{18.A.6}$$

and $\rho = 2\phi_1 = 2\phi_2 = \text{Corr}(\mathbf{U}_1, \mathbf{V}_1) = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\beta}_1$. For $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ not identically zero, the implication is that

$$\begin{vmatrix} -\rho\boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{XY}} \\ \boldsymbol{\Sigma}_{\mathbf{YX}} & -\rho\boldsymbol{\Sigma}_{\mathbf{Y}} \end{vmatrix} = 0. \tag{18.A.7}$$

Note that if $|\Sigma_Y| \neq 0$, from the results of partitioned matrices, we have

$$\begin{vmatrix} -\rho\Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & -\rho\Sigma_Y \end{vmatrix} = |\Sigma_Y||\rho^2\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}| = 0,$$

and therefore the equivalent implication

$$|\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} - \rho^2\Sigma_X| = 0. \tag{18.A.8}$$

Clearly, there are q solutions for ρ^2 in (18.A.8); we choose $\rho(1)$ to be the positive square root of the maximum solution. The corresponding vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are obtained through (18.A.6). The resulting $\mathbf{U}_1 = \boldsymbol{\alpha}_1'\mathbf{X}$ and $\mathbf{V}_1 = \boldsymbol{\beta}_1'\mathbf{Y}$ are called the first canonical variates.

Let $\mathbf{U}_i = \boldsymbol{\alpha}_i'\mathbf{X}$, and $\mathbf{V}_i = \boldsymbol{\beta}_i'\mathbf{Y}$. The ith canonical correlation, $\rho(i)$, is defined to be the maximum correlation between $\mathbf{U}_i$ and $\mathbf{V}_i$ so that $\text{Var}(\boldsymbol{\alpha}_i'\mathbf{X}) = 1$ and $\text{Var}(\boldsymbol{\beta}_i'\mathbf{Y}) = 1$, and $\mathbf{U}_i$ and $\mathbf{V}_i$ are uncorrelated with all previous $(i - 1)$ canonical variates. If we then let the q solutions of (18.A.8) be ordered so that $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_q^2$, it can be shown that $\rho(i)$ is the positive square root of the ith largest solution, ρ_i^2.
Let

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_X & \hat{\Sigma}_{XY} \\ \hat{\Sigma}_{YX} & \hat{\Sigma}_Y \end{bmatrix} \tag{18.A.9}$$

be the sample variance-covariance matrix computed from a sample of size n, i.e.,

$$\hat{\Sigma} = \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{y}_j - \bar{\mathbf{y}})' & \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{y}_j - \bar{\mathbf{y}})' \\ \frac{1}{n}\sum_{j=1}^{n}(\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{x}_j - \bar{\mathbf{x}})' & \frac{1}{n}\sum_{j=1}^{n}(\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})' \end{bmatrix},$$

where $\mathbf{x}_j$ and $\mathbf{y}_j$ are q $\times$ 1 and r $\times$ 1 vector observations of $\mathbf{X}$ and $\mathbf{Y}$.

The sample canonical correlations are $\hat{\rho}(i)$, $i = 1, \ldots, q$, where $\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq \ldots \geq \hat{\rho}_q^2$, are the solutions of

$$|\hat{\Sigma}_{XY}\hat{\Sigma}_Y^{-1}\hat{\Sigma}_{YX} - \hat{\rho}^2\hat{\Sigma}_X| = 0. \tag{18.A.10}$$

Assume that $\mathbf{X}$ and $\mathbf{Y}$ follow a joint multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. To test that $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated, i.e., $\boldsymbol{\Sigma}_{XY} \neq \mathbf{0}$, and hence they are independent, it is equivalent to

test the null hypothesis $H_1 : (\rho(1), \ldots, \rho(q))$ are jointly zero against the alternative that they are not jointly zero. It can be shown that the likelihood ratio statistic is

$$\Lambda_1 = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_X||\hat{\Sigma}_Y|}. \tag{18.A.11}$$

From (18.A.9), when $|\hat{\Sigma}_X| \neq 0$, and $|\hat{\Sigma}_Y| \neq 0$, we have

$$\begin{aligned} |\hat{\Sigma}| &= |\hat{\Sigma}_Y||\hat{\Sigma}_X - \hat{\Sigma}_{XY}\hat{\Sigma}_Y^{-1}\Sigma_{YX}| \\ &= |\hat{\Sigma}_Y||\hat{\Sigma}_X||\mathbf{I} - \hat{\Sigma}_X^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_Y^{-1}\Sigma_{YX}|, \end{aligned}$$

hence,

$$\Lambda_1 = |\mathbf{I} - \hat{\Sigma}_X^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_Y^{-1}\hat{\Sigma}_{YX}|.$$

It is clear from (18.A.10) that $\hat{\rho}^2$ is a characteristic root of $\hat{\Sigma}_X^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_Y^{-1}\hat{\Sigma}_{YX}$. Thus, it follows that

$$\Lambda_1 = \prod_{j=1}^{q}(1 - \hat{\rho}_j^2). \tag{18.A.12}$$

Under the null hypothesis, from the large sample result of Wilks (1938), we have

$$-n \ln \Lambda_1 \xrightarrow{d} \chi^2(rq). \tag{18.A.13}$$

To improve the χ^2 approximation, Bartlett (1941) suggests replacing the multiplicative factor n by the factor $(n - (r + q + 1)/2)$, i.e.,

$$-\left[n - \frac{1}{2}(r + q + 1)\right] \ln \Lambda_1 \xrightarrow{d} \chi^2(rq). \tag{18.A.14}$$

If H_1 is rejected, we next test that $H_2 : (\rho(2), \ldots, \rho(q))$ are jointly zero. We use the test statistic

$$\Lambda_2 = \prod_{j=2}^{q}(1 - \hat{\rho}_j^2). \tag{18.A.15}$$

Under the null hypothesis H_2, we have

$$-\left[n - \frac{1}{2}(r + q + 1)\right] \ln \Lambda_2 \xrightarrow{d} \chi^2((r - 1)(q - 1)). \tag{18.A.16}$$

If H_1 is rejected and H_2 is not rejected, then $\rho(1)$ is the only significant canonical correlation. If H_2 is rejected, however, we can continue the process. At the kth step, we test that H_k: ($\rho(k), \ldots,$ $\rho(q)$) are jointly zero, using the test statistic

$$\Lambda_k = \prod_{j=k}^{q} (1 - \hat{\rho}_j^2). \tag{18.A.17}$$

Under the null hypothesis we have

$$-\left[n - \frac{1}{2}(r + q + 1)\right] \ln \Lambda_k \xrightarrow{d} \chi^2((r - k + 1)(q - k + 1)). \tag{18.A.18}$$

In the canonical correlation analysis of state space modeling, we have $r = (p + 1)m$ and $k = q$.

Canonical correlation analysis is used in testing the independence of two sets of variables. It is an important data-reduction method discussed in most books on multivariate analysis. See, for example, Johnson and Wichern (1998).

EXERCISES

18.1 Find the state space representation for the following models:

(a) $Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$.

(b) $(1 - \phi_1 B - \phi_2 B^2)Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$.

(c) $(1 - \phi_1 B)(1 - B)Z_t = (1 - \theta_1 B)a_t$.

(d) $Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$.

(e) $$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

18.2 (a) Find the state space representation for the model given in Exercise 16.1, part (c).

(b) Is the representation obtained in part (a) unique? Why?

18.3 Consider the so-called autoregressive signal plus noise model in which we observe

$$Z_t = Y_t + b_t$$

and the unobserved process Y_t follows an autoregressive process, say, an AR(1) process

$$Y_t = \phi Y_{t-1} + a_t,$$

where a_t and b_t are independent zero mean Gaussian white noise processes with variances σ_a^2 and σ_b^2, respectively.

(a) Express the model in the state space form.

(b) Find the steady state of the Kalman filter.

18.4 In Exercise 18.3, if $\phi = 1$, the autoregressive signal plus noise model becomes the random walk plus noise model

$$Z_t = Y_t + b_t,$$

where the unobservable process Y_t follows a random walk given by

$$Y_t = Y_{t-1} + a_t.$$

(a) Express the model in the state space form.

(b) Show that the random walk plus noise model is in fact an *ARIMA*(0, 1, 1) model.

18.5 Compare the forecasts obtained from the transfer function model (18.4.4) and the state space model (18.4.7).

18.6 Consider the corn supply data in series W13.

(a) Find a proper univariate ARMA model for the series.

(b) Find a state space model for the series.

(c) Compare your results in parts (a) and (b).

18.7 Use canonical correlation analysis to construct a state space model for the Lydia Pinkham data Series W12.

19

Long Memory and Nonlinear Processes

In earlier chapters, we discussed various methods for analyzing both univariate and multivariate time series. All time series variables (differencing if necessary) discussed so far, however, are short memory processes with autocorrelation functions that decrease fairly fast, and all time series models presented there are linear. To fit these linear models, we sometimes apply transformations so that the transformed series can fit these models well. Many of these commonly used transformations such as the logarithmic transformation and square root transformation are nonlinear. The transformed series following a linear model clearly indicates that the original series may follow a nonlinear model. In fact, many time series exhibit characteristics such as sudden bursts of large amplitude at irregular time epochs that are commonly found in seismological data as well as some financial series, and these linear models cannot properly describe them. In this chapter, we introduce long memory processes and some nonlinear time series models that are in describing these long memory and nonlinear phenomena.

19.1 Long Memory Processes and Fractional Differencing

19.1.1 Fractionally Integrated ARMA Models and Their ACF

The ARIMA(p, d, q) process plays a dominant role in all chapters we have studied. In this process, we have assumed that the value of d is an integer. In other words, the original Z_t process is nonstationary but its dth differences for some integer d is stationary. Clearly, for a process to be stationary, the value of d has to be less than 1. We would, though, like to know more exactly the range of d for the process

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t, \tag{19.1.1}$$

where $\phi_p(B)\theta_q(B) \neq 0$ for $|B| \leq 1$ and a_t is a white noise process with mean zero and constant variance σ_a^2, to be stationary and invertible. With no loss of generality, to answer the question, we consider the simple case

$$(1 - B)^d Z_t = a_t. \tag{19.1.2}$$

If the process in (19.1.2) is stationary, then we should be able to write it as

$$Z_t = (1 - B)^{-d} a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \tag{19.1.3}$$

i.e.,

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \psi_j B^j$$

such that $\{\psi_j\}$ is square summable. From Taylor series expansion, we have the general binomial formula

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \binom{-d}{j} (-B)^j = \sum_{j=0}^{\infty} \psi_j B^j, \tag{19.1.4}$$

where

$$\begin{aligned} \psi_j &= (-1)^j \binom{-d}{j} = (-1)^j \frac{(-d)(-d-1)\cdots(-d-j+1)}{j!} \\ &= \frac{(j+d-1)(j+d-2)\cdots(j+d-j)}{j!} = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \end{aligned}$$

and $\Gamma(\cdot)$ is the gamma function. Using Stirling's formula

$$\Gamma(x) \approx \sqrt{2\pi} e^{-x} x^{x-1/2} \qquad \text{as } x \longrightarrow \infty,$$

we have

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \approx \frac{1}{\Gamma(d)} \frac{\sqrt{2\pi} e^{-j-d}(j+d)^{j+d-1/2}}{\sqrt{2\pi} e^{-j-1}(j+1)^{j+1-1/2}} \approx \frac{1}{\Gamma(d) j^{1-d}}.$$

Clearly, $\{\psi_j\}$ is square summable if and only if $2(1 - d) > 1$, i.e., $d < .5$. Similarly, we see that the process is invertible if and only if $-.5 < d$. Thus, the process in (19.1.2) or more generally the process in (19.1.1) is stationary and invertible if and only if $-.5 < d < .5$. The process in (19.1.1) with $-.5 < d < .5$ is therefore called the autoregressive fractionally integrated moving average model, often denoted the ARFIMA(p, d, q) model. The process in (19.1.2) with $-.5 < d < .5$ is called the fractionally integrated (or differenced) noise.

Clearly, the spectrum of the fractionally differenced noise exists and, from (11.2.5) and (12.2.8b), it equals

$$f(\omega) = \frac{\sigma_a^2}{2\pi}|1 - e^{-i\omega}|^{-2d} = \frac{\sigma_a^2}{2\pi}\left|\frac{2ie^{-i\omega/2}(e^{i\omega/2} - e^{-i\omega/2})}{2i}\right|^{-2d}$$
$$= \frac{\sigma_a^2}{2\pi}\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d}, \qquad -\pi \le \omega \le \pi, \tag{19.1.5}$$

although it diverges at the zero frequency. Now, from (12.1.4),

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\omega} f(\omega)\, d\omega = \frac{\sigma_a^2}{2\pi}\int_0^{\pi} 2\cos(k\omega)\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d} d\omega$$
$$= \frac{(2)^{-2d}\sigma_a^2}{1}\frac{\cos(2k\pi/2)\Gamma(2-2d)(2)^{1+2d-1}}{(1-2d)\Gamma[(1-2d+2k+1)/2]\Gamma[(1-2d-2k+1)/2]}$$
$$= (-1)^k\sigma_a^2\frac{\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-d-k)}, \tag{19.1.6}$$

where we use the following integral result (see Gradshteyn and Ryzhik ([1965])

$$\int_0^{\pi}\cos(k\omega)[\sin(\omega)]^{\alpha-1}d\omega = \frac{\pi\cos(k\pi/2)\Gamma(\alpha+1)2^{1-\alpha}}{\alpha\Gamma[(\alpha+k+1)/2]\Gamma[(\alpha-k+1)/2]}.$$

Thus,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(-1)^k\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-d-k)}\frac{\Gamma(1-d)\Gamma(1-d)}{\Gamma(1-2d)}$$
$$= \frac{\Gamma(1-d)}{\Gamma(k-d+1)}\frac{(-1)^k(1-d-1)(1-d-2)\cdots(1-d-k)\Gamma(1-d-k)}{\Gamma(1-d-k)}$$
$$= \frac{\Gamma(1-d)}{\Gamma(k-d+1)}(-1)^k(-1)^k(d)(d-1)\cdots(d+k-1)$$
$$= \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(k-d+1)\Gamma(d)}. \tag{19.1.7}$$

Using Stirling's formula again, we have

$$\rho_k \sim \frac{\Gamma(1-d)}{\Gamma(d)}\frac{\sqrt{2\pi}e^{-k-d}(k+d)^{k+d-1/2}}{\sqrt{2\pi}e^{-k+d-1}(k-d+1)^{k-d+1-1/2}}$$
$$\sim \frac{\Gamma(1-d)}{\Gamma(d)}e^{-2d+1}(k)^{2d-1}$$
$$\sim c(k)^{2d-1}, \tag{19.1.8}$$

where c is a constant with respect to k.

A stationary process is said to be a short memory process if its ACF is geometrically bounded, i.e., if

$$|\rho_k| \leq c(r)^k, \qquad k = 1, 2, \ldots, \tag{19.1.9}$$

where c is a positive constant and $0 < r < 1$. It is known that the stationary ARMA models that we studied in earlier chapters are short memory processes. On the other hand, a stationary process is said to be a long memory process if its ACF follows an asymptotic hyperbolic decay, i.e., if

$$|\rho_k| \sim c(k)^\alpha, \qquad \text{as } k \to \infty, \tag{19.1.10}$$

where $\alpha < 0$. A long memory process with an ACF which is not absolutely summable, i.e. $\sum_{k=-\infty}^{\infty} |\rho_k| = \infty$, is further said to be persistent.

Because d is known to be less than .5, it is clear that the ACF given in (19.1.8) follows an asymptotic hyperbolic decay. Hence, the fractionally differenced noise process and thus the ARFIMA(p, d, q) process is a long memory process. Furthermore, for the ACF, ρ_k, in (19.1.8), $\sum_{k=-\infty}^{\infty} |\rho_k|$ converges only if $(2d - 1) < -1$, or $d < 0$. The fractionally differenced noise process and thus the ARFIMA(p, d, q) process is in fact persistent when $0 < d < .5$.

19.1.2 Practical Implications of the ARFIMA Processes

An ARFIMA process contains several characteristics that, in finite samples, are similar to those of a nonstationary process. For example, the ACF of a persistent ARFIMA(p, d, q) model decays very slowly, a phenomenon similar to the sample ACF of a nonstationary time series discussed in Chapter 4. Also, realizations of both persistent ARFIMA models and nonstationary ARIMA models have periodograms that diverge at the zero frequency. These similarities often lead to model misspecification. For instance, a stationary ARFIMA model could be misspecified as a nonstationary ARIMA model. The consequence of this overdifferencing has some undesirable effects on parameter estimation and forecasting.

In terms of forecasting, we showed in Chapter 5 that the forecast based on a stationary process converges to the mean value of the process. Thus, the forecast of a long memory process should converge to the mean value of the process, although it will converge at a much slower rate. A misspecified nonstationary model will therefore produce a bias and a forecast error with an inflated variance. For more details, see Crato and Ray (1996).

19.1.3 Estimation of the Fractional Difference

For the ARFIMA model given in (19.1.1), let $W_t = (1 - B)^d Z_t$, and let $f_W(\omega)$ and $f_Z(\omega)$ be the spectral density function of $\{W_t\}$ and $\{Z_t\}$, respectively. Then,

$$f_Z(\omega) = |1 - e^{-i\omega}|^{-2d} f_W(\omega), \qquad 0 \leq \omega \leq \pi, \tag{19.1.11}$$

where

$$f_W(\omega) = \frac{\sigma_a^2}{2\pi}\left|\frac{\theta_q(e^{-i\omega})}{\phi_p(e^{-i\omega})}\right|^2 \tag{19.1.12}$$

is the spectral density of a regular ARMA(p, q) model. Note that $f_Z(\omega) \to \infty$ as $\omega \to 0$.

Taking logarithms on both sides of (19.1.11), we get

$$\begin{aligned}\ln f_Z(\omega) &= d \ln|1 - e^{-i\omega}|^{-2} + \ln f_W(\omega) \\ &= \ln f_W(0) + d \ln|1 - e^{-i\omega}|^{-2} + \ln\left(\frac{f_W(\omega)}{f_W(0)}\right).\end{aligned}$$

Replacing ω by the Fourier frequencies $\omega_j = 2\pi j/n$, $j = 1, \ldots, [n/2]$ and adding $\ln I_Z(\omega_j)$, the periodogram of $\{Z_t\}$, to both sides, gives

$$\ln I_Z(\omega_j) = \ln f_W(0) + d \ln|1 - e^{-i\omega_j}|^{-2} + \ln\left(\frac{f_W(\omega_j)}{\ln f_W(0)}\right) + \ln\left(\frac{I_Z(\omega_j)}{f_Z(\omega_j)}\right).$$

For ω_j near zero, i.e., for $j = 1, \ldots, m << (n/2)$ such that $m/n \to 0$ as $n \to \infty$, we have $\ln(f_W(\omega_j)/f_W(0)) \approx 0$. Thus,

$$Y_j = c + dX_j + e_j, \qquad j = 1, \ldots, m, \tag{19.1.13}$$

where $Y_j = \ln I_Z(\omega_j)$, $c = \ln f_W(0)$, $X_j = \ln|1 - e^{i\omega_j}|^{-2}$, and $e_j = \ln(I_Z(\omega_j)/f_Z(\omega_j))$. Here we recall from Chapter 13 that the sequence $[I_Z(\omega_j)/f_Z(\omega_j)]$ and therefore e_j are approximately i.i.d. random variables. For computation, because

$$\begin{aligned}|1 - e^{-i\omega}|^2 &= \left|\frac{2ie^{-i\omega/2}(e^{i\omega/2} - e^{-i\omega/2})}{2i}\right|^2 \\ &= \left[2\sin\left(\frac{\omega}{2}\right)\right]^2,\end{aligned}$$

we have

$$X_j = \ln\left[\frac{1}{4[\sin(\omega_j/2)]^2}\right].$$

The least square estimator of d is given by

$$\hat{d} = \frac{\sum_{j=1}^{m}(X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{m}(X_j - \bar{X})^2}. \tag{19.1.14}$$

Geweke and Porter-Hudak (1983) showed that

$$\hat{d} \xrightarrow{d} N\left(d, \frac{\pi^2}{6\sum_{j=1}^{m}(X_j - \overline{X})^2}\right). \tag{19.1.15}$$

The 95% confidence interval for d is obtained by

$$\hat{d} \pm 1.96\sqrt{\frac{\pi^2}{6\sum_{j=1}^{m}(X_j - \overline{X})^2}}. \tag{19.1.16}$$

In practice, we often take $m = [\sqrt{n}]$ so that the regression (19.1.13) is performed over the Fourier frequencies $\omega_j = 2\pi j/n, j = 1, \ldots, [\sqrt{n}]$, where $[\cdot]$ is the greatest integer function.

Once d is estimated, we can estimate the parameters, ϕ_j and θ_j, from the fractionally differenced data $W_t = (1 - B)^d Z_t$ using the methods introduced in Chapter 7. The estimated model can then be used to forecast the future values with the results discussed in Chapter 5.

One should beware, however, that because time series aggregates are often used in data analysis and modeling, a generating long memory ARFIMA process could be misidentified as a short memory ARMA process due to the consequence of aggregation as shown in Teles, Wei, and Crato (1999). This misspecification may lead to a very undesirable result in forecasting.

19.2 Nonlinear Processes

With no loss of generality, we now consider a zero mean process. It is known that a nondeterministic linear time series process Z_t can always be written in the form

$$\begin{aligned} Z_t &= \sum_{j=0}^{\infty} \psi_j a_{t-j} \\ &= \psi(B)a_t, \end{aligned} \tag{19.2.1}$$

where $\psi_0 = 1$, a_t is a white noise process with mean 0 and variance σ_a^2, and $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$. If $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, then the process is stationary, and from, section 12.2.2, the spectrum exists and equals

$$f(\omega) = \frac{\sigma_a^2}{2\pi}\psi(e^{-i\omega})\psi(e^{i\omega}) = \frac{\sigma_a^2}{2\pi}|\psi(e^{-i\omega})|^2. \tag{19.2.2}$$

The equation given in (19.2.1) is the general form of a linear process. More generally, a time series process can be written as

$$Z_t = \sum_{i=0}^{\infty}\psi_i a_{t-i} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_{i,j}a_{t-i}a_{t-j} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_{i,j,k}a_{t-i}a_{t-j}a_{t-k} + \cdots. \tag{19.2.3}$$

The process, Z_t, in (19.2.3) is linear if it contains only the first term. The process is nonlinear if it contains more than the first term. Clearly, second-order properties such as autocorrelation functions and spectra are no longer sufficient to describe the general nonlinear process given in (19.2.3). The higher-order moments are called for even with a weakly stationary process. In addition, for the models to be presented in this section, we also require that $\{a_t\}$ in (19.2.3) be a process of i.i.d. $(0, \sigma_a^2)$ random variables.

19.2.1 Cumulants, Polyspectrum, and Tests for Linearity and Normality

Other than regular moments, many higher-order moments are available for describing a process. Because of their many desirable properties, we choose cumulants. Assume that the process in (19.2.3) is an nth order weakly stationary process. Then, from Chapter 2, all its joint regular moments up to order n exist and are time-invariant. In such a case, the nth-order cumulant, $C(k_1, k_2, \ldots, k_{n-1})$, exists and is defined to be the coefficient of $i^n t_1 t_2 \cdots t_n$ in the Taylor expansion about the origin of

$$\varphi(t_1, t_2, \ldots, t_n) = \ln\{E[\exp(it_1 Z_t + it_2 Z_{t+k_1} + \cdots + it_n Z_{t+k_{n-1}})]\}. \quad (19.2.4)$$

As a result, the function, $\varphi(t_1, t_2, \ldots, t_n)$, in (19.2.4) is also known as the cumulant generating function. Clearly, because of the nth-order stationarity, the cumulant generating function is independent of time origin.

Following the similar results from Chapter 12, if the nth-order cumulants, $C(k_1, k_2, \ldots, k_{n-1})$, are absolutely summable, then its Fourier transform exists and equals

$$f(\omega_1, \omega_2, \ldots, \omega_{n-1}) = \left(\frac{1}{2\pi}\right)^{n-1} \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{n-1}=-\infty}^{\infty} C(k_1, \ldots, k_{n-1}) e^{-i(\omega_1 k_1 + \cdots + \omega_{n-1} k_{n-1})}, \quad (19.2.5)$$

where $-\pi \leq \omega_i \leq \pi$ for $i = 1, \ldots, n-1$. The function, $f(\omega_1, \omega_2, \ldots, \omega_{n-1})$, is often known as the nth-order cumulant spectrum or the nth-order polyspectrum. Specifically, the second-order polyspectrum is

$$f(\omega_1) = \left(\frac{1}{2\pi}\right) \sum_{k_1=-\infty}^{\infty} C(k_1) e^{-i\omega_1 k_1}, \qquad -\pi \leq \omega \leq \pi, \quad (19.2.6)$$

and the third-order polyspectrum becomes

$$f(\omega_1, \omega_2) = \left(\frac{1}{2\pi}\right)^2 \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} C(k_1, k_2) e^{-i(\omega_1 k_1 + \omega_2 k_2)}, \qquad -\pi \leq \omega_1, \omega_2 \leq \pi, \quad (19.2.7)$$

which is also known as a bispectrum.

Note that $E[\exp(itZ)]$ is the characteristic function of Z. For a Gaussian random variable with mean μ and variance σ^2, it is known that its characteristic function is given by $\exp[it + (it)^2 \sigma^2/2]$. It also follows that its cumulants of order greater than two are zero. In addition, if

Y and Z are independent, for any constants a and b we have: $E[\exp(it_1aY + it_2bZ)] = E[\exp(it_1aY)]\ E[\exp(it_2bZ)]$. Because a multivariate normal can be defined as a vector of linear combinations of independent univariate normal variables (see, for example, Rao [1965, p. 440]), it follows that all cumulants of order greater than two are also zero; hence, the bispectrum and all higher-order polyspectra of a Gaussian process are identically zero. One can use these higher-order polyspectra to measure the departure of the process from normality.

To illustrate the procedure of deriving cumulants, let us consider the second-order cumulant, $C(k_1)$, which is the coefficient of $i^2t_1t_2$ in the Taylor expansion about the origin of

$$\varphi(t_1, t_2) = \ln\{E[\exp(it_1Z_t + it_2Z_{t+k_1})]\}. \tag{19.2.8}$$

From the Taylor series for a function of several variables, we have

$$\varphi(t_1, t_2) = \varphi(0, 0) + \left[\frac{\partial}{\partial t_1}\varphi(t_1, t_2)\right]_{(0,0)} t_1 + \left[\frac{\partial}{\partial t_2}\varphi(t_1, t_2)\right]_{(0,0)} t_2$$
$$+ \frac{1}{2!}\left\{\left[\frac{\partial^2}{\partial t_1^2}\varphi(t_1, t_2)\right]_{(0,0)} t_1^2 + 2\left[\frac{\partial^2}{\partial t_1 \partial t_2}\varphi(t_1, t_2)\right]_{(0,0)} t_1t_2 + \left[\frac{\partial^2}{\partial t_2^2}\varphi(t_1, t_2)\right]_{(0,0)} t_2^2\right\} + \cdots.$$

Now,

$$\left[\frac{\partial}{\partial t_1 \partial t_2}\varphi(t_1, t_2)\right]_{(0,0)} = \left[\frac{\partial}{\partial t_1}\frac{E\{[\exp(it_1Z_t + it_2Z_{t+k_1})]iZ_{t+k_1}\}}{E[\exp(it_1Z_t + it_2Z_{t+k_1})}\right]_{(0,0)}$$
$$= \left[\frac{\begin{array}{l}E[\exp(it_1Z_t + it_2Z_{t+k_1})]E\{[\exp(it_1Z_t + it_2Z_{t+k_1})]iZ_{t+k_1}iZ_t\} \\ - E[\exp(it_1Z_t + it_2Z_{t+k_1})iZ_{t+k_1}]E[\exp(it_1Z_t + it_2Z_{t+k_1})iZ_t]\end{array}}{\{E[\exp(it_1Z_t + it_2Z_{t+k_1})]\}^2}\right]_{(0,0)}$$
$$= i^2\{E[Z_{t+k_1}Z_t] - E[Z_{t+k_1}]E[Z_t]\}$$
$$= i^2\{E(Z_t - \mu)(Z_{t+k_1} - \mu)\},$$

where $\mu = E(Z_t)$. Hence,

$$C(k_1) = E(Z_t - \mu)(Z_{t+k_1} - \mu) = \gamma_{k_1}, \tag{19.2.9}$$

and the second-order polyspectrum is simply the spectrum

$$f(\omega_1) = \left(\frac{1}{2\pi}\right)\sum_{k_1=-\infty}^{\infty}\gamma_{k_1}e^{-i\omega_1k_1} = \frac{\sigma_a^2}{2\pi}|\psi(e^{-i\omega_1})|^2, \tag{19.2.10}$$

where $-\pi \le \omega_1 \le \pi$. Similarly, the third-order cumulant equals

$$C(k_1, k_2) = E(Z_t - \mu)(Z_{t+k_1} - \mu)(Z_{t+k_2} - \mu). \tag{19.2.11}$$

Let

$$T(\omega_1, \omega_2) = \frac{|f(\omega_1,\omega_2)|^2}{f(\omega_1)f(\omega_2)f(\omega_1+\omega_2)}. \tag{19.2.12}$$

It can be seen that $T(\omega_1, \omega_2)$ can be used to test whether a process is linear because under the null hypothesis the process is linear, we have $Z_t = \sum_{j=0}^{\infty}\psi_j a_{t-j}$, and

$$\begin{aligned} C(k_1, k_2) &= E(Z_t - \mu)(Z_{t+k_1} - \mu)(Z_{t+k_2} - \mu) \\ &= E\left[\sum_{j=-\infty}^{\infty}\psi_j a_{t-j}\sum_{i=-\infty}^{\infty}\psi_i a_{t+k_1-i}\sum_{n=-\infty}^{\infty}\psi_n a_{t+k_2-n}\right] \\ &= \sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\psi_j\psi_i\psi_n E(a_{t-j}a_{t+k_1-i}a_{t+k_2-n}) \\ &= \alpha\sum_{j=\infty}^{\infty}\psi_j\psi_{j+k_1}\psi_{j+k_2}, \end{aligned} \tag{19.2.13}$$

where $\alpha = E(a_t^3)$ and $\psi_j = 0$ for $j < 0$. Hence,

$$\begin{aligned} f(\omega_1, \omega_2) &= \left(\frac{1}{2\pi}\right)^2\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}C(k_1, k_2)e^{-i(\omega_1 k_1+\omega_2 k_2)} \\ &= \frac{\alpha}{(2\pi)^2}\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\psi_j\psi_{j+k_1}\psi_{j+k_2}e^{-i\omega_1 k_1-i\omega_2 k_2} \\ &= \frac{\alpha}{(2\pi)^2}\sum_{j=-\infty}^{\infty}\psi_j e^{i(\omega_1+\omega_2)j}\sum_{k_1=-\infty}^{\infty}\psi_{j+k_1}e^{-i\omega_1(j+k_1)}\sum_{k_2=-\infty}^{\infty}\psi_{j+k_2}e^{-i\omega_2(j+k_2)} \\ &= \frac{\alpha}{(2\pi)^2}\sum_{j=-\infty}^{\infty}\psi_j e^{i(\omega_1+\omega_2)j}\sum_{k=-\infty}^{\infty}\psi_k e^{-i\omega_1 k}\sum_{n=-\infty}^{\infty}\psi_n e^{-i\omega_2 n} \\ &= \frac{\alpha}{(2\pi)^2}\psi(e^{i(\omega_1+\omega_2)})\psi(e^{-i\omega_1})\psi(e^{-i\omega_2}). \end{aligned} \tag{19.2.14}$$

Thus, under the null hypothesis that a process is linear, the function

$$T(\omega_1, \omega_2) = \frac{|f(\omega_1, \omega_2)|^2}{f(\omega_1)f(\omega_2)f(\omega_1+\omega_2)} = \frac{\alpha^2}{2\pi\sigma_a^6} \tag{19.2.15}$$

is constant. Testing for linearity can therefore be based on checking the constancy of the following sample statistic over a grid of frequencies

$$\hat{T}(\omega_1, \omega_2) = \frac{|\hat{f}(\omega_1, \omega_2)|^2}{\hat{f}(\omega_1)\hat{f}(\omega_2)\,\hat{f}(\omega_1 + \omega_2)}, \tag{19.2.16}$$

where

$$\hat{f}(\omega) = \left(\frac{1}{2\pi}\right)\sum_{k=-M}^{M} W(k)\hat{\gamma}_k e^{-i\omega k} = \left(\frac{1}{2\pi}\right)\sum_{k=-M}^{M} W(k)\hat{\gamma}_k \cos(\omega k), \tag{19.2.17}$$

$$\hat{f}(\omega_1, \omega_2) = \left(\frac{1}{2\pi}\right)^2 \sum_{k_1=-M}^{M}\sum_{k_2=-M}^{M} W_2(k_1, k_2)\hat{C}(k_1, k_2)e^{-i(\omega_1 k_1 + \omega_2 k_2)}, \tag{19.2.18}$$

$$\hat{C}(k_1, k_2) = \frac{1}{n}\sum_{t=1}^{n-m} (Z_t - \bar{Z})(Z_{t+k_1} - \bar{Z})(Z_{t+k_2} - \bar{Z}), \tag{19.2.19}$$

M is the truncation point, $W(k)$ is the lag window, which was discussed in Chapter 13, $m = \max(0, k_1, k_2)$, and

$$W_2(k_1, k_2) = W(k_1)W(k_2)W(k_1 - k_2). \tag{19.2.20}$$

If the process is Gaussian, then $\alpha = E(a_t^3) = 0$. Both the bispectrum $f(\omega_1, \omega_2)$ and the function $T(\omega_1, \omega_2)$ equal zero at all frequencies. Thus, testing for normality can also be based on checking whether the sample statistic $\hat{f}(\omega_1, \omega_2)$ or $\hat{T}(\omega_1, \omega_2)$ is statistically zero over a grid of frequencies. For more details of this and the previous tests, we see Subba Rao and Gabr (1980, 1984). A non-Gaussian process could also have a zero bispectrum. From the earlier discussion, we see that, for a comprehensive test, one should really be testing all the higher-order polyspectra.

The principal advantage of the above test is that it can be used for any type of nonlinear processes. As a result, however, the test suffers in terms of its power. Additionally, as pointed out in Chan and Tong (1986), the performance of the test is also affected by the choice of the lag windows, the truncation points, and the placing of the grids of frequencies. For these reasons, great skill is required for its successful use.

19.2.2 Some Nonlinear Time Series Models

The general form of the process given in (19.2.3) contains an infinite number of parameters. Once nonlinearity in the process is confirmed, the next task is to build a suitable finite number parameter nonlinear model. Many nonlinear models have been proposed in the literature with the objective of describing different characteristics exhibited by time series encountered in practice that cannot be appropriately accounted for by linear processes. In addition to the finite order form of (19.2.3) that has been called nonlinear moving average (NLMA) models, we give below some other general finite parametric nonlinear models without details. Interested readers should see excellent references in

Subba Rao and Gabr (1984), Tong (1983, 1990), Priestley (1988), and Fan and Yao (2003). The models are as follows:

1. The bilinear model of order (p, q, r, s) (BL(p, q, r, s) model) is

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t + \sum_{j=1}^{q} \theta_j a_{t-j} + \sum_{i=1}^{r} \sum_{j=1}^{s} \alpha_{ij} Z_{t-i} a_{t-j}.$$

2. The nonlinear autoregressive (NLAR) model of order p (NLAR(p) model) is

$$Z_t = f(Z_{t-1}, \ldots, Z_{t-p}) + a_t,$$

where $f(\cdot)$ is a nonlinear function.

3. The threshold autoregressive (TAR) model is

$$Z_t = \phi_0 + \sum_{i=1}^{p} \phi_i Z_{t-i} + \left(\alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}\right) I\left(\frac{Z_{t-d} - \tau}{\delta}\right) + a_t,$$

where d is the delay parameter, τ and δ are the location and scale parameters, respectively, and $I(\cdot)$ is a smooth function, such as logistic and exponential, or an indicator function.

4. The exponential autoregressive model of order (p, d) (EXPAR(p, d) model) is

$$Z_t = \sum_{i=1}^{p} (\phi_i + \alpha_i e^{-\beta_i Z_{t-d}^2}) Z_{t-i} + a_t,$$

where $\beta_i \geq 0$ for $i = 1, \ldots, p$ and Z_{t-d} is a model-dependent variable.

The series a_t in above models are all i.i.d. $(0, \sigma_a^2)$ random variables.

These models are in the form of general orders. In practice, just as in linear time series modeling, the orders that have been commonly chosen are relatively small. For example, with respect to the NLAR model, NLAR(1) and NLAR(2) models are the most commonly used. When the function $I(\cdot)$ in model 3 above is an indicator function, the model becomes a simple two regime threshold autoregressive model. Because of its many interesting properties and close connection with the models studied in earlier chapters, we discuss the general k regime threshold autoregressive model further in the next section.

19.3 Threshold Autoregressive Models

Threshold autoregressive models were originally proposed by Tong (1978, 1983, 1990) and Tong and Lim (1980). They are piecewise linear models where the linear relationship changes according to the values of the process. Specifically, let us partition the real

line R into k intervals or regimes, $R = \bigcup_{i=1}^{k} R_i$, where $R_1 = (-\infty, r_1)$, $R_i = [r_{i-1}, r_i)$ for $i = 2, \ldots, k-1$, $R_k = [r_k, \infty)$, and $-\infty < r_1 < \cdots < r_{k-1} < \infty$ are the thresholds. A time series process Z_t is a self-exciting threshold autoregressive (TAR) process if it follows the model

$$Z_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Z_{t-i} + a_t^{(j)}, \qquad \text{if } Z_{t-d} \in R_j, \tag{19.3.1}$$

where $j = 1, \ldots, k$, d is a positive integer and is known as the delay parameter, and $a_t^{(j)}$ is a sequence of independently identically distributed random noise with mean zero and variance σ_j^2. For convenience, we will refer to (19.3.1) as a TAR(k; $p_1, \ldots, p_k$; d) model, where k is the number of regimes, separated by $(k-1)$ thresholds r_j's, and p_j denotes the order of the autoregressive model in the jth regime. The process is a linear autoregressive model in each regime and is nonlinear when there are at least two regimes with different linear models. The model contains many interesting features, such as limit cycles, amplitude dependent frequencies, and jump phenomena, which cannot be described by a linear time series process. Moreover, it reduces to a linear model when $k = 1$. It becomes a nonhomogeneous linear AR model when only the noise variances differ for different regimes; it becomes a random level shift model when only the constant terms $\phi_0^{(j)}$ differ for different regimes.

EXAMPLE 19.1 Let k = 2, $R_1 = (-\infty, 0)$, $R_2 = [0, \infty)$, $p_1 = p_2 = 1$, and $d = 1$ for the following simple *TAR*(2; 1, 1; 1) model,

$$\begin{aligned} Z_t &= -1.6Z_{t-1} + a_t^{(1)}, \qquad \text{if } Z_{t-1} < 0, \\ Z_t &= .8Z_{t-1} + a_t^{(2)}, \qquad \text{if } Z_{t-1} \geq 0, \end{aligned} \tag{19.3.2}$$

where $a_t^{(i)}$, $i = 1, 2$, is a sequence of i.i.d. $(0, \sigma_i^2)$ random variables. The series exhibits an asymmetric pattern of increasing and decreasing. At time (t − 1), if the value of Z_{t-1} is negative, then the value of Z_t at the next period tends to be positive. If the value of Z_{t-1} is positive, however, it tends to remain positive for a few periods before it changes to negative. As a consequence, the series has large upward jumps in the negative regime (regime 1), the number of observations in the negative regime is less than that in the positive regime (regime 2), and the variance, σ_1^2, of the error term in the negative regime is normally larger than the variance, σ_2^2, of the error term in the positive regime. The mean of the series, $\mu_t = E(Z_t)$, is a weighted average of the conditional means of the two regimes, which is nonzero, and the weight for each regime is equal to the probability of Z_t in that regime. Many financial time series exhibit phenomena illustrated in this example.

19.3.1 Tests for TAR Models

Let $p = \max\{p_1, \ldots, p_k\}$. With the understanding that $\phi_i^{(j)} = 0$ for $i > p_j$, we can rewrite (19.3.1) as

$$Z_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} Z_{t-i} + a_t^{(j)}, \qquad \text{if } Z_{t-d} \in R_j, \tag{19.3.3}$$

and simply call it the TAR(k; p; d) model. Petruccelli and Davies (1986) proposed a test for threshold nonlinearity using the concept of ordered autoregression. Tsay (1989) further developed the method into a modeling procedure for TAR models. For simplicity, but with no loss of generality, we illustrate the test using the TAR(2; p; d) model.

For any given AR(p) regression with n observations, $Z_t = \phi_0 + \sum_{i=1}^{p}\phi_i Z_{t-i} + a_t$, for $t = (p + 1), \ldots, n$, let us call $(Z_t, 1, Z_{t-1}, \ldots, Z_{t-d}, \ldots, Z_{t-p})$ a case of data for the AR(p) model. Suppose that the threshold variable Z_{t-d} assume values $\{Z_h, \ldots, Z_{n-d}\}$, where h = max $\{1, p + 1 - d\}$. Let (i) be the time index of the ith smallest observation of $\{Z_h, \ldots, Z_{n-d}\}$. We can obtain the following arranged autoregression with cases rearranged according to the ordered values of the delay variable Z_{t-d}:

$$\begin{bmatrix} Z_{(1)+d} \\ \vdots \\ Z_{(r)+d} \end{bmatrix} = \begin{bmatrix} 1 & Z_{(1)+d-1} & \cdots & Z_{(1)} & \cdots & Z_{(1)+d-p} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & Z_{(r)+d-1} & \cdots & Z_{(r)} & \cdots & Z_{(r)+d-p} \end{bmatrix} \begin{bmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} a_{(1)} \\ \vdots \\ a_{(r)} \end{bmatrix}. \tag{19.3.4}$$

With a proper choice of r, the above-rearranged autoregression can be fitted iteratively. For example, Tsay (1989) suggested an iterative fitting with $r = r_{\min}, r_{\min} + 1, \ldots, (n - d - \mathrm{h} + 1)$, where $r_{\min} \approx (n/10) + p$. After each iterative AR fitting of the model in (19.3.4), we can compute the one-step ahead standardized forecast error, $\hat{e}_{r+1}$. Under assumption of linearity, the least squares estimates of an AR model are consistent and the one-step ahead standardized forecast errors $\hat{e}_i$'s are known to be asymptotically independent and identically distributed with zero mean and unit variance. Thus, we can form the cumulative sum

$$S_r = \sum_{i=r_{\min}+1}^{r} \hat{e}_i, \qquad \text{for } r = r_{\min} + 1, \ldots, (n - d - h + 1)$$

and compute the test statistic

$$T = \max_{r_{\min} \le r \le (n-d-h+1)} \frac{|S_r|}{\sqrt{(n - d - h + 1 - r_{\min})}}. \tag{19.3.5}$$

From the invariance principle (Feller, 1971, p. 342), we can compute the asymptotic p-value of the test statistic T by

$$P(T < x) \approx \frac{1}{4\pi} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m + 1)} \exp\left\{-\frac{(2m + 1)^2\pi^2}{8x^2}\right\}. \tag{19.3.6}$$

The linearity is rejected if $P(T < x)$ is less than the significance level α.

It is well known that when the model is linear, the one-step ahead standardized forecast errors $\hat{e}_i$, $(r_{\min} + 1) \le i \le (n - d - h + 1)$, are not only asymptotically independent and

identically distributed but are also orthogonal to the regressors $\{Z_{(i)+d-1}, \ldots, Z_{(i)+d-p}\}$. Thus, Tsay (1989) suggested the least squares regression

$$\hat{e}_i = \omega_0 + \sum_{j=1}^{p} \omega_j Z_{(i)+d-j} + \varepsilon_i \tag{19.3.7}$$

for $(r_{\min} + 1) \leq i \leq (n - d - h + 1)$ and the associated F statistic

$$\hat{F}(p, d) = \frac{\left(\sum \hat{e}_i^2 - \sum \hat{\varepsilon}_i^2\right)/(p + 1)}{\sum \hat{\varepsilon}_i^2/(n - d - h - r_{\min} - p)}, \tag{19.3.8}$$

where the summations are over all the observations in (19.3.7) and $\hat{\varepsilon}_i$ is the least squares residual of (19.3.7). Under the linearity assumption, i.e., when Z_t follows a TAR(k; p; d) model with $k = 1$, the sample statistic $\hat{F}(p, d)$ is asymptotically distributed as an F distribution with $(p + 1)$ and $(n - d - h - r_{\min} - p)$ degrees of freedom. The linearity assumption is rejected if $\hat{F}(p, d) > F_{\alpha, (p+1, n-d-h-r_{\min}-p)}$.

19.3.2 Modeling TAR Models

The AIC Approach Tong and Lim (1980) proposed a TAR modeling procedure based on Akaike's information criterion. We illustrate the procedure for a TAR (2; p_1, p_2; d) model. The other cases follow similarly.

Step 1. Search for the orders of the piecewise AR models.

Given any fixed values d and r, let P be the maximum order to be entertained for the two piecewise linear AR models. The choice of P is subjective. Tong (1990) suggested $P = n^{\alpha}$ with $\alpha < \frac{1}{2}$. Based on the given r, construct an arranged autoregression for each regime. The orders p_j for $j = 1, 2$ are chosen so that

$$\text{AIC}(\hat{p}_j) = \min_{0 \leq p_j \leq P} \left\{ n_j \ln\left[\frac{\|\hat{\mathbf{a}}_j(p_j)\|^2}{n_j}\right] + 2(p_j + 1) \right\}, \tag{19.3.9}$$

where n_j is the number of observations in the arranged autoregression of the jth regime, $\hat{\mathbf{a}}_j(p_j)$ is the vector of residuals from an AR (p_j) fitting, and $\|\mathbf{x}\| = (\mathbf{x}', \mathbf{x})^{1/2}$ is the norm of a vector $\mathbf{x}$. Let

$$\text{AIC}(d, r) = \text{AIC}(\hat{p}_1) + \text{AIC}(\hat{p}_2). \tag{19.3.10}$$

Step 2. Search for the threshold value r.

Let d be fixed and allow r to vary over a set of potential candidates $R = \{r_1, r_2, \ldots, r_s\}$ and minimize the AIC over this set. Thus, we choose the value $\hat{r}$ of r so that

$$\text{AIC}(d, \hat{r}) = \min_{r \in R}\{\text{AIC}(d, r)\}. \tag{19.3.11}$$

Because r can be any value in an interval, this step may take several iterations based on refined grids of potential candidates. To initiate the iteration, Tong (1990) suggested using the sample percentiles as initial candidates of r.

Step 3. Search for the delay parameter d.

Let $D = \{1, 2, \ldots, T\}$ be the set of potential candidates of the delay parameter d. We now let d vary and choose its value $\hat{d}$ by

$$\text{AIC}(\hat{d}, \hat{r}) = \min_{d \in D} \left\{ \frac{\text{AIC}(d, \hat{r})}{(n - n_d)} \right\}, \tag{19.3.12}$$

where $n_d = \max(d, P)$. Because the effective number of observations in the arranged autoregression varies as the value of d changes, we should normalize the AIC by the effective number of observations $(n - n_d)$.

EXAMPLE 19.2 Consider the series of annual sunspot numbers between 1700 and 1983 that is a subseries of Series W2 discussed in earlier chapters. In this example, however, we use the original numbers instead of their square roots. Let $P = 11$, $R = \{r_{20}, r_{30}, \ldots, r_{80}\}$, where r_q is the qth percentile of the data and $D = \{1, 2, \ldots, 6\}$. The AIC approach leads to a TAR(2; 4, 11; 3) model. The estimation results are

$$\begin{aligned} Z_t &= 10.17 + 1.72Z_{t-1} - 1.33Z_{t-2} + .29Z_{t-3} + .27Z_{t-4} + a_t^{(1)}, \quad \text{if } Z_{t-3} < 37.76, \\ Z_t &= 7.41 + .73Z_{t-1} - .05Z_{t-2} - .18Z_{t-3} + .13Z_{t-4} - .23Z_{t-5} + .02Z_{t-6} + .18Z_{t-7} \\ &\quad - .26Z_{t-8} + .29Z_{t-9} + .44Z_{t-10} + .54Z_{t-11} + a_t^{(2)}, \quad \text{if } Z_{t-3} \geq 37.76. \end{aligned} \tag{19.3.13}$$

The numbers of observations are 129 and 144, and the residual variances are 258.66 and 71.44, respectively.

In addition to the formal AIC approach, Tong (1983, 1990) suggested performing a preliminary analysis of a time series using graphical methods such as reverse data plots, directed scatter diagrams, and bivariate histograms to search for evidence of nonlinear time series characteristics such as time irreversibility, cyclical behavior (limit cycles), and nonnormality. For details, we refer readers to Tong's books.

The Arranged Autoregression Approach We now illustrate the arranged autoregression approach proposed by Tsay (1989).

Step 1. Tentatively select the AR order.

Select the maximum order, P, to be entertained for the two piecewise linear AR models. This tentative selection of the AR order can be based on the PACF or on other criteria such as AIC.

Step 2. Search for the delay parameter d and hence the threshold variable Z_{t-d}.

Let D be the set of possible delay lags. Often, we let $D = \{1, 2, \ldots, P\}$. Then fit the arranged autoregressions for the given P and every element d of D, and perform the threshold nonlinearity test. An estimate of the delay parameter, d, is chosen such that

$$\hat{F}(P, \hat{d}) = \max_{d \in D} \{\hat{F}(P, d)\} \tag{19.3.14}$$

or, more precisely,

$$p\text{-value}(\hat{d}) = \min_{d \in D}\{\text{p-value}[\hat{F}(P, d)]\}. \tag{19.3.15}$$

The argument of the above selection is that if a TAR model is called for, then a reasonable starting point for the delay parameter is to begin with the one that gives the most significant result in the threshold nonlinearity test.

Step 3. Search for the threshold values.

Let $\hat{\phi}_j$ be the recursive estimate of the lag-j AR parameter in the arranged autoregression. Under linearity assumption, $\hat{\phi}_j$ is a consistent estimate of ϕ_j, and the t-ratios of $\hat{\phi}_j$ will behave exactly as those of a linear AR model before the recursion reaches the threshold value r. Once the recursion reaches the threshold value r, the estimate $\hat{\phi}_j$ begins to change and the t-ratios start to deviate. Thus, the threshold values can be identified through the change points in the convergence pattern of the t-ratios from a scatterplot of the recursive t ratios of $\hat{\phi}_j$ versus the values of the threshold variable Z_{t-d}.

Step 4. Refine the results.

Refine the AR order and threshold values in each regime using the available linear autoregression techniques, if necessary.

EXAMPLE 19.3 We now analyze the same subseries of 284 annual sunspot numbers using the arranged autoregression method. To be comparable with the AIC approach, we also choose $P = 11$; hence $D = \{1, 2, \ldots, 11\}$. The results of threshold nonlinearity F tests based on the given P and D are summarized in Table 19.1. The value $d = 2$ gives the minimum p-value. Thus, we select Z_{t-2} as our threshold variable. To search for the

TABLE 19.1 The p-values of F tests for the annual sunspot numbers (Series W2).

D	p-value
1	.000650
2	.000000
3	.000120
4	.044741
5	.049350
6	.001621
7	.013555
8	.023646
9	.036691
10	.568104
11	.027104

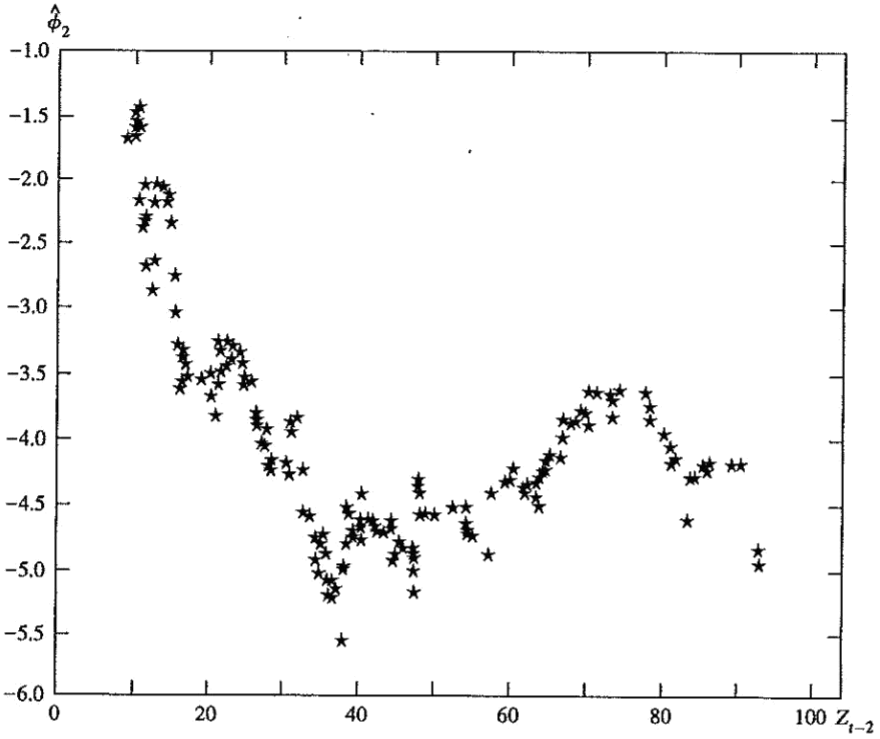


FIGURE 19.1 Scatterplot of recursive t-ratios of the lag-2 AR coefficient $\hat{\phi}_2$ versus ordered Z_{t-2} for the sunspot numbers.

threshold value, we construct a scatterplot of the t-ratios of $\hat{\phi}_2$ versus the ordered values of Z_{t-2} in Figure 19.1. From the plot, we see that the t-ratio changes its direction twice, once at $Z_{t-2} = 35$ and again at around $Z_{t-2} = 70$, suggesting that $r_1 = 35$ and $r_2 = 70$. We then use AIC to refine the AR orders in each regime. The results lead to the following TAR(3; 11, 10, 10; 2) model:

$$
\begin{aligned}
Z_t &= 3.71 + 1.82Z_{t-1} - 1.26Z_{t-2} + .08Z_{t-3} + .06Z_{t-4} + .08Z_{t-5} - .06Z_{t-6} \\
&\quad + .13Z_{t-7} + .03Z_{t-8} + .06Z_{t-9} - .04Z_{t-11} + a_t^{(1)}, \qquad \text{if } Z_{t-2} < 35, \\
Z_t &= 12.33 + 1.07Z_{t-1} - .07Z_{t-2} - .68Z_{t-3} + .34Z_{t-4} - .12Z_{t-5} + .11Z_{t-7} - .17Z_{t-8} \\
&\quad - .02Z_{t-9} + .28Z_{t-10} + a_t^{(2)}, \qquad \text{if } 35 \le Z_{t-2} < 70, \\
Z_t &= 1.47 + .65Z_{t-1} + .18Z_{t-2} - .17Z_{t-3} + .04Z_{t-4} + .19Z_{t-5} + .11Z_{t-6} \\
&\quad + .31Z_{t-7} - .48Z_{t-8} + .23Z_{t-9} + .09Z_{t-10} + a_t^{(3)}, \qquad \text{if } Z_{t-2} \ge 70.
\end{aligned} \tag{19.3.16}
$$

The numbers of observations are 118, 89, and 66, and the residual variances are 174.23, 125.68, and 89.11, respectively.

Chan (1989) gave a detailed analysis of annual sunspot numbers including TAR models in (19.3.13) and (19.3.16). There are many studies on nonlinear models and their applications. Because their ability to capture the asymmetric phenomena, the models or the hybrids of these models such as TAR-GARCH models are found useful in many economic and financial studies. For more examples, we refer readers to Subba Rao and Gabr (1984), Tsay (1989, 1991), Chan (1990), Tong (1990), and Pena, Tiao, and Tsay (2001) among others.

EXERCISES

19.1 Consider the ARFIMA(0, .4, 1) model, $(1 - B)^{.4} Z_t = (1 - .8B)a_t$, where the a_t are i.i.d. $N(0,1)$ random variables.

(a) Find and plot its spectrum.

(b) Find and plot its autocorrelation function.

(c) Comment your findings in part (a) and (b).

19.2 Let Z_t be a stationary Gaussian process. Show that its bispectrum is identically zero.

19.3 Show that the bilinear model, $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-1} a_{t-1} + a_t$, can be expressed in the form of (19.2.3).

19.4 Consider the EXPAR(2, 1) model, $Z_t = \sum_{i=1}^{2} (\phi_i + \alpha_i e^{-b_i Z_{t-1}^2}) Z_{t-i} + a_t$. Discuss the different phenomena of the model when $|Z_{t-1}|$ is large and when $|Z_{t-1}|$ is small, and show that the EXPAR model behaves rather like the TAR model where the coefficients change smoothly between two extreme values.

19.5 Consider Series W2 of the yearly sunspot numbers from 1700 to 2001 given in the appendix.

(a) Test the linearity of the underlying process using the bispectrum test.

(b) Build a TAR model for the series. Compare and comment your TAR model with the standard AR models obtained in Example 6.2.

19.6 Consider Series W7 of the yearly number of lynx pelts sold by the Hudson's Bay Company in Canada from 1857 to 1911 given in the appendix.

(a) Perform the normality test on the series.

(b) Perform the linearity test on the series.

(c) Build a TAR model for the series. Compare and comment your TAR model with the standard AR models obtained in Example 6.7.

20

Aggregation and Systematic Sampling in Time Series

In any scientific investigation, after making a conjecture or a proposition about the underlying model of a phenomenon, a scientist must test his or her proposition against observable data. Suppose that, based on theoretical grounds, a scientist proposes an underlying model for a time series process in terms of a basic time unit t. In some fields, a scientist can test his or her proposition by analyzing the data from a designed experiment in terms of the same time unit t used in his or her proposition. In some cases, however, controlled experiments may not be possible. Although time series observations may be available, the investigator often cannot choose the time interval. Thus, the assumed time unit in the model may not be the same as the time unit for the observed data. In many studies, data often are available only through aggregation or systematic sampling.

Generally speaking, a time series variable is either a flow variable or a stock variable. A flow variable such as industrial production exists only through aggregation over a time interval. A stock variable such as the price of a given commodity exists at every time point. The values of a flow variable are usually obtained through aggregation over equal time intervals, whereas the values of a stock variable are often obtained through systematic sampling. In this chapter, we study the consequences of aggregation and systematic sampling on model structure, parameter estimation, forecasting, and various time series tests.

20.1 Temporal Aggregation of the ARIMA Process

Let z_t be the equally spaced basic series whose dth difference $w_t = (1 - B)^d z_t$ follows a covariance stationary process with mean 0 and autocovariance function

$$\gamma_w(k) = \text{Cov}\{w_t, w_{t+k}\}.$$

Assume that the observed time series Z_T is the m-period nonoverlapping aggregates of z_t defined as

$$\begin{aligned} Z_T &= \sum_{t=m(T-1)+1}^{mT} z_t \\ &= (1 + B + \cdots + B^{m-1}) z_{mT}, \end{aligned} \tag{20.1.1}$$

where T is the aggregate time unit and m is fixed and is called the order of aggregation. Thus, for example, if t represents the time unit of month and m is 3, then the Z_T are quarterly sums of the monthly series z_t. We call Z_T an aggregate series, and we refer to z_t as a basic or nonaggregate series.

20.1.1 The Relationship of Autocovariances between the Nonaggregate and Aggregate Series

To derive the relationship between the autocovariances of the nonaggregate series z_t and its aggregate series Z_T, we first define the following m-period overlapping sum:

$$\begin{aligned} \mathcal{I}_t &= \sum_{j=0}^{m-1} z_{t-j} \\ &= (1 + B + \cdots + B^{m-1}) z_t \end{aligned} \tag{20.1.2}$$

and note that

$$Z_T = \mathcal{I}_{mT}. \tag{20.1.3}$$

Let $\mathcal{B}$ be the backshift operator on the aggregate time unit T such that $\mathcal{B}Z_T = Z_{T-1}$. Now,

$$\begin{aligned} (1 - \mathcal{B}) Z_T &= Z_T - Z_{T-1} \\ &= \mathcal{I}_{mT} - \mathcal{I}_{m(T-1)} \\ &= (1 - B^m) \mathcal{I}_{mT}. \end{aligned} \tag{20.1.4}$$

Thus, by letting

$$U_T = (1 - \mathcal{B})^d Z_T, \tag{20.1.5}$$

we have

$$\begin{aligned} U_T &= (1 - \mathcal{B})^d Z_T \\ &= (1 - B^m)^d \mathcal{I}_{mT} \\ &= [(1 + B + \cdots + B^{m-1})(1 - B)]^d (1 + B + \cdots + B^{m-1}) z_{mT} \\ &= (1 + B + \cdots + B^{m-1})^{d+1} (1 - B)^d z_{mT} \\ &= (1 + B + \cdots + B^{m-1})^{d+1} w_{mT}. \end{aligned} \tag{20.1.6}$$

Hence, the dth differenced aggregate series, $U_T = (1 - B)^d Z_T$, is covariance stationary as it is a finite moving average of a stationary process w_t. Stram and Wei (1986b) show that the autocovariance function $\gamma_w(k)$ for w_t and the autocovariance function $\gamma_U(k)$ for U_T are related as follows:

$$\gamma_U(k) = (1 + B + \cdots + B^{m-1})^{2(d+1)} \gamma_w[mk + (d + 1)(m - 1)], \quad (20.1.7)$$

where B now operates on the index of $\gamma_w(j)$ such that $B\gamma_w(j) = \gamma_w(j - 1)$. Thus, $\gamma_U(k)$ is the linear transformation of the autocovariances $\gamma_w(j)$ from $j = mk - (d + 1)(m - 1)$ to $j = mk + (d + 1)(m - 1)$. The coefficients in the linear transformation are found by expanding the polynomial $(1 + B + \cdots + B^{m-1})^{2(d+1)}$. We can write this transformation in the following matrix form:

$$\begin{bmatrix} \gamma_U(0) \\ \gamma_U(1) \\ \vdots \\ \gamma_U(k) \end{bmatrix} = \mathbf{A} \begin{bmatrix} \gamma_w[-(d + 1)(m - 1)] \\ \gamma_w[-(d + 1)(m - 1) + 1] \\ \vdots \\ \gamma_w(0) \\ \vdots \\ \gamma_w[mk + (d + 1)(m - 1)] \end{bmatrix}. \quad (20.1.8)$$

The coefficient matrix $\mathbf{A}$ is equal to

$$\begin{bmatrix} \mathbf{C} & \mathbf{0}_{mk} & \\ \mathbf{0}_m & \mathbf{C} & \mathbf{0}_{m(k-1)} \\ \vdots & & \\ \mathbf{0}_{mk} & & \mathbf{C} \end{bmatrix},$$

where $\mathbf{0}_n$ is a $1 \times n$ vector of zeros and $\mathbf{C}$ is a $1 \times [2(d + 1)(m - 1) + 1]$ vector of C_i, which is the coefficient of B^i in the polynomial $(1 + B + \cdots + B^{m-1})^{2(d+1)}$. Because $\gamma_w(-k) = \gamma_w(k)$ for all k, we can, in the above matrix, delete the first $(d + 1)(m - 1)$ columns corresponding to $\gamma_w[-(d + 1)(m - 1)], \ldots,$ and $\gamma_w(-1)$, by adding them to the columns corresponding to $\gamma_w[(d + 1)(m - 1)], \ldots,$ and $\gamma_w(1)$, respectively. Thus,

$$\begin{bmatrix} \gamma_U(0) \\ \gamma_U(1) \\ \vdots \\ \gamma_U(k) \end{bmatrix} = \mathbf{A}_m^d \begin{bmatrix} \gamma_w(0) \\ \gamma_w(1) \\ \vdots \\ \gamma_w[mk + (d + 1)(m - 1)] \end{bmatrix}, \quad (20.1.9)$$

where $\mathbf{A}_m^d$ is the resulting matrix formed by this deletion of the first $(d + 1)(m - 1)$ columns and adding them to the proper remaining columns of the matrix $\mathbf{A}$.

EXAMPLE 20.1 To clarify the construction of the matrices $\mathbf{A}$ and $\mathbf{A}_m^d$, we consider an MA(2) model,

$$z_t = (1 - \theta_1 B - \theta_2 B^2) a_t,$$

where $\gamma_z(0) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$, $\gamma_z(1) = (-\theta_1 + \theta_1\theta_2)\sigma_a^2$, $\gamma_z(2) = -\theta_2\sigma_a^2$, and $\gamma_z(j) = 0$, $|j| > 2$. To find the model for its third order aggregate, $Z_T = (1 + B + B^2)z_{3T}$, we have $d = 0$, $m = 3$, and from equation (20.1.7)

$$\begin{aligned}\gamma_Z(k) &= (1 + B + B^2)^2\gamma_z(3k + 2)\\ &= (1 + 2B + 3B^2 + 2B^3 + B^4)\gamma_z(3k + 2).\end{aligned} \tag{20.1.10}$$

Hence, by (20.1.8),

$$\begin{bmatrix}\gamma_Z(0)\\ \gamma_Z(1)\\ \gamma_Z(2)\\ \gamma_Z(3)\end{bmatrix} = \begin{bmatrix}1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1\end{bmatrix}\begin{bmatrix}\gamma_z(-2)\\ \gamma_z(-1)\\ \gamma_z(0)\\ \gamma_z(1)\\ \vdots\\ \gamma_z(11)\end{bmatrix}. \tag{20.1.11}$$

By using (20.1.9), equation (20.1.11) is reduced to

$$\begin{bmatrix}\gamma_Z(0)\\ \gamma_Z(1)\\ \gamma_Z(2)\\ \gamma_Z(3)\end{bmatrix} = \begin{bmatrix}3 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1\end{bmatrix}\begin{bmatrix}\gamma_z(0)\\ \gamma_z(1)\\ \gamma_z(2)\\ 0\\ \vdots\\ 0\end{bmatrix}$$

$$= \begin{bmatrix}3\gamma_z(0) + 4\gamma_z(1) + 2\gamma_z(2)\\ \gamma_z(1) + 2\gamma_z(2)\\ 0\\ 0\end{bmatrix}. \tag{20.1.12}$$

Thus, the third aggregate of an MA(2) model is an MA(1) model,

$$Z_T = (1 - \Theta B)A_T,$$

where $\gamma_Z(0) = (1 + \Theta^2)\sigma_A^2 = 3[1 + \theta_1^2 + \theta_2^2]\sigma_a^2 + 4[-\theta_1 + \theta_1\theta_2]\sigma_a^2 + 2[-\theta_2]\sigma_a^2$, and $\gamma_Z(1) = -\Theta\sigma_A^2 = 2[-\theta_2]\sigma_a^2 + [-\theta_1 + \theta_1\theta_2]\sigma_a^2$. The parameters, Θ and σ_A^2, of the aggregate model for Z_T can be found to be functions of θ_1, θ_2, and σ_A^2 through the relations,

$$\frac{1 + \Theta^2}{-\Theta} = \frac{3[1 + \theta_1^2 + \theta_2^2] + 4[-\theta_1 + \theta_1\theta_2] + 2[-\theta_2]}{2[-\theta_2] + [-\theta_1 + \theta_1\theta_2]}$$

and

$$\sigma_A^2 = \frac{[-\theta_1 + \theta_1\theta_2]\sigma_a^2 + 2[-\theta_2]\sigma_a^2}{-\Theta}.$$

20.1.2 Temporal Aggregation of the IMA(d, q) Process

Suppose that the nonaggregate series z_t follows an IMA(d, q) model

$$(1 - B)^d z_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \tag{20.1.13}$$

where the a_t are white noise series with mean 0 and variance σ_a^2. We would like to know the corresponding model for the mth-order aggregate series Z_T defined in (20.1.1). Let $U_T = (1 - \mathcal{B})^d Z_T$. From (20.1.7), we see that $\gamma_U(l)$ is a weighted sum of those $\gamma_w(k)$ for which $ml - (d + 1)(m - 1) \leq k \leq ml + (d + 1)(m - 1)$. Because $\gamma_w(k) = 0$ for $|k| > q$, $\gamma_U(l)$ will be identically zero if $l > q^* = [d + 1 + (q - d - 1)/m]$. Thus, the only nonzero autocovariances of U_T are at most $\gamma_U(0)$, $\gamma_U(1)$, . . . , $\gamma_U(q^*)$, which implies that the mth-order aggregate series $Z_T = (1 + B + \cdots + B^{m-1})z_{mT}$ follows an IMA(d, N_0) process

$$(1 - \mathcal{B})^d Z_T = (1 - \beta_1 \mathcal{B} - \cdots - \beta_{N_0}\mathcal{B}^{N_0})A_T, \tag{20.1.14}$$

where the A_T is a white noise series with mean 0 and variance σ_A^2,

$$N_0 \leq q^* = \left[d + 1 + \frac{(q - d - 1)}{m}\right], \tag{20.1.15}$$

and $[x]$ is used to denote the integer part of x. The parameters β_i's and σ_A^2 are functions of θ_j's and σ_a^2. They can be derived from solving the system of equations in (20.1.9), which reduces to a much simpler form for an IMA(d, q) model. In fact, equations (20.1.13) and (20.1.15) imply that in studying temporal aggregation for the IMA(d, q) model, one need only consider the first $(q^* + 1)$ rows and $(q + 1)$ columns of the coefficient matrix A_m^d. Letting $A_m^d(q)$ be this submatrix consisting of these first $(q^* + 1)$ rows and $(q + 1)$ columns of A_m^d, we have

$$\begin{bmatrix} \gamma_U(0) \\ \gamma_U(1) \\ \vdots \\ \gamma_U(q^*) \end{bmatrix} = A_m^d(q) \begin{bmatrix} \gamma_w(0) \\ \gamma_w(1) \\ \vdots \\ \gamma_w(q) \end{bmatrix}. \tag{20.1.16}$$

It should be noted that q^* in (20.1.15) is only an upper bound. To see that, let us consider the case when $m = 3$ and the model for z_t is an MA(2) process, $z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$. In this case, $q^* = 1$. Also, by (20.1.7),

$$\gamma_U(k) = (1 + B + B^2)^2 \gamma_w(3k + 2).$$

Thus, from (20.1.8), (20.1.9), and (20.1.16), we have that

$$A_m^d(q) = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \end{bmatrix}$$

and

$$\begin{bmatrix} \gamma_U(0) \\ \gamma_U(1) \end{bmatrix} = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \gamma_w(0) \\ \gamma_w(1) \\ \gamma_w(2) \end{bmatrix}, \tag{20.1.17}$$

where $w_t = z_t$ and $U_T = Z_T$. Now,

$$\begin{cases} \gamma_w(1) = (-\theta_1 + \theta_1\theta_2)\sigma_a^2 \\ \gamma_w(2) = -\theta_2\sigma_a^2 \end{cases} \tag{20.1.18}$$

so that

$$\gamma_U(1) = (\theta_1\theta_2 - \theta_1 - 2\theta_2)\sigma_a^2. \tag{20.1.19}$$

Thus, if $\theta_2 = \theta_1/(\theta_1 - 2)$, then $\gamma_U(1) = 0$ and the aggregates become an uncorrelated white noise series.

20.1.3 Temporal Aggregation of the AR(p) Process

Suppose that the nonaggregate series z_t follows a stationary AR(p) process

$$(1 - \phi_1 B - \cdots - \phi_p B^p)z_t = a_t. \tag{20.1.20}$$

Let $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and δ_i^{-1} for $i = 1, \ldots, p^*$ be the distinct roots of $\phi_p(B)$, each with multiplicity s_i such that $\sum_{i=1}^{p^*} s_i = p$. For any given value of m, let b equal the number of distinct values δ_i^m for $i = 1, \ldots, p^*$. Furthermore, partition the numbers s_i, $i = 1, \ldots, p^*$, into b distinct sets $\mathcal{A}_i$ such that s_k and $s_j \in \mathcal{A}_i$ if and only if $\delta_k^m = \delta_j^m$. Then Stram and Wei (1986b) show that the mth-order aggregate series, $Z_T = (1 + B + \cdots + B^{m-1})z_{mT}$, follows an ARMA($M$, N_1) model

$$(1 - \alpha_1 \mathcal{B} - \cdots - \alpha_M \mathcal{B}^M)Z_T = (1 - \beta_1 \mathcal{B} - \cdots - \beta_{N_1}\mathcal{B}^{N_1})A_T, \tag{20.1.21}$$

where

$$M = \sum_{i=1}^{b} \max \mathcal{A}_i, \tag{20.1.22}$$

$$\max \mathcal{A}_i = \text{the largest element in } \mathcal{A}_i,$$

$$N_1 = \left[p + 1 - \frac{(p+1)}{m}\right] - (p - M)$$

$$= \left[M + 1 - \frac{(p+1)}{m}\right], \tag{20.1.23}$$

the A_T are white noise with mean 0 and variance σ_A^2, and α_i, β_j, and σ_A^2 are functions of ϕ_k's and σ_a^2.

EXAMPLE 20.2 Assume that the nonaggregate series z_t follows the AR(2) model

$$(1 - \phi_1 B - \phi_2 B^2) z_t = a_t \tag{20.1.24}$$

with $\phi_1 = -(.5)^{1/3}$, $\phi_2 = -(.5)^{2/3}$, and $\sigma_a^2 = 1$. To derive the aggregate model for Z_T when $m = 3$, we first find the roots of $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2)$, which are

$$\delta_1^{-1} = 2^{1/3}\left[\cos\left(\frac{2\pi}{3}\right) + i \sin\left(\frac{2\pi}{3}\right)\right]$$

and

$$\delta_2^{-1} = 2^{1/3}\left[\cos\left(\frac{2\pi}{3}\right) - i \sin\left(\frac{2\pi}{3}\right)\right]. \tag{20.1.25}$$

They are distinct, each with multiplicity 1. Thus, $p^* = 2$, $s_1 = s_2 = 1$. But, $\delta_1^3 = \delta_2^3 = \frac{1}{2}$, which implies that $b = 1$ and $\mathcal{A}_1 = \{1\}$. Hence $M = 1$ and $N_1 = [1 + 1 - (2 + 1)/3] = 1$, and the third-order aggregate series, Z_T, follows an ARMA(1, 1) model.

A model is said to have hidden periodicity of order m if $\delta_i \neq \delta_j$ but $\delta_i^m = \delta_j^m$. Clearly, the process given in (20.1.24) has a hidden periodicity of order 3. It is worth noting that the reduction of the AR order on aggregation in the above example is due to the hidden periodicity. For a more detailed discussion, see Stram and Wei (1986b).

20.1.4 Temporal Aggregation of the ARIMA(p, d, q) Process

Now assume that the nonaggregate series follows a mixed ARIMA(p, d, q) model

$$\phi_p(B)(1 - B)^d z_t = \theta_q(B) a_t, \tag{20.1.26}$$

where

$$\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p),$$
$$\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q),$$

and the a_t are white noise. The polynomials $\phi_p(B)$ and $\theta_q(B)$ are assumed to have their roots outside the unit circle and to have no roots in common. Additionally, we also assume that the model has no hidden periodicity of order m in the sense that if δ_i^{-1}, $i = 1, \ldots, p$, are the roots of $\phi_p(B)$, then $\delta_i^m = \delta_j^m$ if and only if $\delta_i = \delta_j$. Letting $\phi_p(B) = \prod_{j=1}^{p}(1 - \delta_j B)$ and multiplying

$$\prod_{j=1}^{p}\left[\frac{(1 - \delta_j^m B^m)(1 - B^m)^{d+1}}{(1 - \delta_j B)(1 - B)^{d+1}}\right]$$

on both sides of (20.1.26), we get

$$\prod_{j=1}^{p}(1 - \delta_j^m B^m)(1 - B^m)^d \mathcal{I}_t = \left[\prod_{j=1}^{p}\left(\frac{1 - \delta_j^m B^m}{1 - \delta_j B}\right)\right]\left(\frac{1 - B^m}{1 - B}\right)^{d+1}\theta_q(B)a_t. \quad (20.1.27)$$

Let

$$X_t = \prod_{j=1}^{p}(1 - \delta_j^m B^m)(1 - B^m)^d \mathcal{I}_t. \quad (20.1.28)$$

It is easily seen that

$$E(X_{mT}X_{mT-mK}) = 0$$

for $K > N_2$, where

$$N_2 = \left[p + d + 1 + \frac{(q - p - d - 1)}{m}\right]. \quad (20.1.29)$$

Thus, the aggregate series $Z_T = \mathcal{I}_{mT}$ follows an ARIMA(p, d, N_2) model

$$\prod_{j=1}^{p}(1 - \delta_j^m \mathcal{B})(1 - \mathcal{B})^d Z_T = (1 - \beta_1 \mathcal{B} - \cdots - \beta_{N_2}\mathcal{B}^{N_2})A_T, \quad (20.1.30)$$

where the A_T are white noise with mean 0 and variance σ_A^2, and the parameter β_j's and σ_A^2 are functions of ϕ_i's, θ_j's, and σ_a^2. Note that in this case, the AR order is unchanged by aggregation and that the roots of the AR polynomial of the aggregate series model are the mth powers of the roots of the nonaggregate series AR polynomial.

If the ARIMA(p, d, q) model of the nonaggregate series z_t has hidden periodicity of order m in the AR polynomial, then by the results of Section 20.1.3, the mth-order aggregate

series, Z_T, will follow a mixed ARIMA(M, d, N) model. The AR order M is determined by (20.1.22), and the MA order N is given by

$$N \leq \left[p + d + 1 + \frac{(q - p - d - 1)}{m}\right] - (p - M). \tag{20.1.31}$$

More generally, aggregation of seasonal ARIMA models was studied by Wei (1978b). Assume that the nonaggregate series z_t follows a general multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ model with a seasonal period s,

$$\Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d z_t = \theta_q(B)\Theta_Q(B^s)a_t, \tag{20.1.32}$$

where

$$\begin{aligned}
\Phi_P(B^s) &= (1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}),\\
\Theta_s(B^s) &= (1 - \Theta_1 B^s - \cdots - \Theta_s B^{Qs}),\\
\phi_p(B) &= (1 - \phi_1 B - \cdots - \phi_p B^p),\\
\theta_q(B) &= (1 - \theta_1 B - \cdots - \theta_q B^q),
\end{aligned}$$

have roots outside of the unit circle and the a_t is a white noise process with mean 0 and variance σ_a^2. Using the results from previous sections, it is easily seen that the mth-order aggregate series, Z_T, when $s = mS$ for some integer S and $m < S$, will follow a multiplicative seasonal ARIMA$(M, d, N) \times (P, D, Q)_S$ model with a seasonal period S, i.e.,

$$\Phi_P(\mathcal{B}^S)\alpha_M(\mathcal{B})(1 - \mathcal{B}^S)^D(1 - \mathcal{B})^d Z_T = \beta_N(\mathcal{B})\Theta_Q(\mathcal{B}^S)A_T, \tag{20.1.33}$$

where

$$\begin{aligned}
\Phi_P(\mathcal{B}^S) &= (1 - \Phi_1 \mathcal{B}^S - \cdots - \Phi_P \mathcal{B}^{PS}),\\
\Theta_Q(\mathcal{B}^S) &= (1 - \Theta_1 \mathcal{B}^S - \cdots - \Theta_Q \mathcal{B}^{QS}),\\
\alpha_M(\mathcal{B}) &= (1 - \alpha_1 \mathcal{B} - \cdots - \alpha_M \mathcal{B}^M),\\
\beta_N(\mathcal{B}) &= (1 - \beta_1 \mathcal{B} - \cdots - \beta_N \mathcal{B}^N),
\end{aligned}$$

which all have roots outside of the unit circle. M is determined by (20.1.22), N by (20.1.31), and the A_T is a white noise process with mean 0 and variance σ_A^2. The parameters α_i's are functions of ϕ_j's; the parameters β_i's and σ_A^2 are functions of ϕ_j's, θ_j's, and σ_a^2. When $m \geq S$, aggregation reduces a seasonal model to a regular ARIMA model.

20.1.5 The Limiting Behavior of Time Series Aggregates

To see the implications of time series aggregation, it is interesting to study the limiting behavior of the time series aggregates. Suppose that z_t follows an ARIMA(p, d, q) model in (20.1.26). When $m \to \infty$, Tiao (1972) shows that the limiting model for the aggregates, Z_T,

exists and equals the IMA(d, d) process, which is independent of p and q. When z_t follows a seasonal ARIMA(p, d, q) $\times$ (P, D, Q)$_s$ model, Wei (1978b) shows that the limiting aggregate model of Z_T becomes an IMA ($D + d$, $D + d$) process. Wei and Stram (1988) show that the results hold also for noninvertible models and that if 1 is a root of the moving average polynomial for the z_t process, then 1 is also a root of the moving average polynomial for the Z_T process. Two interesting consequences of the results are the following:

1. The limiting model of the time series aggregates from a stationary model is a white noise process.
2. The limiting model is unique and is independent of the stationary portions of the nonaggregate model.

As a result, to study the limiting model of a time series aggregate, with no loss of generality, one needs only consider the case where the underlying nonaggregate series z_t follows an IMA(d, d) model. In this case, the autocovariances $\gamma_U(k)$ for $U_T = (1 - \mathcal{B})^d Z_T$ and the autocovariances $\gamma_w(j)$ for $w_t = (1 - B)^d z_t$ are related, by (20.1.16), as

$$\begin{bmatrix} \gamma_U(0) \\ \gamma_U(1) \\ \vdots \\ \gamma_U(d) \end{bmatrix} = A_m^d(d) \begin{bmatrix} \gamma_w(0) \\ \gamma_w(1) \\ \vdots \\ \gamma_w(d) \end{bmatrix}, \tag{20.1.34}$$

where $A_m^d(d)$ becomes a $(d + 1) \times (d + 1)$ square matrix. Wei and Stram (1988) call the matrix $A_m^d(d)$ the aggregation matrix for the linear transformation that maps the autocovariances, $\gamma_w(j)$'s, of the underlying IMA(d, d) model for z_t to the autocovariances, $\gamma_U(j)$'s, of the IMA(d, d) aggregate model for Z_T. From the construction, it is easily seen that the matrix $A_m^d(d)$ is nonsingular. Upon normalization, we have

$$\boldsymbol{\rho}_U = A_m^d(d) \frac{1}{\mathbf{a}_1' \boldsymbol{\rho}_w} \boldsymbol{\rho}_w, \tag{20.1.35}$$

where

$$\begin{aligned} \boldsymbol{\rho}_w &= (\rho_w(0), \rho_w(1), \ldots, \rho_w(d))', \\ \boldsymbol{\rho}_U &= (\rho_U(0), \rho_U(1), \ldots, \rho_U(d))', \\ \rho_w(i) &= \gamma_w(i)/\gamma_w(0), \\ \rho_U(i) &= \gamma_U(i)/\gamma_U(0), \end{aligned}$$

and $\mathbf{a}_1'$ is the first row of $A_m^d(d)$. Note that

$$\mathbf{a}_1' \boldsymbol{\rho}_w = \frac{\mathbf{a}_1' \boldsymbol{\gamma}_w}{\gamma_w(0)}, \tag{20.1.36}$$

where $\boldsymbol{\gamma}_w = (\gamma_w(0), \gamma_w(1), \ldots, \gamma_w(d))'$. Now, $\gamma_w(0)$ is the variance of $w_t = (1 - B)^d Z_t$, and by (20.1.34),

$$\gamma_U(0) = \mathbf{a}_1' \boldsymbol{\gamma}_w, \tag{20.1.37}$$

which is the variance of $U_T = (1 - \mathcal{B})^d Z_T$. Hence, $\mathbf{a}_1' \boldsymbol{\rho}_w > 0$ and the aggregation transformation defined in (20.1.35) is continuous.

Let f^m be the mth-order aggregation transformation given in (20.1.35). If

$$\lim_{m \to \infty} f^m(\boldsymbol{\rho}_w) = \boldsymbol{\rho},$$

then

$$\begin{aligned} \boldsymbol{\rho} &= \lim_{m \to \infty} f^m(\boldsymbol{\rho}_w) \\ &= \lim_{k \to \infty} f^{m^k}(\boldsymbol{\rho}_w) \\ &= \lim_{k \to \infty} f^m[f^{m^{k-1}}(\boldsymbol{\rho}_w)] \\ &= f^m\left[\lim_{m \to \infty} f^{m^{k-1}}(\boldsymbol{\rho}_w)\right] \\ &= f^m(\boldsymbol{\rho}). \end{aligned} \tag{20.1.38}$$

Hence, the autocorrelation vector of a limiting model is unaffected by aggregation. That is, if we let $\boldsymbol{\rho}_U^{(\infty)}$ be the autocorrelation vector of the limiting model, then

$$\boldsymbol{\rho}_U^{(\infty)} = A_m^d(d) \frac{1}{\mathbf{a}_1' \boldsymbol{\rho}_U^{(\infty)}} \boldsymbol{\rho}_U^{(\infty)}$$

or

$$(\mathbf{a}_1' \boldsymbol{\rho}_U^{(\infty)}) \boldsymbol{\rho}_U^{(\infty)} = A_m^d(d) \boldsymbol{\rho}_U^{(\infty)}. \tag{20.1.39}$$

Thus, the limiting autocorrelation vector $\boldsymbol{\rho}_U^{(\infty)}$ is simply the eigen-vector of the aggregation matrix $A_m^d(d)$ with the associated eigenvalue $\mathbf{a}_1' \boldsymbol{\rho}_U^{(\infty)}$.

Let the invertible limiting model of aggregates be

$$(1 - \mathcal{B})^d Z_T = \theta_d^{(\infty)}(\mathcal{B}) A_T, \tag{20.1.40}$$

where $\theta_d^{(\infty)}(\mathcal{B}) = (1 - \theta_1^{(\infty)}\mathcal{B} - \cdots - \theta_d^{(\infty)}\mathcal{B}^d)$, the roots of $\theta_d^{(\infty)}(\mathcal{B}) = 0$ are all outside the unit circle, and the A_T are white noise with mean 0 and variance σ_A^2. Wei and Stram (1988) prove that the autocorrelation vector $\boldsymbol{\rho}_U^{(\infty)}$ for the invertible limiting aggregate model in (20.1.40) is given by the eigenvector associated with the largest eigenvalue of $A_m^d(d)$, where the eigenvector is scaled so that its first element equals 1. Wei and Stram give an interesting note

TABLE 20.1 Sample autocorrelations $\hat{\rho}_z(k)$ of $\{z_t\}$.

k	1	2	3	4	5	6	7	8
$\hat{\rho}_z(k)$	.96	.95	.94	.94	.93	.92	.91	.90

TABLE 20.2 Sample autocorrelations $\hat{\rho}_w(k)$ of $\{(1 - B)z_t\} = \{w_t\}$.

k	1	2	3	4	5	6	7	8
$\hat{\rho}_w(k)$	−.47	.06	−.07	.04	.00	.04	−.04	.06

that because the linear transformation given in (20.1.35) is continuous, when deriving the aggregate model, the m in $A_m^d(d)$ can be any integer larger than or equal to 2. The limit can then be taken through m^k as $k \to \infty$. Specifically, we can choose $m = 2$, which greatly simplifies the calculation of the coefficients of B^j in the polynomial $(1 + B + \cdots + B^{m-1})^{2(d+1)}$ and hence simplifies the construction of $A_m^d(d)$ as illustrated in the following example.

EXAMPLE 20.3 To illustrate the above results, let us consider the number of monthly unemployed young women per 1000 persons between the ages of 16 and 19 in the United States from January 1961 to August 2002, which was examined in Chapters 6 and 7. For convenience, Tables 20.1 and 20.2 repeat the sample autocorrelation functions of the original series and its first differences. It is evident that the original series z_t is nonstationary and that the series of its first differences $(1 - B)z_t$ is stationary following an MA(1) model.

To examine the limiting behavior of the autocorrelation function of the aggregates, we compute the sample autocorrelations of Z_T for $m = 6$ and $m = 12$ corresponding to the semi-annual and annual total of unemployed young women. The autocorrelations for Z_T show nonstationarity. Table 20.3 shows the autocorrelations of the differenced aggregates, i.e.,

$$U_T = (1 - B)Z_T.$$

TABLE 20.3 Sample autocorrelations $\hat{\rho}_U(k)$ of $\{(1 - B)Z_T\} = \{U_T\}$

				$m = 6$				
k	1	2	3	4	5	6	7	8
$\hat{\rho}_U(k)$	.34	.03	.05	−.20	−.12	−.10	−.16	−.01
				$m = 12$				
k	1	2	3	4	5	6	7	8
$\hat{\rho}_U(k)$	.25	−.21	−.22	−.03	.19	.08	.21	−.17

Is the above phenomenon expected? To answer this question, let us derive the invertible autocorrelations $\rho_U^{(\infty)}(0)$ and $\rho_U^{(\infty)}(1)$ of the limiting aggregate model

$$(1 - B)Z_T = \theta_1^{(\infty)}(B)A_T.$$

Take $m = 2$ and consider the limit of m^k as $k \to \infty$. For $m = 2$ and $d = 1$, (20.1.7) gives $\gamma_U(k) = (1 + B)^4\gamma_w(2k + 2)$, and (20.1.8) and (20.1.9) imply that

$$A = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 & 0 & 0 \\ 0 & 0 & 1 & 4 & 6 & 4 & 1 \end{bmatrix},$$

$$A_2^1 = \begin{bmatrix} 6 & 8 & 2 & 0 & 0 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}.$$

Hence, by (20.1.16),

$$A_2^1(1) = \begin{bmatrix} 6 & 8 \\ 1 & 4 \end{bmatrix}.$$

The characteristic equation, $\det(A_2^1(1) - \lambda I) = 0$, is

$$\lambda^2 - 10\lambda + 16 = 0,$$

and the eigenvalues are easily seen to be 2 and 8. Thus, $\lambda_{\max} = 8$ and its corresponding eigenvector is the solution of the system of equations, $(A_2^1(1) - 8I)\mathbf{x} = \mathbf{0}$, i.e.,

$$\begin{bmatrix} -2 & 8 \\ 1 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Because $\operatorname{rank}(A_2^1(1) - 8I) = 1$, there is one free variable, and it can easily be shown that an associated eigenvector is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ \frac{1}{4}x_1 \end{bmatrix}.$$

Thus, the autocorrelation structure of the limiting aggregate model becomes

$$\boldsymbol{\rho}_U^{(\infty)} = \begin{bmatrix} \rho_U^{(\infty)}(0) \\ \rho_U^{(\infty)}(1) \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{4} \end{bmatrix}.$$

Thus, the result shown in Table 20.3 is expected from the limiting behavior of temporal aggregates.

This result may not be evident purely based on empirical values. For example, with 500 observations of z_t, Table 20.2 clearly implies an IMA(1, 1) model for z_t. The result with $m = 6$, as shown in Table 20.3, also implies an IMA(1, 1) model for the aggregates Z_T. When $m = 12$, the

standard deviation of $\hat{\rho}_U(k)$ is approximately equal to .2; hence, as shown in Table 20.3, the sample autocorrelation for this case indicates a white noise phenomenon. Even though the value of $\hat{\rho}_w(1)$ is $-.41$ for $w_t = (1 - B)z_t$, however, the value of $\hat{\rho}_U(1)$ is .34 for $U_T = (1 - B)Z_T$ when $m = 6$ and becomes .25 when $m = 12$. This reduction in $\hat{\rho}_U(1)$ is a direct result of temporal aggregation.

In summary, temporal aggregation may complicate model structure. A simple AR process becomes a mixed ARMA model upon aggregation. As the order of aggregation becomes large, however, it also tends to simplify the model form.

20.2 The Effects of Aggregation on Forecasting and Parameter Estimation

20.2.1 Hilbert Space

A linear vector space $\mathcal{L}$ is a nonempty set that is closed under vector addition and scalar multiplication such that for any elements x, y, z in $\mathcal{L}$ and any real numbers a and b, the operations satisfy the following properties:

I. The operation of addition satisfies
   1. $x + y = y + x$.
   2. $(x + y) + z = x + (y + z)$.
   3. There exists a unique element 0 in $\mathcal{L}$ such that $0 + x = x$.
   4. For each x in $\mathcal{L}$, there exists a unique $-x$ in $\mathcal{L}$ such that $x + (-x) = 0$.

II. The operation of scalar multiplication satisfies
   1. $a(bx) = (ab)x$.
   2. $1x = x$.

III. The operation of addition and scalar multiplication satisfies
   1. $a(x + y) = ax + ay$.
   2. $(a + b)x = ax + bx$.

An inner product is a real-valued function on a space $\mathcal{L}$ denoted by (x, y), such that the following are true:

1. $(x, y) = (y, x)$.
2. $(x + y, z) = (x, z) + (y, z)$.
3. $(ax, y) = a(x, y)$ for any real number a.
4. $(x, x) \geq 0$ and $(x, x) = 0$ if and only if $x = 0$.

Two elements x and y are said to be orthogonal if $(x, y) = 0$. From an inner product, we define $\sqrt{(x, x)}$ as a norm of x, often denoted by $[x[$. It satisfies the following conditions:

1. $[x[\geq 0$ and $[x[= 0$ if and only if $x = 0$.
2. $[ax[= |a|[x[$.
3. $[x + y[\leq [x[+ [y[$.

Then $[x - y[$ is defined as the distance between x and y and is called the metric of the space. It is easy to see that the finite linear combination of zero mean random variables is a linear vector space and $E(XY)$ is an inner product.

A sequence x_n is said to be a Cauchy sequence if

$$\lim_{\substack{n\to\infty \\ m\to\infty}} [x_n - x_m[= 0. \tag{20.2.1}$$

A Hilbert space, $\mathcal{H}$, is a complete inner product linear vector space in the sense that the limit of every Cauchy sequence is also an element in the space so that the space is closed. Suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are subspaces of a Hilbert space $\mathcal{H}$. Then the sum of $\mathcal{H}_1$ and $\mathcal{H}_2$ is defined as the set

$$\mathcal{H}_1 + \mathcal{H}_2 = \{x + y : x \in \mathcal{H}_1 \text{ and } y \in \mathcal{H}_2\}. \tag{20.2.2}$$

If $\mathcal{H}_1$ and $\mathcal{H}_2$ are orthogonal subspaces in the sense that every element $x \in \mathcal{H}_1$ is orthogonal to every element $y \in \mathcal{H}_2$, then $\mathcal{H}_1 + \mathcal{H}_2$ is closed and hence is a Hilbert space. In this case, we will write the sum as $\mathcal{H}_1 \oplus \mathcal{H}_2$. We have the following useful theorem (for the proof see Halmos [1951]):

Projection Theorem If $\mathcal{H}_0$ is a subspace of a Hilbert space $\mathcal{H}$, then $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^{\perp}$, where $\mathcal{H}_0^{\perp}$ is the orthogonal complement of $\mathcal{H}_0$. Every element $x \in \mathcal{H}$ is uniquely expressed as

$$x = p(x) + [x - p(x)], \tag{20.2.3}$$

where $p(x) \in \mathcal{H}_0$, $[x - p(x)] \in \mathcal{H}_0^{\perp}$ and $[x - p(x)[= \min_{y\in\mathcal{H}_0} [x - y[$. The point $p(x)$ is called the orthogonal projection of x on $\mathcal{H}_0$, and $[x - p(x)]$ is the residual of x from $\mathcal{H}_0$. ■

20.2.2 The Application of Hilbert Space in Forecasting

A linear manifold is a nonempty subset of a Hilbert space such that if x and y are in the subset, then $ax + by$ is also in the subset for any real numbers a and b. Let z_t be a zero mean stationary process. Under the inner product $(z_t, x_t) = E(z_t x_t)$, it is known that the closed linear manifold $\mathcal{H}$, denoted by $\mathcal{H} = \mathcal{L}\{z_t : t = 0, \pm 1, \pm 2, \ldots\}$, is a Hilbert space (see Parzen [1961a] Anderson [1971, Section 7.6], or Koopmans [1971, Section 7.4]). $E(z_t - x_t)^2$ is the distance norm between the elements z_t and x_t in the space. In time series analysis, we are interested in forecasting z_{N+l}, for some $l \geq 1$, based on the knowledge of past history $\{z_t : t \leq N\}$. Consider the closed linear manifold $\mathcal{L}\{z_t : t \leq N\}$. It is clear that $\mathcal{L}\{z_t : t \leq N\}$ is a subspace of $\mathcal{H}$. Thus, finding the best linear forecast of z_{N+l}, which is an element in $\mathcal{H}$, based on its past history, is equivalent to finding the element $\hat{z}_N(l)$ in $\mathcal{L}\{z_t : t \leq N\}$ such that the mean squared error

$$E(z_{N+l} - \hat{z}_N(l))^2 = \min_{x\in\mathcal{L}\{z_t:t\leq N\}} E(z_{N+l} - x)^2. \tag{20.2.4}$$

By the above Projection Theorem, the best linear forecast of z_{N+l} is simply its orthogonal projection on the subspace $\mathcal{L}\{z_t : t \leq N\}$. Under the minimum mean square error criterion deduced from the inner product $E(z_t x_t)$, this orthogonal projection is the same as the

conditional expectation $E[z_{N+l}|z_t:t \leq N]$. Also, by the same theorem, the forecast error $(z_{N+l} - \hat{z}_N(l))$ is in the space $\mathcal{L}^{\perp}\{z_t:t \leq N\}$ and hence is orthogonal to the subspace $\mathcal{L}\{z_t:t \leq N\}$.

20.2.3 The Effect of Temporal Aggregation on Forecasting

Suppose that we are interested in forecasting a future aggregate Z_{N+l} at time N. If both nonaggregate and aggregate series are available, then this forecast can be derived from the nonaggregate model of z_t or from the aggregate model of Z_T. It is of interest to examine the relative efficiency between the two forecasts. Let

$$\mathcal{H}_d = \mathcal{L}\{z_t:t = 0, \pm 1, \pm 2, \ldots\}$$

be the closed linear manifold spanned by the nonaggregate series z_t and

$$\mathcal{H}_a = \mathcal{L}\{Z_T:T = 0, \pm 1, \pm 2, \ldots\}$$

be the closed linear manifold spanned by the mth-order aggregate series Z_T. Let

$$\mathcal{H}_d^p = \mathcal{L}\{z_t:t \leq mN\}$$

be the closed linear manifold spanned by the past history of the nonaggregate series $\{z_{mN}, z_{mN-1}, \ldots\}$, and

$$\mathcal{H}_a^p = \mathcal{L}\{Z_T:T \leq N\}$$

be the closed linear manifold spanned by the past history of the aggregate series $\{Z_N, Z_{N-1}, \ldots\}$. It is easy to see that $\mathcal{H}_a$ is a subspace of $\mathcal{H}_d$ and that $\mathcal{H}_a^p$ is a subspace of $\mathcal{H}_d^p$. Thus, the best linear forecasts of Z_{N+l} based on the nonaggregate process z_t and the aggregate series Z_T are simply the projections of Z_{N+l} on the two subspaces $\mathcal{H}_d^p$ and $\mathcal{H}_a^p$. Thus, letting $\hat{Z}_N(l)$ and $\tilde{Z}_N(l)$ be the best linear forecasts of Z_{N+l}, $l \geq 1$, based on the nonaggregate process z_t and the aggregate process Z_T, respectively, we have

$$\begin{aligned} E[Z_{N+l} - \hat{Z}_N(l)]^2 &= \min_{x \in \mathcal{H}_d^p} E(Z_{N+l} - x)^2 \\ &\leq \min_{Y \in \mathcal{H}_a^p} E(Z_{N+l} - Y)^2 \\ &= E(Z_{N+l} - \tilde{Z}_N(l))^2. \end{aligned} \tag{20.2.5}$$

Thus, forecasting from an aggregate model is in general less efficient than forecasting from a nonaggregate model.

Suppose that the nonaggregate series z_t follows a general multiplicative seasonal ARIMA model in (20.1.32) and hence the aggregate series Z_T follows the model in (20.1.33). The

variances of the forecast errors based on the two models can be explicitly calculated. Specifically, the variance of the forecast error based on the aggregate model is given by

$$\text{Var}(Z_{N+l} - \tilde{Z}_N(l)) = \sigma_A^2 \sum_{j=0}^{l-1} \Psi_j^2, \tag{20.2.6}$$

where $\Psi_0 = 1$ and the Ψ_j's can be calculated from the aggregate model (20.1.33) using the method discussed in Section 5.2. To calculate the variance of the forecast error from the nonaggregate model, we first note that

$$\text{Var}(z_{t+h} - \hat{z}_t(h)) = \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2, \tag{20.2.7}$$

where $\psi_0 = 1$ and the ψ_j's are obtained similarly but based on the nonaggregate model (20.1.32). Let

$$\mathcal{I}_t = \left(\sum_{j=0}^{m-1} B^j\right) z_t.$$

It was shown in Wei (1978b) that the variance of the forecast error based on the nonaggregate model is

$$\begin{aligned} \text{Var}(Z_{N+l} - \hat{Z}_N(l)) &= \text{Var}(\mathcal{I}_{mN+ml} - \hat{\mathcal{I}}_{mN}(ml)) \\ &= \sigma_a^2 \sum_{j=0}^{ml-1} \varphi_j^2, \end{aligned} \tag{20.2.8}$$

where

$$\varphi_j = \sum_{i=0}^{m-1} \psi_{j-i}.$$

The relative efficiency of forecasting future aggregates using the aggregate model can be measured by the variance ratio

$$\zeta(m, l) = \frac{\text{Var}(Z_{N+l} - \hat{Z}_N(l))}{\text{Var}(Z_{N+l} - \tilde{Z}_N(l))}, \tag{20.2.9}$$

where m is used to indicate the order of aggregation. Because $\hat{Z}_N(l)$ and $\tilde{Z}_N(l)$ are obviously unbiased forecasts for Z_{N+l}, (20.2.5) implies that $0 \leq \zeta(m, l) \leq 1$. In fact, it was shown by Wei (1978b) that $\zeta(m, l)$ has the following properties:

1. $0 \leq \zeta(m, l) \leq 1$ for all m and l.
2. $\zeta(l) = \lim_{m\to\infty} \zeta(m, l) = 1$ for all l, if $d = 0$.
3. $\zeta(l) << 1$, and $\zeta(l) \to 1$ only when $l \to \infty$, if $d > 0$.
4. $\zeta(m, l) = 1$ for all m and l if $p = d = q = 0$ and $s = mS$ for some integer S.

In summary, the above results imply that insofar as forecasting the future aggregates is concerned, the loss in efficiency through aggregation can be substantial if the nonseasonal component of the model is nonstationary. The loss in efficiency is less serious for long-term forecasting, particularly when the nonseasonal component is stationary. There is no loss in efficiency due to aggregation if the basic model is a purely seasonal process, and the seasonal periods s and S for the basic and aggregate models are related by $s = mS$ where m is the order of aggregation.

20.2.4 Information Loss Due to Aggregation in Parameter Estimation

For a given time series model, let $\boldsymbol{\eta}$ be the $k \times 1$ vector of all the k parameters. It often happens that over the relevant region of the parameter space, the log-likelihood function is approximately quadratic in the elements of $\boldsymbol{\eta}$ so that

$$l(\boldsymbol{\eta}) \equiv \ln L(\boldsymbol{\eta}) \simeq l(\hat{\boldsymbol{\eta}}) + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} l_{ij}(\eta_i - \hat{\eta}_i)(\eta_j - \hat{\eta}_j), \tag{20.2.10}$$

where the partial derivatives

$$l_{ij} = \frac{\partial^2 l(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j} \tag{20.2.11}$$

are constant. For large samples, the variance-covariance matrix $V(\hat{\boldsymbol{\eta}})$ for the maximum likelihood estimator $\hat{\boldsymbol{\eta}}$ is given by

$$V(\hat{\boldsymbol{\eta}}) \simeq \boldsymbol{I}^{-1}(\boldsymbol{\eta}), \tag{20.2.12}$$

where $I(\boldsymbol{\eta}) = -E[l_{ij}]$ is the information matrix for the parameter $\boldsymbol{\eta}$. Using (7.2.10), we have

$$l(\boldsymbol{\eta}) = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S(\boldsymbol{\eta})}{2\sigma_a^2}. \tag{20.2.13}$$

Hence,

$$l_{ij} \simeq -\frac{1}{2\sigma_a^2}\frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j} \tag{20.2.14}$$

and

$$\begin{aligned} V(\hat{\boldsymbol{\eta}}) &\simeq 2\sigma_a^2\left(E\left[\frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}\right]\right)^{-1} \\ &\simeq 2\sigma_a^2\left[\frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}\right]^{-1}, \end{aligned} \tag{20.2.15}$$

where we use that, in the quadratic approximation (20.2.10), the partial derivatives l_{ij} are constant over the region of approximation.

Let $\hat{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\eta}}$ be the maximum likelihood estimators of $\boldsymbol{\eta}$ based on the nonaggregate model and the aggregate model, respectively. Then

$$V(\hat{\boldsymbol{\eta}}) = I_d^{-1}(\boldsymbol{\eta})$$

and

$$V(\tilde{\boldsymbol{\eta}}) = I_a^{-1}(\boldsymbol{\eta}),$$

where $I_d(\boldsymbol{\eta})$ and $I_a(\boldsymbol{\eta})$ are the corresponding information matrices based on the nonaggregate and the aggregate models, respectively. Define

$$\tau(m) = 1 - \frac{\det V(\hat{\boldsymbol{\eta}})}{\det V(\tilde{\boldsymbol{\eta}})} = 1 - \frac{\det I_a(\boldsymbol{\eta})}{\det I_d(\boldsymbol{\eta})}. \tag{20.2.16}$$

We can use $\tau(m)$ to measure the information loss in estimation due to aggregation. Unfortunately, the relationship between the parameters in the aggregate and nonaggregate models is extremely complicated; hence, the derivation of $\tau(m)$ for a general model is difficult. As shown in Section 20.1.4, however, if the nonaggregate model of z_t is a purely seasonal model

$$(1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps})(1 - B^s)^D z_t = (1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs})a_t, \tag{20.2.17}$$

with the a_t being i.i.d. $N(0, \sigma_a^2)$, then the aggregate model for the mth-order aggregate Z_T with $s = mS$ for some integer S becomes

$$(1 - \Phi_1 B^S - \cdots - \Phi_P B^{PS})(1 - B^S)^D Z_T = (1 - \Theta_1 B^S - \cdots - \Theta_Q B^{QS})A_T, \tag{20.2.18}$$

where the A_T's are i.i.d. $N(0, m\sigma_a^2)$. It follows that

$$I_a(\boldsymbol{\eta}) = \frac{1}{m} I_d(\boldsymbol{\eta})$$

and hence

$$\tau(m) = 1 - m^{-(P+Q)} \tag{20.2.19}$$

To see the implication of (20.2.19), assume that $P = 1$, $Q = 0$, and the nonaggregate series is a monthly series with $s = 12$. We have $\tau(2) = \frac{1}{2}$, $\tau(3) = \frac{2}{3}$, $\tau(6) = \frac{5}{6}$, and $\tau(12) = \frac{11}{12}$. Thus, when estimating the parameters in the underlying model, temporal aggregation leads to a tremendous loss in information. In fact, $\tau(m)$ in (20.2.19) is an increasing function of m and the number of parameters in the model. The higher the order of aggregation and the more parameters in the model, the more serious is the information loss in estimation.

20.3 Systematic Sampling of the ARIMA Process

In the previous sections, we discussed temporal aggregation of a flow variable. We now study the phenomenon of a stock variable. The value of a stock variable such as the price of a given commodity exists at every time point. In practice, however, it is rarely the case that every possible value of a stock variable is observed and recorded. Instead, the observed series is often only a subseries obtained by a scheme of systematic sampling. For example, in the stock market, the often-reported series is the daily closing price. In quality control, only every mth item is inspected. We examine the relationship between the underlying process z_t and the model based on the sampled subseries y_T, where $y_T = z_{mT}$ and m is the systematic sampling interval.

With no loss of generality, assume that the underlying process z_t follows the ARIMA(p, d, q) process

$$\phi_p(B)(1 - B)^d z_t = \theta_q(B)a_t, \tag{20.3.1}$$

where $\phi_p(B)$ and $\theta_q(B)$ are stationary and invertible AR and MA polynomials, respectively, and the a_t's are white noise series with mean 0 and variance σ_a^2. Let $\delta_1, \delta_2, \ldots,$ and δ_p be the reciprocal of the roots of $\phi_p(B)$. Then, (20.3.1) can be written as

$$\prod_{j=1}^{p}(1 - \delta_j B)(1 - B)^d z_t = \theta_q(B)a_t.$$

Multiplying both sides of the above equation by

$$\prod_{j=1}^{p}\left(\frac{1 - \delta_j^m B^m}{1 - \delta_j B}\right)\left(\frac{1 - B^m}{1 - B}\right)^d,$$

we get

$$\prod_{j=1}^{p}(1 - \delta_j^m B^m)(1 - B^m)^d z_t = \prod_{j=1}^{p}\left(\frac{1 - \delta_j^m B^m}{1 - \delta_j B}\right)\left(\frac{1 - B^m}{1 - B}\right)^d \theta_q(B)a_t.$$

Let $V_t = \prod_{j=1}^{p}(1 - \delta_j^m B^m)(1 - B^m)^d z_t$, which is now a moving average $MA[(p + d)m + (q - p - d)]$ process. Let

$$\begin{aligned} C_T = V_{mT} &= \prod_{j=1}^{p}(1 - \delta_j^m B^m)(1 - B^m)^d z_{mT} \\ &= \prod_{j=1}^{p}(1 - \delta_j^m \mathcal{B})(1 - \mathcal{B})^d y_T, \end{aligned}$$

where $\mathcal{B} = B^m$. Note that $\mathcal{B}y_T = B^m z_{mT} = z_{mT-m} = z_{m(T-1)} = y_{T-1}$. Hence, $\mathcal{B}$ is the backshift operator on y_T. Now, $E(C_T C_{T-j}) = E(V_{mT}V_{mT-mj})$, which is nonzero only when

$mj \leq \{(p + d)m + (q - p - d)\}$ or $j \leq [(p + d) + (q - p - d)/m]$. Thus, the series y_T follows an ARIMA(p, d, r) process

$$(1 - \alpha_1 \mathcal{B} - \cdots - \alpha_p \mathcal{B}^p)(1 - \mathcal{B})^d y_T = (1 - \beta_1 \mathcal{B} - \cdots - \beta_r \mathcal{B}^r) e_T, \quad (20.3.2)$$

where $r \leq [(p + d) + (q - p - d)/m]$, the e_T's are white noise with mean zero and variance σ_e^2, the α_i's are functions of ϕ_i's, and the β_j's and the σ_e^2 are functions of ϕ_i's, θ_j's, and σ_a^2.

When $d = 0$, (20.3.2) reduces to a result obtained by Brewer (1973). That is, the subseries systematically sampled from a stationary ARMA(p, q) process becomes an ARMA(p, r) process with $r \leq [p + (q - p)/m]$. To study the limiting behavior of the sampled subseries, we consider first the case in which the underlying process z_t follows an integrated moving average IMA(d, q) process

$$(1 - B)^d z_t = \theta_q(B) a_t. \quad (20.3.3)$$

Let $w_t = (1 - B)^d z_t$. The autocovariance generating function, $g_w(B)$, of w_t is given by

$$g_w(B) = \sigma_a^2 \theta_q(B) \theta_q(B^{-1}) = \sum_{j=-q}^{q} \gamma_w(j) B^j,$$

where $\gamma_w(j)$ is the jth autocovariance of the w_t series. Let

$$v_t = (1 - B^m)^d z_t = \left(\frac{1 - B^m}{1 - B}\right)^d \theta_q(B) a_t.$$

We obtain the autocovariance generating function of v_t as

$$\begin{aligned} g_v(B) &= \left(\frac{1 - B^m}{1 - B}\right)^d \left(\frac{1 - B^{-m}}{1 - B^{-1}}\right)^d g_w(B) \\ &= \sum_{j=-q}^{q} \gamma_w(j) \sum_{i=0}^{2d} \binom{2d}{i} (-1)^i B^{-(d-i)m+d+j} \sum_{l=0}^{\infty} \binom{l + 2d - 1}{2d - 1} B^l, \end{aligned}$$

where $\binom{n}{r} = n!/(r!(n - r)!)$. Now let

$$C_T = v_{mT} = (1 - B^m)^d z_{mT} = (1 - \mathcal{B})^d y_T.$$

Then,

$$\begin{aligned} \gamma_C(n) &= \gamma_v(nm) \\ &= \sum_{j=-q}^{q} \gamma_w(j) \sum_{i=0}^{2d} \binom{2d}{i} (-1)^i \binom{(d - n - i)m + d - j - 1}{2d - 1}, \end{aligned} \quad (20.3.4)$$

where $\binom{n}{r} = 0$ for $n < r$.

It is easy to show that for $m > q$

$$\gamma_C(n) = \begin{cases} \sum_{j=-q}^{q} \gamma_w(j) \sum_{i=0}^{d-n-1} \binom{2d}{i} (-1)^i \binom{(d-n-i)m + d - j - 1}{2d-1} + o(m) \\ \qquad\qquad \text{for } n = 0, \ldots, (d-1), \\ 0 \qquad\qquad \text{for } n \geq d. \end{cases}$$

Thus, as m increases, the subseries y_T approaches an IMA(d, $d-1$) process.

Because an ARIMA(p, d, q) process can be approximated with any desired accuracy by an IMA process, the above limiting result holds also for any homogeneous autoregressive integrated moving average process. As expected, when $d = 0$, the limiting model from a stationary process becomes white noise.

In summary, a subseries obtained by systematically taking every mth item from an ARIMA(p, d, q) process follows an ARIMA(p, d, r) model where r is less than or equal to the integer part of $((p + d) + (q - p - d)/m)$. When m becomes large, it approaches an IMA(d, $d-1$) process. In particular, the sampled subseries from an ARIMA(p, 1, q) process approaches a simple random walk model. The random walk model has been widely used to describe the behavior of stock prices. It would be interesting to investigate whether the random walk model is indeed the underlying model for the stock prices or simply a consequence of systematic sampling.

Systematic sampling of a stock variable is an important special case of a more general problem known as missing observations in the literature. The results presented in this section are taken from Wei (1981) and mainly deal with the effect of systematic sampling on model structure. Other references on this topic include Quenouille (1958), Brewer (1973), Werner (1982), and Weiss (1984). Once the resulting model of a sampled subseries is known, the effects of systematic sampling on parameter estimation and forecasting can be examined analogously to the effects of temporal aggregation discussed in Section 20.2; we do not repeat them here. For reference, we refer readers to Zellner (1966) and Jones (1980) for the effect on parameter estimation of missing observations from ARMA models, and to Wei and Tiao (1975) and Lütkepohl (1987) for the effect of systematic sampling on forecasting.

20.4 The Effects of Systematic Sampling and Temporal Aggregation on Causality

20.4.1 Decomposition of Linear Relationship between Two Time Series

The study of causal relationships among a set of time series variables has been one of the most important problems in the literature. In an interesting paper, Geweke (1982) derived a decomposition of linear relationships between two time series variables x_t and y_t. By denoting $F_{x,y}$ as the linear relationship between x_t and y_t, $F_{x \to y}$ as the linear causality from x_t to y_t, $F_{y \to x}$ as the linear causality from y_t to x_t, and $F_{x \cdot y}$ as the instantaneous linear causality between x_t and y_t, he showed that $F_{x,y} = F_{x \to y} + F_{y \to x} + F_{x \cdot y}$.

Consider a bivariate time series x_t and y_t with the invertible moving average representation

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{bmatrix} \begin{bmatrix} a_t \\ b_t \end{bmatrix}, \tag{20.4.1}$$

where $[a_t, b_t]'$ is a bivariate white noise series with mean $\mathbf{0}$ and variance-covariance matrix

$$\sum = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}. \tag{20.4.2}$$

Now, let us consider the linear projections of y_t on the following different subspaces:

1. The linear projection of y_t on $\mathcal{L}\{y_j : j < t\}$:

$$y_t = \sum_{j=1}^{\infty} \pi_{1j} y_{t-j} + u_{1t}, \tag{20.4.3}$$

where u_{1t} is a white noise series with mean 0 and variance σ_1^2. Note that the marginal series y_t derived from (20.4.1) is

$$y_t = \theta_{11}(B)a_t + \theta_{12}(B)b_t = \alpha(B)u_{1t} \tag{20.4.4}$$

for some $\alpha(B)$ such that u_{1t} is white noise. The projection (20.4.3) is obtained by inverting (20.4.4) into an autoregressive form.

2. The linear projection of y_t on $\mathcal{L}\{\{y_j : j < t\} \cup \{x_j : j < t\}\}$:

$$y_t = \sum_{j=1}^{\infty} \pi_{2j} y_{t-j} + \sum_{j=1}^{\infty} \lambda_{2j} x_{t-j} + u_{2t}, \tag{20.4.5}$$

where u_{2t} is a white noise series with mean 0 and variance σ_2^2. By inverting (20.4.1) into an autoregressive representation

$$\begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_t \\ b_t \end{bmatrix}, \tag{20.4.6}$$

where

$$\begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{bmatrix} = \begin{bmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{bmatrix}^{-1},$$

the representation (20.4.5) is obtained from the first component of the joint model in (20.4.6). Hence, $u_{2t} = a_t$ and $\sigma_2^2 = \sigma_a^2$.

3. The linear projection of y_t on $\mathcal{L}\{\{y_j : j < t\} \cup \{x_j : j \leq t\}\}$:

$$y_t = \sum_{j=1}^{\infty} \pi_{3j} y_{t-j} + \sum_{j=0}^{\infty} \lambda_{3j} x_{t-j} + u_{3t}, \tag{20.4.7}$$

where u_{3t} is a white noise series with mean 0 and variance σ_3^2. By premultiplying (20.4.6) by the matrix, which diagonalizes the variance-covariance matrix given in (20.4.2), that is, by premultiplying the matrix

$$\begin{bmatrix} 1 & -\sigma_{ab}/\sigma_b^2 \\ -\sigma_{ab}/\sigma_a^2 & 1 \end{bmatrix},$$

the projection (20.4.7) is obtained from the first component of the resulting system. Hence, $u_{3t} = a_t - (\sigma_{ab}/\sigma_b^2) b_t$.

4. The projection of y_t on $\mathcal{L}\{\{y_j : j < t\} \cup \{x_j : -\infty < j < \infty\}\}$:

$$y_t = \sum_{j=1}^{\infty} \pi_{4j} y_{t-j} + \sum_{j=-\infty}^{\infty} \lambda_{4j} x_{t-j} + u_{4t}, \tag{20.4.8}$$

where u_{4t} is a white noise series with mean 0 and variance σ_4^2. Let

$$\begin{bmatrix} G_{11}(B) & G_{12}(B) \\ G_{21}(B) & G_{22}(B) \end{bmatrix}$$

be the cross-covariance generating function of $[y_t, x_t]'$ given in (20.4.1). In terms of its linear projection on the space $\{x_j : -\infty < j < \infty\}$, y_t can be written as

$$y_t = \left[\frac{G_{12}(B)}{G_{22}(B)}\right] x_t + \psi(B) u_{4t}, \tag{20.4.9}$$

where u_{4t} is a white noise series with mean 0 and variance σ_4^2, and

$$\psi(B)\psi(B^{-1})\sigma_4^2 = G_{11}(B) - \frac{G_{12}(B)G_{21}(B)}{G_{22}(B)}.$$

The projection (20.4.8) is obtained by multiplying (20.4.9) by $\psi^{-1}(B)$. Define

$$\begin{cases} F_{x,y} = \ln(\sigma_1^2/\sigma_4^2), \\ F_{x\to y} = \ln(\sigma_1^2/\sigma_2^2), \\ F_{y\to x} = \ln(\sigma_3^2/\sigma_4^2), \\ F_{x\cdot y} = \ln(\sigma_2^2/\sigma_3^2). \end{cases} \tag{20.4.10}$$

It can be easily seen that

$$F_{x,y} = F_{x\to y} + F_{y\to x} + F_{x\cdot y}. \tag{20.4.11}$$

20.4.2 An Illustrative Underlying Model

For the purpose of illustration and discussion, suppose that on some theoretical grounds the output variable y_t is related to the input variable x_t in terms of the time unit t by

$$\begin{cases} y_t = \nu x_{t-1} + a_t \\ x_t = \phi x_{t-1} + b_t, \end{cases} \tag{20.4.12}$$

where a_t and b_t are zero mean white noise processes with variances σ_a^2 and σ_b^2, respectively, and they are independent of each other. The model can be written in the invertible bivariate moving average form

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & \nu B/(1-\phi B) \\ 0 & 1/(1-\phi B) \end{bmatrix} \begin{bmatrix} a_t \\ b_t \end{bmatrix}, \tag{20.4.13}$$

where $[a_t, b_t]'$ is white noise with mean $\mathbf{0}$ and variance-covariance matrix

$$\begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}.$$

Using (20.4.4), we have

$$y_t = \left[\frac{(1-\alpha B)}{(1-\phi B)}\right] u_{1t},$$

where α is obtained by solving the equation

$$\frac{\alpha}{1+\alpha^2} = \frac{\phi\sigma_a^2}{\nu^2\sigma_b^2 + (1+\phi^2)\sigma_a^2}$$

subject to $-1 < \alpha < 1$, which gives

$$\sigma_1^2 = \frac{\left[\nu^2\sigma_b^2 + (1+\phi^2)\sigma_a^2\right]}{(1+\alpha^2)}.$$

The one-step ahead forecast error variances for the other projections can be easily shown to be $\sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_a^2$. Hence, $F_{y\to x} = F_{x\cdot y} = 0$ and

$$F_{x,y} = F_{x\to y} = \ln\left(\frac{\nu^2\sigma_b^2 + (1+\phi^2)\sigma_a^2}{(1+\alpha^2)\sigma_a^2}\right).$$

TABLE 20.4 Linear causality for the underlying model in Equation (20.4.12).

	$F_{x,y}$	$F_{x\to y}$	$F_{y\to x}$	$F_{x\cdot y}$
$\phi = 0$	ln(2)	ln(2)	0	0
$\phi = .5$	ln(2.13)	ln(2.13)	0	0

That is, the linear relationship between x and y as represented in (20.4.12) contains only the one-sided linear causality from x_t to y_t. For numerical illustration, letting $\nu = 1, \sigma_a^2/\sigma_b^2 = 1$, we obtain Table 20.4 showing causality for the underlying model.

20.4.3 The Effects of Systematic Sampling and Temporal Aggregation on Causality

The Effect of Systematic Sampling on Causality Suppose that both y_t and x_t in (20.4.12) are stock variables and that only the subseries Y_T and X_T are obtained, where $Y_T = y_{mT}$ and $X_T = x_{mT}$, and m is the systematic sampling interval. It can be easily shown that for the model given in (20.4.12), the corresponding model for (Y_T, X_T) is

$$\begin{cases} Y_T = \nu\phi^{m-1}X_{T-1} + A_T \\ X_T = \phi^m X_{T-1} + E_T, \end{cases} \tag{20.4.14}$$

where A_T and E_T are two independent white noise series with mean 0 and variances

$$\sigma_A^2 = \nu^2\sigma_b^2 \sum_{j=0}^{m-2} (\phi^j)^2 + \sigma_a^2$$

and

$$\sigma_E^2 = \sigma_b^2 \sum_{j=0}^{m-1} (\phi^j)^2,$$

respectively. The corresponding bivariate moving average representation is

$$\begin{bmatrix} Y_T \\ X_T \end{bmatrix} = \begin{bmatrix} 1 & \nu\phi^{m-1}\mathcal{B}/(1-\phi^m\mathcal{B}) \\ 0 & 1/(1-\phi^m\mathcal{B}) \end{bmatrix} \begin{bmatrix} A_T \\ E_T \end{bmatrix}.$$

For $\nu = 1, \sigma_a^2/\sigma_b^2 = 1$, and $m = 3$, Table 20.5 gives the measures showing the effect of systematic sampling on causality.

Thus, the linear relationship between X_T and Y_T still contains only the one-sided linear causality from X_T to Y_T. Systematic sampling preserves the direction of the causal relationship among the variables. It weakens the relationship as shown by the magnitude of the coefficient of X_{T-1} given in (20.4.14), however.

TABLE 20.5 Effect of systematic sampling on causality.

	$F_{x,y}$	$F_{x\to y}$	$F_{y\to x}$	$F_{x\cdot y}$
$\phi = 0$	ln(2)	ln(2)	0	0
$\phi = .5$	ln(2.33)	ln(2.33)	0	0

The Effect of Temporal Aggregation on Causality Now suppose that both y_t and x_t in (20.4.12) are flow variables and the observed series are mth-order aggregates

$$Y_T = \left(\sum_{j=0}^{m-1} B^j\right) y_{mT}$$

and

$$X_T = \left(\sum_{j=0}^{m-1} B^j\right) x_{mT}.$$

For the given one-sided causal model in (20.4.12) or (20.4.13), the corresponding model, in terms of a moving average representation, for $\mathbf{Z}_T = [Y_T, X_T]'$, was shown in Tiao and Wei (1976) to be

$$\mathbf{Z}_T = \left\{\frac{1}{(1 - \eta\mathcal{B})}\right\}(I - H\mathcal{B})C_T, \tag{20.4.15}$$

where $\eta = \phi^m$, $\mathbf{C}_T$ is a bivariate white noise process with mean $\mathbf{0}$ and variance-covariance matrix Σ, and H and Σ are related such that

$$(I - H\mathcal{B})\Sigma(I - H'\mathcal{B}^{-1}) = \sigma_a^2\delta^{-1}G(\mathcal{B}),$$

with $\delta = \sigma_b^{-2}\sigma_a^2(1 - \phi)^2(1 - \phi^2)$ and the 2×2 symmetric matrix $G(\mathcal{B})$ having elements

$$\begin{aligned}
G_{12}(\mathcal{B}) &= \nu(\tau_0 + \tau_1\mathcal{B} + \tau_{-1}\mathcal{B}^{-1}),\\
G_{22}(\mathcal{B}) &= k_0 + k_1(\mathcal{B} + \mathcal{B}^{-1}),\\
G_{11}(\mathcal{B}) &= \nu^2 G_{22}(\mathcal{B}) + m\delta(1 - \eta\mathcal{B})(1 - \eta\mathcal{B}^{-1}),\\
\tau_0 &= m(1 - \phi^2)(1 + \eta^2) - (1 + \phi^2)(1 - \eta^2),\\
\tau_1 &= -m\eta(1 - \phi^2) + (1 - \eta)(1 + \phi^2\eta),\\
\tau_{-1} &= -m\eta(1 - \phi^2) + (1 - \eta)(\eta + \phi^2),\\
k_0 &= m(1 - \phi^2)(1 + \eta^2) - 2\phi(1 - \eta^2),
\end{aligned}$$

TABLE 20.6 Effect of temporal aggregation on causality.

	$F_{x,y}$	$F_{x\to y}$	$F_{y\to x}$	$F_{x\cdot y}$
$\phi = 0$	ln(1.42)	ln(1.08)	0	ln(1.31)
$\phi = .5$	ln(2.41)	ln(1.29)	ln(1.01)	ln(1.84)

and

$$k_1 = -m\eta(1 - \phi^2) + \phi(1 - \eta^2).$$

For $v = 1$, $\sigma_a^2/\sigma_b^2 = 1$, and $m = 3$, we obtain the linear relationship and causality between X_T and Y_T upon temporal aggregation for $\phi = 0$ and $\phi = .5$ as shown in Table 20.6.

Thus, as discussed in Tiao and Wei (1976), temporal aggregation turns the one-sided causal relationship from x_t to y_t into a pseudo two-sided feedback system. In fact, from Table 20.6 and Table 20.4, it is clear that upon aggregation the instantaneous linear causality between X_T and Y_T becomes a dominant force in the linear relationship between X_T and Y_T. This result is true even for the case where the input is white noise, i.e., when $\phi = 0$.

As pointed out by Wei (1982), the magnitude of the measures $F_{x,y}$, $F_{y\to x}$, $F_{x\to y}$, and $F_{x\cdot y}$ cannot be used for comparison unless the time units used are the same. For example, for the model given in (20.4.12), it was shown in Table 20.4 that $F_{x\to y} = \ln(2.13)$ for $\phi = .5$. The corresponding model (20.4.14) based on sampled variables Y_T and X_T for $m = 3$ gives $F_{x\to y} = \ln(2.33)$ for $\phi = .5$ as shown in Table 20.5. The latter magnitude of $F_{x\to y}$ is larger. The causal effect from X_T to Y_T in the latter case, however, is easily seen from (20.4.14) to be weaker because $|\phi| < 1$.

20.5 The Effects of Aggregation on Testing for Linearity and Normality

In this section, we study the effects of the use of time series aggregates in performing linearity and normality tests.

20.5.1 Testing for Linearity and Normality

One widely used linearity and normality test is the bispectral test. As described in Section 19.2.2, this test was proposed by Subba Rao and Gabr (1980, 1984). It was later modified by Hinich (1982). In this discussion, we use Hinich's modified bispectral test.

Let Z_t be a third-order stationary process. Recall from Section 19.2.2 that the bispectral density function is

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} C(k_1, k_2) e^{-ik_1\omega_1 - ik_2\omega_2}, \qquad -\pi \leq \omega_1, \omega_2 \leq \pi, \quad (20.5.1)$$

where $C(k_1, k_2)$ is the third-order cumulant of Z_t. The principal domain of $f(\omega_1, \omega_2)$, deduced from the properties of $C(k_1, k_2)$ is the triangular set $\Omega = \{0 \leq \omega_1 \leq \pi, \omega_2 \leq \omega_1, 2\omega_1 + \omega_2 \leq 2\pi\}$, and it is sufficient to consider only the frequencies in this region.

To estimate the bispectral density function based on a third-order stationary time series $Z_1, \ldots, Z_n$, we let $\omega_h = 2\pi h/n$ for $h = 0, 1, \ldots, n$ and for each pair of integers j and s define the set $D = \{0 < j \leq n/2, 0 < s \leq j, 2j + s \leq n\}$. Define also the lattice of points $L = \{(2u - 1)M/2, (2v - 1)M/2 : v = 1, 2, \ldots, u \text{ and } u \leq n/(2M) - v + \frac{3}{4}\}$ in the set D. The estimator of the bispectral density function is

$$\hat{f}(\omega_1, \omega_2) = \frac{1}{M^2} \sum_{j=(u-1)M}^{uM-1} \sum_{s=(v-1)M}^{vM-1} F(j, s), \quad (20.5.2)$$

where

$$F(j, s) = \frac{1}{(2\pi)^2 n} X(\omega_j) X(\omega_s) X^*(\omega_{j+s})$$

$$X(\omega_h) = \sum_{t=1}^{n} (Z_t - \bar{Z}) e^{-i\omega_h t}$$

is the finite Fourier transform of $Z_1, \ldots, Z_n$, $\bar{Z} = \sum_{t=1}^{n} Z_t/n$, and * denotes complex conjugate. Equation (20.5.2) shows that $\hat{f}(\omega_1, \omega_2)$ is constructed by averaging the $F(j, s)$ in a square of M^2 points with the centers of the squares being defined by the lattice of points L. The estimator $\hat{f}(\omega_1, \omega_2)$ is consistent if $M = n^c$ is rounded to the nearest integer with $\frac{1}{2} < c < 1$.

Let

$$\hat{V}_{u,v} = \frac{\hat{f}(u, v)}{[(n^{1-4c} R_{u,v}/2\pi) \hat{f}(\delta_u) \hat{f}(\delta_v) \hat{f}(\delta_{u+v})]^{1/2}}, \quad (20.5.3)$$

where $\delta_x = 2\pi(2x - 1)M/(2n)$, $\hat{f}(\omega)$ denotes the estimator of the spectral density function constructed by averaging M adjacent periodogram ordinates, and $R_{u,v}$ is the number of (j, s) in the square that are in D but not on the boundaries $j = s$ or $(2j + s) = n$ plus twice the number on these boundaries. If the square is in D, then $R_{u,v} = M^2$. Hinich (1982) shows that $2|\hat{V}_{u,v}|^2$ is approximately distributed as $\chi^2(2, \lambda_{u,v})$, i.e., as a noncentral chi-square variable with two degrees of freedom and noncentrality parameter

$$\lambda_{u,v} = \frac{2|f(\delta_u, \delta_v)|^2}{[(n^{1-4c} R_{u,v}/2\pi) f(\delta_u) f(\delta_v) f(\delta_{u+v})]}. \quad (20.5.4)$$

The bispectral test is based on the variables $2|\hat{V}_{u,v}|^2$. Because these variables are independent, the statistic

$$\hat{S} = 2 \sum_{u,\, v \in L} |\hat{V}_{u,\, v}|^2 \tag{20.5.5}$$

is approximately $\chi^2(2Q, \lambda)$ for large n, where $\lambda = \sum_{u,v \in L} \lambda_{u,v}$ and Q, the number of (u, v) in L, is approximately equal to $n^2/(12M^2)$.

Test of Linearity A zero mean stationary linear process z_t can be written in the form

$$z_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j} = \psi(B) a_t, \tag{20.5.6}$$

where a_t is a sequence of independent and identically distributed random variables with zero mean and variance $\mathrm{Var}(a_t) = \sigma_a^2$ and $E(a_t^3) = \mu_{3_a}$. Here, $\psi(x) = \sum_{j=-\infty}^{\infty} \psi_j x^j$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, and B is the backshift operator, $Bz_t = z_{t-1}$. From Sections 12.2.2 and 19.2.1, we have

$$f(\omega_1, \omega_2) = \frac{\mu_{3_a}}{(2\pi)^2} \psi(e^{i(\omega_1+\omega_2)}) \psi(e^{-i\omega_1}) \psi(e^{-i\omega_2}) \tag{20.5.7}$$

and

$$f(\omega) = \frac{\sigma_a^2}{2\pi} |\psi(e^{-i\omega})|^2. \tag{20.5.8}$$

We see from equations (20.5.7) and (20.5.8) that for a linear process z_t, the noncentrality parameter (20.5.4) is constant for all squares in D, i.e.,

$$\lambda_{u,\, v} = 2n^{2c-1} \frac{\mu_{3_a}^2}{\sigma_a^6} = \lambda_0, \tag{20.5.9}$$

where we note that $R_{u,v} = M^2$ and $M = n^c$. Because $2|\hat{V}_{u,\, v}|^2$ is independent and $E(|\hat{V}_{u,\, v}|^2) = 1 + \lambda_{u,\, v}/2$, the parameter λ_0 can be consistently estimated by

$$\hat{\lambda}_0 = \frac{2}{QM^2} \sum_{u,\, v \in L} R_{u,\, v} (|\hat{V}_{u,\, v}|^2 - 1). \tag{20.5.10}$$

Consequently, the distribution $\chi^2\,(2, \hat{\lambda}_0)$ converges to a $\chi^2\,(2, \lambda_0)$ as $n \to \infty$.

Thus, under the null hypothesis of linearity, the noncentrality parameter of the asymptotic distribution of the $2|\hat{V}_{u,v}|^2$ is constant for all u, v in L and squares wholly in D. On the other hand, if the null hypothesis is false, then the noncentrality parameter will be different for different values of u and v. As a result, the sample dispersion of the $2|\hat{V}_{u,v}|^2$ will be larger than that expected under the null hypothesis. Ashley et al. (1986) proposed using the 80% quantile of the $2|\hat{V}_{u,v}|^2$ to measure their dispersion. The sample 80% quantile $\hat{\xi}_{.8}$ has an asymptotic distribution $N(\xi_{0.8}, \sigma_\xi^2)$ with $\sigma_\xi^2 = .8(1 - .8)/[Q_D[g(\xi_{.8})]^2$, where $\xi_{.8}$ is the 80% quantile of the $\chi^2(2, \hat{\lambda}_0)$ distribution with density function g and Q_D denotes the number of squares wholly in D. Linearity is rejected if $\hat{\xi}_{.8}$ is significantly larger than $\xi_{.8}$.

Test of Normality To test a time series for normality, we first note that in (20.5.6), if a_t is Gaussian, then $\mu_{3_a} = 0$. Thus, its bispectral density function $f(\omega_1, \omega_2)$ is identically zero for all ω_1 and ω_2. Thus, testing a linear process for normality can be achieved by checking whether its bispectral density function is zero at all frequencies.

Under the null hypothesis, the statistic $\hat{S}$ defined in (20.5.5) is approximately $\chi^2(2Q, 0)$ for large n, i.e., a central chi-square and the hypothesis is rejected if $\hat{S} > \chi_\alpha^2$, where α is the significance level of the test. Rejection of the null hypothesis implies that the bispectral density function of Z_t is not identically zero. Thus, Z_t cannot be a linear Gaussian process.

20.5.2 The Effects of Temporal Aggregation on Testing for Linearity and Normality

Let Z_T be the nonoverlapping aggregates of z_t, i.e., $Z_T = (1 + B + \cdots + B^{m-1})z_{mT}$, where T is the aggregate time unit and m is the order of aggregation. The time series Z_T is the aggregate time series, and z_t is the unobserved nonaggregate time series.

For the given stationary linear Gaussian process z_t in (20.5.6), it is known from Tiao and Wei (1976) and Wei (1978b) that Z_T is also a stationary linear Gaussian process with the representation

$$Z_T = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{T-j} = \Psi(B)\varepsilon_T \tag{20.5.11}$$

where ε_T is a sequence of i.i.d. random variables with zero mean and variance $\text{Var}(\varepsilon_T) = \sigma_\varepsilon^2$ and $E(\varepsilon_T^3) = \mu_3$. Here, $\Psi(x) = \sum_{j=0}^{\infty} \Psi_j x^j$, $\sum_{j=0}^{\infty} |\Psi_j| < \infty$, and B is the backshift operator, $BZ_T = Z_{T-1}$. The parameters Ψ_j and σ_ε^2 are functions of the parameters ψ_j and σ_a^2 of the nonaggregate process in (20.5.6). Therefore, the tests based on the Z_T series follow also the same lines as for the z_t series.

Some Simple Nonlinear Time Series Models As introduced in Section 19.2.2, many nonlinear time series models have been proposed in the literature. To study the effects of aggregation on the linearity test, we will consider the following simple models:

1. Bilinear BL (1, 0, 1, 1) model,

$$z_t = .4 - .5z_{t-1} + .6z_{t-1}a_{t-1} + a_t.$$

2. NLAR(1) model,

$$z_t = (.004z_{t-1} + a_t)(.55z_{t-1}) + a_t.$$

3. NLMA(2) model,

$$z_t = a_t - .3a_{t-1} + .5a_{t-2} + .6a_t a_{t-2} - .2a_{t-1}^2.$$

4. Threshold TAR(2; 2, 3; 3) model,

$$z_t = \begin{cases} 2 - .3z_{t-1} + .5z_{t-2} + a_t, & \text{if } z_{t-3} \leq 1.5 \\ 2.5 + .2z_{t-1} - .5z_{t-2} - .7z_{t-3} + a_t, & \text{if } z_{t-3} > 1.5. \end{cases}$$

In the above models, a_t is a sequence of i.i.d. standard normal variables.

Aggregation Effects on the Test of Linearity To analyze the effects of aggregation on the test of linearity, Teles and Wei (2000) reported a simulation experiment in which 1000 time series of 12,000 observations each were generated from the above nonlinear models and were temporally aggregated with orders of aggregation $m = 2, 3, 6, 12$.

Recall that the noncentrality parameter $\lambda_{u,v}$ is constant for all u, v (for squares wholly in the principal domain) for linear models; therefore, the test is based on the constancy of this parameter. Because nonlinear models do not generally share this property, we first analyze how aggregation affects the dispersion of this parameter in the lattice in the principal domain. To this end, denoting the (sample) standard deviation of the estimates of the noncentrality parameters computed from the simulation by SD and SD_A for the nonaggregates and the aggregates, respectively, we show the values of $SD_A/(2n_A^{2c-1})$ in Table 20.7 and note that $SD_A = SD$ when $m = 1$. This ratio measures the impact of aggregation while taking into account the decrease in sample size. We use the value $c = 0.55$ for the smoothing parameter.

The results clearly show that the dispersion of the values of the noncentrality parameters decreases as a consequence of aggregation and that this effect generally increases with the order of aggregation. Thus, $\lambda_{u,v}$ tends to be constant as m increases. The effect of aggregation on the test statistic is severe even for low values of m. The reduction of the

TABLE 20.7 Aggregation effects on the dispersion of the noncentrality parameter.

	Order of aggregation						
Model	1	2	3	4	6	8	12
BL	18.020	8.692	4.793	4.835	2.708	2.144	1.593
NLAR	3.162	2.606	2.249	2.008	1.749	1.623	1.409
NLMA	1.626	0.734	0.577	0.556	0.563	0.569	0.568
TAR	3.306	1.601	0.990	0.988	0.863	0.858	0.899

TABLE 20.8 Aggregation effects on the power of the bispectral test.

	Order of aggregation						
Model	1	2	3	4	6	8	12
BL	0.990	0.281	0.148	0.095	0.027	0.019	0.008
NLAR	0.657	0.456	0.249	0.110	0.044	0.024	0.013
NLMA	0.747	0.005	0.001	0.002	0.003	0.000	0.000
TAR	0.798	0.416	0.011	0.013	0.000	0.001	0.000

variability of the values of the noncentrality parameter implies that the sample dispersion of the estimators $2|\hat{V}_{u,v}|^2$ decreases as a consequence of aggregation and therefore the hypothesis of linearity will be less rejected. Teles and Wei (2000) also investigated the power of the test with the results shown in Table 20.8 ($m = 1$ represents the situation of no aggregation, i.e., the test based on the nonaggregate time series).

We can draw two conclusions:

1. Aggregation causes the power of the test to decrease, and the higher the order of aggregation, the greater the loss.
2. The most striking conclusion is the abrupt loss of power that occurs for aggregate data. Even for the lowest order of aggregation, i.e., $m = 2$, the power is very low, reflecting the strong decrease in the variability of the noncentrality parameter noted above. Thus, the test of linearity is very sensitive to aggregation.

A similar study was conducted by Granger and Lee (1989) on some time domain tests. The effects of temporal aggregation we have obtained are much more severe than the results reported in their study.

Aggregation Effects on the Test of Normality Under the null hypothesis of Gaussianity, following (20.5.5), we denote the test statistic based on the aggregate series as $\hat{S}_A$, which is asymptotically $\chi^2(2Q_A, 0)$ for large $n_A = n/m$, where $Q_A \approx n_A^2/(12M_A^2) = m^{2(c-1)}n^2/(12M^2) \approx m^{2(c-1)}Q$ and $M_A = n_A^c$. Because $m \geq 2$ and $\frac{1}{2} < c < 1$, it implies that $Q_A < Q$. The null hypothesis is rejected for large values of $\hat{S}_A$.

Recall that if the process z_t is non-Gaussian, the distribution of the test statistic $\hat{S}$ is asymptotically χ^2 $(2Q, \lambda)$. Similarly, the statistic $\hat{S}_A$ will be asymptotically $\chi^2(2Q_A, \lambda_A)$. Thus, because $Q_A < Q$, aggregation leads to a decrease in the degrees of freedom of the asymptotic distribution of the test statistic as a consequence of the smaller number of observations in the aggregate time series. More important, it causes the noncentrality parameter to change. We now analyze how this parameter is affected.

For a linear non-Gaussian process z_t, we know that the noncentrality parameter of the asymptotic distribution of $2|\hat{V}_{u,v}|^2$ is constant as shown in expression (20.5.9), i.e., $\lambda_{u,v} = 2n^{2c-1}\,\mu_{3_a}^2/\sigma_a^6 = \lambda_0$ for all u, v and for all squares in D, where $\mu_{3_a}^2 \neq 0$. Consequently, $\lambda = \sum_{u,v}\lambda_0 = Q\lambda_0$, assuming for simplicity that all squares lie entirely in D.

Because Z_T is also a linear process, the corresponding noncentrality parameter is also constant:

$$\lambda_{A_0} = 2\left(\frac{n}{m}\right)^{2c-1}\frac{\mu_{3_\varepsilon}^2}{\sigma_\varepsilon^6}, \tag{20.5.12}$$

where $\mu_{3_\varepsilon} = E(\varepsilon_T^3)$ and $\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_T)$; therefore, $\lambda_A = Q_a\lambda_{A_0}$. Bearing in mind that $m \geq 2$ and $\frac{1}{2} < c < 1$, the above expression shows that the relationship between λ_{A_0} and λ_0 depends on the relationships between μ_{3_ε} and μ_{3_a} and between σ_ε and σ_a. For example, if Z_T is the mth-order aggregate series from the series z_t that follows a stationary and invertible ARMA(1, 1) model,

$$(1 - \phi B)z_t = (1 - \theta B)a_t,$$

with $|\phi| < 1$ and $|\theta| < 1$, Teles and Wei (2002) show that the relationship between the noncentrality parameters λ_0 and λ_{A_0} is given by

$$\lambda_{A_0} = \frac{\lambda_0}{m^{2c}}. \tag{20.5.13}$$

Because $\frac{1}{2} < c < 1$ and $m \geq 2$, expression (20.5.13) shows that the noncentrality parameter of the asymptotic distribution of $2|\hat{V}_{u,v}|^2$ decreases at a rate of m^{2c} as the order of aggregation increases. As a consequence of the reduction of λ_{A_0}, the noncentrality parameter of the distribution of the test statistic $\hat{S}_A$

$$\lambda_A = Q_a\lambda_{A_0} = Q_A\frac{\lambda_0}{m^{2c}} = \frac{1}{12}\left(\frac{n}{m}\right)^{2(1-c)}\frac{\lambda_0}{m^{2c}} = \frac{n^{2(1-c)}}{12}\frac{\lambda_0}{m^2} = Q\frac{\lambda_0}{m^2} = \frac{\lambda}{m^2} \tag{20.5.14}$$

also decreases. Thus, $\lambda_A < \lambda$, and the value of λ_A decreases at the rate of the square of the aggregation order. The decrease of the noncentrality parameter implies that temporal aggregation brings the approximate distribution of the test statistic closer to that valid under the null hypothesis, i.e., a central chi-square, for which the noncentrality parameter is zero. It also means that aggregation induces normality. Consequently, the null hypothesis will be less rejected when tested with the aggregate time series than with the basic series. The higher the order of aggregation, the stronger this effect will be, i.e., fewer rejections will happen.

20.6 The Effects of Aggregation on Testing for a Unit Root

To test for a unit root, we consider the AR(1) model

$$z_t = \phi z_{t-1} + a_t \tag{20.6.1}$$

and test the null hypothesis $H_0 : \phi = 1$ against the alternative $H_1 : \phi < 1$, where a_t is assumed to be a sequence of independent and identically distributed variables with zero mean and variance σ_a^2, and $z_0 = 0$.

Recall from Chapter 9 that the test statistic is $n(\hat{\phi} - 1)$, which is also commonly known as the normalized unit root estimator, where

$$\hat{\phi} = \frac{\sum_{t=2}^{n} z_{t-1} z_t}{\sum_{t=2}^{n} z_{t-1}^2}. \tag{20.6.2}$$

The limiting distribution of this statistic under the null hypothesis is given by

$$n(\hat{\phi} - 1) \xrightarrow{d} \frac{\frac{1}{2}[\{W(1)\}^2 - 1]}{\int_0^1 \{W(X)\} dX}, \tag{20.6.3}$$

where $W(u)$ represents the standard Wiener process. The Studentized statistic of $\hat{\phi}$, i.e., its conventional regression t statistic, is, under the null hypothesis,

$$T_{\hat{\phi}} = \frac{\hat{\phi} - 1}{S_{\hat{\phi}}} = \frac{\hat{\phi} - 1}{(\hat{\sigma}_a^2 / \sum_{t=2}^{n} z_{t-1}^2)^{1/2}} = \frac{n^{-1} \sum_{t=2}^{n} z_{t-1} a_t}{(n^{-2} \sum_{t=2}^{n} z_{t-1}^2)^{1/2} (\hat{\sigma}_a^2)^{1/2}}, \tag{20.6.4}$$

where $S_{\hat{\phi}}$ is the estimated standard deviation of $\hat{\phi}$ and $\hat{\sigma}_a^2$ is the OLS estimator of σ_a^2, i.e.,

$$\hat{\sigma}_a^2 = \frac{1}{n-2} \sum_{t=2}^{n} \hat{a}_t^2 = \frac{1}{n-2} \sum_{t=2}^{n} (z_t - \hat{\phi} z_{t-1})^2.$$

The limiting distribution of $T_{\hat{\phi}}$ under the null hypothesis is given by

$$T_{\hat{\phi}} \xrightarrow{d} \frac{\frac{1}{2}[\{W(1)\}^2 - 1]}{\left[\int_0^1 \{W(X)\}\, dX\right]^{1/2}}. \tag{20.6.5}$$

Because of the process of data collection or the practice of researchers, time series are frequently obtained through temporal aggregation. For example, the series commonly used in analysis and modeling are quarterly or annual totals. As a result, the series used in testing for a unit root are often time series aggregates. It is important to know the consequences of unit root testing when nonaggregate series are not available and aggregate data are used.

20.6.1 The Model of Aggregate Series

Under the null hypothesis $H_0 : \phi = 1$, model (20.6.1) becomes a random walk, $(1 - B)z_t = a_t$. The corresponding model for the aggregate series is shown by Teles and Wei (1999) as

$$Z_T = Z_{T-1} + \varepsilon_T - \Theta \varepsilon_{T-1}, \tag{20.6.6}$$

where the ε_T are independent and identically distributed variables with zero mean and variance σ_ε^2 and the parameters Θ and σ_ε^2 are determined as follows:

1. If $m = 1$, then $\Theta = 0$ and $\sigma_\varepsilon^2 = \sigma_a^2$.
2. If $m \geq 2$, then

$$\Theta = -\frac{2m^2 + 1}{m^2 - 1} + \left\{\left(\frac{2m^2 + 1}{m^2 - 1}\right)^2 - 1\right\}^{1/2}$$

$$\sigma_\varepsilon^2 = \sigma_a^2 \frac{m(2m^2 + 1)}{3(1 + \Theta^2)}. \tag{20.6.7}$$

These results imply that the aggregate series follows an ARIMA(0, 1, 1) model, which means that the unit root remains after temporal aggregation, i.e., the aggregate time series is still nonstationary. Moreover, as shown by (20.6.7), the moving average parameter Θ of the aggregate model is negative and it depends exclusively on the order of aggregation m. It is easy to see from that Θ is a decreasing function of m. Because $\Theta < 0$, we conclude that, when m increases, Θ increases in absolute value.

20.6.2 The Effects of Aggregation on the Distribution of the Test Statistics

To examine whether the aggregate series Z_T contains a unit root, from the results (20.6.6) and (20.6.7), we should test the null hypothesis $H_0: \Phi = 1$ against the alternative $H_1: \Phi < 1$ from the model

$$Z_T = \Phi Z_{T-1} + \zeta_T, \tag{20.6.8}$$

where ζ_T follows the MA(1) model $\zeta_T = \varepsilon_T - \Theta\varepsilon_{T-1}$, with the ε_T being i.i.d. with zero mean and variance σ_ε^2. The OLS estimator of Φ is

$$\hat{\Phi} = \frac{\sum_{t=2}^{n_A} Z_{T-1}Z_T}{\sum_{t=2}^{n_A} Z_{T-1}^2} \tag{20.6.9}$$

and, because under the null hypothesis $Z_T = Z_{T-1} + \varepsilon_T - \Theta\,\varepsilon_{T-1}$, we have

$$\hat{\Phi} = \frac{\sum_{t=2}^{n_A} Z_{T-1}(Z_{T-1} + \varepsilon_T - \Theta\varepsilon_{T-1})}{\sum_{t=2}^{n_A} Z_{T-1}^2} = 1 + \frac{\sum_{t=2}^{n_A} Z_{T-1}\varepsilon_T}{\sum_{t=2}^{n_A} Z_{T-1}^2} - \Theta\frac{\sum_{t=2}^{n_A} Z_{T-1}\varepsilon_{T-1}}{\sum_{t=2}^{n_A} Z_{T-1}^2}.$$

Therefore, we obtain the following expression for the normalized unit root estimator:

$$n_A(\hat{\Phi} - 1) = \frac{n_A^{-1}\sum_{t=2}^{n_A} Z_{T-1}\varepsilon_T}{n_A^{-2}\sum_{t=2}^{n_A} Z_{T-1}^2} - \Theta\frac{n_A^{-1}\sum_{t=2}^{n_A} Z_{T-1}\varepsilon_{T-1}}{n_A^{-2}\sum_{t=2}^{n_A} Z_{T-1}^2}. \tag{20.6.10}$$

The Studentized statistic of the OLS estimator $\hat{\Phi}$ for (20.6.6) is, under the null hypothesis,

$$T_{\hat{\Phi}} = \frac{\hat{\Phi} - 1}{S_{\hat{\Phi}}} = \frac{\hat{\Phi} - 1}{(\hat{\sigma}_\xi^2 / \sum_{t=2}^{n_A} Z_{t-1}^2)^{1/2}} = \frac{n_A^{-1} \sum_{t=2}^{n_A} Z_{T-1}\varepsilon_T - \Theta n_A^{-1} \sum_{t=2}^{n_A} Z_{T-1}\varepsilon_{T-1}}{(\hat{\sigma}_\xi^2 n_A^{-2} \sum_{t=2}^{n_A} Z_{t-1}^2)^{1/2}}, \quad (20.6.11)$$

where we applied (20.6.10).

The limiting distributions of the test statistics under the null hypothesis are again shown by Teles and Wei (1999) as

$$n_A(\hat{\Phi} - 1) \xrightarrow{d} \frac{\frac{1}{2}[\{W(1)\}^2 - 1]}{\int_0^1 \{W(X)\} dX} + \frac{(m^2 - 1)/6\, m^2}{\int_0^1 \{W(X)\} dX} \quad (20.6.12)$$

$$T_{\hat{\Phi}} \xrightarrow{d} \left(\frac{3m^2}{2m^2 + 1}\right)^{1/2} \left(\frac{\frac{1}{2}[\{W(1)\}^2 - 1]}{\left[\int_0^1 \{W(X)\} dX\right]^{1/2}} + \frac{(m^2 - 1)/6m^2}{\left[\int_0^1 \{W(X)\} dX\right]^{1/2}} \right). \quad (20.6.13)$$

Expressions (20.6.12) and (20.6.13) show that the limiting distributions of the test statistics for the aggregate time series depend on the order of aggregation m. Because $m \geq 2$, the distributions of the aggregate statistics are shifted to the right relatively to those of the basic statistics given in (20.6.3) and (20.6.5), and the shift increases with the order of aggregation.

20.6.3 The Effects of Aggregation on the Significance Level and the Power of the Test

The shift to the right shown by the distributions of the test statistics implies that the percentiles tabulated by Dickey (1976, pp. 52–53) and by Fuller (1996, pp. 641–642) may lead to improper conclusions of the test when aggregate series are used to test for a unit root. In this section, we investigate more explicitly the impact of the use of aggregate time series on the test in terms of its significance level and power.

We start by studying the effects on the empirical significance levels of the test. In fact, because the rejection region is on the left tail of the distribution, i.e., the null hypothesis of a unit root is rejected when the values of the test statistics are smaller than the appropriate critical points, the distribution shift to the right caused by aggregation will lead to fewer rejections than expected under the null hypothesis. Therefore, the empirical significance levels will be lower than the nominal level.

To assess this effect, we conducted a simulation experiment as follows. In the finite sample case, 50,000 time series of $n = 600$ observations each were generated from a random walk model $z_t = z_{t-1} + a_t$, where $z_1 = 0$ and a_t is a sequence of independent and identically distributed $N(0, 1)$ random variables. The generated time series were aggregated with orders of aggregation $m = 2, 3, 4, 6, 8, 12$. Concerning the limiting case, 100,000 values of each test statistic were generated from the appropriate limiting distributions, as given in equations (20.6.3) and (20.6.5) for the basic series and in equations (20.6.12) and (20.6.13) for the aggregate case with $m = 2, 3, 4, 6, 8, 12$. The observed significance levels

TABLE 20.9 Aggregation effects on the empirical significance level of the test.

	Order of aggregation						
	1	2	3	4	6	8	12
	Finite sample						
NURE	0.0509	0.0211	0.0174	0.0165	0.0153	0.0150	0.0157
Stdt. Stat.	0.0500	0.0213	0.0173	0.0161	0.0154	0.0150	0.0154
	Limiting distribution						
NURE	0.0496	0.0215	0.0174	0.0161	0.0151	0.0148	0.0145
Stdt. Stat.	0.0490	0.0215	0.0172	0.0160	0.0153	0.0149	0.0147

based on the critical points provided by Dickey (1976, pp. 52–53) and by Fuller (1996, pp. 641–642) obtained from the simulation are shown in Table 20.9, where $m = 1$ represents the basic case when there is no aggregation. The nominal significance level used is 5%. In the table, NURE and Stdt. Stat. represent the normalized unit root estimator and the Studentized statistic, respectively.

The simulation results clearly show that, for the basic series ($m = 1$), the critical points provided by Dickey and Fuller are appropriate. In fact, the empirical significance levels are practically coincident with the nominal level of 5%. The observed significance levels decrease when the order of aggregation increases, however, and the impact is very strong. Moreover, we observe that a serious impact occurs even for $m = 2$. It is clear that the effect of aggregation on the test is severe for any order of aggregation.

These conclusions are also supported by several simulation studies like those reported by Schwert (1989), Pantula (1991), Agiakloglou and Newbold (1992), and Cheung and Lai (1995) showing that the Dickey–Fuller test is severely affected by the presence of a moving average component in the model, even for large sample sizes. For negative values of the moving average parameter, which is the situation we have, the observed significance levels of the test in their simulations were also lower than the nominal levels.

The distribution shift of the test statistics also has important implications on the power of the test. Because the null hypothesis is rejected when the test statistics are lower than the critical points, the use of the unadjusted percentiles as critical points for the test based on aggregate series will lead to smaller rejection regions than implied by the significance level adopted for the test. As a result, there will be fewer rejections of the null hypothesis than expected under the alternative hypothesis of no unit roots, i.e., the power of the test will decrease and will be lower than that obtained for basic time series. To avoid the problem, the tables of percentiles provided by Dickey (1976, pp. 52–53) and Fuller (1996, pp. 641–642) need to be adjusted when aggregate time series are used in testing for a unit root. Thus, we constructed Tables J and K in the appendix, which are suitable for the distributions of the test statistics for the most commonly used orders of aggregation $m = 1, 2, 3, 4, 6$, and 12, where $m = 1$ represents no aggregation with values close to those in Tables F and G except for rounding. To obtain the corrected percentiles, we followed the procedure of Dickey (1976), which is based on Monte Carlo simulations. We adapted the procedure

and computed the test statistics for aggregate series. The sample sizes shown in the tables are the sizes of aggregate series. For example, the 5% percentile of the normalized unit root estimator based on an aggregate time series of size $n_A = 100$ resulting from an order of aggregation $m = 2$ is -5.82. Tables J and K are for models without a constant term. We refer readers to Teles (1999) for more detailed tables including tables for models with a constant term.

20.6.4 Examples

To illustrate the effects of aggregation on the tests of a unit root, we analyze two examples: a simulated series and the time series of the U.S. new one-family houses sold.

EXAMPLE 20.4 In this example of a simulated Time Series, we generated a time series of 240 observations from the model $z_t = 0.95z_{t-1} + a_t$, where the a_t are i.i.d. $N(0, 1)$. This series was aggregated with $m = 3$.

First, let us test the basic series z_t (240 observations) for a unit root. Without assuming a known model, we examine its sample autocorrelation function and partial autocorrelation function, both suggest an AR(1) model. The least squares estimation leads to the following result (standard error in parenthesis):

$$\hat{z}_t = \underset{(.0182)}{.9603} z_{t-1}. \tag{20.6.14}$$

To test the hypothesis of a unit root with this series, the values of the test statistics are $n(\hat{\phi} - 1) = 240(0.9603 - 1) = -9.53$ and $T_{\hat{\phi}} = (0.9603 - 1)/0.0182 = -2.18$. From the tables provided by Dickey (1976, pp. 52–53) and by Fuller (1996, pp. 641–642), for $n = 240$ and a 5% significance level, the critical point for the normalized unit root estimator is between -8.0 and -7.9, and for the Studentized statistic it is -1.95. The conclusion of the test is to reject the hypothesis of a unit root, and we conclude that z_t is stationary. The result is certainly consistent with the underlying simulated model.

Let us now consider the aggregate series Z_T for $m = 3$ (80 observations) and use it to test for a unit root. To be realistic with actual data analysis, however, we will not assume a known model and will proceed with a standard model building procedure to identify a best model for the series. The sample autocorrelations and partial autocorrelations (table not shown) suggest an AR(1) model. An overfitting with an ARMA(1, 1) model gives a nonsignificant MA term; hence, we will retain the AR(1) model. The least squares estimation leads to the following model (standard error in parenthesis):

$$\hat{Z}_T = \underset{(.0448)}{.9205} Z_{T-1}. \tag{20.6.15}$$

The values of the test statistics are $n_A(\hat{\Phi} - 1) = 80(0.205 - 1) = -6.36$ and $T_{\hat{\Phi}} = (0.9205 - 1)/0.0448 = -1.77$. Because the test is based on an aggregate time series with $m = 3$ and $n_A = 80$, we will use the adjusted critical points with $m = 3$ from Tables J and K. At a 5% significance level, the critical point for the normalized unit root estimator is between -5.45 and -5.40, and for the Studentized statistic it is -1.60. Consequently, we reject the null hypothesis of a unit root and conclude that the aggregate series Z_T is stationary.

If we derived our conclusion of this test from the unadjusted percentiles given by Dickey (1976, pp. 52–53) and by Fuller (1996, pp. 641–642), then the critical points for $n = 80$ at a 5% significance level would be between -7.9 and -7.7 for the normalized unit root estimator and -1.95 for the Studentized statistic. We would then retain the hypothesis of a unit root and conclude that the series Z_T is nonstationary. This wrong conclusion would lead us to overdifference Z_T and therefore distort its analysis.

There have been extensive studies on the aggregation of stationary processes, including those by Amemiya and Wu (1972), Brewer (1973), Wei (1978a,b), Weiss (1984), Stram and Wei (1986b), and Lütkepohl (1987). From these studies, it is known that the aggregate series of a stationary time series is stationary. Thus, this example shows that using the unadjusted critical points to test for a unit root with an aggregate series may lead to the conclusion that a stationary process contains a unit root and to overdifferencing. To avoid this problem, the adjusted percentiles should be used when the test is based on an aggregate series.

EXAMPLE 20.5 We now consider the time series of the monthly U.S. new one-family houses sold in the United States from January 1968 through December 1982 (180 observations) listed as W16 in the appendix. The source of the data is the Current Construction Reports published by the Bureau of the Census, U.S. Department of Commerce. The data have been seasonally adjusted, i.e., the series has been adjusted by subtracting the corresponding monthly average seasonal values.

Suppose first that we are interested in modeling the quarterly U.S. new one-family house sales, Z_T (60 observations). To test Z_T for a unit root, we first note that as a seasonally adjusted series, it has a zero mean. Next, the sample autocorrelation and partial autocorrelation functions of Z_T given in the following table lead to the identification of an AR(1) model:

Sample autocorrelations of Z_T									
1–9	.89	.75	.62	.53	.41	.22	.07	−.08	−.18
St.E.	.13	.21	.25	.27	.29	.30	.30	.30	.30
Sample partial autocorrelation of Z_T									
1–9	.89	−.20	.01	.06	−.23	−.38	.11	−.22	−.01
St.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13

An overfitting with an ARMA(1, 1) model gives the nonsignificant MA term

$$\hat{Z}_T = \underset{(.0765)}{.8612}\, Z_{T-1} + \varepsilon_T + \underset{(.1476)}{.1610}\, \varepsilon_{T-1},$$

and we will retain the AR(1) model. The least squares estimation (standard error in parenthesis) gives

$$\hat{Z}_T = \underset{(.0590)}{.8915}\, Z_{T-1}. \tag{20.6.16}$$

The values of the test statistics are $n_A(\hat{\Phi} - 1) = 60(.8915 - 1) = -6.51$ and $T_{\hat{\Phi}} = (.8915 - 1)/.059 = -1.84$. Because the test is based on an aggregate time series with $m = 3$ and $n_A = 60$, we should use the adjusted critical points from Tables J and K. At a 5% significance level, the critical point for the normalized unit root estimator is between -5.45 and -5.40, and for the Studentized statistic it is -1.60. Consequently, we reject the null hypothesis of a unit root, i.e., we conclude that the aggregate series Z_T is stationary.

If we based the conclusion of this test on the unadjusted percentiles given by Dickey (1976, pp. 52–53) and by Fuller (1996, pp. 641–642), then the critical points for $n = 60$ at a 5% significance level would be between -7.9 and -7.7 for the normalized unit root estimator and -1.95 for the Studentized statistic. As in Example 20.4, we would then retain the hypothesis of a unit root and conclude that the series Z_T is nonstationary.

We note that the data were collected monthly. In fact, we can consider the month as the appropriate time unit for the series of the new one-family houses sold because the demand for new homes depends on the capacity of home buyers to pay for the houses they buy. The overwhelming majority of home buyers rely on mortgages to finance their homes. Therefore, the decision of buying a home is strongly dependent on mortgage payment plans. In the period we are studying, most payment plans are monthly, and they reflect the way the demand for new homes is formed.

Consequently, we may take the month as the basic time unit and therefore the monthly series z_t as the basic series. We now test z_t for a unit root. Based on its sample autocorrelation function and partial autocorrelation function (table not shown), the series z_t is identified as an AR(1) model. We then fit an AR(1) model to z_t with least squares. The estimation results are (standard error in parenthesis)

$$\hat{Z}_t = \underset{(.0252)}{.9412}\, z_{t-1}. \tag{20.6.17}$$

To test the hypothesis of a unit root with this series, we compute the test statistics $n(\hat{\phi} - 1) = 180(.9412 - 1) = -10.58$ and $T_{\hat{\phi}} = (.9412 - 1)/.0252 = -2.33$. Because z_t is the basic series, we will use the tables provided by Dickey and Fuller. The critical point for the normalized unit root estimator is between -8.0 and -7.9, and for the Studentized statistic it is -1.95 for $n = 180$ and a 5% significance level. Therefore, we reject the hypothesis of a unit root and conclude that z_t is stationary. Consequently, as in the previous example, concluding that Z_T is nonstationary, as implied by the use of the unadjusted critical points, clearly leads to a wrong conclusion.

20.6.5 General Cases and Concluding Remarks

We analyzed how the use of aggregate time series affects testing for a unit root. The presentation is based on the model originally proposed in Dickey and Fuller (1979). For a nonzero mean process, we can simply compute the mean adjusted series before applying the testing procedure.

For a general process, we consider the AR (p) approximation of the aggregate series, $(1 - \Phi_1 B - \cdots - \Phi_p B^p)Z_T = \varepsilon_T$. To test for a unit root, we assume that $\Phi_p(B) = \varphi_{p-1}(B)(1 - B)$

where $\varphi_{p-1}(B) = (1 - \varphi_1 B - \cdots - \varphi_{p-1} B^{p-1})$ has roots lying outside the unit circle. Thus, testing for a unit root is equivalent to testing $\Phi = 1$ in the model

$$Z_T = \Phi Z_{T-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{T-j} + \varepsilon_T. \tag{20.6.18}$$

where $\Delta Z_{T-j} = (Z_{T-j} - Z_{T-j-1})$. Following the results in Section 9.4, the normalized unit root estimator becomes

$$n_A(\hat{\Phi} - 1)\psi_A(1), \tag{20.6.19}$$

where $\psi_A(B) = 1/\varphi(B)$, and $\varphi(B) = (1 - \varphi_1 B - \cdots - \varphi_{p-1} B^{p-1})$, and the corresponding t statistic is

$$T_{\hat{\Phi}} = \frac{(\hat{\Phi} - 1)}{S_{\hat{\Phi}}}. \tag{20.6.20}$$

The limiting distributions of these test statistics in (20.6.19) and (20.6.20) are the same as those in (20.6.12) and (20.6.13). Aggregation causes the distributions shift to the right, leads to empirical significance levels lower than the nominal level, and significantly reduces the power of the tests. Similar to the results in Section 9.4, however, when aggregate data are used, we can use the same adjusted Table J and Table K for the normalized unit root estimator in (20.6.19) and the t statistic in (20.6.20), respectively. For details and examples, see Teles (1999).

To correct these problems, we constructed adjusted tables of critical points suitable for testing for a unit root based on aggregate time series. Both simulation and empirical examples clearly show that the use of the adjusted tables is necessary for a proper test whenever aggregate time series are used.

A similar study was reported by Pierse and Snell (1995). Because their study does not provide the explicit distribution of the test statistic used in testing aggregate series, one could not examine the underlying causes of the effects as a result of order of aggregation illustrated in this paper. One may argue that many tests are available for a series with correlated errors, for example, the augmented Dickey–Fuller test, the Phillips (1987a,b) test, and the Phillips and Perron (1988) test. Although aggregation induces correlation, however, the induced correlation is often not strong enough to show significance in the process of empirical model fitting, which is clear from both the simulated and real-life examples in Section 20.6.4. As a result, the standard Dickey–Fuller test instead of other desirable tests may still be used to make the inference. The essence of the results in this study is that although the standard Dickey–Fuller test can still be used, for a proper inference, the adjusted tables based on the order of aggregation should be used as long as an aggregate time series is used in the analysis. The Dickey–Fuller test is possibly one of the most widely used tests in time series analysis, so it is hoped that the results presented in this study will be useful to data analysts.

20.7 Further Comments

Temporal aggregation and systematic sampling pose important problems in time series analysis. In working with time series data, one must first decide on the observation time unit to be used in the analysis. If the model for a phenomenon under study is regarded appropriate in terms of a certain time unit t, then the proper inferences about the underlying model should be drawn from the analysis of data in terms of this time unit. As shown in the preceding sections, improper use of data in some larger time unit could lead to a substantial information loss in parameter estimation and forecasting. More important, the improper use of aggregate data could distort results in terms of lag structure and causality and hence may lead to very serious consequences in terms of final policy and decision making.

Most time series procedures have been established under the assumptions such as linearity, normality, and stationarity. Many time series encountered in practice are, however, often nonlinear, nonnormal, or nonstationary. To check whether an observed series confirms the assumption, tests have been introduced to test linearity, normality, and nonstationarity. Because of the process of data collection or the practices of researchers, time series used in analysis and modeling are frequently obtained through temporal aggregation. As a result, the series used in testing are often time series aggregates. In this chapter, we study the effects of the use of aggregate time series in performing various tests. Through the study of its impacts on sampling distributions of the test statistics, we show that the use of time series aggregates results in serious loss of power of the tests and hence often leads to erroneous conclusions. To avoid missing the nature of the underlying process, special care is required in interpreting the testing results when time series aggregates are used in the tests.

We presented only some results which are directly related to the materials covered in this book. References on the effect of temporal aggregation on model structure include Engle and Liu (1972), Tiao (1972), Brewer (1973), Tiao and Wei (1976), Teräsvirta (1977), Wei (1978b), Harvey (1981), Werner (1982), Engle (1984), Lütkepohl (1984), Niemi (1984), Weiss (1984), Stram and Wei (1986b), and Wei and Stram (1988). The consequence of temporal aggregation on parameter estimation has been studied by Zellner and Mont-marquette (1971), Engle and Liu (1972), Tiao and Wei (1976), Wei (1978a,b), Wei and Mehta (1980), and Ahsanullah and Wei (1984b). The effect of temporal aggregation on forecasting has been investigated by Amemiya and Wu (1972), Tiao and Wei (1976), Wei (1978b), Abraham (1982), Ahsanullah and Wei (1984a), and Lütkepohl (1987). The problems on aggregation and seasonal adjustment are also studied by Lee and Wei (1979, 1980).

A related problem that has not been discussed in this chapter is the contemporaneous aggregation of time series. It deals with aggregation of several component series in a system. For example, the total money supply is the aggregate of a demand deposit component and a currency component. National housing starts is the aggregate of component housing starts in the northeast, north central, south, and west regions. An interested reader is referred to the work by Rose (1977), Tiao and Guttman (1980), Wei and Abraham (1981), Kohn (1982), Lütkepohl (1984), Engel (1984), and Lütkepohl (1987), among others. The book by Lütkepohl (1987) deals with temporal aggregation of vector ARMA processes and contains extensive references.

To recover information loss, several disaggregation procedures have been introduced. Due to space limitations, we will not discuss them here and refer interested readers to Boot, Feibes, and Lisman (1967), Chow and Lin (1971), Cohen, Müller, and Padberg (1971), Denton (1971), Stram and Wei (1986a), Al-Osh (1989), Guerrero (1990), Wei and

Stram (1990), Chan (1993), Gudmundsson (1999), Hotta and Vasconcellos (1999), Hodgess and Wei (1996, 2000a,b), and Di Fonzo (2003), among others.

EXERCISES

20.1 Consider the nonaggregate model $z_t = (1 - .1B - .2B^2)a_t$, where a_t is a zero mean Gaussian process with $\sigma_a^2 = 2$.

(a) Let $Z_T = (1 + B + B^2)z_{3T}$. Find the autocovariance function, $\gamma_Z(k)$, of Z_T for $k = 0, 1, 2$, and 3, and find the model for Z_T.

(b) Let $Z_T = z_{3T}$. Find the autocovariance function, $\gamma_Z(k)$, of Z_T for $k = 0, 1, 2$, and 3, and the model for Z_T.

20.2 Consider the nonaggregate model, $(1 - .9B)z_t = a_t$, where a_t is a Gaussian $N(0, 1)$ white noise process.

(a) Let $Z_T = (1 + B + B^2)z_{3T}$. Find the autocovariance function, $\gamma_Z(k)$, of Z_T for $k = 0, 1, 2$, and 3, and the model for Z_T.

(b) Let $Z_T = z_{3T}$. Find the autocovariance function, $\gamma_Z(k)$, of Z_T for $k = 0, 1, 2$, and 3, and the model for Z_T.

20.3 Consider Series W4 of the monthly unemployed young women between ages 16 and 19 in the United States from January 1961 to August 2002.

(a) Build a monthly model using the data from January 1961 to December 2001. Forecast the next nine-month data for unemployed young women, and obtain the corresponding quarterly forecasts for quarters 1, 2, and 3 of 2002.

(b) Build a quarterly model for the quarterly unemployed data from the first quarter of 1961 to the fourth quarter of 2001. Forecast the next three quarters.

(c) Derive the implied quarterly aggregate model of the monthly model obtained in part (a), and compute forecasts for the next three quarters.

(d) Compare and comment on your findings in parts (a), (b), and (c).

20.4 (a) Let $H_t = (\sum_{j=0}^{m-1} B^j)x_t$. Prove that

$$E(H_t H_{t-\ell}) = mE(x_t x_{t-\ell}) + \sum_{s=1}^{m-1}(m - s)E(x_t x_{t-(s-\ell)} + x_t x_{t-(s+\ell)}).$$

(b) In part (a), consider $x_t = (1 - \theta B)a_t$, where a_t is a white noise process with mean 0 and variance σ_a^2. Show that

$$E(H_t H_{t-\ell}) = \begin{cases} [m(1 - \theta)^2 + 2\theta]\sigma_a^2, & \ell = 0, \\ (m - \ell)(1 - \theta)^2\sigma_a^2, & \ell = 1, \ldots, (\mathrm{m} - 1), \\ -\theta\sigma_a^2, & \ell = \mathrm{m}, \\ 0, & \text{otherwise.} \end{cases}$$

(c) Consider the one-sided causal model,

$$y_t = \alpha x_{t-1} + e_t,$$

where $x_t = (1 - \theta B)a_t$, e_t and a_t are independent white noise processes with mean 0 and variance σ_e^2 and σ_a^2, respectively. Show that the mth order temporal aggregation turns the one-sided causal relationship into a two-sided feedback system,

$$Y_T = \frac{\alpha(1 - \theta)^2[(m - 1) + B]}{[m(1 - \theta)^2 + 2\theta] - \theta(B + F)} X_T + U_T,$$

where $F = B^{-1}$, X_T is an MA(1) process with the autocovarince generating function (AGF) given by $\gamma_X(B) = \sigma_a^2\{[m(1 - \theta)^2 + 2\theta] - \theta(B + F)\}$, and the U_T is an error process independent of X_T with the AGF given by

$$\alpha^2\sigma_a^2\left\{[m(1 - \theta)^2 + 2\theta] - \theta(B + F) - \frac{(1 - \theta)^4[(m - 1) + B][(m - 1) + F]}{[m(1 - \theta)^2 + 2\theta - \theta(B + F)]}\right\} + m\sigma_e^2.$$

20.5 Let Z_T be the mth order aggregate of z_t where $z_t = a_t$, and a_t is a sequence of i.i.d. random variable with mean 0 and variance σ_a^2. Let λ_0 and λ_{A_0} be the noncentrality parameter for z_t and Z_T, respectively. Show that $\lambda_{A_0} = \lambda_0/m^{2c}$, where $\frac{1}{2} < c < 1$.

20.6 (a) Perform the unit root test on Series W4.

(b) Perform the unit root test on the third-order aggregates of Series W4.

(c) Compare and comment on your findings in parts (a) and (b).

(d) Repeat the above for Series W14.

References

Abraham, B. (1982). Temporal aggregation and time series, *International Statistical Review*, **50**, 285–291.

Abraham, B., and Box, G. E. P. (1979). Bayesian analysis of some outlier problems in time series, *Biometrika*, **66**, 229–236.

Abraham, B., and Chuang, A. (1989). Outlier detection and time series modeling, *Technometrics*, **31**, 241–248.

Abraham, B., and Ledolter, J. (1983). *Statistical Methods for Forecasting*, John Wiley, New York.

Abraham, B., and Ledolter, J. (1984). A note on inverse autocorrelations, *Biometrika*, **71**, 609–614.

Abraham, B., and Wei, W. W. S. (1984). Inferences about the parameters of a time series model with changing variance, *Metrika*, **31**, 183–194.

Agiakloglou, C., and Newbold, P. (1992), Empirical evidence on Dickey–Fuller-type tests, *Journal of Time Series Analysis*, **13**, 471–483.

Ahsanullah, M., and Wei, W. W. S. (1984a). Effects of temporal aggregation on forecasting in an ARMA(1, 1) process. *ASA Proceedings of Business and Economic Statistics Section*, 297–302.

Ahsanullah, M., and Wei, W. W. S. (1984b). The effects of time aggregation on the AR(1) process, *Computational Statistics Quarterly*, **1**, No. 4, 343–352.

Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting, *Ann. Inst. Statist. Math.*, **21**, 407–419.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Proc. 2nd International Symposium on Information Theory* (Eds. B. N. Petrov and F. Csaki), 267–281, Akademiai Kiado, Budapest.

Akaike, H. (1974a). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Ann. Inst. Statist. Math.*, **26**, 363–387.

Akaike, H. (1974b). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716–723.

Akaike, H. (1976). Canonical correlations analysis of time series and the use of an information criterion, in *System Identification: Advances and Case Studies* (Eds. R. Mehra and D. G. Laniotis), Academic Press, New York and London.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, **30A**, 9–14.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, **66**, 237–242.

Al-Osh, M. (1989). A dynamic linear model approach for disaggregating time series data, *Journal of Forecasting*, 8, 85–96.

Ali, M. M. (1977). Analysis of ARMA models—estimation and prediction, *Biometrika*, **64**, 1–11.

Amemiya, T., and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model, *J. Amer. Statist. Assoc.* No. **339**, 628–632.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley, New York.

Anderson, T. W. (1980). Maximum likelihood estimation for vector autoregressive moving average models, in *Directions in Time Series* (Eds. D. R. Brillinger and G. C. Tiao), *Institute of Mathematical Statistics*, 49–59.

Andrews, P. E., and Herzberg, A. M. (1985). *The Data: A Collection of Problems from Statistics*, Springer-Verlag, Berlin.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process, *Biometrika*, **66**, 59–65.

Ansley, C. F., and Newbold, P. (1979a). Multivariate partial autocorrelations, *ASA Proceedings of Business and Economic Statistics Section*, 349–353.

Ansley, C. F., and Newbold, P. (1979b). On the finite sample distribution of residual autocorrelations in autoregressive-moving average models, *Biometrika*, **66**, 547–554.

Ashley, R. A., Patterson, D. M., and Hinich, M. J. (1986). A diagnostic test for nonlinear serial dependence in time-series fitting errors, *Journal of Time Series Analysis*, **7**, 165–178.

Automatic Forecasting Systems, Inc. (1987). *MTS User's Guide*, Hatboro, PA.

Bartlett, M. S. (1941). The statistical significance of canonical correlation, *Biometrika*, **32**, 29–38.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series, *J. Royal Stat. Soc.*, **B8**, 27–41.

Bartlett, M. S. (1947). Multivariate analysis, *Journal of the Royal Statistical Society Supplement*, B9, 176–197.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra, *Biometrika*, **37**, 1–16.

Bartlett, M. S. (1955). *Stochastic Processes*, Cambridge University Press, Cambridge.

Bartlett, M. S. (1966). *An Introduction to Stochastic Processes with Reference to Methods and Application*, 2nd ed., Cambridge University Press, London.

Beguin, J. M., Gourieroux, C., and Monfort, A. (1980). Identification of a mixed autoregressive-moving average process: The corner method, *Time Series* (Ed. O. D. Anderson), 423–436, North-Holland, Amsterdam.

Bell, W. R. (1983). A computer program for detecting outliers in time series, *ASA Proceedings of Business and Economic Statistics Section*, 634–639.

Bell, W. R., and Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of economic time series, *Journal of Business and Economic Statistics*, **2**, 291–320.

Blackman, R. B., and Tukey, J. W. (1959). *The Measurement of Power Spectrum from the Point of View of Communications Engineering*, Dover, New York (Originally published in *Bell Systems Tech. Journal*, 37, 185–282 and 485–569[1958]).

Bloomfield, P. (2000). *Fourier Analysis of Time Series: An Introduction*, 2nd ed., Wiley Interscience, New York.

Bohlin, T. (1976). Four cases of identification of changing systems, in *System Identification: Advances and Cases Studies* (Eds. R. Mehra and D. G. Laniotis). Academic Press, New York and London.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Boot, J. C. G., Feibes, W., and Lisman, J. H. C. (1967). Further methods of derivation of quarterly figures from annual data, *Allied Statistics*, **16**, 65–67.

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations, *J. Roy. Stat. Soc.* Ser. B. **26**, 211–252.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control*, 2nd ed., Holden-Day, San Franscisco.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ.

Box, G. E. P., and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *J. Amer. Statist. Assoc.*, **65**, 1509–1526.

Box, G. E. P., and Tiao, G. C. (1965). A change in level of a non-stationary time series, *Biometrika*, **52**, 181–192.

Box, G. E. P., and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems, *J. Amer. Statist. Assoc.*, **70**, 70–79.

Box, G. E. P., and Tiao, G. C. (1977). A canonical analysis of multiple time series, *Biometrika*, **64**, 355–365.

Brewer, K. R. W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models, *Journal of Econometrics*, **1**, 133–154.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York.

Brillinger, D. R., Guckenheimer, J., Guttorp, P. E., and Oster, G. (1980). Empirical modeling of population time series data: The case of age and density dependent vital rates, *Lectures on Mathematics in the Life Science*, (Amer. Math. Soc.) **13**, 65–90.

Brillinger, D. R., and Rosenblatt, M. (1967). Asymptotic theory of *k*th order spectra, *Spectral Analysis of Time Series* (Ed. B. Harris), 153–188, John Wiley, New York.

Bryant, P. G., and Smith, M. A. (1995). *Practical Data Analysis: Case Studies in Business Statistics*, Richard Irwin Inc., Burr Ridge, Illinois.

Burr, I. W. (1976). *Statistical Quality Control Methods*, Marcel Dekker, New York.

Buys-Ballot, C. H. D. (1847). *Leo Chaements Pèriodiques de Tempèrature*, Kemink et Fils, Utrecht.

Campbell, M. J., and Walker, A. M. (1977). A survey of statistical work on the MacKenzie River series of annual Canadian lynx trappings for the years 1821–1934, and a new analysis, *J. Roy. Statist. Soc.*, Ser. A, **140**, 411–431; Discussion 448–468.

Chan, W. S. (1989). *Some Robust Methods for Time Series Modeling*, unpublished Ph.D. Dissertation, Temple University.

Chan, W. S. (1990). On tests for no linearity in Hong Kong stock returns, *Hong Kong Journal of Business Management*, **8**, 1–11.

Chan, N. H., and Wei, C. Z. (1988). Limiting distribution of least squares estimates of unstable autoregressive processes, *Annals of Statistics*, **16**, 367–401.

Chan, W. S., and Tong, H. (1986). On tests for non-linearity in time series, *Journal of Forecasting*, **5**, 217–228.

Chan, W. S., and Wei, W. W. S. (1992). A comparison of some estimators of time series autocorrelations. *Computational Statistics and Data Analysis*, **14**, 149–163.

Chan, W. S. (1993). Disaggregation of annual time series data to quarterly figures: a comparative study, *Journal of Forecasting*, 12, 677–688.

Chang, I., and Tiao, G. C. (1983). Estimation of time series parameters in the presence of outliers, Technical Report 8, University of Chicago, Statistics Research Center.

Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers, *Technometrics*, **30**, No. 2, 193–204.

Chen, C., and Liu, L. M. (1991). Forecasting time series with outliers, *Journal of Forecasting*, **12**, 13–35.

Chen, C., and Liu, L. M. (1993). Joint estimation of model parameters and outliers effect in time series. *Journal of the American Statistical Association*, **88**, 284–297.

Cheung, Y. W., and Lai, K. S. (1995). Estimating finite sample critical values for unit root tests using pure random walk processes: a note, *Journal of Time Series Analysis*, 16, 493–498.

Chow, G. C., and Lin, A. (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series, *Review of Economic Statistics*, **53**, 372–375.

Clarke, D. G. (1976). Econometric measurement of the duration of advertising effect on sales, *Journal of Marketing Research*, **13**, 345–357.

Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications, *Technometrics*, **14**, 277–293.

Cohen, K. J., Müller, W., and Padberg, M. W. (1971). Autoregressive approaches to disaggregation of time series data, *Applied Statistics*, **20**, 119–129.

Cooley, J. W., and Tukey, J. W. (1965). An algorithm, for the machine calculation of complex Fourier series, *Math. Comp.*, **19**, 197–301.

Crato, N., and Ray, B. K. (1996). Model selection and forecasting for long-range dependent processes, *Journal of Forecasting*, **15**, 107–125.

Cupingood, L. A., and Wei, W. W. S. (1986). Seasonal adjustment of time series using one-sided filters, *Journal of Business and Economic Statistics*, **4**, 473–484.

Dagum, E. B. (1980). The X-11-ARIMA seasonal adjustment method, Catalogue 12–564, Statistics Canada, Ottawa, Ontario.

Daniell, P. J. (1946). Discussion on symposium on autocorrelation in time series, *J. Roy. Statist. Soc.*, Suppl. **8**, 88–90.

Davidson, J. E. H., Henry, D. F., Srba, F., and Yeo, S. (1978). Econometric modeling of the aggregate time series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal*, **88**, 661–692.

Day, T. E., and Lewis, C. M. (1992). Stock market volatility and the information content of stock index options, *Journal of Econometrics*, **52**, 267–287.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization, *J. Amer. Statist. Assoc.*, **66**, 99–101.

Di Fonzo, T. (2003). Temporal disaggregation of economic time series: towards a dynamic extension, Working Paper #12 of Eurostate, Office for Official Publications of the European Communities, Luxembourg, 1–41.

Dickey, D. A. (1976). *Estimation and Hypothesis Testing in Nonstationary Time Series*, unpublished PhD. dissertation, Iowa State University.

Dickey, D. A., Bell, B., and Miller, R. (1986). Unit Roots in time series models: Tests and implications, *The American Statistician*, **40**, No. 1, 12–26.

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.*, **74**, 427–431.

Dickey, D. A., Hasza, D. P., and Fuller, W. A. (1984). Testing for unit roots in seasonal time series, *J. Amer. Statist. Assoc.*, **79**, 355–367.

Draper, N., and Smith, H. (1981). *Applied Regression Analysis*, 2nd ed., Wiley-Interscience, New York.

Dunsmuir, W., and Robinson, R. M. (1981). Asymptotic theory for time series containing missing and amplitude modulated observations, Sankhya, A, 260–281.

Durbin, J. (1960). The fitting of time series models, *Review of the Institute of International Statistics*, **28**, 233–244.

Engel, E. M. R. A. (1984). A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes, *Journal of Time Series Analysis*, **5**, No. 3, 159–171.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom Inflation, *Econometrica*, **50**, 987–1007.

Engle, R. F., and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica*, **55**, 251–276.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model, *Econometrica*, **55**, 391–407.

Engle, R. F., and Liu, T. C. (1972). Effects of aggregation over time on dynamic characteristics of an econometric model, in *Econometric Models of Cyclical Behaviors*, Vol. 2 (Ed. B. G. Hickman), 673–737, Columbia University Press, New York.

Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-arch covariance structure: empirical estimates for Treasury bills, *Journal of Econometrics*, **45**, 213–237.

Fan, J., and Yao, Q. (2003). *Nonlinear Time Series*, Springer-Verlag, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed., John Wiley, New York.

Findley, D. F. (1985). On the unbiased property of AIC for exact or approximating linear stochastic time series models, *Journal of Time Series Analysis*, **6**, 229–252.

Fisher, R. A. (1929). Tests of significance in harmonic analysis, *Proc. Roy. Soc.*, Ser. A. **125**, 54–59.

Fox, A. J. (1972). Outliers in time series, *J. Roy. Statist. Soc.*, Ser. B, **43**, 350–363.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed., John Wiley, New York.

Gantmacher, F. R. (1960). *The Theory of Matrices*, Chelsea, New York.

Geweke, J. M. (1982). Measurement of linear dependence and feedback between multiple time series, *J. Amer. Statist. Assoc.*, **77**, 304–313.

Geweke, J., and Porter-Hudak, S. (1983). The estimation and application of long-memory time series models, *Journal of Time Series Analysis*, **4**, 221–238.

Gnedenko, B. V. (1962). *The Theory of Probability* (Trans. B. D. Seckler), Chelsea, New York.

Godfrey, M. D. (1974). Computational methods for time series, *Bull Inst. Math Appl.*, **10**, 224–227.

Good, I. J. (1958). The interaction algorithm and partial Fourier series, *J. Roy. Statist. Soc.*, Ser. B, **20**, 361–372. Addendum (1960), **B22**, 372–375.

Gradshteyn, I. S., and Ryzhik, I. M. (1965). *Tables of Integrals, Series and Products*, 4th ed., Academic Press, New York.

Graham, A. (1981). *Kronecker Products and Matrix Calculus: With Applications*, John Wiley, New York.

Granger, C. W. J. (1986). Developments in the study of co-integrated economic variables, *Oxford Bulletin of Economics and Statistics*, **48**, 213–228.

Granger C. W. J., and Lee, T. H. (1989). The effects of aggregation on nonlinearity, tech. report, (University of California, San Diego); also in R. Mariano (Ed.), *Advances in Statistical Analysis and Statistical Computing*, JAI Press, Greenwich, CT.

Granger, C. W. J., and Newbold, P. (1986). *Forecast Economic Time Series*, 2nd ed., Academic Press, New York.

Gray, H. L., Kelley, G. D., and McIntire, D. D. (1978). A new approach to ARMA modeling, *Communications in Statistics*, **87**, 1–77.

Gudmundsson, G. (1999). Disaggregation of annual flow data with multiplicative trends, *Journal of Forecasting*, **18**, 33–37.

Guerrero, V. M. (1990). Temporal disaggregation of time series: an ARIMA based approach, *International Statistical Review*, **58**, 29–46.

Halmos, P. R. (1951). *Introduction to Hilbert Space*, Chelsea, New York.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hannan, E. J. (1969). The identification of vector mixed autoregressive moving average system, *Biometrika*, **56**, 223–225.

Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.

Hannan, E. J. (1976). Review of multiple time series, *SIAM Reviews*, **18**, 132.

Hannan, E. J. (1979). The statistical theory of linear systems, in *Developments in Statistics* (Ed. P. R. Krishnaiah), Vol. 2, 81–121, Academic Press, New York and London.

Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist*, **8**, 1071–1081.

Hannan, E. J., and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York.

Hannan, E. J., Dunsmuir, W. T. M., and Deistler, M. (1980). Estimation of vector ARMAX models, *Journal of Multivariate Analysis*, **10**, 275–295.

Hannan, E. J., and Quinn, B. G. (1979). The determination of the order of an autoregression, *J. Roy. Statist. Soc.*, Ser. B, **41**, 190–195.

Harrison, P. J., and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc.*, Ser. B, **38**, 205–247.

Hartley, H. O. (1949). Tests of significance in harmonic analysis, *Biometrika*, **36**, 194–201.

Harvey, A. C. (1981). *Time Series Models*, Halsted Press, New York.

Helmer, R. M., and Johansson, J. K. (1977). An exposition of the Box-Jenkins transfer function analysis with an application to the advertising-sales relationship, *Journal of Marketing Research*, **14**, 227–239.

Heyse, J. F., and Wei, W. W. S. (1984). Partial process autocorrelation for vector time series, tech. report # 30, Department of Statistics, Temple University.

Heyse, J. F., and Wei, W. W. S. (1985a). Inverse and partial lag autocorrelation for vector time series, *ASA Proceedings of Business and Economic Statistics Section*, 233–237.

Heyse, J. F., and Wei, W. W. S. (1985b). The partial lag autocorrelation function, tech. report # 32, Department of Statistics, Temple University.

Heyse, J. F., and Wei, W. W. S. (1986). Modelling the advertising-sale relationship through use of multiple time series techniques, *Journal of Forecasting*, **4**, 165–181.

Hillmer, S. C., Bell, W. R., and Tiao, G. C. (1983). Modelling considerations in the seasonal adjustment of economic time series, in *Applied Time Series Analysis of Economic Data* (Ed. A. Zellner), 74–100, U.S. Bureau of the Census, Washington, DC.

Hillmer, S. C., and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models, *J. Amer. Statist. Assoc.*, **74**, 652–660.

Hillmer, S. C., and Tiao, G. C. (1982). An approach to seasonal adjustment, *J. Amer. Statist. Assoc.*, **77**, 63–70.

Hinich, M. J. (1982). Testing for Gaussianity and linearity of a stationary time series, *Journal of Time Series Analysis*, **3**, 169–176.

Hodgess, E. M., and Wei, W. W. S. (1996). Temporal disaggregation of time series, *Applied Statistical Science I*, 33–44.

Hodgess, E. M., and Wei, W. W. S. (2000a). An autoregressive disaggregation method, *Applied Statistical Science V*, 221–235.

Hodgess, E. M., and Wei, W. W. S. (2000b). Temporal disaggregation of stationary bivariate time series, *Linear Algebra and Its Applications*, **321**, 175–196.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic, *J. Amer. Statist. Assoc.*, **75**, 602–607.

Hotta, L. K., and Vasconcellos, K. L. (1999). Aggregation and disaggregation of structural time series models, *Journal of Time Series Analysis*, **20**, 155–171.

Izenman, A. J. (1985). J. R. Wolf and the Zurich sunspot relative numbers, *The Mathematical Intelligencer*, **7**, No. 1, 27–33.

Izenman, A. J., and Zabell, S. L. (1981). Babies and the blackout: The genesis of a misconception, *Social Science Research*, **10**, 282–299.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York and London.

Jenkins, G. M., and Watts, D. G. (1968). *Spectral Analysis and Its Applications*, Holden-Day, San Francisco.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vector in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551–1580.

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration with applications to the demand for money *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Johnson, R. A., and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th ed., Prentice Hall, Upper Saddle River, New Jersey.

Jones, R. H. (1980). Maximum likelihood filtering of ARMA models to time series with missing observations, *Technometrics*, **22**, 389–395.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Engrg.*, Ser. D, **82**, 35–45.

Kalman, R. E., and Bucy, R. S. (1961). New results in linear filtering and prediction problems, *Trans. ASME. J. Basic Engrg.*, Series D, **83**, 95–108.

Kalman, R. E., Falb, P. L., and Arbib, M. A. (1969). *Topics in Mathematical System Theory*, McGraw-Hill, New York.

Kohn, R. (1982). When is an aggregate of a time series efficient forecast by its past? *Journal of Econometrics*, **18**, 337–349.

Kohn, R., and Ansley, C. F. (1983). Exact likelihood of vector autoregressive moving average process with missing or aggregated data, *Biometrika*, **70**, 225–278.

Kolmogorov, A. (1939). Sur l'interpolation et l'extrapolation des suites stationnaires, *C. R. Acad. Sci. Paris*, **208**, 2043–2045.

Kolmogorov, A. (1941). Interpolation und extrapolation von stationären Zufälligen Folgen, *Bull. Acad. Sci.* (Nauk), USSR, Ser. Math., **5**, 3–14.

Koopmans, L. H. (1974). *The Spectral Analysis of Time Series*, Academic Press, New York and London.

Kremers, J. M. (1989). U.S. federal indebtedness and the conduct of fiscal policy, *Journal of Monetary Economics*, **23**, 219–238.

Lamoureux, C. G., and Lastrapes, W. D. (1993). Forecasting stock return variance: toward an understanding of stochastic implied volatilities, *Review of Financial Studies*, **5**, 293–326.

Ledolter, J. (1990). Outlier diagnostics in time series models, *Journal of Time Series Analysis*, **11**, 317–324.

Lee, J. H., and Wei, W. W. S. (1995). A model-independent outlier detection procedure, *Journal of Applied Statistical Science*, **2**, No. 4, 345–359.

Lee, R. M., and Wei, W. W. S. (1979). The Census X-11 program and quarterly seasonal adjustments, *ASA Proceedings of Business and Economic Section*, 366–370.

Lee, R. M., and Wei, W. W. S. (1980). Model based seasonal adjustment and temporal aggregation, *ASA Proceedings of Business and Economic Section*, 427–431.

Li, W. K. (2003). *Diagnostic Checks in Time Series*, Chapman and Hall/CRC, *London*.

Li, W. K., and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA times series models, *J. Roy. Statist. Soc.*, Ser. B, **43**, 231–239.

Lin, K. S. (1987). A comparative study of various univariate time series models for Canadian lynx data, *Journal of Time Series Analysis*, **8**, No. 2, 161–176.

Liu, L. M. (1987). Sales forecasting using multi-equation transfer function models, *Journal of Forecasting*, **6**, 223–238.

Liu, L. M., and Hanssen, D. M. (1982). Identification of multiple input transfer function models, *Communication in Statistics*, **11**, 297–314.

Ljung, G. (1993). On outlier detection in time series, *J. Roy. Stat. Soc.*, **B55**, 559–567.

Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297–303.

Ljung, G. M., and Box, G. E. P. (1979). The likelihood function of stationary autoregressive moving average models, *Biometrika*, **66**, 265–270.

Lütkepohl, H. (1984). Forecasting contemporaneously aggregated vector ARMA processes, *Journal of Business and Economic Statistics*, **2**, 201–214.

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.

Marquardt, P. W. (1963). An algorithm for least squares estimation of non-linear parameters, *J. Soc. Ind. Appl. Math.*, **2**, 431–441.

Martin, R. D. (1980). Robust estimation of autoregressive models, in *Directions in Time Series* (Eds. D. R. Brillinger and G. C. Tiao), 254–288, Institute of Mathematical Statistics, Hagwood, CA.

Masarotto, G. (1987). Robust identification of autoregressive moving average models, *Applied Statistics*, **36**, 214–220.

Mcleod, A. I., and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis*, **4**, 269–273.

Mehra, R. K. (1979). Kalman filters and their applications to forecasting, in *Forecasting* (Eds. S. Makridakis and S. C. Wheelwright), *TIMS Studies in the Management Science*, **12**, 74–894, North-Holland, Amsterdam.

Meinhold, R. J., and Singpurwalla, N. D. (1983). Understanding the Kalman filter, *American Statistician*, **57**, 123–127.

Miller, R., and Wichern, D. W. (1977). *Intermediate Business Statistics*, Holt, Rinehart and Winston, New York.

Montgomery, D. C., and Weatherby, G. (1980). Modeling and forecast time series using transfer function and intervention methods, *AIIE Transactions*, 289–307.

Müller, D., and Wei, W. W. S. (1997). Iterative least squares estimation and identification of the transfer function model, *Journal of Time Series Analysis*, **18**, 579–592.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset return: a new approach, *Econometrica*, **59**, 347–370.

Newbold, P. (1974). The exact likelihood function for a mixed autoregressive moving average process, *Biometrika*, **61**, 423–426.

Newbold, P. (1982). Causality testing in economics, *Time Series Analysis: Theory and Practice 1* (Ed. O. D. Anderson), 701–716, North Holland, Amsterdam.

Newton, H. J., and Pagano, M. (1984). Simultaneous confidence bands for autoregressive spectra, *Biometrika*, **71**, 197–202.

Nicholls, D. F. (1976). The efficient estimation of vector linear time series models, *Biometrika*, **63**, 381–390.

Nicholls, D. F. (1977). A comparison of estimation methods for vector linear time series models, *Biometrika*, **64**, 85–90.

Nicholls, D. F., and Hall, A. D. (1979). The exact likelihood function of multivariate autoregressive moving average models, *Biometrika*, **66**, 259–264.

Nicholson, A. J. (1950). Population oscillations caused by competition for food, *Nature*, London, **165**, 476–477.

Niemi, H. (1984). The invertibility of sampled and aggregated ARMA models, *Metrika*, **31**, 43–50.

Noble, B. (1969). *Applied Linear Algebra*, Prentice Hall, Englewood Cliffs, NJ.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction, *J. Roy. Statist. Soc.*, Ser. B, **40**, 1–42.

Osborn, R. R. (1977). Exact and approximate maximum likelihood estimators for vector moving average processes, *J. Roy. Statist. Soc.*, Ser. B, **39**, 114–118.

Pack, D. J. (1979). Forecasting time series affected by identifiable isolated events and an explanatory variable, tech. report, Computer Science Division, Oak Ridge National Laboratory, Oak Ridge, TN.

Palda, K. S. (1964). *The Measurement of Cumulative Advertising Effects*, Prentice Hall, Englewood Cliffs, NJ.

Pantula, S. G. (1991). Asymptotic distributions of unit-root tests when the process is nearly stationary, *Journal of Business and Economic Statistics*, **9**, 63–71.

Parzen, E. (1961a). An approach to time series analysis, *Ann. Math Statist.*, **32**, 951–989.

Parzen, E. (1961b). Mathematical considerations in the estimation of spectra, *Technometrics*, **3**, 167–190.

Parzen, E. (1963). Notes on Fourier analysis and spectral windows, in *Time Series Analysis Papers* (Ed. E. Parzen), Holden-Day, San Francisco.

Parzen, E. (1974). Some recent advances in time series modelling, *IEEE Trans. Automatic Control*, **AC-19**, 723–729.

Parzen, E. (1977). Multiple time series modeling: Determining the order of approximating autoregressive schemes, in *Multivariate Analysis IV* (Ed. P. Krishnaiah), 283–295, North-Holland, Amsterdam.

Pena, D., and Box, G. E. P. (1987). Identifying a simplifying structure in time series, *J. of Amer. Statist. Assoc.*, **82**, 836–843.

Pena, D., Tiao, G. C., and Tsay, R. S. (2001). *A Course in Time Series Analysis*, John Wiley, New York.

Petruccelli, J. D., and Davies, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series, *Biometrika*, **73**, 687–694.

Phadke, M. S. (1981). Quality audit using adaptive Kalman filtering, *ASQC Quality Congress Transactions—San Francisco*, 1045–1052.

Phadke, M. S., and Kedem, G. (1978). Compútation of the exact likelihood function of multivariate moving average models, *Biometrika*, **65**, 511–519.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311–340.

Phillips, P. C. B. (1987a). Time series regression with a unit root, *Econometrica*, **55**, 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression, *Biometrika*, **74**, 535–547.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*, **75**, 335–346.

Pierce, D. A. (1968). Distribution of residual correlations in dynamic/stochastic time series models, tech. report 173, Dept. of Statistics, Univ. of Wisconsin-Madison.

Pierce, D. A. (1980). A survey of recent developments in seasonal adjustment, *The American Statistician*, **34**, 125–134.

Pierse, R. G., and Snell, A. J. (1995). Temporal aggregation and the power of tests for a unit root, *Journal of Econometrics*, **65**, 333–345.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vols. I and II, Academic Press, London.

Priestley, M. B. (1988). *Non-Linear and Non-stationary Time Series Analysis*, Academic Press, London.

Priestley, M. B., and Subba Rao, T. (1975). The estimation of factor scores and Kalman filtering for discrete parameter stationary processes, *Int. J. Control.* **21**, 971–975.

Quenouille, M. H. (1949). Approximate tests of correlation in time series, *J. Roy. Stat. Soc.*, **B11**, 68–84.

Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*, Griffin, London.

Quenouille, M. H. (1958). Discrete autoregressive schemes with varying time-intervals, *Metrika*, **1**, 21–27.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, John Wiley, New York.

Reinsel, G. C., and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, **13**, 353–375.

Rose, D. E. (1977). Forecasting aggregates of independent ARIMA process, *Journal of Econometrics*, **5**, 323–346.

Said, S., and Dickey, D. A. (1985). Hypothesis testing in ARIMA(p, 1, q) models, *J. Amer. Statist. Assoc.*, **80**, 369–374.

SAS Institute, Inc. (1999). *SAS/ETS User's Guide*, Version 8, Cary, NC.

SCA Corp. (1992). *Forecasting and Time Series Analysis Using the SCA Statistical System*, Vol. 1, DeKalb, IL.

SCA Corp. (1997). *Forecasting and Time Series Analysis Using the SCA Statistical System*, Vol. 2, DeKalb, IL.

Schmidt, A. M., and Gamerman, D. (1997). Temporal aggregation in dynamic linear models, *Journal of Forecasting*, **16**, 293–310.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena, *Terr. Mag. Atmos. Elect.*, **3**, 13–41.

Schwartz, G. (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461–464.

Schwert, G. W. (1989). Tests for unit roots: a Monte Carlo investigation, *Journal of Business and Economic Statistics*, **7**, 147–159.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*, John Wiley, New York.

Shaman, P. (1983). Properties of estimates of the mean square error of prediction in autoregressive models, in *Studies in Econometrics, Time Series, and Multivariate Statistics* (Eds. S. Karlin, T. Amemiya, and L. A. Goodman), 331–342, Academic Press, New York and London.

Shen, S. Y., and Wei, W. W. S. (1995). A note on the representation of a vector ARMA model, *Journal of Applied Statistical Science*, **2**, 311–318.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117–126.

Slutzky, E. (1927). The summation of random causes as the source of cyclic processes, *Econometrica*, **5**, 105–146 (1937). Translated from the earlier paper of the same title in *Problems of Economic Conditions* (Ed. Conjuncture Institute Moscow).

Stock, J. H., and Watson, M. W. (1988). Testing for common trends, *J. Amer. Statist. Assoc.*, **83**, 1097–1107.

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwartz, *J. Roy. Statist. Soc.*, Ser. B, **41**, 276–278.

Stram, D. O., and Wei, W. W. S. (1986a). A methodological note on the disaggregation of time series totals, *Journal of Time Series Analysis*, **7**, No. 4, 293–302.

Stram, D. O., and Wei, W. W. S. (1986b). Temporal aggregation in the ARIMA process, *Journal of Time Series Analysis*, **7**, No. 4, 279–292.

Subba Rao, T., and Gabr, M. M. (1980). A test for linearity of stationary time series, *Journal of Time Series Analysis*, **1**, 145–158.

Subba Rao, T., and Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Springer-Verlag, Berlin.

Teles, P. (1999). *The Effects of Temporal Aggregation on Time Series Tests*, unpublished Ph.D. dissertation, Temple University.

Teles, P., and Wei, W. W. S. (1999). The use of aggregate series in time series testing, *1999 NBER/NSF Time Series Conference Proceedings IX & X*, 1–29, Academia Sinica, Taipai, Taiwan.

Teles, P., and Wei, W. W. S. (2000). The effects of temporal aggregation on tests of linearity of a time series, *Computational Statistics and Data Analysis*, **34**, 91–103.

Teles, P., and Wei, W. W. S. (2002). The use of aggregate time series in testing for Gaussianity, *Journal of Time Series Analysis*, **23**, 95–116.

Teles, P., Wei, W. W. S., and Crato, N. (1999). The use of aggregate time series in testing for long memory, Bulletin of the International Statistical Institute, **3**, 341–342.

Teräsvirta, T. (1977). The invertibility of sums of discrete MA and ARMA processes, *Scandinavian Journal of Statistics*, **4**, 165–170.

Tiao, G. C. (1972). Asymptotic behavior of temporal aggregates of time series, *Biometrika*, **59**, 525–531.

Tiao, G. C., and Ali, M. M. (1971). Analysis of correlated random effects: linear model with two random components, *Biometrika*, **58**, 37–51.

Tiao, G. C., and Box, G. E. P. (1981). Modeling multiple time series with applications, *J. Amer. Statist. Assoc.*, **76**, 802–816.

Tiao, G. C., and Guttman, I. (1980). Forecasting contemporal aggregates of multiple time series, *Journal of Econometrics*, **12**, 219–230.

Tiao, G. C., and Tsay, R.S. (1983). Multiple time series modeling and extended sample cross-correlations, *Journal of Business and Economic Statistics*, **1**, No. 1, 43–56.

Tiao, G. C., and Tsay, R. S. (1989). Model specification in multivariate time series (with discussion), *J. Roy. Statist. Soc.*, **B51**, 157–213.

Tiao, G. C., and Wei, W. W. S. (1976). Effect of temporal aggregation on the dynamic relationship of two time series variables, *Biometrika*, **63**, 513–523.

Tong, H. (1977). Some comments on the Canadian lynx data with discussion, *J. Roy. Statist. Soc.*, Ser. A, **140**, 448–468.

Tong, H. (1978). On a threshold model, in *Pattern Recognition and Signal Processing* (Ed. C. H. Chen), Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Springer-Verlag, New York.

Tong, H. (1990). *Non-linear Time Series: A Dynamic System Approach*, Oxford University Press, New York.

Tong, H., and Lim, K. S. (1980). Threshold autoregression, limit cycles, and cyclical data (with discussion), *J. Roy. Statist. Soc.*, **B42**, 245–292.

Tsay, R. S. (1986). Time series model specification in the presence of outliers, *J. Amer. Statist. Assoc.*, **81**, 132–141.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, **7**, 1–22.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *J. Amer. Statist. Assoc.*, **84**, 231–240.

Tsay, R. S. (1991). Detecting and modeling no linearity in univariate time series analysis, *Statistica Sinica*, **1**, 431–451.

Tsay, R. S., and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and non-stationary ARIMA models, *J. Amer. Statist. Assoc.*, **79**, 84–96.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis, in *Advanced Seminar on Spectral Analysis* (Ed. B. Harris), 25–46, John Wiley, New York.

Wegman, E. J. (1986). Another look at Box-Jenkins forecasting procedure, *Commun. Statist. Simula.*, **15(2)**, 523–530.

Wei, S., and Wei, W. W. S. (1998). Comparison of some estimates of autoregressive parameters, *Journal of Applied Statistical Science*, **8**, 51–58.

Wei, W. W. S. (1978a). The effect of temporal aggregation of parameters estimation in distributed lag model, *Journal of Econometrics*, **8**, 237–246.

Wei, W. W. S. (1978b). Some consequences of temporal aggregation in seasonal time series models, in *Seasonal Analysis of Economic Time Series* (Ed. A. Zellner), 433–444, U.S. Department of Commerce, Bureau of the Census, Washington, DC.

Wei, W. W. S. (1981). Effect of systematic sampling on ARIMA models, *Communications in Statistics—Theor. and Math.*, **A10(23)**, 2389–2398.

Wei, W. W. S. (1982). The effect of systematic sampling and temporal aggregation on causality—A cautionary note, *J. Amer. Statist. Assoc.*, **77**, 316–319.

Wei, W. W. S., and Abraham, B. (1981). Forecasting contemporal time series aggregates, *Communications in Statistics—Theor. and Meth.*, **A10(13)**, 1335–1344.

Wei, W. W. S., and Mehta, J. (1980). Temporal aggregation and information loss in a distributed lag model, in *Analyzing Time Series* (Ed. O. D. Anderson), 391–400, North-Holland, Amsterdam.

Wei, W. W. S., and Stram, D. O. (1990). Disaggregation of time series models, *J. Roy. Statist. Soc.*, **B52**, 453–467.

Wei, W. W. S., and Tiao, G. C. (1975). The effect of temporal aggregation on forecasting in the dynamic model, *ASA Proceedings of Business and Economic Section*, 613–617.

Wei, W. W. S., and Stram, D. O. (1988). An eigenvalue approach to the limiting behavior of time series aggregates, *Ann. Inst. Statist. Math.*, **40**, 101–110.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models, *Journal of Econometrics*, **26**, 255–278.

Werner, H. J. (1982). On the temporal aggregation in discrete dynamical systems, in *System Modeling and Optimization* (Eds. R. F. Drenick and F. Kozin), 819–825, Springer, New York.

Whittle, P. (1952). The simultaneous estimation of a time series' harmonic covariance structure, *Trabajos. Estadist.*, **3**, 43–57.

Whittle, P. (1954). A statistical investigation of sunspot observations with special reference to H. Alfven's sunspot model, *The Astrophs. J.*, **120**, 251–260.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*, 2nd ed., Oxford, Blackwell.

Wichern, D. W., Miller, R. B., and Hsu, D. A. (1976). Changes of variance in first-order autoregressive time series models with an application. *Appl. Stat.*, **25** (3), 248–256.

Wiener, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, John Wiley, New York.

Wilk, S. S. (1932). Certain generalization in the analysis of variance, *Biometrika*, **24**, 471–494.

Wold, H. O. A. (1938). *A Study in the Analysis of Stationary Time Series*, Almqvist and Wiksell, Uppsala.

Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*, Prentice Hall, Englewood Cliffs, NJ.

Young, P. C. (1974). Recursive approaches to time series analysis, *Bull. Inst. Math. Appl.*, **10**, 209–224.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, *Philos. Trans. Roy. Soc. London*, Ser. A, **226**, 267–298.

Zellner, A. (1966). On the analysis of first order autoregressive models with incomplete data, *International Economic Reviews*, **7**, 72–76.

Zellner, A. (1978). *Seasonal Analysis of Economic Time Series*, (Ed. A. Zellner), U.S. Department of Commerce, Bureau of the Census, Washington DC.

Zellner, A., and Montmarquette, C. (1971). A study of some aspects of temporal aggregation problems in econometric analyses, *Review of Economics and Statistics*, **53**, 335–342.

Appendix

Time Series Data Used for Illustrations (pp. 566–580)

Series W1	Daily Average Number of Truck Manufacturing Defects
Series W2	Wolf Yearly Sunspot Numbers from 1700 to 2001
Series W3	Blowfly Data
Series W4	Monthly Unemployed Young Women between Ages 16 and 19 in the United States from January 1961 to August 2002
Series W5	Yearly Cancer (All Forms) Death Rate (per 100,000) for Pennsylvania between 1930 and 2000
Series W6	Yearly U.S. Tobacco Production from 1871 to 1984
Series W7	Yearly Number of Lynx Pelts Sold by the Hudson's Bay Company in Canada from 1857 to 1911
Series W8	Simulated Seasonal Time Series
Series W9	Monthly Employment Figures for Males between Ages 16 and 19 in the United States from January 1971 to December 1981
Series W10	Quarterly U.S. Beer Production from the First Quarter of 1975 to the Fourth Quarter of 1982
Series W11	Daily Closing Stock Price of Duke Energy Corporation between January 3, 2002 and August 31, 2002
Series W12	Lydia Pinkham Annual Advertising and Sales Data from 1907 to 1960
Series W13	The Five Series: U.S. Hog Data from 1867 to 1948
Series W14	Number of Monthly Airline Passengers in the United States from January 1995 to March 2002
Series W15	Oklahoma and Louisiana Spot Prices for Natural Gas between January 1988 and October 1991
Series W16	U.S. New One-Family Houses Sold (thousands) from January 1968 to December 1982

Statistical Tables (pp. 581–599)

Table A	Tail Areas of the Standard Normal Distribution
Table B	Critical Values of the t-Distribution
Table C	Critical Values of the Chi-Square (χ^2) Distribution
Table D	Critical Values of the F-Distribution for $F_{0.05}\,(\nu_1, \nu_2)$
Table E	Critical Values of the F-Distribution for $F_{0.01}\,(\nu_1, \nu_2)$
Table F	Empirical Cumulative Distribution of $n(\hat{\phi} - 1)$ for $\phi = 1$

Table G Empirical Cumulative Distribution of T for $\phi = 1$

Table H Percentiles of $n(\hat{\Phi} - 1)$ for the Zero Mean Seasonal Models

Table I Percentiles of the Studentized T Statistic for the Zero Mean Seasonal Models

Table J Empirical Percentiles of the Normalized Unit Root Estimator for Aggregate Time Series: AR(1) Model without Intercept $n_A(\hat{\Phi} - 1)$

Table K Empirical Percentiles of the Studentized Statistic for Aggregate Time Series: AR(1) Model without Intercept $T_{\hat{\Phi}} = (\hat{\Phi} - 1)/S_{\hat{\Phi}}$

Time Series Data Used for Illustrations

Series W1 Daily Average Number of Truck Manufacturing Defects (Read Across)

1.20	1.50	1.54	2.70	1.95	2.40	3.44	2.83	1.76	2.00	2.09	1.89	1.80
1.25	1.58	2.25	2.50	2.05	1.46	1.54	1.42	1.57	1.40	1.51	1.08	1.27
1.18	1.39	1.42	2.08	1.85	1.82	2.07	2.32	1.23	2.91	1.77	1.61	1.25
1.15	1.37	1.79	1.68	1.78	1.84							

Series W2 Wolf Yearly Sunspot Numbers from 1700 to 2001

1700	5.00	1718	60.00	1736	70.00	1754	12.20
1701	11.00	1719	39.00	1737	81.00	1755	9.60
1702	16.00	1720	28.00	1738	111.00	1756	10.20
1703	23.00	1721	26.00	1739	101.00	1757	32.40
1704	36.00	1722	22.00	1740	73.00	1758	47.60
1705	58.00	1723	11.00	1741	40.00	1759	54.00
1706	29.00	1724	21.00	1742	20.00	1760	62.90
1707	20.00	1725	40.00	1743	16.00	1761	85.90
1708	10.00	1726	78.00	1744	5.00	1762	61.20
1709	8.00	1727	122.00	1745	11.00	1763	45.10
1710	3.00	1728	103.00	1746	22.00	1764	36.40
1711	0.00	1729	73.00	1747	40.00	1765	20.90
1712	0.00	1730	47.00	1748	60.00	1766	11.40
1713	2.00	1731	35.00	1749	80.90	1767	37.80
1714	11.00	1732	11.00	1750	83.40	1768	69.80
1715	27.00	1733	5.00	1751	47.70	1769	106.10
1716	47.00	1734	16.00	1752	47.80	1770	100.80
1717	63.00	1735	34.00	1753	30.70	1771	81.60

(Continued)

Series W2 ***(continued)***

1772	66.50	1809	2.50	1846	61.50	1883	63.70
1773	34.80	1810	0.00	1847	98.50	1884	63.50
1774	30.60	1811	1.40	1848	124.70	1885	52.20
1775	7.00	1812	5.00	1849	96.30	1886	25.40
1776	19.80	1813	12.20	1850	66.60	1887	13.10
1777	92.50	1814	13.90	1851	64.50	1888	6.80
1778	154.40	1815	35.40	1852	54.10	1889	6.30
1779	125.90	1816	45.80	1853	39.00	1890	7.10
1780	84.80	1817	41.10	1854	20.60	1891	35.60
1781	68.10	1818	30.10	1855	6.70	1892	73.00
1782	38.50	1819	23.90	1856	4.30	1893	85.10
1783	22.80	1820	15.60	1857	22.70	1894	78.00
1784	10.20	1821	6.60	1858	54.80	1895	64.00
1785	24.10	1822	4.00	1859	93.80	1896	41.80
1786	82.90	1823	1.80	1860	95.80	1897	26.20
1787	132.00	1824	8.50	1861	77.20	1898	26.70
1788	130.90	1825	16.60	1862	59.10	1899	12.10
1789	118.10	1826	36.30	1863	44.00	1900	9.50
1790	89.90	1827	49.60	1864	47.00	1901	2.70
1791	66.60	1828	64.20	1865	30.50	1902	5.00
1792	60.00	1829	67.00	1866	16.30	1903	24.40
1793	46.90	1830	70.90	1867	7.30	1904	42.00
1794	41.00	1831	47.80	1868	37.60	1905	63.50
1795	21.30	1832	27.50	1869	74.00	1906	53.80
1796	16.00	1833	8.50	1870	139.00	1907	62.00
1797	6.40	1834	13.20	1871	111.20	1908	48.50
1798	4.10	1835	56.90	1872	101.60	1909	43.90
1799	6.80	1836	121.50	1873	66.20	1910	18.60
1800	14.50	1837	138.30	1874	44.70	1911	5.70
1801	34.00	1838	103.20	1875	17.00	1912	3.60
1802	45.00	1839	85.70	1876	11.30	1913	1.40
1803	43.10	1840	64.60	1877	12.40	1914	9.60
1804	47.50	1841	36.70	1878	3.40	1915	47.40
1805	42.20	1842	24.20	1879	6.00	1916	57.10
1806	28.10	1843	10.70	1880	32.30	1917	103.90
1807	10.10	1844	15.00	1881	54.30	1918	80.60
1808	8.10	1845	40.10	1882	59.70	1919	63.60

(Continued)

Series W2 ***(continued)***

1920	37.60	1941	47.50	1962	37.50	1983	66.60
1921	26.10	1942	30.60	1963	27.90	1984	45.90
1922	14.20	1943	16.30	1964	10.20	1985	17.90
1923	5.80	1944	9.60	1965	15.10	1986	13.40
1924	16.70	1945	33.20	1966	47.00	1987	29.40
1925	44.30	1946	92.60	1967	93.80	1988	100.20
1926	63.90	1947	151.60	1968	105.90	1989	157.60
1927	69.00	1948	136.30	1969	105.50	1990	142.60
1928	77.80	1949	134.70	1970	104.50	1991	145.70
1929	64.90	1950	83.90	1971	66.60	1992	94.30
1930	35.70	1951	69.40	1972	68.90	1993	54.60
1931	21.20	1952	31.50	1973	38.00	1994	29.90
1932	11.10	1953	13.90	1974	34.50	1995	17.50
1933	5.70	1954	4.40	1975	15.50	1996	8.60
1934	8.70	1955	38.00	1976	12.60	1997	21.50
1935	36.10	1956	141.70	1977	27.50	1998	64.30
1936	79.70	1957	190.20	1978	92.50	1999	93.30
1937	114.40	1958	184.80	1979	155.40	2000	119.60
1938	109.60	1959	159.00	1980	154.60	2001	111.00
1939	88.80	1960	112.30	1981	140.40		
1940	67.80	1961	53.90	1982	115.90		

Series W3 Blowfly Data (Read Across)

1676	3075	3815	4639	4424	2784	5860	5781	4897	3920	3835	3618
3050	3772	3517	3350	3018	2625	2412	2221	2619	3203	2706	2717
2175	1628	2388	3677	3156	4272	3771	4955	5584	3891	3501	4436
4369	3394	3869	2922	1843	2837	4690	5119	5838	5389	4993	4446
4651	4243	4620	4849	3664	3016	2881	3821	4300	4168	5448	5477
8579	7533	6884	4127	5546	6316	6650	6304	4842	4352	3215	2652
2330	3123	3955	4494	4780	5753	5555	5712	4786	4066		

Series W4 Monthly Unemployed Young Women between Ages 16 and 19 in the United States from January 1961 to August 2002 (in Thousands)

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1961	347	348	365	300	311	353	379	387	406	354	347	262
1962	342	331	332	350	299	277	312	308	291	303	335	334
1963	354	375	359	334	418	391	407	335	403	414	398	371
1964	374	374	366	369	398	398	343	385	392	422	349	471
1965	432	444	432	436	375	381	360	338	440	380	393	396
1966	396	364	388	414	420	419	428	423	399	397	384	372
1967	374	410	347	339	355	379	409	414	394	423	412	410
1968	356	417	399	386	428	444	474	397	400	379	393	406
1969	365	383	391	406	422	422	417	427	429	463	389	413
1970	489	447	461	489	449	487	459	522	538	560	622	567
1971	554	543	577	547	552	557	623	570	578	560	599	576
1972	588	593	602	613	523	621	637	627	605	599	585	588
1973	507	623	567	624	585	572	566	545	604	586	613	606
1974	612	620	620	560	640	682	744	596	698	737	738	739
1975	797	765	809	758	850	777	810	832	802	811	775	837
1976	773	790	781	774	741	776	767	876	753	765	779	787
1977	833	793	818	808	774	842	770	768	804	780	790	690
1978	777	773	780	756	777	759	801	777	780	726	762	766
1979	731	727	711	751	738	758	709	751	750	785	731	773
1980	756	772	761	690	801	772	802	775	742	723	743	719
1981	771	788	805	794	792	786	762	757	818	863	865	798
1982	870	907	811	862	893	839	933	906	901	905	911	899
1983	868	815	823	853	822	952	852	850	794	778	757	743
1984	699	720	719	717	693	688	660	703	715	657	606	678

(*Continued*)

Series W4 ***(continued)***

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1985	697	659	689	631	685	640	710	581	629	683	642	698
1986	679	695	679	722	674	702	658	640	668	668	677	641
1987	630	656	641	600	618	606	612	551	582	686	635	575
1988	606	600	585	605	571	502	515	584	566	520	506	542
1989	543	487	497	511	512	611	569	561	561	513	530	536
1990	535	543	542	523	544	529	514	561	562	533	565	574
1991	659	578	587	608	579	586	620	618	586	620	594	661
1992	594	601	600	575	621	715	659	614	655	552	652	621
1993	629	621	614	584	661	601	560	571	576	591	589	567
1994	558	587	590	654	566	596	561	587	550	614	531	560
1995	575	565	581	630	600	594	658	584	622	585	610	625
1996	621	576	568	584	577	556	531	578	542	553	584	603
1997	604	646	572	551	618	551	638	546	582	516	545	554
1998	483	492	526	505	523	562	505	547	519	581	516	465
1999	567	551	528	546	502	522	498	517	587	530	515	472
2000	465	512	556	482	496	383	494	506	489	488	507	482
2001	499	473	527	515	465	513	549	523	531	542	540	584
2002	612	547	561	591	570	594	587	532				

Series W5 Yearly Cancer (All Forms) Death Rate (per 100,000) for Pennsylvania between 1930 and 2000

1930	100.4	105.1	109.4	110.5	116.9	118.2	122.9	126.5	125.8	130.7	131.7
1941	131.7	134.8	136.1	137.8	142.8	144.5	144.4	150.5	152.2	155.2	157.2
1952	157.4	161.9	165.0	164.0	168.2	170.8	169.9	169.7	173.3	173.8	175.8
1963	177.8	179.9	180.0	182.0	184.2	187.7	187.4	186.5	188.3	190.6	190.8
1974	196.4	199.1	203.8	205.7	215.9	216.3	219.5	219.6	222.8	226.3	230.8
1985	235.2	235.8	241.0	235.9	241.4	249.6	248.6	251.4	250.2	249.5	248.7
1996	251.1	250.1	247.6	251.3	244.2						

Series W6 Yearly U.S. Tobacco Production from 1871 to 1984 (in Millions of Pounds)

1871	327	1876	466	1881	426	1886	609	1891	747
1872	385	1877	621	1882	579	1887	469	1892	757
1873	382	1878	455	1883	509	1888	661	1893	767
1874	217	1879	472	1884	580	1889	525	1894	767
1875	609	1880	469	1885	611	1890	648	1895	745

(Continued)

Series W6 *(continued)*

1896	760	1914	1037	1932	1018	1950	2030	1968	1710
1897	703	1915	1157	1933	1372	1951	2332	1969	1804
1898	909	1916	1207	1934	1085	1952	2256	1970	1906
1899	870	1917	1326	1935	1302	1953	2059	1971	1705
1900	852	1918	1445	1936	1163	1954	2244	1972	1749
1901	886	1919	1444	1937	1569	1955	2193	1973	1742
1902	960	1920	1509	1938	1386	1956	2176	1974	1990
1903	976	1921	1005	1939	1881	1957	1668	1975	2182
1904	857	1922	1254	1940	1460	1958	1736	1976	2137
1905	939	1923	1518	1941	1262	1959	1796	1977	1914
1906	973	1924	1245	1942	1408	1960	1944	1978	2025
1907	886	1925	1376	1943	1406	1961	2061	1979	1527
1908	836	1926	1289	1944	1951	1962	2315	1980	1786
1909	1054	1927	1211	1945	1991	1963	2344	1981	2064
1910	1142	1928	1373	1946	1315	1964	2228	1982	1994
1911	941	1929	1533	1947	2107	1965	1855	1983	1429
1912	1117	1930	1648	1948	1980	1966	1887	1984	1728
1913	992	1931	1565	1949	1969	1967	1968		

Series W7 Yearly Number of Lynx Pelts Sold by the Hudson's Bay Company in Canada from 1857 to 1911

1857	23362	1871	45686	1885	27187	1899	26761
1858	31642	1872	7942	1886	51511	1900	15185
1859	33757	1873	5123	1887	74050	1901	4473
1860	23226	1874	7106	1888	78773	1902	5781
1861	15178	1875	11250	1889	33899	1903	9117
1862	7272	1876	18774	1890	18886	1904	19267
1863	4448	1877	30508	1891	11520	1905	36116
1864	4926	1878	42834	1892	8352	1906	58850
1865	5437	1879	27345	1893	8660	1907	61478
1866	16498	1880	17834	1894	12902	1908	36300
1867	35971	1881	15386	1895	20331	1909	9704
1868	76556	1882	9443	1896	36853	1910	3410
1869	68392	1883	7599	1897	56407	1911	3774
1870	37447	1884	8061	1898	39437		

Series W8 Simulated Seasonal Time Series (Read Across)

97.323	99.386	100.581	95.671	102.601	101.751	103.672
100.169	104.475	105.712	109.055	103.344	109.132	110.472
110.752	108.949	111.713	113.635	115.535	111.439	115.798
119.851	120.605	115.752	119.380	122.423	122.320	118.610
121.799	127.540	126.888	122.581	124.873	130.316	129.496
125.721	130.557	133.554	134.013	130.221	133.251	137.640
136.152	134.049	136.034	141.636	140.264	137.761	140.608
145.765	145.491	143.401	145.873	151.227	149.660	148.268
152.012	156.293	156.306	151.636	156.866	161.653	158.831
157.871	159.769	166.352	163.872	162.214	165.239	170.421
169.188	167.274	170.048	175.631	174.562	172.450	174.456
181.318	179.960	177.068	179.141	183.945	183.056	181.098
183.128	191.285	188.676	185.391	187.751	194.459	192.726
189.861	191.439	200.030	196.391	196.406	198.099	204.684
203.344	197.990	202.407	210.073	207.403	203.713	208.789
213.626	211.808	209.169	211.793	219.515	216.688	213.734
216.468	223.600	219.971	216.864	221.199	227.439	224.872
222.034	225.794	231.908	228.497	225.202	229.665	236.350
233.512	230.428	233.996	240.955	237.421	234.174	236.984
245.057	241.983	237.094	242.367	249.358	245.153	240.533
245.504	251.616	250.315	244.670	249.887	256.824	254.345
251.052	253.507	260.501				

Series W9 Monthly Employment Figures for Males between Ages 16 and 19 in the United States from January 1971 to December 1981 (in Thousands)

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1971	707	655	638	574	552	980	926	680	597	637	660	704
1972	758	835	747	617	554	929	815	702	640	588	669	675
1973	610	651	605	592	527	898	839	614	594	576	672	651
1974	714	715	672	588	567	1057	949	683	771	708	824	835
1975	980	969	931	892	828	1350	1218	977	863	838	866	877
1976	1007	951	906	911	812	1172	1101	900	841	853	922	886
1977	896	936	902	765	735	1234	1052	868	798	751	820	725
1978	821	895	851	734	636	994	990	750	727	754	792	817
1979	856	886	833	733	675	1004	956	777	761	709	777	771
1980	840	847	774	720	848	1240	1168	936	853	910	953	874
1981	1026	1030	946	860	856	1190	1038	883	843	857	1016	1003

Series W10 Quarterly U.S. Beer Production from the First Quarter of 1975 to the Fourth Quarter of 1982 (Millions of Barrels)

	Quarter			
Year	I	II	III	IV
1975	36.14	44.60	44.15	35.72
1976	36.19	44.63	46.95	36.90
1977	39.66	49.72	44.49	36.54
1978	41.44	49.07	48.98	39.59
1979	44.29	50.09	48.42	41.39
1980	46.11	53.44	53.00	42.52
1981	44.61	55.18	52.24	41.66
1982	47.84	54.27	52.31	42.03

Series W11 Daily Closing Stock Price of Duke Energy Corporation between January 3, 2002 and August 31, 2002 (Read Across)

38.75	37.91	38.53	37.74	38.15	38.12	37.76	38.02	38.06	37.74
35.45	35.67	35.19	35.05	35.19	35.49	34.91	34.01	33.71	34.33
34.13	33.58	32.49	31.63	32.06	32.78	33.64	33.97	34.57	34.44
33.27	33.27	33.68	33.69	33.55	33.55	34.15	34.80	35.04	35.93
36.24	36.43	37.24	36.68	35.87	35.68	35.52	34.93	34.91	34.98
35.66	35.78	35.19	37.75	37.73	37.30	36.79	37.32	37.52	37.22
37.67	37.47	38.26	38.22	38.91	38.80	38.86	38.37	38.21	37.18
36.92	38.32	37.82	37.74	38.05	38.60	38.73	37.72	36.92	37.84
38.04	37.58	37.32	37.22	36.74	36.06	36.12	36.21	35.98	36.03
36.69	36.19	34.70	33.52	34.35	34.19	35.01	35.34	34.70	34.59
33.51	31.78	32.01	30.70	31.70	31.32	30.68	30.00	29.09	29.12
30.13	31.00	31.01	31.71	32.92	32.83	31.00	31.76	31.66	31.70
31.29	30.84	31.10	30.36	30.00	29.75	30.30	30.31	29.06	26.00
27.95	24.75	23.70	22.26	22.42	20.72	19.95	19.30	19.00	22.07
22.54	22.45	22.40	24.93	25.49	25.44	24.26	24.13	25.00	25.75
26.09	26.33	26.71	26.62	27.75	26.63	27.51	27.55	28.36	28.00
27.10	27.86	26.76	27.29	27.25	26.83				

Series W12 Lydia Pinkham Annual Advertising and Sales Data from 1907 to 1960

Year	Advertising (in $1000)	Sales (in $1000)	Year	Advertising (in $1000)	Sales (in $1000)
1907	608	1016	1934	1504	1770
1908	451	921	1935	807	1518
1909	529	934	1936	339	1103
1910	543	976	1937	562	1266
1911	525	930	1938	745	1473
1912	549	1052	1939	749	1423
1913	525	1184	1940	862	1767
1914	578	1089	1941	1034	2161
1915	609	1087	1942	1054	2336
1916	504	1154	1943	1164	2602
1917	752	1330	1944	1102	2518
1918	613	1980	1945	1145	2637
1919	862	2223	1946	1012	2177
1920	866	2203	1947	836	1920
1921	1016	2514	1948	941	1910
1922	1360	2726	1949	981	1984
1923	1482	3185	1950	974	1787
1924	1608	3351	1951	766	1689
1925	1800	3438	1952	920	1866
1926	1941	2917	1953	964	1896
1927	1229	2359	1954	811	1684
1928	1373	2240	1955	789	1633
1929	1611	2196	1956	802	1657
1930	1568	2111	1957	770	1569
1931	983	1806	1958	639	1390
1932	1046	1644	1959	644	1387
1933	1453	1814	1960	564	1289

Series W13 The Five Series: U.S. Hog Data from 1867 to 1948

Year	Hog Numbers	Hog Prices	Corn Prices	Corn Supply	Wage Rates
1867	538	597	944	900	722
1868	522	509	841	964	719
1869	513	663	911	893	716
1870	529	751	768	1051	724
1871	565	739	718	1057	732
1872	594	598	634	1107	740
1873	600	556	735	1003	748
1874	584	594	858	1025	756
1875	554	667	673	1161	748
1876	553	776	609	1170	740
1877	595	754	604	1181	732
1878	637	689	457	1194	744
1879	641	498	612	1244	756
1880	647	643	642	1232	778
1881	634	681	849	1095	799
1882	629	778	733	1244	799
1883	638	829	672	1218	799
1884	662	751	594	1289	799
1885	675	704	559	1313	801
1886	658	633	604	1251	803
1887	629	663	678	1205	806
1888	625	709	571	1352	806
1889	648	763	490	1361	806
1890	682	681	747	1218	810
1891	676	627	651	1368	813
1892	655	667	645	1278	810
1893	640	804	609	1279	806
1894	668	782	705	1208	771
1895	678	707	453	1404	771
1896	692	653	382	1427	780
1897	710	639	466	1359	789
1898	727	672	506	1371	799

(*Continued*)

Series W13 ***(continued)***

Year	Hog Numbers	Hog Prices	Corn Prices	Corn Supply	Wage Rates
1899	712	669	525	1423	820
1900	708	729	595	1425	834
1901	705	784	829	1234	848
1902	680	842	654	1443	863
1903	682	886	673	1401	884
1904	713	784	691	1429	906
1905	726	770	660	1470	928
1906	729	783	643	1482	949
1907	752	877	754	1417	960
1908	766	777	813	1409	971
1909	720	810	790	1417	982
1910	682	957	712	1455	987
1911	743	970	831	1394	991
1912	743	903	742	1469	1004
1913	730	995	847	1357	1013
1914	723	1022	850	1402	1004
1915	753	998	830	1452	1013
1916	782	928	1056	1385	1053
1917	760	1073	1163	1464	1149
1918	799	1294	1182	1388	1248
1919	808	1346	1180	1428	1316
1920	779	1301	805	1487	1384
1921	770	1134	714	1467	1190
1922	777	1024	865	1432	1179
1923	841	1090	911	1459	1228
1924	823	1013	1027	1347	1238
1925	746	1119	846	1447	1246
1926	717	1195	869	1406	1253
1927	744	1235	928	1418	1253
1928	791	1120	924	1426	1253
1929	771	1112	903	1401	1255
1930	746	1129	777	1318	1223

(Continued)

Series W13 *(continued)*.

Year	Hog Numbers	Hog Prices	Corn Prices	Corn Supply	Wage Rates
1931	739	1055	507	1411	1114
1932	773	787	500	1467	982
1933	793	624	716	1380	929
1934	768	612	911	1161	978
1935	592	800	816	1362	1013
1936	633	1104	1019	1178	1045
1937	634	1075	714	1422	1100
1938	649	1052	687	1406	1097
1939	699	1048	754	1412	1090
1940	786	891	791	1390	1100
1941	735	921	876	1424	1188
1942	782	1193	962	1487	1303
1943	869	1352	1050	1472	1422
1944	923	1243	1037	1490	1498
1945	774	1314	1104	1458	1544
1946	787	1380	1193	1507	1582
1947	754	1556	1334	1372	1607
1948	737	1632	1114	1557	1629

Series W14 Number of Monthly Airline Passengers in the United States from January 1995 to March 2002

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1995	40878	38746	47103	45282	45961	48561	49883	51443	43480	46651	44712	45068
1996	41689	43390	51410	48335	50856	51317	52778	54377	45403	49473	44585	48935
1997	44850	43133	53305	49461	50856	52925	55366	55868	46826	50216	47190	49134
1998	44705	43742	53050	52255	52692	54702	55841	56546	47356	52024	49461	50483
1999	45972	45101	55402	53256	53334	56457	59881	58424	49816	54684	52754	50874
2000	46242	48160	58459	55800	57976	60787	62404	61098	51954	56322	54738	52212
2001	49390	47951	58824	56357	56677	59515	61969	62654	34365	43895	44442	45316
2002	42947	42727	53553									

Series W15 Oklahoma and Louisiana Spot Prices for Natural Gas between January 1988 and October 1991

Year	Month	Oklahoma Spot Price	Louisiana Spot Price
1988	1	1.875	2.065
1988	2	1.898	1.988
1988	3	1.643	1.818
1988	4	1.332	1.493
1988	5	1.262	1.383
1988	6	1.240	1.378
1988	7	1.265	1.433
1988	8	1.310	1.543
1988	9	1.467	1.713
1988	10	1.500	1.688
1988	11	1.633	1.908
1988	12	1.780	2.207
1989	1	1.803	2.173
1989	2	1.472	1.740
1989	3	1.247	1.458
1989	4	1.273	1.515
1989	5	1.373	1.642
1989	6	1.408	1.687
1989	7	1.378	1.643
1989	8	1.375	1.575
1989	9	1.308	1.513
1989	10	1.315	1.555
1989	11	1.447	1.765
1989	12	1.750	2.102
1990	1	2.203	2.420
1990	2	1.623	1.982
1990	3	1.263	1.487
1990	4	1.252	1.468
1990	5	1.252	1.450
1990	6	1.277	1.460
1990	7	1.252	1.418
1990	8	1.220	1.340
1990	9	1.240	1.353
1990	10	1.412	1.568

(Continued)

Series W15 *(continued)*

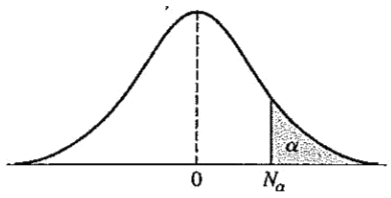
Year	Month	Oklahoma Spot Price	Louisiana Spot Price
1990	11	1.807	2.052
1990	12	1.903	2.237
1991	1	1.627	1.830
1991	2	1.223	1.340
1991	3	1.208	1.318
1991	4	1.208	1.320
1991	5	1.205	1.298
1991	6	1.165	1.258
1991	7	1.020	1.117
1991	8	1.065	1.137
1991	9	1.287	1.368
1991	10	1.613	1.732

Series W16 U.S. New One-Family Houses Sold (Thousands) from January 1968 to December 1982

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1968	35	43	46	46	43	41	44	47	41	40	32	32
1969	34	40	43	42	43	44	39	40	33	32	31	28
1970	34	29	36	42	43	44	44	48	45	44	40	37
1971	45	49	62	62	58	59	64	62	50	52	50	44
1972	51	56	60	65	64	63	63	72	61	65	51	47
1973	55	60	68	63	65	61	54	52	46	42	37	30
1974	37	44	55	53	58	50	48	45	41	34	30	24
1975	29	34	44	54	57	51	51	53	46	46	46	39
1976	41	53	55	62	55	56	57	59	58	55	49	47
1977	57	68	84	81	78	74	64	74	71	63	55	51
1978	57	63	75	85	80	77	68	72	68	70	53	50
1979	53	58	73	72	68	63	64	68	60	54	41	35
1980	43	44	44	36	44	50	55	61	50	46	39	33
1981	37	40	49	44	45	38	36	34	28	29	27	29
1982	28	29	36	32	36	34	31	36	39	40	39	33

Statistical Tables

Table A Tail Areas of the Standard Normal Distribution



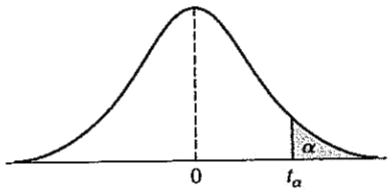
N	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064

(*Continued*)

Table A *(continued)*

N	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Table B Critical Values of the *t*-Distribution



ν \ α	.25	.10	.05	.025	.01	.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7649	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467

(Continued)

Table B *(continued)*

ν \ α	.25	.10	.05	.025	.01	.005
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896

(Continued)

Table B *(continued)*

ν \ α	.25	.10	.05	.025	.01	.005
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6459
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387

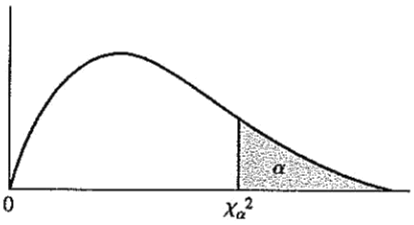
(Continued)

Table B *(continued)*

ν \ α	.25	.10	.05	.025	.01	.005
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
∞	0.6745	1.2816	1.6449	1.96	2.3263	2.5758

D. B. Owen, *Handbook of Statistical Tables*, © 1962, Addison-Wesley Publishing Co. Copyright renewal © 1990. Reprinted by permission of Pearson Education, Inc. Publishing as Pearson Addison Wesley.

Table C Critical Values of the Chi-Square (χ^2) Distribution



ν \ α	.995	.99	.975	.95	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188

(Continued)

Table C *(continued)*

ν \ α	.995	.99	.975	.95	.05	.025	.01	.005
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	23.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.641	48.278	50.993
29	13.121	14.257	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	44.985	48.232	51.191	55.003
32	15.134	16.362	18.291	20.072	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	52.192	55.668	59.892	62.883
38	19.289	20.691	22.878	24.884	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766

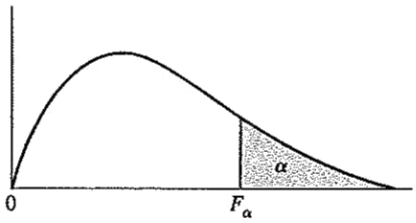
(Continued)

Table C *(continued)*

ν \ α	.995	.99	.975	.95	.05	.025	.01	.005
41	21.421	22.906	25.215	27.326	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	61.656	65.410	69.957	73.166
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.545	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215
80	51.712	53.540	57.153	60.392	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.930	124.342	129.561	135.807	140.169

For large values of ν, one can use approximation $\chi^2_\alpha = (\sqrt{2\nu - 1} + N_\alpha)^2/2$, where N_α is the α% upper tail standardized normal deviate given in Table A. For example, when $\nu = 100$,
$\chi^2_{.05} \simeq (\sqrt{2(100) - 1} + 1.645)^2/2 = 124.059$.

Table D Critical Values of the F-Distribution for $F_{0.05}(\nu_1, \nu_2)$



ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

(Continued)

Table D *(continued)*

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

(Continued)

Table D *(continued)*

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, 1958, by permission of E. S. Pearson and the Biometrika Trustees.

Table E Critical Values of the F-Distribution for $F_{0.01}(\nu_1, \nu_2)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

(Continued)

Table E *(continued)*

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, 1958, by permission of E. S. Pearson and the Biometrika Trustees.

Table F Empirical Cumulative Distribution of $n(\hat{\phi} - 1)$ for $\phi = 1$

	Probability of a Smaller Value								
n	.01	.025	.05	.10	.50	.90	.95	.975	.99
					(a) $\hat{\phi}$				
25	−11.8	−9.3	−7.3	−5.3	−.82	1.01	1.41	1.78	2.28
50	−12.8	−9.9	−7.7	−5.5	−.84	.97	1.34	1.69	2.16
100	−13.3	−10.2	−7.9	−5.6	−.85	.95	1.31	1.65	2.09
250	−13.6	−10.4	−8.0	−5.7	−.86	.94	1.29	1.62	2.05
500	−13.7	−10.4	−8.0	−5.7	−.86	.93	1.29	1.61	2.04
∞	−13.7	−10.5	−8.1	−5.7	−.86	.93	1.28	1.60	2.03
					(b) $\hat{\phi}_\mu$				
25	−17.2	−14.6	−12.5	−10.2	−4.22	−.76	.00	.64	1.39
50	−18.9	−15.7	−13.3	−10.7	−4.29	−.81	−.07	.53	1.22
100	−19.8	−16.3	−13.7	−11.0	−4.32	−.83	−.11	.47	1.13
250	−20.3	−16.7	−13.9	−11.1	−4.34	−.84	−.13	.44	1.08
500	−20.5	−16.8	−14.0	−11.2	−4.35	−.85	−.14	.42	1.07
∞	−20.6	−16.9	−14.1	−11.3	−4.36	−.85	−.14	.41	1.05
					(c) $\hat{\phi}_t$				
25	−22.5	−20.0	−17.9	−15.6	−8.49	−3.65	−2.51	−1.53	−.46
50	−25.8	−22.4	−19.7	−16.8	−8.80	−3.71	−2.60	−1.67	−.67
100	−27.4	−23.7	−20.6	−17.5	−8.96	−3.74	−2.63	−1.74	−.76
250	−28.5	−24.4	−21.3	−17.9	−9.05	−3.76	−2.65	−1.79	−.83
500	−28.9	−24.7	−21.5	−18.1	−9.08	−3.76	−2.66	−1.80	−.86
∞	−29.4	−25.0	−21.7	−18.3	−9.11	−3.77	−2.67	−1.81	−.88

Reproduced with minor notation adjustment from Table 10.A.1 in Fuller (1996). Copyright © 1996 by John Wiley & Sons, Inc. This material is used by permission of John Wiley & Sons, Inc.

Table G Empirical Cumulative Distribution of T for $\phi = 1$

	Probability of a Smaller Value								
n	.01	.025	.05	.10	.50	.90	.95	.975	.99
				(a) T					
25	−2.65	−2.26	−1.95	−1.60	−.47	.92	1.33	1.70	2.15
50	−2.62	−2.25	−1.95	−1.61	−.49	.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	−.50	.90	1.29	1.64	2.04
250	−2.58	−2.24	−1.95	−1.62	−.50	.89	1.28	1.63	2.02
500	−2.58	−2.23	−1.95	−1.62	−.50	.89	1.28	1.62	2.01
∞	−2.58	−2.23	−1.95	−1.62	−.51	.89	1.28	1.62	2.01
				(b) T_μ					
25	−3.75	−3.33	−2.99	−2.64	−1.53	−.37	.00	.34	.71
50	−3.59	−3.23	−2.93	−2.60	−1.55	−.41	−.04	.28	.66
100	−3.50	−3.17	−2.90	−2.59	−1.56	−.42	−.06	.26	.63
250	−3.45	−3.14	−2.88	−2.58	−1.56	−.42	−.07	.24	.62
500	−3.44	−3.13	−2.87	−2.57	−1.57	−.44	−.07	.24	.61
∞	−3.42	−3.12	−2.86	−2.57	−1.57	−.44	−.08	.23	.60
				(c) T_t					
25	−4.38	−3.95	−3.60	−3.24	−2.14	−1.14	−.81	−.50	−.15
50	−4.16	−3.80	−3.50	−3.18	−2.16	−1.19	−.87	−.58	−.24
100	−4.05	−3.73	−3.45	−3.15	−2.17	−1.22	−.90	−.62	−.28
250	−3.98	−3.69	−3.42	−3.13	−2.18	−1.23	−.92	−.64	−.31
500	−3.97	−3.67	−3.42	−3.13	−2.18	−1.24	−.93	−.65	−.32
∞	−3.96	−3.67	−3.41	−3.13	−2.18	−1.25	−.94	−.66	−.32

Reproduced with minor notation adjustment from Table I0.A.2 in Fuller (1996). Copyright © 1996 by John Wiley & Sons, Inc. This material is used by permission of John Wiley & Sons, Inc.

Table H Percentiles of $n(\hat{\Phi} - 1)$ for the Zero Mean Seasonal Models

		Probability of a Smaller Value								
$n = ms$		.01	.025	.05	.10	.50	.90	.95	.975	.99
	20	−12.20	−9.66	−7.70	−5.62	−.67	1.83	2.43	3.02	3.74
	30	−12.67	−9.93	−7.83	−5.71	−.71	1.72	2.28	2.79	3.43
	40	−12.98	−10.11	−7.94	−5.78	−.73	1.67	2.21	2.69	3.30
$s = 2$	100	−13.63	−10.50	−8.23	−5.93	−.76	1.58	2.10	2.54	3.07
	200	−13.88	−10.65	−8.34	−5.99	−.77	1.56	2.06	2.50	3.01
	400	−14.01	−10.73	−8.41	−6.02	−.77	1.54	2.04	2.48	2.97
	∞	−14.14	−10.81	−8.47	−6.06	−.78	1.53	2.02	2.46	2.94
	40	−13.87	−10.89	−8.67	−6.40	−.61	2.86	3.69	4.41	5.29
	60	−14.10	−11.09	−8.86	−6.51	−.65	2.68	3.47	4.13	4.97
	80	−14.31	−11.23	−8.94	−6.54	−.67	2.60	3.38	4.02	4.82
$s = 4$	200	−14.82	−11.57	−9.08	−6.59	−.70	2.50	3.23	3.85	4.55
	400	−15.04	−11.70	−9.12	−6.59	−.71	2.47	3.19	3.80	4.47
	800	−15.15	−11.77	−9.14	−6.59	−.71	2.46	3.17	3.78	4.43
	∞	−15.27	−11.85	−9.16	−6.59	−.71	2.45	3.15	3.75	4.39
	120	−18.10	−14.25	−11.51	−8.76	−.53	5.54	7.02	8.09	9.70
	180	−18.04	−14.31	−11.55	−8.74	−.62	5.25	6.66	7.82	9.17
	240	−18.02	−14.33	−11.56	−8.73	−.65	5.13	6.52	7.69	8.97
$s = 12$	600	−18.00	−14.35	−11.57	−8.72	−.69	4.97	6.34	7.49	8.71
	1200	−18.00	−14.35	−11.58	−8.72	−.69	4.93	6.30	7.42	8.65
	2400	−18.00	−14.35	−11.58	−8.72	−.69	4.92	6.28	7.39	8.63
	∞	−17.99	−14.35	−11.58	−8.72	−.69	4.90	6.27	7.36	8.61

Reproduced with Minor Notation Adjustment from Table 2 in Dickey, Hasza, and Fuller (1984), by permission of the American Statistical Association. Copyright 1984 by the American Statistical Association. All rights reserved.

Table I Percentiles of the Studentized T Statistic for the Zero Mean Seasonal Models

		Probability of a Smaller Value								
$n = ms$		.01	.025	.05	.10	.50	.90	.95	.975	.99
	20	−2.69	−2.27	−1.94	−1.57	−.29	1.07	1.48	1.85	2.28
	30	−2.64	−2.24	−1.93	−1.58	−.31	1.04	1.43	1.79	2.21
	40	−2.62	−2.23	−1.93	−1.58	−.32	1.03	1.42	1.76	2.18
$s = 2$	100	−2.58	−2.22	−1.92	−1.59	−.34	1.01	1.39	1.72	2.12
	200	−2.57	−2.22	−1.92	−1.59	−.35	1.00	1.38	1.71	2.10
	400	−2.56	−2.22	−1.92	−1.59	−.35	1.00	1.38	1.70	2.09
	∞	−2.55	−2.22	−1.92	−1.60	−.35	.99	1.38	1.69	2.08
	40	−2.58	−2.20	−1.87	−1.51	−.20	1.14	1.53	1.86	2.23
	60	−2.57	−2.20	−1.89	−1.52	−.21	1.10	1.49	1.82	2.20
	80	−2.56	−2.20	−1.89	−1.53	−.22	1.09	1.47	1.80	2.18
$s = 4$	200	−2.56	−2.20	−1.90	−1.53	−.23	1.07	1.45	1.78	2.17
	400	−2.56	−2.20	−1.90	−1.53	−.24	1.07	1.44	1.77	2.16
	800	−2.56	−2.20	−1.90	−1.53	−.24	1.07	1.44	1.77	2.16
	∞	−2.56	−2.20	−1.90	−1.53	−.24	1.07	1.44	1.77	2.16
	120	−2.50	−2.08	−1.77	−1.39	−.10	1.20	1.55	1.87	2.24
	180	−2.49	−2.09	−1.77	−1.41	−.12	1.17	1.53	1.85	2.22
	240	−2.49	−2.10	−1.77	−1.42	−.13	1.16	1.53	1.85	2.21
$s = 12$	800	−2.49	−2.10	−1.79	−1.43	−.14	1.15	1.52	1.84	2.20
	1200	−2.49	−2.10	−1.79	−1.43	−.14	1.15	1.52	1.34	2.20
	2400	−2.49	−2.10	−1.80	−1.43	−.14	1.15	1.52	1.84	2.20
	∞	−2.49	−2.10	−1.80	−1.44	−.14	1.15	1.52	1.84	2.20

Reproduced with Minor Notation Adjustment from Table 3 in Dickey, Hasza, and Fuller (1984), by permission of the American Statistical Association. Copyright 1984 by the American Statistical Association. All rights reserved.

Table J Empirical Percentiles of the Normalized Unit Root Estimator for Aggregate Time Series: AR(1) Model without Intercept $n_A(\hat{\Phi} - 1)$

Order of aggregation $m = 1$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−11.87	−9.35	−7.32	−5.32	1.01	1.40	1.78	2.28
50	−12.82	−9.91	−7.69	−5.52	0.97	1.34	1.69	2.15
100	−13.30	−10.19	−7.88	−5.61	0.95	1.31	1.65	2.08
250	−13.59	−10.36	−7.99	−5.67	0.93	1.29	1.62	2.04
500	−13.69	−10.42	−8.03	−5.69	0.93	1.29	1.61	2.03
750	−13.72	−10.44	−8.04	−5.70	0.93	1.28	1.61	2.03
∞	−13.78	−10.48	−8.07	−5.71	0.93	1.28	1.60	2.02
Order of aggregation $m = 2$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−9.36	−7.31	−5.56	−4.04	1.23	1.66	2.09	2.64
50	−9.72	−7.49	−5.77	−4.07	1.17	1.58	1.97	2.48
100	−9.90	−7.58	−5.82	−4.11	1.15	1.54	1.91	2.40
250	−10.01	−7.64	−5.85	−4.13	1.13	1.51	1.87	2.35
500	−10.05	−7.66	−5.87	−4.13	1.13	1.50	1.86	2.33
750	−10.06	−7.66	−5.87	−4.14	1.13	1.50	1.86	2.33
∞	−10.09	−7.68	−5.88	−4.14	1.12	1.49	1.85	2.32
Order of aggregation $m = 3$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−8.79	−6.83	−5.29	−3.78	1.27	1.70	2.13	2.72
50	−9.15	−7.02	−5.40	−3.82	1.21	1.62	2.02	2.56
100	−9.33	−7.12	−5.45	−3.84	1.19	1.58	1.96	2.47
250	−9.44	−7.18	−5.48	−3.85	1.17	1.56	1.93	2.42
500	−9.47	−7.20	−5.49	−3.86	1.17	1.55	1.92	2.41
750	−9.48	−7.20	−5.49	−3.86	1.17	1.55	1.91	2.40
∞	−9.51	−7.21	−5.50	−3.86	1.16	1.54	1.91	2.39

(*Continued*)

Table J *(continued)*

Order of aggregation $m = 4$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−8.70	−6.73	−5.19	−3.67	1.30	1.73	2.16	2.76
50	−9.00	−6.89	−5.29	−3.71	1.24	1.65	2.05	2.59
100	−9.16	−6.96	−5.34	−3.74	1.21	1.61	1.99	2.50
250	−9.25	−7.01	−5.37	−3.75	1.19	1.58	1.95	2.45
500	−9.28	−7.02	−5.37	−3.75	1.18	1.57	1.94	2.43
750	−9.29	−7.03	−5.38	−3.76	1.18	1.57	1.94	2.42
∞	−9.31	−7.04	−5.38	−3.76	1.18	1.57	1.93	2.41
Order of aggregation $m = 6$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−8.62	−6.53	−5.11	−3.62	1.30	1.74	2.16	2.76
50	−8.87	−6.76	−5.19	−3.65	1.25	1.65	2.05	2.60
100	−8.99	−6.83	−5.22	−3.67	1.22	1.61	1.99	2.51
250	−9.06	−6.87	−5.25	−3.68	1.20	1.59	1.96	2.45
500	−9.09	−6.88	−5.25	−3.68	1.19	1.58	1.95	2.44
750	−9.10	−6.88	−5.26	−3.68	1.19	1.57	1.94	2.44
∞	−9.11	−6.89	−5.26	−3.68	1.19	1.57	1.93	2.43
Order of aggregation $m = 12$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−8.44	−6.51	−5.02	−3.55	1.31	1.75	2.18	2.77
50	−8.76	−6.68	−5.12	−3.60	1.25	1.66	2.07	2.61
100	−8.92	−6.76	−5.17	−3.62	1.22	1.62	2.01	2.52
250	−9.02	−6.81	−5.19	−3.64	1.21	1.60	1.97	2.48
500	−9.05	−6.83	−5.20	−3.64	1.20	1.59	1.96	2.46
750	−9.06	−6.83	−5.21	−3.64	1.20	1.59	1.96	2.45
∞	−9.08	−6.85	−5.21	−3.65	1.19	1.58	1.95	2.44

This table was constructed by Paulo J. F. C. Teles using the Monte Carlo method. Details are given in Teles (1999).

Table K Empirical Percentiles of the Studentized Statistic for Aggregate Time Series: AR(1) Model without Intercept $T_{\hat{\Phi}} = (\hat{\Phi} - 1)/S_{\hat{\Phi}}$

Order of aggregation $m = 1$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.07
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.24	−1.95	−1.62	0.89	1.28	1.63	2.01
500	−2.58	−2.24	−1.95	−1.62	0.89	1.28	1.62	2.00
750	−2.58	−2.24	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	1.99
Order of aggregation $m = 2$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.28	−1.95	−1.67	−1.36	1.24	1.69	2.11	2.61
50	−2.24	−1.92	−1.66	−1.36	1.21	1.66	2.06	2.52
100	−2.21	−1.91	−1.65	−1.37	1.20	1.64	2.03	2.48
250	−2.20	−1.90	−1.65	−1.37	1.19	1.63	2.01	2.46
500	−2.19	−1.90	−1.65	−1.37	1.19	1.63	2.01	2.45
750	−2.19	−1.90	−1.65	−1.37	1.19	1.63	2.01	2.45
∞	−2.19	−1.90	−1.65	−1.37	1.19	1.63	2.00	2.44
Order of aggregation $m = 3$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.18	−1.87	−1.60	−1.31	1.30	1.77	2.20	2.71
50	−2.16	−1.85	−1.60	−1.31	1.27	1.74	2.15	2.63
100	−2.14	−1.85	−1.60	−1.31	1.26	1.72	2.12	2.59
250	−2.14	−1.84	−1.60	−1.32	1.25	1.71	2.11	2.57
500	−2.13	−1.84	−1.60	−1.32	1.25	1.71	2.10	2.56
750	−2.13	−1.84	−1.60	−1.32	1.25	1.70	2.10	2.56
∞	−2.13	−1.84	−1.60	−1.32	1.25	1.70	2.10	2.55

(*Continued*)

Table K *(continued)*

Order of aggregation $m = 4$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.16	−1.85	−1.58	−1.29	1.33	1.80	2.22	2.73
50	−2.13	−1.83	−1.58	−1.29	1.31	1.77	2.18	2.66
100	−2.12	−1.82	−1.58	−1.29	1.29	1.75	2.15	2.62
250	−2.11	−1.82	−1.58	−1.30	1.28	1.74	2.14	2.60
500	−2.11	−1.82	−1.59	−1.30	1.28	1.74	2.14	2.59
750	−2.11	−1.82	−1.58	−1.30	1.28	1.74	2.13	2.59
∞	−2.11	−1.82	−1.58	−1.30	1.28	1.74	2.13	2.58
Order of aggregation $m = 6$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.14	−1.82	−1.56	−1.27	1.34	1.83	2.26	2.80
50	−2.11	−1.81	−1.56	−1.28	1.32	1.79	2.21	2.70
100	−2.10	−1.80	−1.56	−1.28	1.31	1.77	2.18	2.65
250	−2.09	−1.80	−1.56	−1.28	1.30	1.76	2.16	2.62
500	−2.09	−1.80	−1.56	−1.28	1.30	1.76	2.16	2.62
750	−2.09	−1.80	−1.56	−1.28	1.29	1.75	2.15	2.61
∞	−2.09	−1.80	−1.56	−1.28	1.29	1.75	2.15	2.61
Order of aggregation $m = 12$								
n_A	.01	.025	.05	.10	.90	.95	.975	.99
25	−2.12	−1.80	−1.55	−1.26	1.36	1.84	2.27	2.80
50	−2.10	−1.80	−1.55	−1.27	1.33	1.81	2.23	2.72
100	−2.09	−1.80	−1.55	−1.27	1.32	1.79	2.21	2.68
250	−2.08	−1.79	−1.55	−1.27	1.31	1.78	2.19	2.65
500	−2.08	−1.79	−1.55	−1.28	1.31	1.78	2.19	2.65
750	−2.08	−1.79	−1.55	−1.28	1.31	1.78	2.19	2.64
∞	−2.08	−1.79	−1.55	−1.28	1.31	1.77	2.18	2.64

This table was constructed by Paulo J. F. C. Teles using the Monte Carlo method. Details are given in Teles (1999).

Author Index

Abraham, B., 76, 128, 223, 235, 549, 553, 563
Agiakloglou, C., 544, 553
Ahn, S. K., 439, 440, 561
Ahsanullah, M., 549, 553
Akaike, H., 133, 156, 157, 319, 464, 471, 473, 553
Al-Osh, 549, 553
Ali, M. M., 145, 553, 562
Amemiya, T., 546, 549, 553
Anderson, T. W., 415, 521, 553, 554
Andrews, P. F., 121, 554
Ansley, C. F., 145, 153, 411, 413, 483, 554, 559
Arbib, M. A., 464, 559
Ashley, R. A., 537, 554

Bartlett, M. S., 19, 20, 112, 308, 330, 402, 425, 473, 486, 554
Beguin, J. M., 133, 554
Bell, W. R., 109, 164, 223, 554, 556, 558
Blackman, R. B., 305, 310, 311, 554
Bloomfield, P., 261, 554
Bohlin, T., 483, 554
Bollerslev, T., 370, 375, 554
Boot, T. C. G., 549, 554
Box, G. E. P., 25, 26, 70, 85, 114, 129, 132, 143, 145, 153, 212, 218, 219, 223, 322, 401, 402, 403, 408, 410, 411, 417, 429, 435, 443, 449, 476, 553, 554, 555, 559, 560, 562
Brewer, K. R. W., 527, 528, 546, 549, 555
Brillinger, D. R., 112, 115, 261, 358, 555
Bryant, P. G., 376, 555
Bucy, R. S., 464, 483, 559
Burr, I. W., 112, 555
Buys-Ballot, C. H. D., 160, 555

Campbell, M. J., 121, 555
Chan, N. H., 186, 555
Chan, W. S., 233, 235, 498, 506, 550, 555
Chang, I., 223, 226, 555
Chen, C., 223, 230, 235, 555
Cheung, Y. W., 544, 555
Chow, G. C., 549, 555
Chuang, A., 235, 553
Clarke, D. G., 336, 555
Cleveland, W. S., 128, 275, 555
Cohen, K. J., 549, 555
Cooley, J. W., 259, 556
Cox, D. R., 85, 554
Crato, N., 492, 494, 556, 562
Cupingood, L. A., 164, 556

Dagum, E. B., 164, 556
Daniell, P. J., 314, 556
Davidson, J. E. H., 434, 556
Davies, N., 501, 561
Day, T. E., 380, 556
Deistler, M., 416, 453, 557
Denton, F., 549, 556
Dickey, D. A., 109, 190, 191, 193, 194, 195, 201, 206, 207, 208, 209, 543, 544, 545, 546, 547, 556, 561
Di Fonzo, 550, 556
Draper, N., 148, 556
Dunsmuir, W. T. M., 234, 416, 556, 557
Durbin, J., 411, 556

Engel, E. M. R. A., 549, 556
Engle, R. F., 368, 369, 380, 429, 436, 437, 549, 556

Falb, P. L., 464, 559
Fan, J., 499, 556
Feibes, W., 549, 554
Feller, W., 501, 556
Findley, D. F., 157, 557
Fisher, R. A., 294, 557

Fox, A. J., 223, 557
Fuller, W. A., 20, 109, 186, 190, 191, 193, 194, 195, 201, 207, 208, 209, 543, 544, 545, 546, 547, 556, 557

Gabr, M. M., 498, 499, 506, 534, 562
Gamerman, D., 483, 561
Gantmacher, F. R., 391, 397, 557
Geweke, J. M., 494, 528, 557
Gnedenko, B. V., 20, 557
Godfrey, M. D., 261, 557
Good, I. J., 259, 557
Gourieroux, C., 133, 554, 557
Gradshteyn, 491, 557
Graham, A., 409, 444, 557
Granger, C. W. J., 332, 368, 429, 436, 437, 539, 556, 557
Gray, H. L., 133, 557
Guckenheimer, J., 115, 555
Gudmundsson, G., 483, 550, 557
Guerrero, 549, 557
Guttman, I., 549, 562
Guttorp, P. E., 115, 555

Hall, A. D., 145, 416, 560
Halmos, P. R., 521, 557
Hamilton, J. D., 201, 557
Hannan, E. J., 20, 157, 358, 387, 402, 408, 414, 416, 448, 453, 557
Hanssen, D. M., 362, 559
Harrison, P. J., 483
Hartley, H. O., 295, 557
Harvey, A. C., 549, 557
Hasza, D. P., 207, 208, 209, 556
Helmer, R. M., 336, 339, 349, 558
Henry, D. F., 434, 556
Herzberg, A. M., 121, 554
Heyse, J. F., 336, 408, 410, 417, 442, 447, 558
Hillmer, S. C., 145, 164, 223, 416, 554, 558
Hinich, M. J., 534, 535, 537, 554, 558
Hodgess, E. M., 550, 558
Hosking, J. R. M., 416, 558
Hotta, 550, 558
Hsu, 235, 563

Izenman, A. J., 112, 212, 221, 222, 558

Jazwinski, A. H., 483, 558
Jenkins, G. M., 25, 26, 70, 114, 132, 143, 313, 322, 358, 411, 443, 476, 554, 558
Johansen, S., 439, 440, 441, 558
Johansson, J. K., 336, 339, 349, 558
Johnson, R. A., 487, 558
Jones, R. H., 528, 558
Juselius, K., 441, 558

Kalman, R. E., 88, 464, 483, 558, 559
Kedem, G., 416, 561
Kelley, G. D., 133, 557
Kohn, R., 483, 549, 559
Kolmogorov, A., 88, 559
Koopmans, L. H., 521, 559
Kremers, J. M., 429, 559

Lai, K. S., 544, 555
Lamoureux, 380, 559
Lastrapes, W. D., 380, 559
Ledolter, J., 76, 128, 230, 553, 559
Lee, J. H., 230, 559
Lee, R. M., 549, 559
Lee, T. H., 539, 557
Lewis, C. M., 380, 556
Li, W. K., 235, 373, 416, 559
Lilien, D. M. 369, 556
Lim, K. S., 499, 502, 563
Lin, A., 549, 555
Lin, K. S., 121, 559
Lisman, J. H. C., 549, 554
Liu, L. M., 230, 235, 349, 362, 555, 559
Liu, T. C., 549, 556
Ljung, G. M., 145, 153, 230, 559
Lütkepohl, H., 528, 546, 549, 559

Marquardt, P. W., 148, 559
Martin, R. D., 223, 559
Masarotto, 235, 559
McIntire, D. D., 133, 557
McLeod, A. I., 373, 416, 559
Mehra, R. K., 483, 560
Mehta, J., 549, 563
Meinhold, R. J., 483, 560
Miller, R., 109, 146, 235, 556, 560, 563
Monfort, A., 133, 554
Montgomery, D. C., 212, 220, 560
Montmarquette, C., 549, 563
Müller, W., 549, 555
Müller, D., 333, 560

Nelson, D. B., 371, 560
Newbold, P., 145, 153, 332, 368, 411, 413, 454, 544, 553, 554, 557, 560
Newton, H. J., 319, 560
Ng, V. K., 380, 556
Nicholls, D. F., 145, 415, 416, 560
Nicholson, A. J., 115, 560
Niemi, H., 549, 560
Noble, B., 391, 466, 560

O'Hagan, A., 483, 560
Osborn, R. R., 416, 560
Oster, G., 115, 555

Pack, D. J., 362, 560
Padberg, M. W., 549, 555
Pagano, M., 319, 560
Palda, K. S., 336, 421, 560
Pantula, S. G., 544, 560
Parzen, E., 19, 157, 311, 319, 521, 560
Patterson, D. M., 537, 554
Pena, D., 435, 506, 560, 561
Perron, P., 200, 201, 206, 548, 561
Petruccelli, 501, 561
Phadke, M. S., 416, 483, 561
Phillips, P. C. B., 200, 201, 206, 368, 548, 561
Pierce, D. A., 153, 164, 554, 561
Pierse, R. G., 548, 561
Porter-Hudak, S., 494, 557
Priestley, M. B., 121, 312, 358, 483, 499, 561

Quenouille, M. H., 22, 235, 413, 414, 449, 528, 561
Quinn, B. G., 157, 557

Rao, C. R., 496, 561
Ray, B. K., 492, 556
Reinsel, G. C., 114, 132, 143, 322, 411, 439, 440, 443, 476, 554, 561
Robins, 369, 556
Robinson, R. M., 234, 556
Rose, D. E., 549, 561
Rosenblatt, M., 112, 555
Rothschild, M., 380, 556
Ryzhik, 491, 557

Said, S., 206, 561
Schmidt, A. M., 483, 561
Schuster, A., 290, 561
Schwartz, G., 157, 561
Schwert, G. W., 544, 561
Searle, S. R., 392, 466, 561
Shaman, P., 92, 562
Shen, S. Y., 454, 562
Shibata, R., 156, 562
Singpurwalla, N. D., 483, 560
Slutzky, E., 47, 562
Smith, H., 148, 556
Smith, M. A., 376, 555
Snell, A. J., 548, 561
Srba, F., 434, 556
Stevens, C. F., 483, 557
Stock, J. H., 435, 562
Stone, M., 157, 562
Stram, D. O., 509, 512, 513, 516, 517, 546, 549, 550, 562, 563
Subba Rao, T., 483, 498, 499, 506, 534, 561, 562

Teles, P., 494, 538, 539, 540, 541, 543, 548, 562
Teräsvirta, T., 549, 562
Tiao, G. C., 129, 130, 132, 145, 164, 212, 218, 219, 223, 226, 363, 401, 402, 403, 408, 410, 416, 417, 429, 449, 457, 506, 515, 528, 533, 534, 537, 549, 554, 555, 558, 561, 562, 563
Tong, H., 115, 121, 498, 499, 502, 503, 506, 555, 562, 563
Tsay, R. S., 130, 132, 223, 230, 235, 449, 457, 501, 502, 503, 506, 561, 562, 563
Tukey, J. W., 26, 259, 261, 305, 310, 311, 554, 556, 563

Vasconcellos, 550, 558

Walker, A. M., 121, 555
Watson, M. W., 435, 562
Watts, D. G., 313, 358, 558
Wegman (1986), E. J., 91, 563
Weatherby, G., 212, 220, 560
Wei, C. Z., 186, 555
Wei, S., 235, 563
Wei, W. W. S., 164, 230, 233, 235, 333, 336, 363, 408, 410, 417, 442, 447, 454, 494, 509, 512, 513, 515, 516, 517, 523, 528, 533, 534, 537, 538, 539, 540, 541, 543, 546, 549, 550, 553, 555, 556, 558, 559, 562, 563

Weiss, A. A., 528, 546, 549, 563
Werner, H. J., 528, 549, 563
Whittle, P., 88, 112, 295, 563
Wichern, D. W., 146, 235, 487, 558, 560, 563
Wiener, N., 88, 563
Wilks, S. S., 486, 563
Wold, H. O. A., 24, 26, 160, 271, 563
Wu, R. Y., 546, 549, 553

Yaglom, A. M., 88, 563
Yao, G., 499, 556
Yeo, S., 434, 556
Young, P. C., 483, 563
Yule, G. U., 33, 42, 112, 563

Zabell, S. L., 212, 221, 222, 558
Zellner, A., 164, 528, 549, 563

Subject Index

Additive outlier, 223
Aggregate Series, 508
  limiting behavior, 515
Aggregation, 507
  of the ARIMA process, 507
  of seasonal model, 515
  matrix, 516
  order of, 508, 523
  (see also effect of aggregation, aggregate series, and temporal aggregation)
Akaike's information criterion (AIC), 156, 407
Aliasing, 285, 286
Amplitude, 70
Amplitude, spectrum, 250, 354
Analysis of variance for periodogram, 290
ARCH models, 366, 368
ARFIMA model, 490, 492
ARMA (See Autoregressive moving average), 57
ARIMA (See Autoregressive integrated moving average), 71
ARMAX model, 322
Autocorrelation function (ACF), 10
  covariance of sample, 20
  for stationary processes, 109
  generating, 25
  inverse, 122
  matrix, 384
  of AR(1) process, 34
  of AR(2) process, 41
  of AR(p) process, 45
  of MA(1) process, 48
  of MA(2) process, 51
  of MA(q) process, 52
  of ARMA(1, 1) process, 60, 64
  of ARMA(p, q) process, 58
  partial, 11
  properties of, 10
  sample, 20
  α-trimmed sample, 233
  standard error of sample, 21, 35
Autocorrelation matrix function, 384
Autocovariance function, 10
  definition of, 10, 272
  generating, 25
  sample, 18
  spectral representation of, 267
Autocovariance generating function, 25
  of ARMA(p, q) model, 275
  inverse, 123
Autocovariance matrix function, 384
Autoregressive integrated moving average (ARIMA) processes, 71
  ARIMA model, 72, 271
  ARIMA(0, 1, 1) process, 74
  forecasting of, 91
  IMA(1, 1) process, 74
  seasonal, 164
Autoregressive moving average processes, 26, 57
  AR representation, 23
  AR(1) process, 34
  AR(2) process, 39
  AR(p) process, 45
  ARMA(1, 1) process, 59
  ARMA(p, q) process, 57
  estimation of, 136
  exact likelihood function, 143
  MA presentation, 23
  MA(1) process, 48
  MA(2) process, 51
  MA(q) process, 52
  spectrum of, 274
  state space representation of, 466
  variance-covariance matrix of estimates, 148
  vector ARMA process, 386

Autoregressive processes, 33
  first-order, 34
  pth-order, 45
  second-order, 39
  stationarity of, 33, 38
  vector, 384
Autoregressive representation, 23
Autoregressive signal plus noise model, 487
Autoregressive spectral estimation, 319

Backcasting method, 140, 141
Backshift operator, 23, 27, 384
Bandwidth, of window, 302, 307
Bayesian information criterion (BIC), 156
Bilinear model, 499
  BL(p, q, r, s) model, 499
  BL(1, 0, 1, 1) model, 537
Bispectrum, 495
Blackman–Tukey window, 310
Box–Cox's transformation, 85, 109
Brownian motion, 187 , 440
Buys–Ballot table, 160, 161, 165

CAT criterion, 157
Canonical correlations, 483
Causal model, 323, 340, 354
Causality, 340, 363
  linear, 528
Cayley–Hamilton theorem, 466, 467
Central limit theorem for stationary processes, 199
Coherence, 350, 361
  estimation of, 358
Cointegrated processes, 428
  AR representation of, 431
  decomposition of, 434
  error-correction representation of, 433
  MA representation of, 430
  representation of, 430
Cointegrating vectors, 428
Cointegration, 428
  testing and estimating, 435
Complex conjugate, 238, 272, 351
Complex exponentials, 240
Conditional sum of squares function, 139
Confidence interval, 313
  for spectral ordinates, 313
Continuous time series, 1
Convolution, 262
Correlation function, 7
Correlation matrix function, 382
  partial lag, 408
  sample, 401
Correlogram, 10
  sample, 20
Cospectrum, 350
Covariance matrix function, 382, 383, 391
Covariance matrix generating function, 442
Cramer's representation, 273, 351
Cross-correlogram, 326
Cross-correlation function (CCF), 325, 363
  and the transfer function, 325
  sample, 329
Cross-covariance function, 326, 383
Cross-covariance generating function, 349
Cross-spectrum, 349, 359
  and transfer function models, 359
  estimation of, 357
  interpretation of, 351
Cumulants, 495
Cyclical model, 268

Daniell's estimator, 315
Data, appendix, 565
  blowfly data, 2, 115, 223, 153, 475, 569
  contaimated blowfly data, 2, 223
  Colorado ski resort data, 183
  daily average number of truck manufacturing defects, 2, 111, 153, 228, 566
  daily closing stock price of Duke Energy Corporation, 216, 573
  Lydia Pinkham annual advertising and sales data, 335, 416, 437, 441, 574
  Monthly airline passengers in the U.S., 231, 578
  monthly average of hourly readings of ozone in downtown Los Angeles, 218
  monthly U.S. electricity consumption, 220
  monthly employment figures for young man, 172, 185, 572
  monthly unemployed young women, 68, 116, 153, 161, 518, 569
  monthly U.S. housing sales and housing starts, 364
  monthly U.S. new one-family houses sold, 546, 580
  Oklahoma and Louisiana spot prices for natural gas, 376, 579
  Quarterly U.S. beer production, 2, 178, 228, 573

Data *Continued*
series C of Box, Jenkins, and Reinsel, 132
series M of Box, Jenkins, and Reinsel, 476
U.S. hog data, 448, 575
simulated AR(1) series, 34, 36, 543
simulated AR(2) series, 43
simulated ARMA(1, 1) series, 34, 36, 545
simulated ARIMA series, 134, 135
simulated cointegrated series, 430
simulated MA(1) series, 50
simulated MA(2) series, 52
simulated series with outliers, 236
simulated random walk series, 78
simulated seasonal time series, 153, 170, 572
weekly births in New York City, 222
Wolf yearly sunspot numbers, 112, 153, 566
yearly cancer death rate for Pennsylvania, 117, 119, 153, 570
yearly number of lynx pelts sold, 121, 124, 153, 296, 317, 320, 571
yearly U.S. personal consumption of gasoline and oil, 220
yearly U.S. tobacco production, 2, 69, 120, 153, 570
yearly U.S. lumber production, 159
Data space, 471
Delta function, 263
Deterministic trends, 69
Deterministic trend term, 72, 110
test for, 110
Diagnostic checking, 152, 172
of ARMA models, 153
of transfer function models, 334
of vector time series models, 416
Difference equations, 26
auxiliary equation, 27
general solution of, 27
homogeneous, 27
nonhomogeneous, 27
particular solution of, 27
Differencing, 70, 109
overdifferencing, 109
Dirac's delta function, 263
Dirichlet conditions, 255
for periodic functions, 255
for nonperiodic functions, 258
Disaggregation, 483
Discrete time series, 1
Dual relationship, between AR(p) and MA(q) processes, 54
Effect(s)
of aggregation, on forecasting and parameter estimation, 520
of aggregation, on testing for a unit root, 540
of aggregation, on testing for linearity and normality, 534, 537
of differencing, 284
of moving average, 283
of systematic sampling and temporal aggregation on causality, 528
of systematic sampling on causality, 532
of temporal aggregation on causality, 533
of temporal aggregation, on forecasting, 522
Efficiency, relative, for forecasting future aggregates, 523
Energy spectrum, 250
Equivalent degrees, of freedom for smoothed spectrum, 316
Equivalent representation, of a vector ARMA model, 387, 451
Ergodicity, 19, 20
Ensemble, 6
average, 18
Estimation
ARMA spectral, 318
of autocorrelations, 16
of autocovariances, 16
autoregressive spectral, 319
conditional maximum likelihood, 138
of cross-spectrum, 357
of GARCH models, 373
least squares, 139, 148, 151
maximum likelihood, 140
of mean, 16
nonlinear, 145
parameter, 136
of spectrum, 289
of transfer function models, 332
unconditional maximum likelihood, 140
of vector time series models, 414
Euler relation, 238
Even function, 266
Eventual forecast functions, 100
Exact likelihood function, 143
Exponential autoregressive model, EXPAR(p, d) model, 499
Exponential smoothing, 97
Extended sample autocorrelation function (ESACF), 128, 130
asymptotic variance of, 130

Fast Fourier transform (FFT), 258, 259
Filter, 281
  Kalman, 463
  linear, 281
  moving average, 283
  transfer function, 283
Filtered output series, 329, 331
Fisher's test, 294
Flow variable, 507
Forcing function, 27
Forecast, 88
  computation of, 94
  error, 88, 89, 181, 343, 375, 416
  eventual functions of, 100
  limits of, 90
  minimum mean square error, 88, 342, 416
  updating, 99
  variance of, 90, 181, 343, 375, 416
Forecasting, 25, 88
  application of Hilbert space in, 521
  ARIMA models, 91
  effects of temporal aggregation on, 522
  seasonal models, 170, 175
  using transfer function models, 341
  vector ARMA models, 414
Fourier
  analysis, 237
  coefficients, 241, 242, 289
  frequencies, 241, 290
  inverse transform, 257
  series, 241
Fourier representation, 241, 289
  of finite sequences, 241
  of periodic sequences, 242
  of nonperiodic sequences (discrete-time Fourier transform), 247
  of periodic functions, 254
  of continuous-time functions (continuous-time Fourier transform), 254, 256
Fourier series, of the sequence, 241
Fourier transform, 247
  discrete-time, 247
  continuous-time, 256
  integral, 258
  inverse, 249, 257
  pair(s), 249, 258
Fractional difference, 489, 492
  estimation of, 492

Frequency, 70, 286
  fundamental, 244
  spacing of, sample, 248
Frequency domain analysis, 3, 237
  bivariate, 349
Frequency response function, of linear system, 359

Gain function, 352, 358
GARCH models, 366, 368
  estimation of, 373
  computation of forecast error variance, 375
Gaussian process, 10, 16
General ARIMA models, 168
Generalized function, 263
Generalized least squares (GLS), 367

Harmonic components, 244
Heteroscedasticity, 368
Hidden periodic components, 292
Hidden periodicity, of order m, 513
High pass filter, 284
Hilbert space, 520
  application in forecasting, 521
Homogeneous equation, 27

Identifiability, for vector ARMA models, 387
Identification
  of ARIMA models, 108
  of noise model, 331
  of transfer function models, 331
  of seasonal models, 172
  of Series W1–W7, 111
  of vector time series models, 401
Impulse response function, 324
Impulse response functions, typical, 324
Impulse response weights, 324–325
Information criterion (AIC), 133, 156
Information loss, 524
  due to aggregation in parameter estimation, 524
Inner product, 520
Innovation, of input, 464
Innovational outlier, 223
Input matrix, 464
Input series
  prewhitened, 329
  nonstationary, 343
Interpretation, of the cross-spectral functions, 351
Intervention analysis, 212
Intervention analysis, examples of, 215
Intervention models, 212

Interventions, 212
Inverse autocorrelation function (IACF), 122, 275
Inverse autocorrelation generating function, 125
Inverse autocovariance function, 275
Inverse Fourier transform (inverse Fourier integral), 257
Inverse process, 126
Invertibility, 25, 33, 385
  of AR models, 33
  of ARMA models, 47
  of MA models, 47
  of vector ARMA models, 386

Kalman filter, 463, 478
Kalman gain, 482
Kernel, 302
Kronecker product, 392

Lag window, 304, 305, 316, 358, 498
Leakage, 310
Least squares estimation, 139
  conditional, 139
  nonlinear, 146
  ordinary, 151
  unconditional, 140
Likelihood ratio test, 425, 439
Linear difference equations, 26
Linear manifold, 521
Linear time-invariant system, 463
Linear vector space, 520
Linearity test, 495
Long memory processes, 489
Low pass filter, 284

Matrix,
  absolutely summable, 385
  Coefficient, 384
  Correlation, 383
  Covariance, 382
  equation, 397
  polynomial, 385
  square summable, 385
Maximum likelihood estimation, 138
  GARCH models, 373
  ARMA models, 138
  conditional, 138
  unconditional, 140
  vector ARMA models, 415
m-dimensional white noise random vectors, 384

Mean function, 7
  sample, 17
  variance of sample, 18
Mean absolute error, 182
Mean absolute percentage error, 182
Mean percentage error, 181
Mean square error, 19, 88, 91, 181
Measurement equation, 464
Method of moments, 136
Metric, of the space, 520
Minimum mean square error forecasts, 89
  for ARMA models, 89
  for ARIMA models, 91
  for stationary input and output series, 342
  for nonstationary input and output series, 343
  for vector time series models, 416
Model identification, 108, 230, 331, 401, 417
Model selection, 156
  Akaike's AIC, 156
  Akaike's BIC, 156
  based on forecast errors, 181
  method of, 156
  Parzen's CAT, 157
  Schwartz's SBC, 157
Models
  autoregressive moving average (ARMA), 57
  autoregressive integrated moving average (ARIMA), 69, 71
  construction of transfer function, 329
  deterministic trend, 69
  GARCH, 366, 368
  general ARIMA, 168
  intervention, 212
  multiple-input transfer function, 361
  nonlinear time series, 498
  nonstationary time series, 69
  nonstationary vector autoregressive moving average, 400
  realizable, 323
  seasonal, 160, 164, 169
  state space, 463
  stationary time series, 33
  threshold autoregressive, 499
  time series regression, 366
  transfer function, 322
  vector, 382, 387, 394, 396, 397, 398
Moving average (MA)
  and autoregressive representations of time series process, 23

Moving average (MA) *Continued*
and autoregressive representations of vector processes, 384
representation, 24, 89
Moving average processes, 47
autocorrelation function of, 52
invertibility of, 47
MA(1) process, 48
MA(2) process, 51
MA(q) process, 52
stationarity of, 47
vector MA processes, 397
Multiple-input transfer function models, 361
Multiplicative model, 166
Multivariate linear regression models, 423
Multivariate partial autocorrelation matrix, 413

Noise model, 331
Noise spectrum, 360
Noise-to-signal ratio, 361
Nonaggregate series, 508
Nonhomogeneous equation, 27
Nonlinear autoregressive model, 499
NLAR(p) model, 499
NLAR(1) model, 499
Nonlinear moving average (NLMA) model, 498
Nonlinear processes, 489, 494
Nonlinear time series models, 498
Normality test, 495
Nonperiodic sequence, 247
Nonstationarity
homogeneous, 71
in the mean, 2, 69
in the variance and the autocovariance, 2, 80
Nonstationary time series models, 68
Nonstationary vector autoregressive moving average models, 400
nth order weakly stationary, 8
Nyquist frequency, 286

Observation equation, 464
Order of aggregation, 508, 523
Ordinary least squares (OLS)
estimation, 129, 151, 189, 366
Orthogonal basis, 241
Orthogonal functions, 238
Orthogonal subspaces, 522
Outliers, 212, 223
additive (AO), 223
detection of, 226
innovational, 223
level shift, 230
time series, 223
temporary change, 230
Outlier model, 224
Output equation, 464
Output (or observation) matrix, 464
Overdifferencing, 109, 401

PACF, IACF, and ESACF, for seasonal models, 169
Parseval's relation, 244, 250, 258, 290
Parsimony, 27
Partial autocorrelation function (PACF), 11, 15
for seasonal models, 169
of AR(1) process, 34
of AR(2) process, 42
of AR(p) process, 47
of ARMA(1, 1) process, 59
of ARMA(p, q) process, 59
of MA(1) process, 49
of MA(2) process, 52
of MA(q) process, 54
sample, 22
standard error of sample, 35
Partial autoregression matrices, 402
sample, 405
Partial autoregression matrix function, 404
Partial covariance matrix generating function, 443
Partial lag autocorrelation matrix, 408, 410
Partial lag correlation matrix, 413
sample, 414
Partial lag correlation matrix function, 408
Partial process, 442, 447
correlation matrix, 442, 447
correlation matrix function, 447
covariance matrix generating function, 443
sample correlation matrix function, 447
Parzen's CAT criterion, 157
Period, 160, 242
fundamental, 242
seasonal, 160
Periodic components, hidden, 292
Periodicity,
hidden, 513
test for, 292
Periodogram, 289
analysis, 289
sampling properties of, 290
Phase, 70

Phase spectrum, 354, 358
Polyspectrum, 495, 496
Portmanteau test, 334
  lack of fit test, 153
Positive semidefinite, 10
Power, of sequence, 245
Power spectrum, 245
Power transformation, 85
Prediction, 88
Predictor space, 471
Prewhitened input series, 329
Prewhitening, 331
  double, 332
Process
  completely stationary, 7
  stationary, spectral theory of, 264
  strictly stationary, 7
  strongly stationary, 7
Projection theorem, 521
Pulse function, 213

Quadrature spectrum, 350

Random walk model, 72, 189
  with drift, 73
Random walk plus noise model, 488
Rational spectrum, 275
Realizable model, 323
Realization, 6
Real-valued process, 7
Rectangular window, 306
Recursive algorithm, for computing $P(s)$, 411
Relationship
  between cross-correlation and transfer functions, 328
  of autocovariances between nonaggregate and aggregate series, 508
  between state space and ARMA models, 464

Sample
  autocorrelation function, 20
  autocovariance function, 18
  correlation matrix function, 401
  correlogram, 20
  cross-correlation function, 329
  cross-covariance function, 329
  inverse autocorrelation function, 126
  mean, 17
  partial autocorrelation function, 22
  partial autoregression matrices, 405
  spacing of frequencies, 248
  spectrum, 298
Sample function, 6
Sampling properties, of periodogram, 290
SBC criterion, 157
Schwartz's Bayesian criterion, 157
Seasonal adjusted series, 164
Seasonal models, 160
  ARIMA, 164
  forecasting, 170
  IACF for, 169
  model building for, 170
  PACF for, 169
  testing for unit root, 206
Seasonal dummy, 162
Seasonal period, 160
Seasonal time series, 3, 160, 206
  moving average method, 163
  regression method, 162
  seasonal adjustment, 164
  seasonal ARIMA models, 164
Second order stationary, in distribution, 7
Sequence
  absolutely summable, 250
  finite, 242
  nonperiodic, 247
  periodic, 242
  power of, 245
  square summable, 250
Short memory process, 492
Signal, of the system, 361
Signal-to-noise ratio, 361
Single-input transfer function models, 322
Smoothed spectrum, 301, 357
Smoothing
  in frequency domain, 301
  in time domain, 304
Space
  Hilbert, application of in forecasting, 521
  predictor, 471
Space representation, 464
Spectral density, 271
Spectral density matrix function, 422
Spectral distribution function, 267
Spectral distribution matrix function, 422
Spectral estimation, ARMA, 318
Spectral representation
  of covariance matrix function, 422
  of autocovariance functions, 267, 268
  of stationary processes, 272

Spectral window, 301, 305
Spectrum, 264, 274. *See also* Cospectrum; Cross spectrum
  amplitude, 250
  and autocovariance generating function, 274
  discrete, 269
  energy, 250
  estimation of, 289, 298
  of ARMA models, 274
  of linear filters, 281
  line, 270
  matrix, 422
  noise, 361
  phase, 354, 358
  power, 245
  properties of, 264
  quadrature, 358
  rational, 275
  sample, 298
  of seasonal models, 279
  smoothed, 301, 315, 357
  of sum of two independent processes, 278
  of white noise process, 276
Spectrum matrix, 422
Squared coherency, 350
Square summable, 250, 385
State, of the process, 464
State equation, 464
State space model, 463
  canonical correlation analysis and, 470
  data space, 471
  fitting, 470
  input equation, 464
  measurement equation, 464
  output equation, 464
  predictor space, 471
  system equation, 464
  transition equation, 464
State space representation of ARMA models, 464
State vector, 464
Stationarity, 7
  of autoregressive moving average processes, 57
  of autoregressive processes, 33, 38
  complete, 7
  covariance, 8
  in distribution, 7
  of vector ARMA processes, 386
  strict, 7
  strong, 7
  weak, 8
  wide sense, 8
Stationary process
  short memory process, 492
  spectral theory of, 264
Stationary time series models, 33
Steady state, 482
Step function, 213
Stochastic integral, 272
Stochastic process, 6
Stochastic trend, 71
Stochastic trend models, and differencing, 70
Stock variable, 507
Strongly stationary, 7
Subspaces, orthogonal, 521
Summary statistics, for partial autoregression, 406
Sunspot numbers, 112, 113, 504, 506
System, 463
  linear, 463
System equation, 464
Systematic sampling
  in time series, 507
  of ARIMA process, 526
  effect of, on causality, 528
Systematic sampling interval, 526

Tables, statistical, 581–599
  analysis of variance table for periodogram analysis, 290
  Buys-Ballot table for seasonal data, 165
  chi-square distribution, 585
  characteristics of ACF and PACF for stationary processes, 109
  cointegration and unit root test on residuals, 437
  equivalent degrees of freedom for various windows, 317
  F-distribution, 587
  Fisher's test for the largest periodogram ordinate, 295
  likelihood ratio test for cointegration rank, 440, 441
  some typical impulse response functions, 324–326
  standard normal distribution, 581
  summary of models for series W1-W7, 155
  t-distribution, 582
  unit root test
    for aggregate series, 596–599
    for nonaggregate series, 592–593
    for seasonal series, 594–595

TAR-GARCH models, 506
Temporal aggregation, 507
  of AR(p) process, 512
  of ARIMA process, 507
  of IMA(d, q), process, 511
  of ARIMA(p, d, q) process, 513
Threshold autoregressive models, 499
  modeling TAR models, 502
  self-exciting TAR models, 500
  TAR(k; $p_1, \ldots, p_k$; d) model, 500
  TAR(2; 1, 1; 1) model, 500
  TAR(2; 4, 11; 3) model, 503
  TAR(3; 11, 10, 10; 2) model, 505
  test for, 500
Time delay, 354
Time domain analysis, 3, 237
Time series, 1
Time series, homogeneous, nonstationary, 71
Time series aggregates, limiting behavior of, 515
Time series models
  nonlinear, 498
  nonstationary, 68
  seasonal, 160
  stationary, 33
  vectors, 382, 401
Time series outliers, 223
Time series regression, 162, 366
Time-invariant system, 463
Time-varying regression model, 483
Traditional methods, 162
Transfer function, 283, 322
Transfer function model, 322, 331
  and cross-spectrum, 359
  and vector ARMA model, 388
  construction of, 329
  diagnostic checks of, 334
  estimation of, 332
  forecasting of, 341
  identification of, 331
  multiple input, 361
  single input, 322
  stable, 322
Transition equation, 464
Transition matrix, 464
Trend
  deterministic, 69
  stochastic, 70
Trend-cycle, 162
Triangular window, 308
Trigonometric functions, 238
  orthogonality of, 240
Trigonometric identities, 240
Truncation point, 313, 357, 498
Tukey window, 311
Tukey–Hamming window, 311
Typical impulse response functions, 324

Unconditional least squares estimators, 140
Unconditional maximum likelihood estimation, 140
Unconditional sum of squares, 140
Unit root, 186
  AR(1) model, 189
  AR(p) model, 201
  a general model, 196
  seasonal model, 206
  testing for, 186, 196, 201, 540
  tables, 592–599
Updating
  forecasts, 99
  procedure for a state vector, 482

Variance
  of forecast errors, 90
  of sample autocorrelations, 21
  of sample autocovariances, 19, 20
  of sample cross-correlations, 330
  of sample mean, 17
  of sample partial autocorrelations, 22
  of sample spectrum, 300
  of smoothed spectrum, 315
Variance function, 7
Variance stabilizing transformations, 83
Variance-covariance matrix, 383
Vector time series models, 382
  AR(1) models, 387
  ARMA(1, 1) models, 398
  ARMA and transfer function models, 388
  AR representation of, 384
  AR(p) models, 394
  autoregressive moving average process, 386
  estimation of, 414
  forecasting of, 414
  identifiability of, 387
  identification of, 401
  invertibility of, 385, 389
  MA(1) models, 396
  MA(q) models, 397
  MA representation of, 384

Vector time series models *Continued*
m-dimensional white noise random, 384
nonstationarity, 400
stationarity of, 385, 389
unit roots and cointegration in, 428
Vector processes
autoregressive presentations of, 384
spectral properties of, 421

Weakly stationary, 8
Weighting function, 304, 305
White noise processes, 15, 152, 384
spectrum of, 276
Wiener process, 187
Window carpentry, 313
Window closing, 313
Window
bandwidth of, 302, 307
Bartlett's, 308
Blackman–Tukey, 310
hamming, 311
hanning, 311
lag, 304, 305, 358
leakage, 310
Parzen, 311
rectangular, 306
spectral, 301, 305, 311
triangular, 308
Tukey, 311
Tukey–Hamming, 311
Wold's decomposition, of stationary process, 271
Wold's representation, 24, 384

X-12 method, 164

Yule process, 42
Yule–Walker equations, 136, 403
matrix, generalized, 396
Yule–Walker estimators, 137